SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Department of Psychology

PhD program in Psychology, Linguistics, and Cognitive Neuroscience

Curriculum: Social, Cognitive, and Clinical Psychology

Cycle XXXIII

TRUST

in

BORDERLINE PERSONALITY DISORDER

PhD Candidate: **Anita Poggi**

Registration no.: 749738

Supervisors: Emanuele Preti, Juliette Richetin

Coordinator: Marco Perugini

ACADEMIC YEAR: 2019/20

CONTENTS

ACKNOWLEDGMENTS **6**

ABSTRACT **7**

INTRODUCTION **9**

CHAPTER 1. A SYSTEMATIC REVIEW OF TRUST PROCESSES IN BORDERLINE PERSONALITY DISORDER **14**

Overview 14

1. Introduction 16

1.1 Trust: Conceptualization and Operationalization 19

1.2 Trust processes and borderline personality disorder 22

2. Methods 24

3. Results 32

3.1. Trust processes in BPD: distal antecedents 32

3.1.1. Distal antecedents of trust appraisal: developmental factors 32

3.2. Trust processes in BPD: proximal antecedents 35

3.1.1. Prior beliefs and dispositions 35

3.1.2. Perception of the situation 40

3.1.3. The impact from trustors' emotional state on trust processing 42

3.3. Trust Appraisal 46

3.4. Behavioral manifestations: interpersonal cooperation and repair of ruptures 48

3.5. Trust Learning 52

4. Discussion 55

4.1. Implications for further research 59

4.2. Clinical implications 63

CHAPTER 2. THE EMOTIONAL COMPONENTS OF REJECTION SENSITIVITY AS A MEDIATOR BETWEEN BORDERLINE PERSONALITY DISORDER AND BIASED APPRAISAL OF TRUST IN FACES **67**

Overview 67

1. Introduction 69

2. Methods 73
Participants and Procedure 73
Materials 74
3. Results 75
Correlations 75
Mediations 77
4. Discussion 80
CHAPTER 3. BIASED APPRAISAL OF TRUST IN BORDERLINE PERSONALITY DISORDER: THE ROLE OF SENSITIVITY TO JUSTICE 84
Overview 84
1. Introduction 86
2. Main Aims 89
3. Study 1 90
3.1. Methods 91
Sample determination 91
Participants and Procedure. 91
3.2. Materials 91
3.3. Results 93
Exploratory Analyses 96
3.4. Discussion 99
4. Study 2 101
4.1. Methods 102
Sample determination 102
Participants, Procedure, and Materials. 102
4.2. Results 102
4.3. Discussion 104
5. General discussion 105
CHAPTER 4. SUSPICIOUSNESS AND SELF-CONCEPT IN BORDERLINE PERSONALITY DISORDER 110

Overview 110

1. Introduction 112

2. Main Aims & Hypotheses 115

3. Methods 116

Participants and Procedure 116

Materials 116

4. Results 118

5. Discussion 122

CHAPTER 5. THE INFLUENCE OF DIRECT AND INDIRECT CUES OF TRUSTWORTHINESS ON BORDERLINE PERSONALITY DISORDER'S BIASED APPRAISAL OF TRUST. 129

Overview 129

1. Introduction 131

2. Aims of the contribution 134

3. Study 1 135

3.1. Methods 135

Sample size determination 135

Participants and Procedure 136

3.2. Materials 136

3.3. Results 137

3.4. Discussion 139

4. Study 2 139

4.1. Methods 140

Sample size determination 140

Participants and Procedure 140

4.2. Materials 141

4.3. Results 142

4.4. Discussion 143

5. General discussion 145

CHAPTER 6. BORDERLINE PERSONALITY DISORDER AND TRUST DISPOSITIONS DURING THE COVID-19 OUTBREAK 150

Overview 150

1. Introduction 152

2. Aims of the contribution 155

3. Preliminary study 156

3.1. Methods 156

Participants and Procedure 156

Statistical analyses and Results 157

4. Main Study 159

4.1. Methods 160

Participants and Procedure 160

4.2. Materials 160

4.3. Results 162

4.3.1. Test of the two-factor structure of ITSQ in the Italian version 162

4.3.2. Associations between BPD features and self-report measures of trust. 164

4.3.3. Different trust disposition between high and low BPD features samples. 166

5. Discussion 167

GENERAL DISCUSSION 172

1. Implications for research 178

2. Clinical implications 183

3. Conclusions 187

REFERENCES 189

APPENDIX I 221

APPENDIX II 223

APPENDIX III 226

APPENDIX IV 228

APPENDIX V 229

APPENDIX VI 232

ACKNOWLEDGMENTS

Three years ago, when I started this Ph.D. program, I was confident that through my work I would have found answers to many questions. Now, three years later, I am aware that some questions cannot be answered. I would like to extend a deep and sincere “thank you” to all those who taught me to accept and enjoy the uncertainty of not having all of the answers. Such a valuable legacy will inform my practice in the therapy room and beyond.

“The way to make people trustworthy is to trust them.”

Ernest Hemingway

ABSTRACT

The present work's main aim is to shed light on the processes underlying the lower propensity of individuals with Borderline Personality Disorder features to trust others (i.e., untrustworthiness bias). The first chapter proposes a novel heuristic model for studying interpersonal trust processes in Borderline Personality Disorder (BPD). We recommend considering trust as an iterative, circular, and multi-step process from a systematic review of the previous literature on trust. Hence, we stress the importance of considering trust expressions as a dynamic process in continuous evolution rather than a static personality disposition. Throughout the thesis, we examine and provide support for atypical processing at some stages of the model (i.e., prior dispositions, the influence of the situation, and trust appraisal) among individuals with BPD features. The following chapters present empirical studies showing the peculiar functioning of individuals with BPD features in some of the stages suggested by the model. In chapters 2, 3, and 4, we investigate the influence of individual differences in Rejection Sensitivity, Justice Sensitivity, and Suspiciousness on the BPD's untrustworthiness bias. In particular, we empirically demonstrate these dispositions have a distinctive influence on the interpersonal trust among individuals with BPD features. Chapter 5 explores the effects of different signals of others' trustworthiness (i.e., direct and indirect cues) on the untrustworthiness bias of individuals with BPD features. Our findings suggest that the association between BPD features and untrustworthiness bias is not stable. Finally, in chapter 6, we focus on the influence of Covid-19 circumstances on interpersonal trust among individuals with different levels of BPD features, developing a scenarios-based measure. Surprisingly, independently from their BPD levels, individuals did not show differences on this measure. In the last

chapter, we discuss our findings and the implications of using the model proposed in the first chapter to study interpersonal trust in BPD from an empirical and clinical perspective.

Keywords: Interpersonal Trust, Untrustworthiness bias, Borderline Personality Disorder

INTRODUCTION

Given the importance of social interactions for humans (Ainsworth, 1989; Baumeister & Leary, 1995), scholars from various academic fields, including ethologists, anthropologists, sociologists, psychologists, and more recently, neuroscientists, have investigated interpersonal relationships both theoretically and empirically. Several interrelated theories informed the empirical investigation of interpersonal functioning. The present contribution is situated within psychodynamic theories, such as Object Relations Theory (ORT; Kernberg, 1975), attachment, and mentalization theory (Ainsworth & Bowlby, 1991; Fonagy & Target, 2006). These theories generally explore the role of problems in early relationships and caregiving experiences on the subsequent development of dysfunctional interpersonal behaviors. According to the ORT perspective, the quality of the representations of self and others that the child forms in early relationships with the primary caregiver plays a central role in developing interpersonal abilities (Kernberg, 1996). In a similar vein, attachment theorists posit that children develop internal models of the self and others based mostly on interactions with primary caregivers (Bowlby, 1973). Such internal models guide expectations and beliefs in future interpersonal relationships. Thus, disturbed early attachment experiences and representations of the self and others that are negative and polarized may reduce mentalization ability (i.e., the ability to perceive and interpret behavior based on underlying mental states) and difficulties in interpersonal experiences (Fonagy & Target, 2006).

Over the last decades, the growing recognition of interpersonal functioning's importance for human behavior promoted an interpersonal approach to psychopathology (Pincus & Hopwood, 2012). The contemporary interpersonal theory

assumes that the “most important expressions of personality and psychopathology occur in phenomena involving more than one person” (Pincus & Ansell, 2003). In line with such a standpoint is the definition of Personality Disorders in the current psychiatric diagnostic system as pervasive, inflexible, and stable patterns of thinking, feeling, behaving, and interacting with others (American Psychiatric Association, 2013). The novel conceptualization of Personality Disorders suggested by the Alternative Model for the diagnosis of Personality Disorders, presented in Section III of the DSM-5, posits disturbances in interpersonal functioning as a core dimension of personality pathology (American Psychiatric Association, 2013). Given the central role of interpersonal dysfunction for Personality Disorders, recent empirical contributions tried to disentangle key interpersonal features in specific Personality Disorders.

Interpersonal functioning is a major concern for patients with Borderline Personality Disorder (BPD) compared to other Personality Disorders. The diagnostic criteria for BPD comprise a wide range of dysfunctions, including behavioral dysregulation, affective dysregulation, and disturbances in interpersonal relatedness (Clarkin et al., 1993). However, disturbed interpersonal relationships are increasingly recognized as central for understanding BPD expressions (Gunderson, 2007). According to various empirical contributions, disturbances in interpersonal relatedness are a unique feature of BPD that diagnostically distinguishes BPD from other disorders (e.g., Blais et al., 1999; Sanislow et al., 2002). Behavioral (i.e., impulsive or suicidal behavior) and affective dysregulation (i.e., affective instability, sensitivity to abandonment, irritability) are also common features of other clinical conditions (Stanley & Siever, 2010). Given the importance of dysfunctional interpersonal

functioning to understand BPD, the pace of empirical research examining the social-cognitive processes behind problems with interpersonal functioning in BPD has accelerated in the last decades. The existing literature in the social cognition field mostly focused on investigating emotion recognition abilities in BPD (for a review, see Domes et al., 2009). Although effective interpersonal functioning requires gathering information about other people and accurately appraising their characteristics, including personality traits and intentions, the empirical research on dysfunctional trait appraisal among individuals with Personality Disorders is still deficient (Roepke et al., 2012). Within the wide range of traits, the appraisal of others' trustworthiness has been acknowledged as one of the most relevant features for Personality Disorders diagnosis and, therefore, for BPD (Poggi et al., 2019).

Trust is a cornerstone of social relationships and social order (Zucker, 1986). Trust refers to a general assumption about others' good nature. Conversely, mistrust implies negative expectations regarding others (Evans & Revelle, 2008; Lewicki et al., 1998). Nonetheless, a clear-cut definition of trust and a comprehensive model of trust processes is still lacking. One key reason is that any conceptual model heavily depends on the underlying theoretical framework and approaches to the topic. Traditionally authors referred to trust as a stable personality trait (Costa & McCrae, 2008). However, recent conceptualizations of trust as a "behavioral decision" to rely on others enabled researchers to consider it a dynamic process and investigate cognitive biases linked to the tendency to judge others as trustworthy or untrustworthy (Ping Li, 2012). In line with the latter assumption, individual differences in trust dispositions and accuracy in the appraisal of others' trustworthiness shape behaviors in interpersonal situations. It is now well-established that BPD individuals tend toward a generalized mistrust of

others resulting in a greater attribution of and sensitivity to others' untrustworthiness, namely untrustworthiness bias (e.g., Fertuck et al., 2013; King-Casas et al., 2008). Several clinical and empirical contributions emphasize that one core interpersonal dysfunction among individuals with BPD is a failure to maintain fulfilling relationships based upon mutual trust (Poggi et al., 2019). However, a comprehensive and shared understanding of the multiple and mixed motives that shape untrustworthy attitudes among individuals with BPD features is still lacking. Given the lack of a consensual understanding of trust impairments among individuals with BPD, the present work aims to shed light on the processes leading individuals with BPD features to express an untrustworthiness bias in social interactions.

To this aim, we start with a systematic review of the theoretical and empirical investigations of trust impairments and BPD features. Within the first chapter, we suggest a heuristic model delineating the temporal stages of trust processes. Such a model describes the cognitive steps that unfold when individuals (regardless of any clinical condition) trust others, from the distal and proximal antecedents of trust appraisal towards behavioral manifestations and potential learning of others' trustworthiness based on past and actual experiences. Such a model provides a useful framework for identifying trust impairments (for a graphical representation of such a model, see Figure 1, p. 23). In the following empirical section, we present a series of empirical studies aiming to deepen the knowledge of the processes underlying untrustworthiness bias among individuals with BPD features. These studies investigate BPD impairments in some of the stages suggested by the framework. The second, third, and fourth chapters are situated within the prior belief and dispositions stages, exploring the influence of some individual differences on the untrustworthiness bias.

The second Chapter focuses on Rejection Sensitivity, the cognitive and affective disposition to anxiously or angrily expect, readily perceive, and overreact to rejection (Downey & Feldman, 1996). The third Chapter explores the influence of Justice Sensitivity on untrustworthiness bias at varying BPD features levels. Justice sensitivity reflects individual differences in the perceptions and reactions to observed, suffered, or committed injustice (Schmitt et al., 2010). In the fourth Chapter, we focus on suspiciousness. Suspiciousness consists of a pattern of expectations and sensitivity to signs of interpersonal ill-intent or harm, which implies doubts about the loyalty of others and persecution feelings (Skodol et al., 2011). More precisely, we disentangle the role of suspiciousness on trust impairments at different levels of BPD features. The fifth Chapter focuses on the trust appraisal stage. More precisely, this chapter presents a contribution showing the effect of indirect and direct manipulations of trust on BPDs' untrustworthiness bias. Finally, in the sixth Chapter, particular attention is granted to an explorative investigation of how the Covid-19 situation affected the association between BPD and trust impairments. Given that all the mentioned chapters are based on published papers or in preparation for submission manuscripts, some overlaps could occur in their introductions. The last Chapter discusses the studies' main implications in light of the suggested model, proposes some future developments for research, and delineates some clinical implications.

CHAPTER 1. A SYSTEMATIC REVIEW OF TRUST PROCESSES IN BORDERLINE PERSONALITY DISORDER[1]

Overview

Although impaired trust appraisal is nowadays recognized as a core dimension of Borderline Personality Disorder, this topic has been less investigated than others (e.g., impulsivity and affective instability). Despite both researchers and clinicians could benefit from a comprehensive conceptualization of the phenomenon, a systematic review of empirical findings available is still missing up to now.

The present chapter aims to provide a framework for understanding how individuals with BPD differ from the general population concerning trust processes. To this end, we propose a heuristic model delineating the stages that characterize normative trust processes. We argue that this novel model could help researchers and clinicians localize at a specific stage of the process the enacted trust impairments of participants in empirical studies and patients in the clinical setting. We will refer to this model in each of the next chapters' overview to help readers navigate the present dissertation.

We present the impairments occurring at each stage of the model among individuals with BPD. We first conducted a systematic review of the literature related to trust and BPD in two primary electronic databases (i.e., PsycINFO and PubMed). We then summarize the main findings for each of the twenty-one articles resulting from the selection process. Interestingly, the empirical findings suggest that

---

[1]This chapter is based on Poggi, Richetin, Fertuck and Preti. (2021). Trust Processes in Borderline Personality Disorder: A Systematic Review. The manuscript is currently submitted under review in Psychological Bulletin.

individuals with BPD show impairments across all the model's stages. Nonetheless, such an in-depth analysis of literature highlighted significant research gaps and some inconsistencies across studies. The empirical chapters of this dissertation represent an attempt to address some of the inconsistencies and gaps found.

1. Introduction

The ability to determine accurately who is trustworthy – and, in the face of new information, adjust judgments of others' trustworthiness accurately, flexibly, and efficiently – is critical in navigating unpredictable and dynamic human social systems. Individuals with borderline personality disorder (BPD), in particular, exhibit extreme distress and confusion while navigating social environments and display an array of behaviors that indicate impairments in appraising others' trustworthiness. BPD is a multifaceted clinical condition characterized by marked impulsivity and a pattern of instability that influences self-image, interpersonal relationships, and affects (American Psychiatric Association, 2013). BPD also significant impacts public health for several reasons, including the high prevalence of BPD diagnoses in the general population (from 0.5% to 3.54%; see ten Have et al., 2016) and the greater use of clinical services by BPD patients compared with those who have other mental health conditions (Jackson & Burgess, 2004).

Although disturbed interpersonal relationships historically have been recognized as a core dimension of BPD (e.g., Kernberg, 1967; Stern, 1938), empirical research on BPD traditionally has focused on impulsivity and affective instability. In the past 10 years, there has been a sharp increase in empirical research focusing on interpersonal difficulties in BPD (Gunderson, 2007; Seres et al., 2009). In this sense, research on BPD's social cognitive bases has started to investigate the processes underlying interpersonal dysfunction (for reviews, see Herpertz & Bertsch, 2014; Roepke et al., 2012). There are two reasons for the heavy interest in interpersonal dynamics among BPD patients. First, interpersonal issues in BPD often lead to extreme consequences (i.e., suicide attempts; Black et al., 2004). Second, many BPD symptoms

(e.g., self- and other-aggressive behaviors, emotional instability) primarily occur in the context of turbulent interpersonal relationships (Brodsky et al., 2006). Furthermore, interpersonal difficulties with BPD manifest themselves as several clinical challenges (e.g., misplaced trust in others, confusion over interpersonal boundaries, extreme suspiciousness of others, including clinicians), and adverse countertransference reactions; Gunderson, 2007; Gunderson & Lyons-Ruth, 2008). Along these lines, different theoretical and clinical approaches have conceptualized the interpersonal domain as a particular focus of psychotherapeutic intervention (e.g., Bateman & Fonagy, 2010; Yeomans et al., 2015; Young et al., 2006).

BPD patients exhibit an array of trust impairments, including a lower appraisal of trustworthiness in others, fewer cooperative behaviors with others, difficulty with updating assessments of and accurately learning about others' trustworthiness, and unstable trust appraisal (Fertuck et al., 2018; King-Casas et al., 2008; Unoka et al., 2009). There are various rationales for investigating trust processes in BPD. Clinically, impairments in trust appraisal among BPD patients can make it difficult for them to maintain secure and trusting bonds with their therapists. These include problems in developing and maintaining solid therapeutic alliances (Richardson-Vejlgaard et al., 2013). On a related note, premature treatment termination may result from impaired trust concerning a clinician (Fertuck et al., 2012). Some mental health professionals may even withdraw from and stigmatize these individuals, reinforcing BPD patients' trust biases and alienating them, thereby limiting their access to effective intervention (Sansone & Sansone, 2013). Treatments effectively reduce self-destructive behaviors in BPD, but improvements in interpersonal functioning are difficult to achieve. Zanarini and colleagues (2007) examined the course of interpersonal dysfunctions in

a sample of treated BPD patients. The authors showed that dysfunctional interpersonal features persisted over 10 years of follow-up and influenced both treatment relationships and social functioning. Thus, addressing the core, persistent, interpersonal dysfunctions in BPD may be crucial to long-term recovery (Sinnaeve et al., 2015). In light of this, considering that trust impairments are a specific expression of interpersonal difficulties for BPD, efficient management of trust impairments might be central to long-term improvements in BPD. Developing valid assays of trust processing is scientifically timely, considering that psychotherapy research has shifted away from "horse race" studies that directly compare manualized treatments. Instead, the contemporary approach to psychotherapy research is to conceptualize the mechanisms of a disorder like BPD, validly assess these mechanisms, and demonstrate changes in these mechanisms as a consequence of an intervention (Kazdin, 2007). Thus, empirically dissecting the components and stages of trust impairment in BPD could inform treatments focused on improving social functioning beyond symptom reduction.

Finally, in the fields of social, cognitive, and affective neurosciences, there has been increasing interest in unraveling the complex mechanism subtending interpersonal trust (e.g., Krueger & Meyer-Lindenberg, 2019). However, a conceptual framework that integrates separate findings into a comprehensive model of interpersonal trust remains missing. Along these lines, empirical contributions on trust-impairment processes in BPD may improve the general understanding of trust's normative neuropsychological underpinnings. Thus, in the present paper, we integrate research findings from several fields (i.e., neuroeconomics, psychology, neuroscience, etc.), as well as investigations of trust impairments in BPD. Such a transdisciplinary

approach leverages cutting-edge, empirical social neuroscience findings to unravel the mechanisms of one of the most critical factors in beneficial human relationships: trust.

1.1 Trust: Conceptualization and Operationalization

Different theoretical and methodological frameworks in social science have proposed definitions and conceptualizations of interpersonal trust (for a summary, see Table 1). It is possible to assimilate trust impairments in BPD into attributional style framework, i.e., the "*tendency to make particular kinds of causal inference, rather than others, across different situations and across time*" (Metalsky & Abramson, 1981, p. 38). Attributional styles have been articulated and investigated using varying measurement approaches, including self-report questionnaires on personality (e.g., Bach et al., 2016; Giesen-Bloo et al., 2006), experimental behavioral measures (e.g., Wagner & Linehan, 1999), longitudinal measurements (e.g., Russell et al., 2007), and neural activity measures (e.g., Fertuck et al., 2019; Minzenberg et al., 2007). These diverse conceptualizations may derive from the salience of trust in several disciplines, such as social psychology, sociology, anthropology, economics, and political sciences (Hosmer, 2011; Lewis & Weigert, 1985; Shapiro, 1987). Across the different definitions, several common parameters emerge.

Trustors are the agents who depend on others' behaviors for their goals and objectives. *Trustees* are the agents who have the power to react in a trustworthy or untrustworthy way to trustors' actions. Trust is relevant in situations in which trustors depend on trustees' action(s) to achieve their goals and objectives (Hosmer, 2011; Lane & Bachmann, 1998; Whitener et al., 1998). When acting on their trust, trustors make themselves vulnerable to trustees' actions (Hosmer, 2011).

Trustors, relying on trustees to act in a particular way, must view trustees as free agents, and that they depend on trustees' behavior (Hosmer, 2011).

Trust is a choice that entails voluntary, not forced, cooperation on the trustor's part (Hosmer, 2011).

Trust is related to optimistic expectations that the trustee will not take advantage of the trustor's vulnerability (Hosmer, 2011; Whitener et al., 1998).

Trust, consistent with the variety of definitions across disciplines, has been assessed differently. First, self-report questionnaires investigate one person's explicit attitude toward others (e.g., "*Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people*?"; Verba & Almond, 1963). A second approach entails using economic games to study trust. For example, the "trust game" is a multiphase game with a trustor and trustee interacting. In the first phase, the trustor's task is to send some, all, or none of their endowment to the trustee. During this phase, the money sent is tripled, then the trustee chooses how much of the money (comprising the initial endowment plus the tripled money sent to themselves) to return to the trustor (Berg et al., 1995). An increase in the amount of money sent by trustors to trustees over time is the measure of trust. Finally, in social-cognition research, trust has been investigated mainly as a process involving an individual's appraisal of others' trustworthiness (i.e., trust appraisal) based on social cues (e.g., facial morphology) in different social contexts (e.g., social exclusion). In this context, trust is the trustee's appraisal of the trustor's honesty, benevolence, and potential for cooperation.

The integration of elements across diverse trust's conceptualization informed the formulation of a conceptual model of trust processes. Such a model provides a

more integrative picture of the phenomenon. Based on interpersonal psychology conceptualizations, we argue that trust emerges in interpersonal relationships over time. In accordance with personality psychology, we assume that the ability to develop and maintain trust changes across individuals and personality features. Building on learning theory, we suggest that individuals learn best-action patterns based on rewards or punishments in interactions. From economic psychology, we acknowledge individuals' intrinsic motivation to pursue self-regarding interests during interpersonal exchanges. Finally, according to the social psychology framework, we emphasize the social environment's effect on individuals' trust behavior.

Future investigations about trust's psychological underpinnings in mental health disorders such as BPD may benefit from such an interdisciplinary model, including contributions from diverse fields.

Table 1

Definitions of constructs related to trust across several disciplines

Construct	Definition	Citation	Discipline
Trust	Optimistic expectations are held by an individual or a group that he/she can rely on the word, promise, or verbal or written statement of another individual or group.	Rotter (1980)	Interpersonal psychology
Trust	Trust implies reliance on confidence in some events, processes, or people. It comprises a "set of optimistic expectations on the part of an individual strongly linked to confidence in and overall optimism about desirable events taking place."	Golembiewski & McConkie (1975, p. 134)	Personality psychology
Trust learning	If a person found that a group of people he or she relied on in the past had acted according to informal, or "embedded," obligations of the society in the past, he or she would be more likely to trust these group members in the future.	Granovetter (1985)	Learning theory
Trust	The expectation that another individual or group will (1) make a good-faith effort to behave by any commitments, both explicit or implicit; (2) be honest in whatever negotiations preceded these commitments; and (3) not take excessive advantage of others even when the opportunity is available.	Bromiley & Cummings (1995)	Economic psychology
Trust	Strong positive expectations based on the relationships between people in a social system. Trust is essentially social and normative, rather than individual, and, therefore, requires prior experience of "fair" social relationships and generally accepted "rights" to develop.	Zucker (1986)	Social psychology

1.2 Trust processes and borderline personality disorder

Deciding whether to trust others, take action following such a decision, then re-evaluate the degree of trust in a relationship implies an iterative, multi-step process. Different disciplines recognize the need for a multi-stage conceptualization of trust

(Golembiewski & McConkie, 1975). Drawing on the strengths of the multiple models of trust processing that we have reviewed, we propose a heuristic for a multi-stage model of trust processing. We will utilize this model to characterize normative and atypical BPD trust processing at each normative stage (see Figure 1).

We organized our review according to the different stages of the model proposed. First, we synthesized the literature on the first stage: *distal and proximal antecedents of trust attribution* (i.e., factors that help build and support the appraisal of others' trustworthiness). We then analyzed contributions in the model's second stage: *real-time trust appraisals and decision-making in social contexts* (e.g., cooperation). Finally, we examined the iterative process of updating trustworthiness appraisals and social behaviors according to novel trust-relevant interactions and stimuli, i.e., *the trust learning process.*

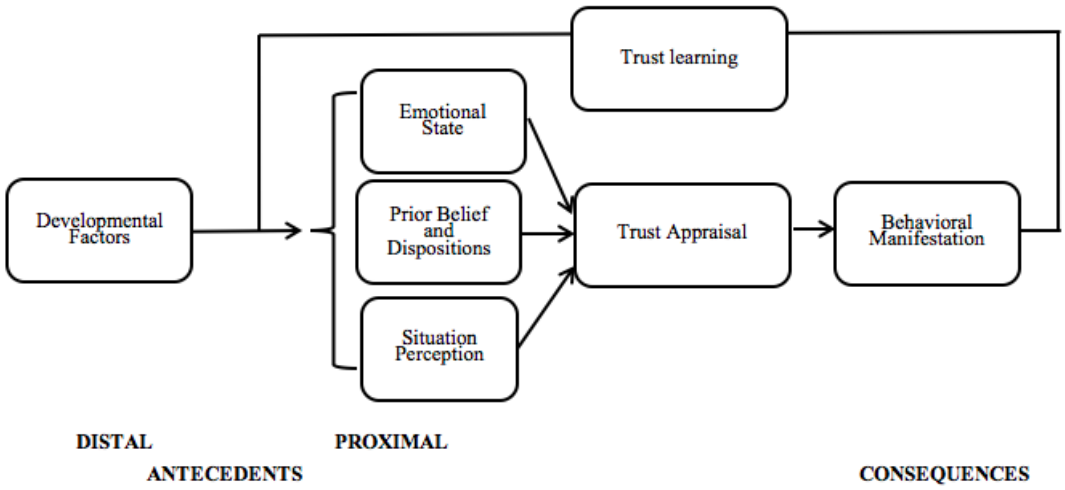


Figure 1. A graphic representation of interpersonal trust's formation and change over time. The model identifies both distal (i.e., developmental factors) and proximal (i.e., factors occurring moments before trust-relevant exchanges) psychological antecedents of trust. Moreover, the model recognizes the behaviors that

trust predicts (such as cooperation) and how trust evolves according to the outcomes of transactions (i.e., whether the trustee cooperated or betrayed the trust).

2. Methods

We conducted a systematic review of the literature related to trust and BPD, registered our review protocol with PROSPERO (CRD42019125457), and reported our methods and results, while following PRISMA recommendations. We used the electronic databases PsycINFO and PubMed to locate studies that address the topic, searching for specific keywords in the title or abstract ["trust" OR "trustworthiness" AND "borderline personality disorder"]. We conducted the literature search of databases in July 2020 and identified 107 publications in PsycINFO and 71 in PubMed (for the selection flow diagram, see Figure 2). We removed duplicates (k = 114) and screened the remaining papers' abstracts (k = 64).

Our inclusion criteria were: (1) studies using self-report questionnaires or structured clinical interviews for BPD diagnoses or traits (e.g., a structured clinical interview for DSM-IV axis II; First et al., 1995; Borderline Personality Disorder Checklist; Giesen-Bloo et al., 2006; Personality Assessment Inventory; Morey, 2004); and (2) studies published or accepted for publication in peer-reviewed journals. Our initial exclusion criteria were: (1) studies coming from fields other than psychology (e.g., economics); (2) studies that did not apply valid and reliable assessment of BPD or BPD traits (i.e., self-report questionnaires or clinical interviews); (3) studies focusing on organizational (and not interpersonal) trust; (4) studies reported in a language other than English; and (5) studies that did not report original quantitative empirical data. During this step, we excluded reviews (k = 5), qualitative studies (k =

2), dissertation abstracts (k = 5), clinical illustrations (k = 2), commentaries (k = 1), pilot studies (k = 2), and non-English articles (k = 3).

We screened the remaining 44 articles' content and excluded studies that addressed phenomena linked to trust issues unrelated to our scope. More precisely, we excluded: a) studies reporting on users, families, or care providers' experience in health services and their trust in them (k = 11); b) studies focusing on how to increase trust in care providers other than psychotherapists, such as nurses or social assistants (k = 5); and c) studies focusing on trust in care institutions (k = 2). We also excluded papers that mentioned "trust" within the names of associations for the recruitment of specialists or participants to study (e.g., National Health Service Trust) (k = 5). We ended up with a final set of 21 research reports.

In Table 2, for all 21 articles, we reported the stage of the model to which the study refers, the sample size, the assessment of BPD features, the task that participants performed, and the key findings. Moreover, we reported Cohen's *d*, the most

frequently used index of effect size for experimental and intervention studies in social and behavioral sciences (Hunter & Schmidt, 2004).

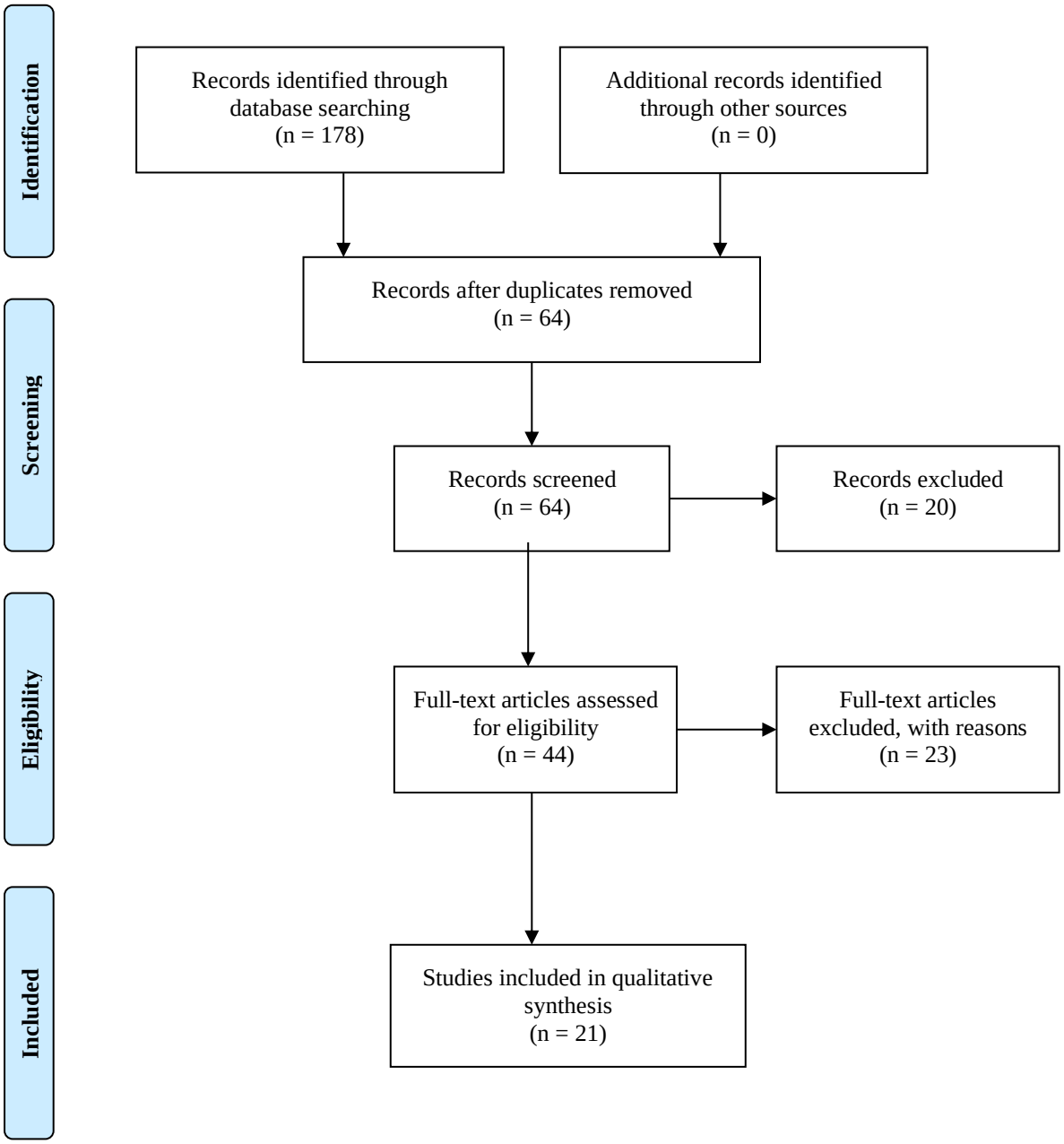


Figure 2. Flow diagram of study selection

Table 2.

Empirical Studies Addressing Trust Impairment in Borderline Personality Disorder

Study	**Stage**	**Sample**	**BPD Assessment**	**Task**	**Key Findings**	**Effect Sizes** (Cohen's *d*)
Ebert et al. (2013)	Distal antecedents: Developmental factors	*N* = 26 BPD *n* = 13 Control *n* = 13 Random assignment to placebo or oxytocin condition.	Structured clinical interview for DSM-IV (SCID; German version by Wittchen et al., 1997)	Trust game assessment of childhood trauma	Only in BPD participants and only in the oxytocin condition did childhood trauma scores correlate with trust behaviors	$d = 0.98$[1]
Orme et al. (2019)	Distal antecedents: Developmental factors	*N* = 322 inpatient adolescents	Borderline personality features scale for children, child report (BPFS-C; Crick et al., 2005) Borderline personality features scale for children, parent report (BPSF-P; Chang et al., 2011) Child interview for DSM-IV borderline personality disorder (CIBPD; Zanarini, 2003)	Inventory of Parent and Peer Attachment (IPPA)	A negative association was found between BPD measures and adolescent trust in mothers and fathers	In mothers $d = 0.39$[1] In fathers $d = 0.25$[1]

Butler et al. (2002)	Proximal antecedents: prior beliefs and dispositions	*N* = 288 BPD *n* = 84 Other personality disorder *n* = 102	Structured clinical interview for the DSM-III-R (SCID; Spitzer et al., 1990)	Personality Belief Questionnaire	"I cannot trust other people" is the most discriminative belief, significantly higher in BPD compared with OPD participants.	$d = 0.94$[1]
Botsford et al. (2019)	Proximal antecedents: prior beliefs and dispositions	*N* = 338 BPD *n* = 41 Major depression disorder *n* = 30 Seasonal affective disorder *n* = 31 Healthy controls *n* = 236,	SCID-II	Interpersonal Trust Scenario Questionnaire	BPD patients display lower levels of interpersonal trust compared with non-clinical controls and also patients with MDD.	$d = 0.41$[1]
Miano et al. (2013)	Proximal antecedents: prior beliefs and dispositions	*N* = 95 Non-clinical population	SCID II Screener for Personality Disorders, Version 2.0 (First et al., 1997).	Facial appraisal on 17 trait dimensions (including untrustworthiness and trustworthiness)	The High-BPD group shows higher untrustworthiness bias than Low-BPD. RS as a mediator of the effect of BPD features on trust appraisal.	$d = 0.58$[1]
Richetin et al. (2018)	Proximal antecedents: prior beliefs and dispositions	*N* = 125 non-clinical Non-clinical population	Borderline Personality Disorder Checklist (BPDCL; Giesen-Bloo et al., 2006)	Facial trust appraisal	Only emotional components of RS (anger and anxiety) mediate the effect from BPD features on trust appraisal No mediation of cognitive component (expectation)	$d = 0.43$[1]
Bartz et al. (2011)	Proximal antecedents: prior beliefs and dispositions	*N* = 27 BPD *n* = 14 Healthy controls *n* = 13 Random assignment to placebo or oxytocin condition.	SCID-II	Economic game	oxytocin produced more divergent effects in BPD participants than controls: decrease in trust and cooperation. Divergent effects were driven by anxious attachment and rejection sensitivity (peculiar to BPD)	$d = 0.88$[1]

Miano et al. (2017)	Proximal antecedents: situation perception	*N* = 67 couples BPD *n* = 31 (women diagnosed with BPD) Controls *n* = 36	SCID-II	Evaluation of partners' trustworthiness after discussion of neutral, personal, or relationship-threatening topics	BPD couples were significantly different from controls just after relationship-threatening discussions.	*d* = 0.55[1]
Preuss et al. (2016)	Proximal antecedents: situation perception	*N* = 77 BPD *n* = 17 Healthy controls *n* = 36 Major depression controls *n* = 24	Zanarini Rating Scale for Borderline Personality Disorders (ZAN-BPD; Zanarini, 2003)	Social trust game Non-social trust game Punishment game Dictator game Cooperation game	BPD participants had more inconsistent behavior than controls only in the social conditions social trust game and punishment game	PG *d* = 0.64[1] TG *d* = 0.69[1]
Hula et al. (2017)	Proximal antecedents: emotional state	*N* = 93 BPD *n* = 55 Controls *n* = 38	DIPD-IV	Multiple-round trust game	BPD participants were less aware of guilt and irritation than controls.	[2]
Roberts et al. (2018)	Proximal antecedents: emotional state	*N* = 284 Non-clinical population	Personality Assessment Inventory Borderline Features Scale (PAI-BOR; Morey, 2004).	Economic trust game	Acetaminophen reduces behavioral mistrust at high levels of BPD features.	*d* = 1.20[1]
Masland & Hooley (2019)	Proximal antecedents: emotional state	*N* = 77 non-clinical High BPD features *n* = 30 Low BPD features (controls) *n* = 47	Schedule for Nonadaptive and Adaptive Personality (SNAP-2; Simms & Clark, 2006)	Facial trust appraisal after positive, negative, or neutral affective priming	BPD participants show an untrustworthiness bias, i.e., lower trust ratings regardless of priming	neutral *d* = 0.57 negative *d* = 0.81 positive *d* = 0.48.

Fertuck et al. (2013)	Trust appraisal	*N* = 36 BPD = 17 Control = 19	SCID-II	Fear and trust facial appraisal	BPD participants show an increased response bias in trustworthiness appraisal compared with control (no significant differences in sensitivity and discriminability).	*d* = 0.72
Nicol et al. (2013)	Trust appraisal	Total *N* = 40 females BPD *n* = 20 Control *n* = 21	SCID-II	Childhood Trauma Questionnaire (CTQ) Facial trust appraisal	BPD group judged faces as being less trustworthy. Furthermore, there was a correlation between CTQ and bias only in the BPD group.	*d* =1.31[1]
Fertuck et al. (2019)	Trust appraisal	*N* = 33 BPD *n*= 16 Control *n*= 17	SCID-II	Fear and trust appraisal	Lower trustworthiness appraisal in BPD participants. Lower BOLD activity while trustworthiness appraisal in prefrontal cortex related to bias intensity	*d* =0.84[1]
King-Casas et al. (2008)	Behavioral manifestations	*N* = 93 BPD *n* = 55 Controls *n* = 38	DIPD-IV	Multiple round trust game	BPD participants more likely to cause cooperation ruptures BPD participants sustain lower rates of coaxing behaviors	[2]
Unoka et al. (2009)	Behavioral Manifestations	*N* = 75 BPD *n* = 25 Healthy controls *n* = 25 Major depression controls *n* = 25	ZAN-BPD	Single trust game Risk lottery game	BPD participants transferred less money during the trust game: untrustworthiness bias. No behavioral differences in the lottery game.	*d* = 0.76[1] *d* = 0.80[1]
Saunders et al. (2016)	Behavioral manifestations	*N* = 40 females BPD *n* = 20 Controls *n* = 20	International Personality Disorder Examination (IPDE; Loranger, 1997)	Iterated form of the prisoner's dilemma	Only BPD participants failed in build cooperative relationships.	*d* = 0.93[1]
Franzen et al. (2011)	Trust learning	*N* = 60 BPD *n* = 30 Healthy control *n* = 30	SCID-II	Multi-round trust-game	Both BPD participants and control participants adapted their investment	*d* = 0.30[1]

					behaviors according to counterparts' facial expressions.	
Fineberg et al. (2018)	Trust learning	*N* = 43 BPD *n* = 20 Controls *n* = 23	- SCID-II - Borderline Symptom List 23 (BSL-23; Bohus et al., 2009).	Reward learning task.	BPD participants had lower trust learning scores even if they weighted more social cues compared with controls.	*d* = 0.81[1]
Abramov et al., (2020)	Trust learning	*N* = 234 non-clinical	McLean Screening Instrument for BPD (MSI-BPD; Zanarini et al., 2003).	Multi-round trust-game	Individuals with a higher number of BPD traits show a greater decline in investments during the trust-formation phase.	[2]

Note. [1] No Cohen's *d* reported in the original paper. We estimated Cohen's *d* post hoc using available information such as *t*, *F*-values, or Beta scores, according to Fritz et al. (2012).

[2] No effect size was reported in the original paper with the impossibility to estimate Cohen's *d*.

3. Results

To assess the strength of each study's findings and allow for comparisons between them, we report Cohen's *d* from commented effects. If authors did not report Cohen's *d* in the original paper, we either calculated it from the available data (when the necessary information was available) or converted the reported effect size indices (i.e., *r*, odds ratio). Generally, a positive index indicates that the assessed criterion is more impaired in the BPD sample (or individuals with high BPD features) than controls (or individuals with low BPD features).

3.1. Trust processes in BPD: distal antecedents

3.1.1. Distal antecedents of trust appraisal: developmental factors

In typical development, children experience trusting, attuned behavior from their caregivers, implying that the trustee (i.e., the trusted person, caregiver) generally does not violate trust expectations. Attachment theory posits that such favorable experiences lay the foundation for positive internal working models of the infant's relationship with others that later lead to an accurate appraisal of trustworthiness in others (Bowlby, 1988). According to Erikson (1993), trust emerges from responsive and consistent caregiving experiences. Conversely, when infants are mistreated, abused, or must wait extensively for the attachment figure to comfort them, they can develop legitimate mistrust. As a consequence of positive caregiving experiences, infants with secure attachments differ from infants with insecure attachments (i.e., avoidant or anxious-ambivalent) because they expect to trust attachment figures during stressful times (Ainsworth & Bowlby, 1991). Such attachment styles extend into adolescence and adult years (Hazan & Shaver, 1987), driving securely attached infants to develop trusting relationships in adulthood.

Within the framework of mentalization, the imaginative mental activity that enables us to perceive and interpret human behavior in terms of intentional mental states develops during early childhood (Allen et al., 2008). From this perspective, epistemic trust comprises the ability to "appraise incoming information from the social world as accurate and reliable, allowing for the incorporation of the information into existing knowledge" (Fonagy et al., 2015, p. 2). Such an ability is likely to emerge in positive caregiving experiences. Conversely, impairments in this ability are common when children face traumatic or negative experiences with primary caregivers.

This section will outline these studies, suggesting risk factors that may facilitate impairments in trust processes in individuals with BPD.

A theory posited by Orme ad colleagues (2019) asserts that the establishment of epistemic mistrust in childhood (i.e., the misattribution of intentions and the assumption of malevolent motives behind another person's actions) may contribute to the development of personality pathology in general – and BPD in particular. Mentalizing theorists recently addressed the importance of epistemic trust during early development stages. Early attachment relationships' quality is of extreme importance for the subsequent development of epistemic mistrust or trust (Fonagy et al., 2015; Fonagy & Allison, 2014). Mentalization theorists define epistemic trust as the "trust in the authenticity and personal relevance of interpersonally transmitted knowledge about how the social environment works and how best to navigate it" (Fonagy, Luyten, et al., 2017, p. 177). To test this theory, the authors compared levels of BPD symptoms between admission, treatment, and discharge in an adolescent sample. They found a significant negative correlation between levels of BPD symptoms at admission and self-reported trust toward participants' mothers and fathers. These data support the

hypothesis that lower trust disposition toward parents increases the likelihood of developing stable and rigid dysfunctional beliefs and untrustworthy dispositions toward others in general during adolescence and adulthood. However, given the study's retrospective nature, a clear direction for this relationship cannot be determined.

In a previous study, Ebert and colleagues (2013) tested the hypothesis that experiencing trauma during childhood may predict an adult's expectations regarding other people's emotional availability and trustworthiness. They hypothesized that early negative emotional experiences are a risk factor for developing dysfunctional expectations about others' disposition and consequent dysfunctional behavioral manifestations during a trust game (TG) procedure. Additionally, Ebert and colleagues (2013) were interested in oxytocin's role because abnormal activity by this neuropeptide is associated with lower interpersonal trust (Theodoridou et al., 2009). The authors assessed self-reported childhood trauma in BPD patients and controls, randomly assigning participants to oxytocin or placebo administration conditions before having them play a TG. The authors hypothesized an association between oxytocin activity's influence on one's behavior during the TG and a personal history of attachment and early emotional experiences. More precisely, they expected childhood trauma in BPD patients to be associated with fewer investments after receiving oxytocin, but not after a placebo. Consistent with their prediction, there was no correlation between childhood trauma scores and trust behavior in the control group, either in the oxytocin or placebo condition. In contrast, in BPD patients, higher emotional experiences with neglect were associated with fewer monetary units transferred to the trustee (i.e., trusting behavior), but only in the oxytocin condition.

Increased oxytocin seems to reduce trusting behaviors in BPD patients (but not controls), and such a decrease is more important for individuals who experienced greater early parental neglect.

Ebert and colleagues (2013) found a large effect size (d = 0.98) in a small clinical sample compared with a non-clinical adult sample. Their findings also suggest that the number of reported emotional experiences of neglect in BPD correlates with trust behavior. Orme and colleagues (2019) found small effect sizes from correlations occurring between BPD symptoms and self-reported trust in mothers and fathers (d = 0.39 and 0.25, respectively), with a large sample of inpatient adolescents indicating that BPD features correlate negatively with trust toward parents. Although Orme and colleagues (2019) and Ebert and colleagues' (2013) studies have methodological differences, they both suggest that developmental factors, such as emotional-neglect experiences and lack of trust in parents, are stronger distal risk factors for trust issues in adults with BPD compared with the controls.

3.2. ***Trust processes in BPD: proximal antecedents***

3.1.1. Prior beliefs and dispositions

Early findings on typical individual tendencies to trust others are inconsistent. According to Botsford and colleagues (2019), individuals typically assume others to be potentially reliable and trustworthy, leading to mutually beneficial interpersonal exchanges. Nonetheless, approximately 70% of participants in the European and World Values Surveys did not respond that "most people can be trusted" (European Values Study Group & World Values Survey Association, 1999–2002). Although general trust of others is a positive predictor of self-rated health (Barefoot et al., 1998; Kim et al., 2008; Mohseni & Lindstrom, 2007; Schneider et al., 2011), according to

the adaptative-evolutionary approach, mistrustful beliefs can serve a protective function. Suspecting a trustee increases one's awareness in vulnerable situations in which individuals fear others' motives, intentions, and future actions (Koslow, 2000; Shapiro, 1987). Despite such mixed findings on trust dispositions in the general population, the development, maintenance, and activation of dysfunctional prior beliefs about others' untrustworthiness comprise one plausible proximal cause of trust impairments in the BPD population.

According to investigations on core dysfunctional beliefs in personality disorders (Beck et al., 2001), mistrust represents BPD's specific feature. Butler and colleagues (2002) administered the Personality Belief Questionnaire (PBQ; Beck et al., 2001), a self-report questionnaire that assesses 126 dysfunctional beliefs associated with personality disorders, to 288 patients categorized as either BPD or other personality disorder (OPD). They found that the item that best distinguished BPD from OPD individuals was, "*I cannot trust other people*." Such a belief represents dysfunctional expectations of others' behaviors. A strong endorsement of this belief may lead to misreading others' ambiguous actions to confirm BPD-related expectations about others' untrustworthiness.

Botsford and colleagues (2019) supported Butler et al.'s (2002) findings. The authors asked 41 BPD patients, 30 patients with major depressive disorder (MDD), 31 patients with seasonal affective disorder (SAD), and 236 healthy control individuals to fill out the Interpersonal Trust Scenario Questionnaire (ITSQ, Botsford et al., 2019), which measures individual differences in interpersonal trust from the trustor to a trustee. BPD patients reported lower interpersonal trust levels than non-clinical controls and patients with MDD or seasonal affective disorder.

Both Butler et al. (2002) and Botsford et al.'s (2019) findings demonstrated an association between two variables: BPD features and dysfunctional expectations of others' behavior. However, some studies investigated the role of other dispositions, such rejection sensitivity (RS), whose high levels may influence the positive association between BPD and impaired trust processes.

Rejection sensitivity (RS) is a cognitive-affective disposition to expect anxiously, readily perceive, and overreact to social rejection (Downey & Feldman, 1996). Individuals with strong BPD features might be less inclined to trust others because of their concerns and anxiety about the possibility of being rejected or abandoned. Miano and colleagues (2013) tested such a hypothesis, in which 95 undergraduate students completed RS and BPD features questionnaires and rated unfamiliar faces on 17 different dimensions, including trustworthiness. Consistent with the predictions, there was a negative correlation between BPD features and facial appraisal of trust. Moreover, RS fully mediated this relationship. Richetin and colleagues (2018) replicated and extended these findings by investigating the specific contributions of cognitive (expectations) and emotional (anger and anxiety) components of RS (see Preti et al., 2018; Zimmer-Gembeck et al., 2013). They found that only emotional RS components (anger and anxiety for rejection), not the cognitive one (expectations), mediated the association between BPD features and trust appraisal.

To investigate trust impairment's neural mechanisms, more and more researchers have used measures of neural activity during trust-relevant interpersonal exchanges to tap into automatic and unconscious neural processes engaged in trust impairment. Although the effect's direction is unknown, in general, greater circulating blood levels of oxytocin are linked to increased trustworthiness appraisal of trustees

in non-clinical samples (Kosfeld et al., 2005). Some studies show that oxytocin administration results in increased trusting behavior during economic games (Kosfeld et al., 2005) and increased trustworthiness appraised in community samples (Theodoridou et al., 2009). However, one study suggests that oxytocin increased negative social emotions, such as envy, in a non-clinical sample (Shamay-Tsoory et al., 2009).

Bartz et al. (2010) hypothesized that individuals' chronic concerns about abandonment and trust, as well as difficulties with cooperation, might influence oxytocin activity during trust-relevant BPD patients' interpersonal exchanges, but not those of controls. BPD and control participants completed the Experience in Close Relationship Scale (ECR, a self-report instrument for assessing attachment anxiety and avoidance in adults; Brennan et al., 1998). Participants then were administered intranasal oxytocin or a placebo, then played the assurance game, a variation of the classic prisoner's dilemma paradigm. In the assurance game, cooperation is the best strategy for everybody, but being unsure of other players' cooperative dispositions may drive participants to rationally "defect" (i.e., the choice of sharing a small amount of monetary units with counterparts, an untrustworthy choice), although it does not confer the maximum benefit for the trustor (Kollock, 1998). During the game, participants indicated their preference (cooperation or defection) for each round and the strategy they thought the trustee might choose. Oxytocin administration in BPD participants resulted in significantly more trustee expectations of less cooperation than the BPD placebo condition group. Conversely, healthy controls showed higher trusting expectations following oxytocin doses compared with placebo controls. The authors conducted further analysis of the overall sample and found that mainly anxiously

attached and rejection-sensitive participants showed these divergent results. The authors found that oxytocin does not uniformly facilitate trust and pro-social behavior in humans and generally activates positive affects in trust exchanges, but paradoxically, it seems that such positive affects trigger attachment insecurities and personality traits of rejection sensitivity among BPD patients.

In summary, five different studies support the model's stage about the role of dysfunctional prior beliefs regarding others' trustworthiness for trust impairments in BPD. The effect sizes range from medium (d = 0.41 in Botsford et al., 2019; d = 0.58 Miano et al., 2013; and d = 0.43 in Richetin et al., 2018) to high (d = 0.9 in Butler et al., 2002, and Bartz et al., 2013), indicating greater mistrustful expectations from trustees toward BPD trustors compared with the controls (or high vs. low BPD features in individuals). The differences in the study designs might explain the range of effect sizes. While Butler and colleagues (2002), Botsford and colleagues (2019), and Bartz and colleagues (2013) tested their hypotheses with clinical samples, Miano and colleagues (2013) and Richetin and colleagues (2018) utilized non-clinical samples. Furthermore, Butler and colleagues (2002) compared BPD patients with other personality disorder patients, Botsford and colleagues (2019) compared BPD patients with MDD and SAD patients and healthy controls, and Bartz and colleagues (2013) compared BPD patients with healthy controls. Despite such methodological differences, different studies found alterations in prior beliefs and dispositions in BPD patients' features. Studies' effect sizes, ranging from medium to high, may suggest that such alterations are a stable feature of BPD and could be relevant for trust impairments.

Interestingly, accumulated empirical evidence suggests that the general population exhibits a common mistrust of others (Macko et al., 2014) and is aware of the possibility of being misled (DePaulo et al., 1996). However, BPD patients show even more severe dysfunctional beliefs about others' untrustworthiness than the general population and other psychiatric groups.

3.1.2. Perception of the situation

A fundamental assumption in personality and social psychology is that social behavior cannot be dissociated from the situation in which it unfolds (Fleeson & Noftle, 2012). Personality traits can determine behavior, but what people do also depends critically on their circumstances (Funder et al., 2012). From this perspective, the perception of a trust-relevant social situation might modulate trust behavior. Thus, trust impairment may result from the interaction between dysfunctional personality characteristics, dysfunctional beliefs, dispositions, and modulating social contexts. Regarding BPD, the following studies investigated whether trust impairments could relate to dysfunctions in the appraisal of trust-relevant interpersonal situations.

To investigate social context's role in the perception of the situation and appraisal of others' trustworthiness in ecological interactions, Miano and colleagues (2017) investigated interpersonal trust in romantic relationships involving BPD patients compared with controls in romantic relationships. They hypothesized that the situation might change the partner's trustworthiness appraisal while discussing different topics for BPD patients. The authors asked heterosexual couples in which the women were diagnosed with BPD and control couples with no psychiatric history to discuss three topics: neutral (i.e., favorite films); personal (i.e., personal fears); and relationship-threatening (i.e., possible reasons for separation from partners). After

each discussion, participants rated their partners' trustworthiness. The authors expected a lower appraisal of the partner's trustworthiness in BPD couples than controls, especially after relationship-threatening discussions. Women with BPD did not differ from healthy controls on their partners' perceived trustworthiness after a neutral conversation but did so after personal or relationship-threatening discussions. The result supports the idea that an untrustworthiness appraisal may not be a stable feature in BPD. Still, untrustworthiness appraisal is state-dependent upon situations that posit romantic relationship-related threats may activate trust impairments that are otherwise dormant.

Preuss and colleagues (2016) tested whether social, compared with non-social, situations could activate untrustworthiness appraisal bias more easily in BPD patients. BPD patients, individuals with major depressive disorder (MDD), and healthy controls completed (i) a trust game with a social condition, (ii) a trust game with a non-social condition (interacting with a computer), (iii) a punishment game (social), (iv) a dictator game (non-social), and (v) a cooperation game (social). BPD patients demonstrated significantly less-consistent behavior (i.e., volatile and unpredictable reactions) than healthy controls and MDDs in the social conditions (trust game and punishment game). However, the BPD group did not exhibit such volatility in the non-social conditions. If exposed to social situations, those with BPD again indicated a proneness to perceiving counterparts as threatening and activated untrustworthy behaviors with trustees.

Two studies using different methodologies demonstrate that BPD patients show state-dependent trust impairments connected to specific situations. Both studies reported medium effect sizes (d = 0.58 in Miano et al., 2017; d from 0.64 to 0.69 in

Preuss et al., 2017). However, in a less socially demanding context, BPD patients showed enough coping resources to face demands being placed on them in social situations. In relationship-threatening or socially demanding contexts, trust impairments arose in BPD patients.

3.1.3. The impact from trustors' emotional state on trust processing

The last antecedent of trust appraisal in our model is related to the trustor's emotional state. Besides early emotional experiences, a trustor's emotional state in the here-and-now also may interfere with trustees' trust-related interactions. For example, Jones and George (1998) showed that one's present moods and emotions contribute to the ongoing experience of trust. Positive emotions in the trustor may cause more positive appraisals and heightened attribution of trust in trustees. Conversely, negative emotions in the trustee may result in a bias to appraise trustors as less trustworthy.

Dunn and Schweitzer (2005) found that state activation of intense negative emotions decreased trust evaluations, even when the cause of the negative emotions was distinct from the trustees' behavior and attributes. Furthermore, in the context of economic games, the trustor's emotional state during the game changed the interpretation of the feedback coming from the trustee, either as a reward or a punishment (Rekosh & Feigenbaum, 1966).

Psychodynamic models regarding the development of dysfunctional interpersonal behaviors in BPD, such as mentalization theory (Fonagy, Campbell, et al., 2017) and object relations theory (Kernberg, 1975), and cognitive-behavioral theories (e.g., Biosocial Model, Linehan, 2018) converge on the critical role of positive and negative emotions on trust appraisal. Nevertheless, only a few studies have

examined the role of both the trustor and trustee's emotional state on trust impairment in BPD.

King-Casas et al. (2008) used a translational functional magnetic resonance imaging (fMRI) approach to examine differences between BPD and control participants behaviorally and neurally during a trust game. Traditionally, the quantitative analysis of trust games relies on the computational model of the Interactive Partially Observable Markov Decision Process (I-POMDP; Gmytrasiewicz & Doshi, 2004) or the Bayes-Nash model. Both models assume that players in a trust game will behave as rational players seeking to maximize each player's benefits. These models do not consider emotions' role in behavior. Hula an colleagues (2017) pointed out many structural limitations in the Bayes-Nash model. They analyzed the King-Casas' data set, focusing on two novel phenomena – social risk aversion and irritation/anger – by adopting an alternative computational model that allows for inferences about three experiences relevant to trust: risk aversion (determined by the value of money kept over potential money gained); irritation (inferred from the tendency to retaliate after negative partner actions); and guilt (inferred from the tendency to increase investments toward partners after unequal advantageous payoffs). Hula et al. (2018) re-analyzed King-Casas et al.'s (2008) data set and showed that BPD trustors experienced less guilt and awareness of their irritation than healthy control trustors during the economic game. The authors labeled trustees with low guilt proneness and unawareness of irritation as "perilous individuals" who deliberately may exploit the investor and create problematic interactions. They found perilousness to be more common in the BPD sample compared with controls. Furthermore, perilous individuals were more likely to interpret cooperative situations negatively and were

less prone to establish cooperative interchanges or repair cooperation ruptures. Moreover, BPD patients, like perilous individuals, showed increased irritation from unpleasant interactions during economic exchanges. Interestingly, BPD patients were unaware of the irritation coming from uncooperative interactions. According to the authors, BPD patients' unawareness of interactions quickly transforms drops in investment toward economic partners into complete cooperation breaks.

In another study, Roberts and colleagues (2018) hypothesized that administration of acetaminophen, a pain reliever, would reduce behavioral mistrust (i.e., low investment) exhibited by participants with high levels of BPD features during a TG procedure. Before the TG, participants with high *vs.* low BPD features received either acetaminophen or a placebo. The authors also asked participants to express their expectations about counterparts' behavior before interacting with them in a TG. The procedure entailed four conditions: participants with high BPD features after receiving acetaminophen; participants with high BPD features after receiving a placebo; controls (low BPD features) after receiving acetaminophen; and controls after receiving a placebo. The findings confirmed their hypotheses.

Interestingly, the authors found similar rates of untrustworthy expectations in individuals with high BPD features and individuals with low BPD features regardless of acetaminophen or placebo administration. For this reason, the authors speculated that the decrease in behavioral mistrust in the acetaminophen condition among individuals with high BPD features was due to a reduction in negative emotional affect related to possible unpleasant outcomes in interpersonal interactions (and not due to changes in expectations). In other words, acetaminophen might reduce betrayal aversion, described as the emotional reaction to others' untrustworthy behaviors, in

individuals with high BPD features, independent of their expectations about others' untrustworthiness.

Another study, by Masland and Hooley (2019), examined the influence of an emotional prime for the trustor on the trustee's trustworthiness appraisal. Non-clinical participants with high *vs.* low borderline features rated unfamiliar faces' trustworthiness after an affective priming paradigm that exposed them to negative, neutral, or positive images. The authors found that high-BPD-features individuals showed significantly lower trust appraisal after exposure to negative, neutral, and positive primes relative to the low-BPD group. However, low-BPD and high-BPD groups showed a significant decrease in trust appraisal after negative emotional primes. Compared with the low-BPD group, negative affective primes influenced appraisal more in the high-BPD group. If taken together, these results support the negativity bias theory (Arntz, et al., 2004; Arntz & Veen, 2001). This study suggests that individuals with high BPD features exhibit negatively biased trust appraisal of the trustee. Moreover, according to the results, the influence from negative emotional states on trust appraisal performance is stronger for individuals with high BPD features, compared with those with low BPD features.

Three studies with different methodologies and samples indicate that BPD patients who are in trustor roles show impairments regarding their emotions' role in trust appraisal processes. All studies reported medium to high effect sizes (Masland & Hooley, 2019, d from .50 to .80; Roberts et al., 2018, d = 1.20). This consistency in effect sizes supports the hypothesis that the effect from emotions' state intensity (especially if negative) on negative trustworthiness appraisal of trustees is more substantial among BPD patients than controls. If non-BPD trustors usually show

appropriate emotional reactions to potentially unpleasant outcomes in interpersonal interactions, BPD patients tend to exhibit more intense emotional responses. Such volatility and intensity in the emotional experience of interpersonal situations among BPD patients may facilitate a negative appraisal of trustees' trustworthiness.

3.3. Trust Appraisal

The human face is a salient source of interpersonal information. The appraisal of others' trustworthiness is such a relevant judgment for interpersonal exchanges that people, on average, make initial trust appraisals of others based on visual facial morphology after only 100 milliseconds (Willis & Todorov, 2006). Research on facial trustworthiness shows that individuals make decisions about whom to trust based on various elements. Among such features are: facial attractiveness (Eckel & Wilson, 2006); similarity to kin (DeBruine, 2002); appraised trustworthiness (van 't Wout & Sanfey, 2008); and facial expression (Oosterhof & Todorov, 2008). Several studies investigated the ability to appraise emotional facial expressions in BPD, with mixed results (for a review, see Domes et al., 2009). Nonetheless, there is evidence that emotion appraisal (e.g., fear, joy, happiness appraisal) and trait appraisal (e.g., trust appraisal) are different processes (Oosterhof & Todorov, 2009).

Three empirical contributions that examine the ability to appraise trustworthiness in BPD are presented below.

Fertuck and colleagues (2013) compared facial trustworthiness and fear appraisal in BPD and healthy controls. Participants rated the degree of trustworthiness and fearfulness of faces morphed from neutral to fearful or trustworthy to untrustworthy expressions. Reaction times and three parameters of psychophysical performance were assessed: discriminability (of trust- and fear-related facial features);

sensitivity (to the presence of trust and fear-related facial features); and response bias (mean untrustworthiness or fear appraised across all morphs). Compared with controls, BPD participants rated the faces as more untrustworthy across all morph levels, indicating the presence of response bias. No significant differences in sensitivity, discriminability, or bias in fear appraisal emerged. Moreover, BPD participants showed slower trust ratings than controls, especially toward more ambiguous faces, while there were no differences in RTs for fear ratings between groups.

In a different study, Nicol and colleagues (2013) found similar results. Participants (BPD vs. control) assessed whether the presented faces were "high" or "low" in age, distinctiveness, attractiveness, intelligence, approachability, and trustworthiness. BPD participants showed a significantly larger negativity bias effect in the appraisal of social dimensions, such as unknown faces' approachability and trustworthiness compared with controls. No differences in non-social aspects of the appraisal of others, such as age or intelligence, emerged. In other words, BPD patients, compared with controls, perceived less trustworthiness and approachability, whereas no differences occurred in perceptions of age and intelligence.

Fertuck and colleagues (2019), in a different study, asked participants to perform a trustworthiness and fearfulness appraisal of faces during fMRI scanning. The authors replicated behavioral findings on the occurrence of untrustworthiness bias among BPD patients, who showed more biased trustworthiness (but not fearfulness) appraisal (i.e., more untrustworthiness appraisal) and less discriminability than the controls. The BPD group also exhibited slower reaction times when appraising ambiguous trustworthiness compared with the controls. Furthermore, the neural fMRI activation of BPD patients during trustworthiness ratings evidenced less activity in the

anterior insula and lateral prefrontal cortex than the controls. Such a decrease was proportional to the degree of trustworthiness bias and impaired discriminability demonstrated by participants behaviorally (in both BPD patients and the controls). BPD patients did not show amygdala hyperactivation during trustworthiness appraisal relative to healthy controls. Thus, impaired probabilistic reasoning (linked to prefrontal cortex activity) might be more relevant than hypersensitivity to threatening stimuli (traditionally linked to hyperactivity in the amygdala) in playing a role in trustworthiness appraisal impairments in BPD.

In summary, three studies consistently documented lower trust appraisals of trustors' facial stimuli among BPD patients compared with healthy controls with large effect sizes (*d* ranging from 0.70 to 1.30). While the general population usually does not show systematic bias in trust appraisal of neutral faces, BPD patients show significantly lower evaluation of others' trustworthiness based on facial stimuli.

3.4. Behavioral manifestations: interpersonal cooperation and repair of ruptures

Researchers often have used game theory procedures to investigate the behavioral manifestations of trust in controlled experimental conditions. Game theory is a branch of behavioral economics that considers the roles of emotions, mistakes, limited foresight, doubts about how intelligent others are, etc., in the study of decision making (Cameron, 2003).

Personality dimensions are significant predictors of behavioral trust shown in TG procedures (for a general overview, see Johnson & Mislin (2011) meta-analysis involving 162 studies with more than 23,000 participants playing the TG). For instance, Kurzban and Houser (2001) showed that individual differences in self-monitoring, self-esteem, neuroticism, and conscientiousness predicted different levels

of exhibited trust in the trust game. While self-monitoring, self-esteem, and conscientiousness predicted increases in exhibited trust, neuroticism significantly predicted decreases in exhibited trust.

Several researchers have investigated atypical trust manifestations in the BPD population using game theory procedures. King-Casas et al. (2008) used a trust game procedure and concurrently recorded neural activation in BPD and control individuals. The authors focused on two trust behaviors: the capacity to sustain a mutually rewarding, cooperative social exchange (*vs.* cooperation ruptures) and the ability to repair non-cooperative interactions. They measured the ability to repair non-cooperative interactions through rates of "coaxing" behaviors (i.e., when a trustor repays a large part of the investment to the trustee to signal their trustworthiness and gather more substantial investments on subsequent rounds from trustees). BPD patients, compared with the controls, were more likely to initiate cooperation ruptures by sending rejecting social signals, i.e., reduced investment in the trustee. Moreover, BPD trustors had lower rates of coaxing behavior to repair the cooperation ruptures compared with the controls. Furthermore, in the control group, the anterior insula response was associated with the amount of money received from the trustee and sent back during coaxing exchanges. The amount of money sent back to the trustee was associated with anterior insula activity in the BPD sample. No associations were detected with the trustee's offers. Usually, anterior insula activity is related to the violation-of-social-norms perception, but in the BPD sample, this was not the case. Considering that the BPD group showed no insula activation, the authors attributed BPD patients' low investment behaviors to a lack of sensitivity (assessed via insensitive insula activity) from social norms violations. Furthermore, they suggested

that a lack of insula activation may occur because of dysfunctional beliefs' top-down influence, such as holding negative expectations about social partners. To sum up, according to the authors, the rise of dysfunctional beliefs and dispositions in BPD patients in the face of previous negative social experiences might be associated with a more robust top-down inhibition of brain areas responsible for the processing and evaluation of social signals. Such association may cause higher cooperation ruptures and lower rates of coaxing behaviors.

Unoka and colleagues (2009) replicated the effect found by King-Casas and colleagues (2008). The authors investigated the actions during a single-trial TG and a risk game in three groups: BPD; major depression disorder (MDD) individuals; and controls. Additionally, before playing both the trust game and risk game, participants shared their expectations about the games' outcomes. In the single trial TG, the trustee can share a fair number of monetary units with the trustor (i.e., the participant) or an unfair amount (violating the investor's trust). By contrast, in the risk game, the number of monetary units returned by the trustees to the trustor is determined randomly. The BPD group evidenced lower investment rates in the TG procedure than MDD and controls, but comparable investment rates to the other groups in the risk game. Moreover, BPD had more skeptical forecasts about the TG outcomes, as well as more accurate estimates about the risk game, compared with MDD individuals and controls. Unoka et al. (2009) interpreted the results as being a consequence of dysfunctional beliefs in BPD patients regarding others' cooperativeness. Even if they did not strictly test for dysfunctional beliefs, they viewed their results as proof that BPD and MDD individuals rely on opposite patterns of ideas. On one hand, BPD participants made a better estimation of the lottery game outcome compared with MDD participants and

believed in “impersonal luck,” but not in the cooperativeness of other people during the trust game. On the other hand, MDD individuals reported skeptical forecasts about the risk game’s outcome, but not the trust game, and believed in other people’s cooperativeness, but not in luck. The authors’ interpretation of the results further supports previous findings regarding dysfunctional beliefs (see Paragraph 3.2.1.) as a plausible cause of cooperation impairments in BPD (Butler et al., 2002; Botsford et al., 2019; Miano et al., 2013; Richetin et al., 2018; Bartz et al., 2011).

Saunders et al. (2015) found cooperation impairments in BPD patients with an iterated form of the prisoner’s dilemma game (Rilling et al., 2002). The original prisoners’ dilemma comprises participants faced with the choice to cooperate or defect (i.e., keep all monetary units for themselves) for their sole or joint benefit. The iterated version allows for measuring how individuals acquire and maintain reciprocal altruistic behavioral patterns in multiple exchanges. In the iterated version, the rational strategy is to seek cooperation that maximizes both players’ gains. To get the maximum mutual benefit, the trustor should systematically repeat the trustee’s last choice, undertaking a “tit-for-tat” approach to elicit cooperation from social partners (Axelrod & Dion, 1988). Saunders and colleagues (2016) showed that BPD patients were less able to form reciprocally cooperative relationships with social partners (i.e., they did not assume a tit-for-tat strategy) than the controls. The authors speculated that, due to frequent failures to sustain gratifying relationships with their social partners, BPD patients might assign a diminished reward value to mutual cooperation compared with controls. Due to the lack of perceiving cooperation as rewarding, BPD patients may not sustain future cooperation exchanges.

In summary, three studies proved that BPD patients, compared with controls, show significant alterations in the model's stage regarding behavioral manifestations of trust. More specifically, these studies consistently found that regardless of the game paradigm used, BPD patients evidenced a lack of cooperation with social partners. The studies, wherever possible to compute, show large effect sizes (d = 0.76 and d = 0.93). In the general population, trustors usually behaviorally "trust 'too much' in relation to their pessimistic beliefs (i.e., they send big sums to the trustee) and accept the risk of loss during the game" (Macko et al., 2014, p. 44). However, BPD patients do not trust enough social encounters' counterparts, leading to cooperation rupture.

3.5. Trust Learning

Trust learning refers to "*learning whom to trust and when to revise trust attributions*" (Fineberg et al., 2018, p. 838). A manifestation of successful trust learning processes is the occurrence of adaptive updates in the appraisal of others' trustworthiness due to exposure to trust-relevant interpersonal interactions. Conversely, rigidity and inflexibility in trustworthiness appraisal may signal a failure in trust learning processes. For instance, in economic game procedures, accurate trust learning corresponds to an increase in the likelihood of positive and cooperative interactions in future transactions after the trustors' investment is reciprocated with cooperation. When trustees violate trustors' trust by defecting, the probability of future trust decreases significantly (Jones & George, 1998).

Three studies addressed the investigation of trust learning in BPD. In a study by Franzen et al. (2011), trustees expressed happiness or anger during a trust game. In the game, trustee fairness was manipulated to display fair behavior (high MU amount transferred to the trustor) or unfair behavior (low amount of monetary units transferred

to the trustor). BPD patients typically demonstrated stable impairments in the appraisal of others' emotions, such as less accuracy in recognizing facial displays of anger and disgust than non-clinical controls (e.g., Daros et al., 2013). Thus, the authors expected the rates of cooperative behaviors among BPD patients to be lower than those of controls regardless of trustees' facial expressions.

Against their expectations, both BPD patients and controls adapted their investment behavior to trustees' facial expressions. There were no group differences in influence from the valence and intensity of trustees' emotional expressions on cooperation rates endorsed by trustors. In both groups, the amount of cooperation decreased as the trustee's anger increased. Both BPD patients and controls used information about others' emotional states to set their behavior in social interactions accordingly. In response to trustees' happiness, both groups engaged in cooperative actions more often than in response to trustees' anger. The lack of difference in performance demonstrates that BPD patients did not show any impairment in cooperation.

In a second study, Fineberg and colleagues (2018) investigated the different weight of social and non-social cues in a learning task for BPD participants compared with controls. The authors asked participants to carry out a decision task. During each trial, participants saw a blue and green card, then received a social cue (confederate's advice) or a non-social cue (computer's advice). Finally, participants had to choose which card (blue or green) to pick. The task design included five different subphases that varied in terms of volatility and reliability of cues: low volatility and high accuracy of social signals; high volatility and low accuracy of social cues; low volatility and low accuracy of social cues; low volatility of non-social cues; and high volatility of

non-social cues. The participants' task was to learn the reward probability of social (i.e., partner's advice during the game) and non-social (i.e., computer's advice) cues. Learning rates were modeled based on the number of trials occurring between the start of the phase and the engagement of choices consistent with the ongoing condition (i.e., following the advice during stable and reliable stages and not following the advice during volatile or unreliable phases). Immediately after the task, participants answered a few questions about the task experience. The authors, examining the transcripts of participants' answers to debriefing questions, counted the number of times each participant mentioned the confederate. Compared with controls, BPD patients mentioned the confederate more frequently, suggesting more attention paid to and reliance on social cues. Looking at learning rates during the task, BPD patients learned more slowly than control subjects during all three phases of the task, providing social cues. Furthermore, BPD patients showed lower learning rates (more extended time for learning) in volatility conditions than the controls.

Finally, in a third study, Abramov and colleagues (2020) implemented a 15-round trust game manipulating trustees' investment rates with three separated phases: formation of trust (high investment of trustee toward trustor); dissolution of trust (after a trust violation, i.e., low investment of trustee toward trustor); and trust restoration (back to the high investment of trustee toward trustor). Against the authors' expectations, individuals with high BPD features showed declining trust only during formation of the trust phase (i.e., when interacting with an unknown partner showing cooperative behavior). Surprisingly, following trust violation and during the restoration phase, individuals with high BPD features showed higher investment rates than individuals with low BPD features. According to such findings, BPD patients

may have slower responses to cooperative decision making as they use more demanding and sophisticated cognitive evaluation of surroundings' social cues than the controls.

To sum up, three studies investigated trust learning in BPD. In terms of effect size, Franzen et al.'s (2011) results suggested that no difference existed in emotional information's influence on investment decisions between BPD patients and controls with a small effect size (d = 0.30). However, Fineberg et al.'s (2018) results indicated that considerably more attention was paid to social cues, but that learning rates were lower in BPD patients compared with controls with a larger effect size (d = 0.81). In conclusion, healthy controls tend to evaluate cues on surroundings accurately, such as social (others' advice) and non-social (computers' advice), as well as adapt their trust-relevant behaviors accordingly. Individuals with high BPD features adjusted their behavior according to their counterparts' emotional experience accurately. Furthermore, they overestimated social cues' value when deciding whether or not to behave cooperatively and exhibited lower learning rates, compared with the controls, in a reward probability task.

4. Discussion

With this review, we aimed to propose a multi-stage model of trust processes (Figure 1) to dissect BPD-specific impairments at each stage. The model outlines the antecedents (both distal and proximal) of trust appraisal and attribution, which influence behavioral manifestations in interpersonal exchanges and trust learning (the update of information about others' trustworthiness resulting from real social experiences). We examined the variations from typical processes related to BPD, a

clinical condition strongly associated with trust impairments, and found support for significant atypical processing at each stage of the model in BPD.

Regarding distal antecedents, BPD patients reported more early adverse experiences and attachment insecurity than non-BPD populations. Developmental adversities are associated with trust impairments in BPD patients. This corresponds with research on the impact from developmental factors in the emergence of BPD (e.g., Chanen et al., 2008; Schmeck et al., 2013).

Concerning proximal antecedents, dysfunctional beliefs and dispositions (such as "I cannot trust other people"), increased sensitivity to situational cues, and greater negative emotional states (such as irritation) in BPD trustors predict trust impairments in BPD.

While healthy individuals can build trust relationships by accurately appraising those whom they can depend on safely, BPD patients usually show a fixed and pervasive social expectation of others' untrustworthiness that is resistant to modification by positive and cooperative interpersonal experiences. In a vicious cycle, these negative expectations sustain negative interpersonal exchanges that further reinforce BPD patients' untrustworthiness assumptions and expectations about others. More precisely, BPD patients, due to their untrustworthiness expectations, demonstrate less cooperative behaviors, thereby eliciting negative interpersonal encounters with others that strengthen their initial untrustworthiness dispositions. In other words, relying on an increased expectation of others' untrustworthiness can affect how future social information is perceived, processed, and updated in BPD patients.

We propose three main processes as proximal antecedents of trust processing: prior beliefs and dispositions; perception of the situation; and the trustor's emotional state. Considering a synergistic relationship between stable (i.e., personality traits, prior beliefs, and dispositions), unstable (or volatile, such as the perception of the situation), state (emotional state of the trustor), and contextual factors, we assert an interactionist perspective, which fits well with recent approaches in personality pathology accounting in terms of the significant role of situations in translating personality factors into behaviors, from both social cognitive and emotional perspectives (i.e., interpretation of and emotional experiences with the situation, respectively; Shoda et al., 2013). We suggest that situation-behavior contingencies can intensify baseline pathological traits such as less-cooperative behavior.

Regarding trust appraisal, all studies on the ability to appraise trust showed a bias toward untrustworthiness among BPD patients, who express maladaptive behaviors during social encounters, such as increased ruptures in cooperation during economic games. According to the authors, BPD patients might rely on enhanced top-down inhibitory activity in areas specializing in the processing and evaluation of social signals to face early adversity (King-Casas et al., 2008) and manage dysfunctional beliefs (Unoka et al., 2013). Increasing the strength of top-down inhibitory processes, BPD patients might try to avoid distressing and negative emotional experiences with untrustworthy others. This explanation further strengthens the idea of a link between distal antecedents, proximal antecedents, and trust-appraisal stages of the model that we suggested. Furthermore, this explanation further supports evidence provided by studies on trust appraisal concerning the need for more cognitive efforts to appraise others' trustworthiness by BPD patients compared with others. The extra reasoning

needed could serve the scope to compensate for the robust inhibitory brain activity that BPD patients endorse when facing trust-relevant decision-making.

In our heuristic model, we suggest that an efficient process of updating our expectations regarding others' trust (i.e., trust learning) is relevant for building stable, trusting relationships over time. Considering that we suggest a circular model, we argue that the quality of trust learning responds to the emotional state, the perception of the situation, the disposition, appraisals of trust, and individuals' behavior, and it influences these aspects. Such a statement is consistent with the contingency approach to personality disorders that argues for the importance of considering both distal mechanisms (such as neglecting parents or child abuse) and proximal mechanisms as triggers of the expression of symptoms (Miskewicz et al., 2015). The economic game studies provided initial evidence of the relevance of a match between personal dispositions and triggering situations (i.e., contingencies) through which BPD patients can show dysfunctional features.

We thus suggest that in the general population, if dysfunctional beliefs and dispositions about others do not meet confirmation in social situations (i.e., positive interpersonal experiences), these can be updated and adjusted to more positive beliefs and dispositions about others. This flexibility does not seem to occur in BPD patients, who demonstrated lower learning rates than non-BPD patients.

Furthermore, we suggest that in moving from early adverse experiences and adversity, BPD patients might develop a robust set of dysfunctional beliefs about others (e.g., "Others would take advantage of me if they could"). BPD patients firmly apply such a set of dysfunctional beliefs to every person, group of people, institutions, etc., revealing low capacities to discriminate among situations. Such defensive

attention to every feature of individual and social experiences, somehow being trust-relevant, increases BPD patients' sensitivity to others' untrustworthiness cues, making them react abruptly. BPD patients tend to respond to every betrayal signal, including the most insignificant ones, with impulsive and punitive behaviors, even if it means renouncing a profit (i.e., lacking coaxing behavior and having few investments in the economic game procedures). Moving upon a rigid set of assumptions about others' untrustworthiness, BPD patients' actions result in rigidity. In the general population, individuals update their prior beliefs and dispositions to perceive the social situation and read others' emotional expressions to trust more (or less) others according to present positive or negative experiences. BPD patients could show impairments in updating their expectations, fear of others' untrustworthiness, and behaviors according to present interactions. Their inability to update their trust decisions based on the quality of real-time interactions culminates in negative experiences with social encounters that further reinforce the rigid set of beliefs and biases about others' untrustworthiness.

To sum up, in our model, we indicated that contextual elements, personality features, and BPD patients' social style might impede updating information about others' trustworthiness according to present experiences. BPD features' complexity might foster a self-fulfilling prophecy. A person who expects to encounter untrustworthy behavior in others perceives every action as potentially untrustworthy and further fuels the belief that trustworthiness is not an option.

4.1. Implications for further research

Our main aim was to provide a framework to review how BPD patients differ from the general population concerning trust processing. To this end, we proposed a

heuristic model delineating trust impairment's temporal stages. Such a conceptual model may guide future research in identifying the specific steps in which impairments in trust occur across different diagnoses. Referring to a single model to systematize trust impairments across other diagnoses may allow for comparing and contrasting at each stage across diagnoses.

In the present contribution, by collecting and reviewing the empirical contribution on trust impairments among BPD patients, we have exemplified an application of such a framework to the specific BPD diagnosis. Our results demonstrate that BPD patients show specific impairments across all stages of the suggested model.

By dissecting the components of the trust impairment process into discrete, sequential steps, as we did in our model, we may not have fully valued the mutual influences occurring between each stage. Considering that real-life interpersonal trust dynamics occur continuously, we recommend that readers consider each step of the model presented as potentially influencing the whole process, both backward and forward.

In light of the wide range of effect sizes and the variety of unreplicated methodologies reported in this review, a need exists for replication studies. For example, differences exist in the instruments used to assess BPD features (self-reports: k = 6; semi-structured interviews: k= 13; and both self-reports and semi-structured interviews: k = 2) in the population examined (clinical population: k = 16; non-clinical population: k = 5), and in the comparison groups used (healthy controls: k = 17; individuals with other diagnoses and healthy controls: k = 3; and one study with other diagnoses without healthy controls). Moreover, future research could focus on

overcoming several limitations. For example, the studies reported here mainly used visual stimuli from unknown others. It would be interesting to attempt to replicate original findings obtained with visual stimulation using alternative stimuli such as immersive virtual environment procedures. Furthermore, there was no examination of whether familiarity with the visual target during the procedures might moderate trust judgment. Apart from Miano et al. (2013), no study has tested untrustworthiness bias toward significant others. Moreover, all the empirical contributions analyzed in the present work were laboratory studies. We argue that the use of controlled tasks and stimuli in laboratory research significantly limited findings' ecological validity and generalizability to real-life incautious (Hoc, 2001). For this reason, we invited future researchers to try and bridge the gap between laboratory settings and daily life. This could increase our understanding of the processes through which trust impairments impact BPD patients' real lives. To this aim, the implementation of ecological momentary assessment (EMA) procedures may be useful in testing whether dysfunctional trust appraisal in BPD is linked to uncooperative behavioral outcomes in real life. EMA procedures could help reach a better understanding of several dimensions: the nature and quality of situational cues triggering lower appraisal of trust; the volatility of trust appraisal; and the dynamics of trust learning with related cognitions and behaviors (see Berenson et al. (2011) for an exemplification of the use of EMA procedures).

Trust learning is a crucial stage of trust impairment in BPD that merits continued emphasis, as it carries implications for prevention and intervention development. Whereas learning models state that people update expected outcomes of future interactions according to simple reinforcement-learning mechanisms, Franzen

et al. (2011) and Fineberg et al.'s (2018) studies show that BPD trustors' untrustworthiness bias may interfere with these learning mechanisms. For this purpose, future research could test whether BPD patients rely on learning strategies other than those the controls employ. For example, implicit learning and explicit learning about others' trust might be partially independent processes. Implicit learning comprises an automatic learning process that occurs without people's consciousness of the access to their procedural skills, while explicit learning is a conscious process implying operational attempts to derive and test hypotheses (Reber et al., 1991). We believe that future research could investigate whether trust impairments among BPDs specifically affect explicit trust learning processes, implicit learning processes, both, or none.

Finally, evaluating the degree to which these impairments shown are specific to BPD or could apply to other PDs and psychiatric disorders would be essential. Considering that comorbidity rates among PDs are higher than other diagnoses (Clark, 2005), the isolation of an element that solely characterizes BPD diagnoses could be useful for differential diagnoses. Few empirical studies have investigated the link between trust impairment and PDs other than BPD, suggesting that trust impairments might be relevant in other disorders as well (e.g., psychotic features in antisocial personality disorder, paranoid personality disorder, or schizotypal personality disorder; see Poggi et al., 2019). More precisely, with narcissistic personality disorder (NPD), appraising others as less trustworthy links to increased aggressiveness or hostility toward trustors (Kwiatkowska et al., 2019). This result may suggest that NPD patients with highly hostile and aggressive motives trust others less as well. Concerning paranoid personality disorder (PPD), Furnham and Crump (2015) studied

a non-clinical sample of 4,100 British adults. The authors found that the more skeptical the subjects were about other people (subclinical paranoid), the lower they scored on trust and compliance. This result may suggest significant and specific impairments among PPD individuals regarding their prior beliefs and our heuristic model's dispositions section: The more PPD individuals have skeptical beliefs about others, the less they trust others. It could be interesting to examine whether prior beliefs and dispositions toward others influence individuals' appraisal of others' trust, behaviors, and trust learning accuracy differently between BPD and PPD.

4.2. Clinical implications

To our knowledge, no extant study has investigated empirically the differences in trust processes across different forms of psychopathology. The framework that we outlined in this systematic review may facilitate investigations of trust processes across various clinical disorders. Nonetheless, such knowledge might inform a more accurate evaluation and clinical management of trust impairments across conditions by suggesting how to tailor treatments according to specific individual impairments. For example, taking trust processes into account in clinical work with BPD patients might inform the management of suicide risk. BPD is associated with elevated suicide rates, and a large percentage of BPD patients have a history of at least one suicide attempt (46%-92%; Black et al., 2004). It has been found that low interpersonal trust is associated with higher suicide ideation in an adolescent inpatient sample (Hill et al., 2019). Building on such evidence, one could hypothesize that interventions aimed at increasing interpersonal trust may help prevent suicide ideation. In this sense, we believe that shedding light on the processes that lead to BPD patients developing less

interpersonal trust also could inform the development of specific clinical and therapeutic techniques to prevent suicide ideation among BPD patients.

Regarding clinical management, future investigations should evaluate whether trust impairment may predict BPD patients' non-compliance with treatment. Trust impairments easily could influence BPD patients' non-compliance with medications and inconsistent therapy engagement (Sinnaeve et al., 2015). Both clinicians and BPD patients view trust as a crucial therapeutic relationship element that is predictive of positive clinical outcomes (Langley & Klopper, 2005).

Many of the studies that we analyzed in this paper suggest that BPD patients tend to enact "maladaptive transaction circles" (Hopwood et al., 2013). Consequently, BPD patients may feel distressing and dysregulated affects even after both neutral or positive interpersonal encounters. Such an atypical emotional response could culminate in pathological misperceptions and misattribution of malicious motives to others, including those of therapists. In clinical practice, these maladaptive transaction circles usually dominate early phases of therapy in the form of paranoid transference. According to the object-relation theory of personality pathology, these are expressions of the massive use of splitting defense that involves "all bad" and "all good" representations of the object that do not correspond to reality. TFP interventions work to integrate extreme positive and negative representations of the object (and self) into a more realistic image, including positive and negative aspects. To this purpose, TFP therapists suggest acknowledging both positive and negative elements of internal patients' world projected into those of therapists since the earliest stages of therapy (Yeomans et al., 2015). Also, given the role of proximal antecedents of trust appraisal in our model, we argue that therapists should address trust issues starting with the early

phases of therapy to improve therapeutic alliances and avoid patient dropout. TFP therapists suggest that negative transference interpretations may serve this scope (Yeomans et al., 2013). Additionally, therapists may benefit from negotiating contracts stating the scopes, goals, objectives, and responsibilities of patients and therapists in the therapeutic process.

Feeling safe within the therapeutic environment is essential for a solid therapeutic alliance (regardless of patients' condition). Even if not empirically tested, it is very plausible that BPD patients' trust impairments inhibit therapeutic alliances; therefore, management of trust impairments during treatment could help break trust learning's vicious cycles and at best enable the restoration of positive interpersonal interactions. Although psychotherapy research has identified the therapeutic alliance as a central and generic factor of change (Lambert, 1992), and major treatment approaches for BPD stress the need for building a solid therapeutic alliance, it remains unclear which strategies and methods are most effective.

More precisely, five major treatments have been established as evidence-based treatments (EBTs) for BPD: dialectical behavior therapy (DBT); mentalization-based therapy (MBT); transference focused psychotherapy (TFP); schema-focused therapy (SFT); and systems training for emotional predictability and problem-solving (STEPPS) (Storebø et al., 2020). According to a recent contribution, these EBTs for BPD share a common factor: building a solid therapeutic alliance and creating conditions under which patients can begin trusting others anew (Bateman et al., 2018). While EBTs for BPD develop, the proposed framework model might guide future investigations toward a better understanding of best practices to engender patients' perception of therapists' trustworthiness since the earliest stages of treatment.

In conclusion, we outlined a multi-stage, heuristic model of trust processes as a framework for dissecting how BPD patients express trust impairments. We found evidence of atypical phenomena in the BPD population (or individuals with high BPD features, in the case of non-clinical studies) for each stage of the model that we hypothesized. We found robust literature with empirical evidence that BPD patients: (1) experience adversities in trusting their caregivers since the earliest developmental stages; (2) rely on a robust set of dysfunctional beliefs about others' trustworthiness; (3) develop negative trustworthiness bias that applies to others without discrimination; (4) show high sensitivity to signs of trust ruptures in relationships; (5) do not act to fix cooperation ruptures; (6) show impaired flexibility and do not update information about others' trustworthiness depending on new, real-time interactions; and (7) show pervasive negative activations when they perceive a trust rupture.

Finding empirical support for each stage of the model may suggest that it is a fruitful model. Nonetheless, we believe that the model still needs to be more thoroughly elaborated and investigated. For this reason, we encourage future researchers to examine the validity of the framework into other clinical conditions, going beyond the application to BPD diagnosis.

CHAPTER 2. THE EMOTIONAL COMPONENTS OF REJECTION SENSITIVITY AS A MEDIATOR BETWEEN BORDERLINE PERSONALITY DISORDER AND BIASED APPRAISAL OF TRUST IN FACES

Overview

Considering the heuristic model in chapter 1, here we explore the stage of Prior beliefs and Disposition. More precisely, we examine the role of Rejection Sensitivity (RS) on the association between BPD features and decreases in trust appraisal[2].

Our study starts from mixed findings of the RS's mediating role in the trust impairments-BPD features link. Whereas some studies (Miano and colleagues, 2013) found that RS mediated the effect of BPD features on trust appraisal, others (Masland and Hooley, 2017) did not replicate such a result. Following Preti and colleagues' (2018) suggestion, we explore the emotional and cognitive components of RS separately. We argue that considering the RS components as distinct would provide additional insights on the unique role of anxiety, anger, and expectation of rejection on the BPD features-trust appraisal link.

A hundred and twenty-five undergraduate women complete the Adult Rejection Sensitivity Questionnaire, the Borderline Personality Disorder checklist, and appraise trust of 16 neutral facial stimuli (eight male and eight female). Results show a significant mediation effect of anger and anxiety for rejection on the BPD features-trust appraisal link. BPD traits are connected to lower appraisal of trust through anxiety

[2] This chapter was published as: Richetin, J., Poggi, A., Ricciardelli, P., Fertuck, E., & 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*(2).

and anger for rejection and not through the expectation of being rejected. We discuss the implications of these findings.

1. Introduction

Borderline Personality Disorder (BPD) is a complex and severe diagnosis, defined by impulsivity and instability, especially in the domain of interpersonal relationships (American Psychiatric Association, 2013). Most of the typical symptoms of BPD occur within interpersonal contexts, and patients with BPD usually display severe difficulties in interpersonal relationships. Therefore, in the last decade, many researchers focused on understanding the possible deficits in the perception, processing, and emission of social signals that sustain BPD patients' impaired social relations, i.e., their social cognition deficits (Adolphs, 1999). Several studies have been conducted, and recent reviews (Herpertz & Bertsch, 2014; Roepke et al., 2012) have underlined, among other problematic areas, a general impairment in cooperation as a core deficit of the disorder (Thielmann et al., 2014). Cooperative behavior relies on trustworthiness judgments about others or, in other words, on "second-order trust", (i.e., the belief that someone can be trusted; Jansson & Eriksson, 2015). Different studies suggested that BPD patients have a generalized mistrust of others resulting in an appraisal of greater untrustworthiness in neutral faces, greater sensitivity to others' untrustworthiness, and a behavioral untrustworthiness bias (Fertuck et al., 2013; Miano et al., 2013).

Another feature that impairs the elaboration of social stimuli of patients with BPD is their inclination toward Rejection Sensitivity (RS). RS is a cognitive and affective disposition to anxiously or angrily expect, readily perceive, and overreact to social rejection (Downey & Feldman, 1996). RS leads to negative and hurtful dispositional attributions and interpretations of the interactions with others, and one typical manifestation of RS is hypervigilance toward rejection cues. Several studies in

literature support the evidence of a strong connection between RS and BPD traits. First, individuals high in RS have interpersonal difficulties similar to those of patients with BPD (Downey & Feldman, 1996), such as concerns over abandonment and conflicts in romantic and social relationships. Second, compared to control participants, BPD patients believe to a greater extent that they will be abandoned and rejected (Arntz et al., 1999; Arntz, Dreessen, et al., 2004; Ayduk et al., 2008). More specifically, RS is higher in individuals with BPD compared to both healthy controls and Social Anxiety Disorder (Staebler et al., 2011). Finally, experimental studies have even demonstrated that BPD patients perceive rejection even when actually included (e.g., De Panfilis et al., 2015). In sum, BPD patients may show strong impairments in cooperative behavior partly because of their untrustworthiness bias and high levels of rejection sensitivity.

Despite the interest toward both untrustworthiness bias and rejection sensitivity in BPD, only a few studies focused on the possible connections between these two main features. Moreover, these few investigations obtained divergent results. On the one hand, Miano and colleagues (2013) hypothesized that individuals with BPD features might be more negative in the trust appraisal of others because of their anxiety about the possibility of being rejected or abandoned. They found that participants with high BPD traits scored significantly higher on the untrustworthy facial appraisal as compared to participants with low BPD features. Moreover, RS mediated the effect of BPD features on trust ratings. On the other hand, Masland and Hooley (2019) investigated the influence of irrelevant emotional information on trustworthiness appraisal in a BPD sample versus a control group. Participants rated unfamiliar faces on trustworthiness after an affective priming paradigm that exposed

them to negative, neutral, or positive information. Results confirmed the untrustworthiness bias of the BPD sample. BPD participants made more untrusting appraisals regardless of the prime condition, and they were more influenced by negative primes relative to the control group. However, more central to our concerns, both effects were not mediated by RS.

It may be useful to focus on the differences in the two procedures used by Miano and colleagues (2013) and by Masland and Hooley (2017) to clarify discrepancies between their findings. First, Masland and Hooley (2017) used a priming procedure to introduce a trust appraisal context, whereas Miano and colleagues (2013) did not. The influence of contextual factors on the expression of the untrustworthiness bias in BPD (see also Miano et al., 2017) suggests that trustworthiness is not a stable feature in BPD. The context might also affect the mediating role of RS. Second, Miano and colleagues (2013) assessed trust appraisal together with 16 other dimensions, whereas Masland and Hooley (2017) only measured trust. Considering that Miano and colleagues (2013) pointed out that, among the 16 dimensions, some were valenced and related to BPD-specific views of others and self, trust ratings might have also been influenced by the context set by the other dimensions of evaluation (i.e., halo effect). Finally, it is not clear whether there are differences in the stimuli each study used. Miano and colleagues (2013) used 12 Caucasian male and female neutral faces in equal proportion, whereas Masland and Hooley (2017) used 50 male and female neutral faces without indication about the race. All these differences in the procedure confer generalizability of the untrustworthiness bias in BPD. Still, at the same time, they could explain the mixed results concerning the mediating role of RS.

Moreover, it should be noted that even in the presence of a mediating role of RS (Miano et al., 2013), the specific contribution of the emotional and cognitive components of RS has not been investigated. RS is commonly measured through the Rejection Sensitivity Questionnaire (RSQ; Downey & Feldman, 1996). The standard way of scoring the RSQ responses is to consider the anxious and angry rejection expectation as the sum of the cross products between anxiety/anger and expectation responses. Such a scoring method only considers the interaction between rejection expectation and emotional activation and does not consider the distinctive role of the two components separately. Recent work has challenged this scoring method. For example, Zimmer-Gembeck and colleagues (2013) showed that expectations and anxiety were associated with withdrawal responses in a young adult sample, whereas anger was not.

Moreover, when considered separately, the cognitive and emotional components of RS play different roles for predicting interpersonal problems among adolescents (Preti et al., 2018), whereas their interaction had no predictive validity. Furthermore, both studies (Miano et al., 2013; Masland & Hooley, 2016) used only one emotional component: anxiety. The original model of RS considers both anxiety and anger as anticipatory defensive effects accompanying rejection expectations and posits that specific behavioral responses can be expected according to the specific emotion activated (London et al., 2007). Given the crucial role of anger in BPD (Trull, 1995; Zanarini et al., 2005) and the inverse relation between negative emotion such as anger and trust (e.g., Dunn & Schweitzer, 2005), it would be interesting to investigate the specific role of the angry component of RS in relation to BPD.

The present study aims to clarify the potential mediating role of RS on the untrustworthiness bias, according to which high BPD shows lower trust. To do so, we examine whether BPD traits are connected to an untrustworthiness bias toward neutral male and female faces in a non-clinical sample of young female adults. As in Miano and colleagues (2013) study, we considered BPD traits a continuous variable and not as a basis to create two groups as in Masland and Hooley's (2017) study. Moreover, to disentangle the potential role of the components of RS, we measure both anxiety and anger as emotional components of RS, and we compute three single scores, namely Rejection expectation, Anxiety for rejection, and Anger for rejection. We thus aim at testing different mediation models in which the distinctive emotional and cognitive components of RS mediate the association between BPD traits and trust ratings of neutral faces. We hypothesize that previous mixed findings might be due to RS having been considered and measured without considering its cognitive and emotional components separately. We thus anticipate that RS might mediate the relation between BPD traits and trust appraisal only considering one of its components (cognitive or emotional).

2. Methods

Participants and Procedure

A hundred and twenty-five undergraduate women (*M* age =22.13, *SD* = 2.69) took part in a one-session study (approximately 45 minutes). Participants rated the trustworthiness of neutral face stimuli and completed the Adult-Rejection Sensitivity Questionnaire (RSQ, Ayduk et al., 2003; Downey & Feldman, 1996) and the Borderline Personality Disorder Checklist (BPDCL, Giesen-Bloo et al., 2006). At the end of the study, participants were thanked, debriefed, and received course credit for

their participation, if applicable. The study has received approval from the Università degli Studi di Milano Bicocca ethics committee.

Materials

Trust evaluations. The participants rated on a 7-point Likert scale the trustworthiness of each of 48 black and white photographs of unfamiliar Caucasian faces presented in random order. We used 16 different identities (8 male and 8 female) selected from a pretest. The pretest consisted of 30 undergraduate female students (*M* Age = 23.2, *SD* = 2.2) selecting the stimulus they detected as "mildly happy" along a continuum of 21 pictures created through morphing from neutral expression to happy expression for 18 different identities (9 female, 9 male) from the NimStim database. For each identity, we used the average frame chosen and selected 16 identities. In addition to the original stimulus with the straightforward look, we created a left gaze and a right gaze version of each identity by moving the pupils on the right and on the left using Photoshop[3]. We computed three different scores considering all faces, only male faces, and only female faces, respectively.

Borderline Personality Disorder Checklist (BPDCL, Giesen-Bloo, Arntz, & Schouten, 2006). This self-report questionnaire comprises 47 items that assess the current severity of specific BPD manifestations during the last month on 9 dimensions (abandonment, relationships, identity disturbance, impulsivity, mutilation, affective instability, anger, dissociation, and emptiness). Using 5-point Likert scales, ranging

[3] We developed averted gaze faces for using them in one of the following tasks. We chose to analyse all faces independently from the gaze direction because the faces were presented in a random order and thus faces with direct gaze were not systematically the first one presented.

from ‘not at all’ to ‘extremely’, participants indicated how much they were troubled by the 47 different BPD complaints during the last month. We compute a total BPD score (α = .95).

Adult-Rejection Sensitivity Questionnaire (RSQ, Ayduk et al., 2003; Downey & Feldman, 1996). The questionnaire consists of 9 situations in which participants are asked to imagine making a request to a significant other. Participants indicated whether they would be concerned or anxious about the response to their request on a 6-point Likert scale ranging from “not concerned” to “very concerned” and whether they would expect the other person to honor or reject the request on a 6-point Likert scale ranging from “very unlikely” to “very likely”. Besides, following London et al.’s (2007) suggestion, participants indicated whether they would be angry about the response to their request on a 6-point Likert scale ranging from “not angry at all” to “very angry”. We computed the scores for the three scales separately 4 : Anxiety for rejection, Anger for rejection, and Expectation of rejection (α = .85, α = .90, and α = .73, respectively).

3. Results

We discarded the data from 15 participants because of the random pattern of responses to the questionnaire and the other tasks. The final sample consisted of 110 undergraduate women (*M* age = 22.21, *SD* = 2.75). Three participants (2.73%) were not Italian citizens. The BPD traits distribution was representative (*M* = 82.43, *SD* = 24.02, range = 47-155) (see Table 1 for the descriptive statistics of all measures).

Correlations

Table 1 reports the correlations between all constructs. First, there was a significant correlation between BPD traits and overall trust ratings and trust ratings

toward male faces but not toward female faces. The higher the participants were on BPD traits, the less trustworthy they judged the faces (untrustworthiness bias). Second, the cognitive component of rejection (i.e., expectation) did not correlate with BPD traits or the three trust ratings. On the contrary, the emotional component of rejection (i.e., anxiety and anger for rejection) correlated positively with BPD traits indicating that the higher the participants were on BPD traits, the more they would be angry and anxious at the perspective of being rejected. Moreover, the correlations between the emotional component of rejection (i.e., anxiety and anger for rejection) and trust ratings were all significant, except for anger for rejection and trust ratings toward female faces. In general, the more angry and anxious participants would be at the perspective of being rejected, the lower they rated the faces as trustworthy.

Table 1

Descriptive statistics and Correlations.

	M	*DS*	1	2	3	4	5	6	7
1. BPD traits	82.36	23.97	1						
2. Expectation of Rejection	2.20	0.58	.12	1					
3. Anxiety for Rejection	3.85	1.01	.36***	.04	1				
4. Anger for Rejection	2.78	1.11	.47***	.04	.50***	1			
5. Trust Ratings (Total)	3.66	0.88	-.21*	-.07	-.28**	-.23*	1		
6. Trust Ratings (Female)	3.94	0.95	-.18	-.07	-.26**	-.17	.96***	1	
7. Trust Ratings (Male)	3.37	0.88	-.23*	-.07	-.27**	-.27**	.95***	.82***	1

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

Mediations

We hypothesized that the significant effects of BPD traits on trustworthiness ratings could be mediated by the emotional and cognitive components of rejection sensitivity. Considering the lack of significant correlations with any of the constructs with the cognitive component, we can already rule out its mediating role. Moreover, given the non-significant correlations between the trust ratings toward female faces and BPD traits, we only investigated the mediation effects of the relation between BPD traits and overall trust ratings as well as trust ratings toward male faces.

We ran a series of mediation analyses using PROCESS for SPSS (Hayes, 2013, Model #4) to test the significance of indirect or mediated effect with the bootstrap method considering the two emotional components of rejection separately. For overall trust ratings, anger for rejection appeared not to be a mediator of the relation between BPD traits and overall trust ratings because the effect of anger for rejection on overall trust ratings was not significant anymore when controlling for the effect of BPD traits. Anxiety was a significant mediator with a significant indirect effect, *B* = -.08, *SE* = .04, 95% CI: [-.18, -.02] and a non significant direct effect indicating a full mediation. In other words, Rejection Anxiety fully mediated the effect of BPD traits on trust ratings (see Figure 1, panel A). For the trust ratings toward male faces, both Anxiety and Anger for rejections significantly and fully mediated the effect of BPD traits, *B* = -.08, *SE* = .04, 95% CI: [-.16, -.01] and *B* = -.10, *SE* = .04, 95% CI: [-.20, -.03], respectively. Anxiety and Anger mediated to a similar extent the effect of BPD traits on trust ratings toward male faces (see Figure 1, panels B & C). To sum up, there are three mediations of RS for the untrustworthiness bias: Anxiety for rejection on overall trust ratings and toward only male faces, whereas Anger for rejection on trust ratings toward only males. One should note that none of the mediations was significant when considering the standard product score of RS.

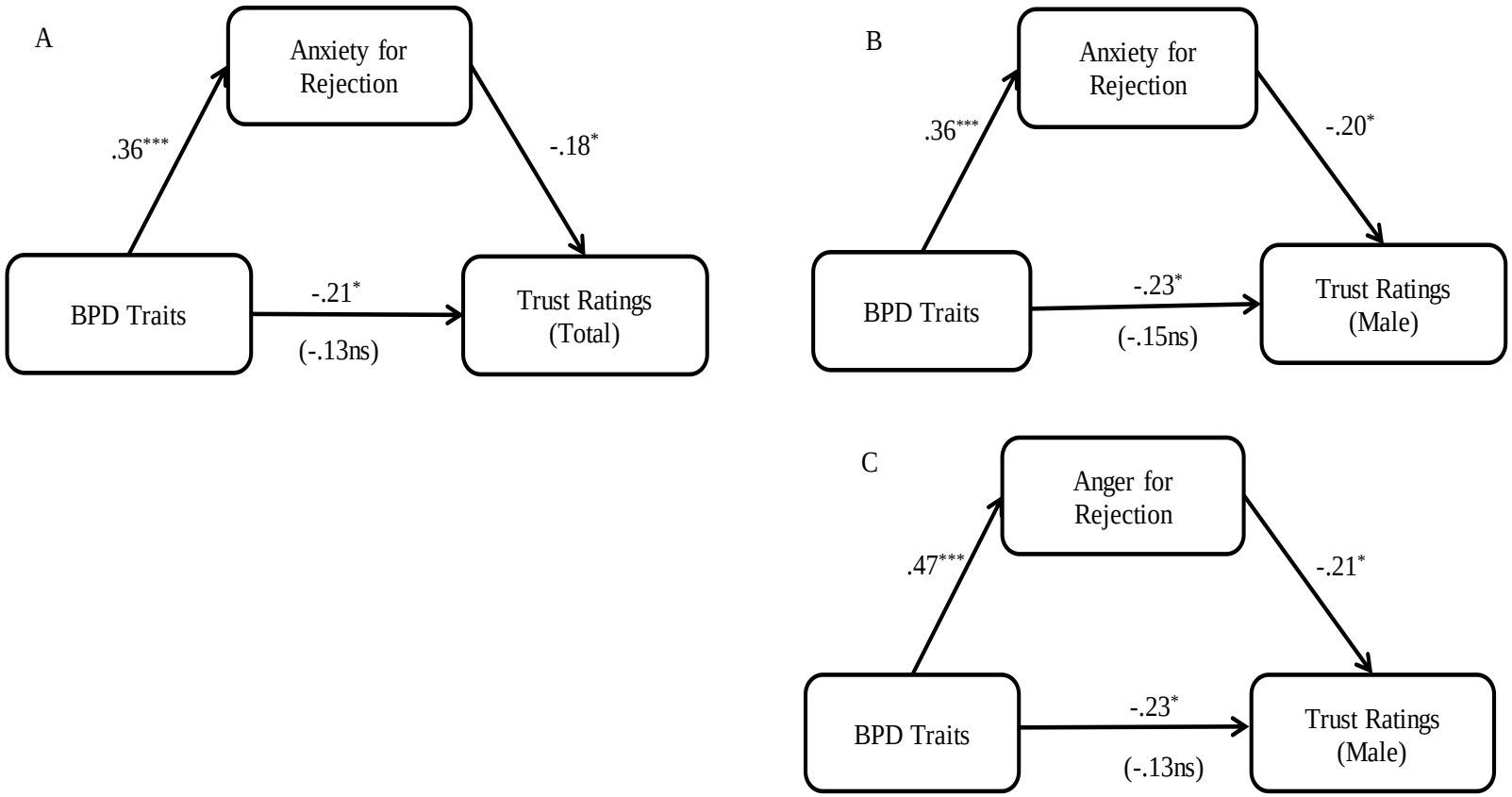


Figure 1. Mediations by Anger and Anxiety for Rejection of the relations between BPD traits and trust ratings.

3. Discussion

Starting from the discrepancies between two previous research reports (i.e., Miano et al., 2013; Masland & Hooley, 2017), this study aimed to clarify whether rejection sensitivity mediates the relation between BPD traits and trust appraisal toward faces. Moreover, based on previous evidence demonstrating the usefulness of considering the cognitive and emotional components of RS separately (e.g., Preti et al., 2018; Zimmer-Gembeck & Nesdale, 2013) and of including anger as an emotional component (e.g., London et al., 2007), we tested the potential mediating role of all three components of RS. Our main results are manifold. First, BPD traits are associated with a decrease in trust evaluation of neutral faces but not when considering only female faces. Second, only the emotional components of RS (i.e., anxiety and anger) and not the cognitive component (i.e., expectation) proved to be significant mediators.

The first result provides additional evidence to a demonstrated association between BPD traits and a general untrustworthiness bias (e.g., Fertuck et al., 2013; Masland & Hooley, 2017; Miano et al., 2013) confirming this social-cognitive impairment in BPD. This impairment is further demonstrated in impaired cooperative behavior when, for example, BPD patients are engaged in economic games (e.g., King-Casas et al., 2008). Furthermore, considering therapy with BPD patients, a lack of trust toward the other (i.e., the therapist) that takes the form of intense "paranoid transference" is common, especially in the early phases of treatment (Yeomans et al., 2015). From a theoretical point of view, recent theories on the development of BPD have pointed at the ability to trust others as a major issue (Fonagy & Allison, 2014). Moreover, the object relation approach to personality pathology underlines the

presence of a polarized negative representation of others as untrustworthy (Yeomans, Clarkin, & Kernberg, 2015). However, our results showed a non-significant association between BPD traits and trust ratings toward female faces. Overall, trust ratings of male stimuli (M = 3.35, SD = .90) were significantly lower than the female ones (M = 3.92, SD = .98), $t(1,110) = 11.00$, $p < .001$, effect probably due to the exclusively female sample. Our results could thus point to a particular form of mistrust in interpersonal relationships, namely the evaluation of the trustworthiness of opposite-sex individuals. One could hypothesize that such a process is particularly relevant in the romantic relationship dysfunctions that are specifically associated with BPD (Hill et al., 2008; Miano, Grosselli, et al., 2017). However, it may not be the case for a male sample for whom females are usually perceived as less dominant and thus more trustworthy, suggesting an asymmetrical relationship in trustworthiness (Buchan et al., 2008). However, future research should test this interpretation to exclude alternative explanations to our results, such as the ceiling effect in trust ratings toward female targets or insufficient power for a small effect.

According to our results, the expectation of rejection did not act as a mediator. From our results, it seems that RS is connected to BPD traits exclusively through the emotional activation (both anxious and angry) that the idea of being rejected elicits. Even though previous literature documented an association between BPD traits and RS (e.g., Staebler et al., 2011), to our knowledge, this is the first study that demonstrates that the cognitive component of RS does not play a role in this association. On the contrary, both the emotional components, anxiety and anger for rejection, mediated fully such association. In other words, our results suggest that the process through which BPD is connected with an untrustworthiness bias has to do with

a robust emotional activation when facing situations that could imply social rejection. Such a mechanism points to the prevalence of strong, polarized emotional activations (over cognitions) in explaining the clinical manifestations of BPD and is in line with an object relations conceptualization of the disorder. According to such a theoretical framework, in fact, split and polarized representations of self and others, imbued with extreme negative affectivity, maintain the pathological personality structure characteristic of BPD that interferes with healthy interpersonal relationships (Yeomans, Clarkin, & Kernberg, 2015).

Our results are not easily comparable with the ones from the two previous studies (i.e., Masland & Hooley, 2017; Miano et al., 2013) because they used a composite score of rejection sensitivity, that is the product between the cognitive component of the construct (i.e., the expectation of rejection) and the emotional component (i.e., anxiety) and did not assess anger for rejection. However, without this alternative approach, we could not have disentangled the cognitive and emotional components of RS, especially considering that using the composite scores, none of the mediation models were significant. Future research should systematically test for mediation effect using both approaches. It would help to clarify the mediating role of rejection sensitivity and its different components. Because our sample was only composed of women and because we did not obtain any untrustworthiness bias for female faces, future research should investigate possible gender effects and specific biases depending on the congruence between the perceiver and the target (same vs. opposite-sex).

In conclusion, our results support the idea of an untrustworthiness bias related to BPD. Moreover, we confirmed the role of RS in such bias but more important; only the emotional activation related to possible rejection is involved in such an interplay.

CHAPTER 3. BIASED APPRAISAL OF TRUST IN BORDERLINE PERSONALITY DISORDER:

THE ROLE OF SENSITIVITY TO JUSTICE

Overview

While in chapter 2, we examined the role of Rejection Sensitivity on the BPD features-trust appraisal link, here we focus on a different individual disposition: Justice Sensitivity, continuing to explore the stage of Prior beliefs and Disposition of our model.

This chapter develops within the framework of the Sensitivity-to-mean-intentions Model (SeMi; Gollwitzer & Rothmund, 2009). According to SeMI, individuals afraid of being exploited in social situations develop a suspicious mindset and, in turn, behave uncooperatively and show mistrust of others. Justice Sensitivity is a construct that reflects the individual's concern for justice. Schmitt and colleagues (2010) developed a JS measure that distinguishes individuals' justice-related sensitivity depending on the individuals' perspective on unjust acts: victim, perpetrator, beneficiary, or observer. Interestingly, RS and JS are two theoretically overlapping constructs: both tap individuals' sensitivity to potentially problematic interpersonal situations. For this reason, and in light of the findings in chapter 2, we expected JS to weigh on the BPD features-trust appraisal link likewise RS. Furthermore, moving from the opposite gender effect found in chapter 2, we questioned whether gender could weigh on the BPD traits and biased trust appraisal link. Besides our previous findings (Richetin et al., 2018), Buchan, Croson, and Solnick (2008) found significant differences in the tendency to trust across genders.

This chapter presents two empirical studies. The first study explores the potential role of JS on the link between BPD traits and biased trust appraisal. More precisely, drawing upon the SeMi model, we expected JS to mediate the effect of BPD features on trust ratings. To this aim, one hundred and-eighty-one students completed an assessment of Borderline Features, Justice Sensitivity, Rejection Sensitivity, and appraised trust of 16 neutral facial stimuli (eight male and eight female). Results suggest that JS from the beneficiary's perspective mediates the BPD features-trust appraisal link. JS from the beneficiary's perspective suppressed the negative effect of BPD features on the trust appraisal of male faces.

The second study formally tests the potential role of gender on the effect of BPD features on trust appraisal. To this purpose, we collected a second sample (N = 408; 180 males) and asked participants to fill the assessment of Borderline Features, of Justice Sensitivity, and appraise trust of 16 neutral facial stimuli (eight male and eight female). Results show no significant effect of gender on the BPD features-trust appraisal link. In the discussion, we comment on the main empirical and clinical implications of these findings.

1. Introduction

Borderline Personality Disorder (BPD) is a severe and chronic psychiatric condition. Individuals with BPD show chronic instability in multiple areas such as emotional dysregulation, self-harm, impulsivity, and identity disturbances (American Psychiatric Association, 2013). Individuals with BPD often struggle with interpersonal dysfunctions (for a review, see Lazarus et al., 2014), and such impairments often persist even beyond symptom remission (Gunderson et al., 2011). Due to the relevance and endurance of dysfunctions in interpersonal functioning for individuals with BPD, there is now a general trend in the empirical investigation of the processes that may support and trigger interpersonal dysfunctions among BPD individuals.

Generalized mistrust of others is one of such processes. It culminates in the incapability to maintain healthy and satisfying relationships based upon mutual trust (for a general overview, see Poggi et al., 2019). Whereas trust refers to a general assumption about others' good nature, mistrust means negative expectations regarding others' behaviors, with the ensuing tendency to judge others as untrustworthy and results in untrusting behaviors (Evans & Revelle, 2008). Previous investigations found a link between BPD traits and generalized mistrust of others (Fertuck et al., 2013; Miano et al., 2013; Nicol et al., 2013). More precisely, individuals with BPD show a higher attribution of and sensitivity to others' untrustworthiness (i.e., untrustworthiness bias, Fertuck et al., 2013). Individuals with high BPD features show lower levels of trust than individuals with low BPD features, even when asked to rate neutral faces (Miano, Fertuck, et al., 2017; Nicol et al., 2013; Richetin et al., 2018). Such untrustworthiness towards others impacts BPD individuals' behaviors too. For example, in a trust game procedure, BPD individuals, compared to controls, are more

likely to initiate cooperation ruptures and sustain lower rates of generous gestures to repair such cooperation ruptures (King-Casas et al., 2008). Since BPD patients may show impairments in interpersonal functioning partly because of their untrustworthiness bias, it is significant to understand further the subtending processes (Poggi, et al., 2019).

Researchers have recently attempted to identify potential mediators and moderators of the tendency for BPD individuals to mistrust others. For example, Rejection Sensitivity is the disposition to anxiously or angrily expect, readily perceive, and strongly react to the possibility of interpersonal rejection (RS, Downey & Feldman, 1996). Empirical evidence shows that only one emotional component of RS, anxiety for rejection, mediates the link between BPD traits and lower trust appraisal (Miano et al., 2013; Richetin et al., 2018).

RS has a theoretical overlap with other constructs such as Justice Sensitivity (JS; Schmitt et al., 1995), i.e., individuals' concern for justice. Thus, JS could be another potential mediator of the association between BPD traits and biased appraisal of trust. High JS predicts justice-related emotions and behaviors. However, the individual experience of injustice depends much upon the standpoint from which the unjust act is experienced, as the victim, perpetrator, beneficiary, or observer (Schmitt et al., 2010). JS-Victim indicates the sensitivity to being mistreated by others, JS-Observer the sensitivity to observing unjust events, JS-Beneficiary the sensitivity to profiting from unfair events, JS-Perpetrator the sensitivity to imposing injustice upon a victim. Thus JS refers to the emotional and behavioral reactions that may arise from a potentially problematic interpersonal situation viewed from different perspectives.

In this sense, JS differs from RS because the latter refers to the emotional and cognitive consequences of potential rejection from the victim's sole perspective.

Žitný & Halama (2011) found a positive association between JS and neuroticism and antagonism personality traits. Little is known, however, about how JS may weigh on personality pathology. To our knowledge, only one study demonstrated that BPD individuals are prone to develop cognitive, emotional, and behavioral reactions to experiences of injustice and suggested that increased JS may be of particular importance for the development of interpersonal dysfunctions among BPD individuals (Lis et al., 2018). More precisely, Lis and colleagues (2018) investigated the relation between BPD features and JS, finding evidence of a positive correlation between victim and observer JS and BPD features. Furthermore, the authors showed that JS-Victim partially mediated the link between BPD features and the frequency of aggressive behavior, one of the significant behavioral manifestations of interpersonal problems among BPD individuals.

To explain the processes that translate JS attitudes into detrimental interpersonal effects such as uncooperative behavior and low trust of others, Gollwitzer and Rothmund elaborated the Sensitivity-to-mean-intentions Model (SeMi; Gollwitzer et al., 2009). According to SeMI, a suspicious mindset suggesting the intentional meanness of others characterizes victim-sensitive persons. Hostile interpretations and expectancies of others' injustice are the core components of the suspicious mindset and make highly victim sensitive individuals afraid of being exploited (Gollwitzer et al., 2013; Gollwitzer & Rothmund, 2011). Such a theoretical assumption was confirmed by Maltese an colleagues (2016), who found that high victim sensitivity was associated with expectancies of injustice in ambiguous

situations. Furthermore, the authors found that these expectancies mediated the relationship between victim sensitivity and cooperation behavior in a trust game. To prevent being exploited in social situations, people with such a suspicious mindset are very likely to mistrust others and act as uncooperative and antisocial (Gollwitzer et al., 2009; Gollwitzer & Rothmund, 2011).

BPD individuals showed a similar tendency in suspiciousness, mistrust, and uncooperative behaviors (Poggi et al., 2019). Moreover, BPD individuals tend to base their interpretations of social interactions upon their own suspicious and hostile expectations rather than actual objective situational events (Colle et al., 2019). Several studies support the evidence that in BPD individuals, this tendency is connected with high RS. BPD patients, compared to controls, report stronger beliefs that others will abandon and reject them (Arntz et al., 1999; Arntz, et al., 2004; Ayduk et al., 2003, 2008). Such concerns about others' meanness may turn into impairments in cooperative behavior and trust. BPD individuals' inclination to rely upon their own suspicious and hostile expectations for understanding social contexts presents similarities with the SeMi model's suggestion for high JS individuals and empirical research on high RS individuals. Such inclination towards a suspicious mindset could lead to the perception of others' untrustworthiness or uncooperative behaviors.

2. Main Aims

Drawing upon SeMi, we expect that JS dispositions would weigh on BPD individuals' propensity to lower appraisal of others' trustworthiness. In a first study, we aimed to test such a hypothesis and determine whether JS would mediate the link between BPD traits and biased trust appraisal. Additionally, we designed a second study to explore the role of gender on the association between BPD traits and biased

trust appraisal. Previous studies in the economic literature show men's tendency to trust more than women and different behavioral styles across gender (Buchan et al., 2008). Moreover, relationships are more core dimensions of self-construal for women than men (Cross et al., 2000; Cross & Madson, 1997). Regarding the link between BPD traits and biased trust appraisal, in a previous contribution, Richetin and colleagues (2018) found an opposite-gender effect on a female sample. More precisely, BPD features correlated positively with trust appraisal of male faces and not female faces in a female sample. To our knowledge, because there is no published contribution systematically testing it, in the second study, we aimed to explore the role of gender as a moderator of the association between BPD traits and trust appraisal.

3. Study 1

In this first study, we tested four main hypotheses. First, we expected to replicate the negative association between BPD traits and trust appraisal of neutral faces (Miano et al., 2016; Nicol et al., 2013; Richetin et al., 2018). Second, based on previous findings, we expected to find BPD traits positively associated with victim and observer JS (Lis et al., 2018). Third, we expected to replicate previous reports of a positive correlation between RS and BPD features (e.g., Staebler et al., 2011). Finally, JS and RS are similar conceptualizations. Both JS and RS describe sensitivities to negative social cues affecting an individual's cognitions, emotions, and behavioral responses during social encounters. Because RS has been shown to mediate the relation between BPD trait and trust appraisal (Miano et al., 2013; Richetin et al., 2018), we thus expected JS to mediate the predicted negative relation between BPD features and trust appraisal in neutral faces.

The first, second, and third hypotheses will be tested by examining correlations among constructs. To test the fourth hypothesis, we will run a mediation model with the PROCESS macro in SPSS (Hayes, 2013)to inspect the total, direct, and indirect effects.

3.1.Methods

Sample determination

We conducted a power analysis via G*Power, 3.1.6 (Erdfelder et al., 2009) to determine the minimum sample size. To detect a significant mediation effect among continuous variables ($f^2 = 0.8$) with a power of 0.95 and alpha error probability at 0.05, 121 participants would be sufficient. To account for missing data or participants' inattentiveness, we decided to aim for a sample of 180. We pre-registered our procedure, hypothesis, and analysis plan on the Open Science Framework portal (OSF Registration DOI: 10.17605/OSF.IO/AXZ36).

Participants and Procedure.

One hundred and-eighty-one students (*M* age = 21.73, *SD* = 2.86; Male = 38, Female = 143) were recruited on a voluntary basis. Participants took part in a single laboratory session and received, when applicable, credit courses. All participants had sufficient cognitive and language capabilities to give their written informed consent to participate. After consenting to the study, participants filled in the Personality Assessment Inventory Borderline section (PAI-BOR, Morey, 2004), the Justice Sensitivity Inventory (JSI, Schmitt et al., 2010), and the Adult Rejection Sensitivity Questionnaire (RSQ, Downey & Feldman, 1996). Then, participants completed a trust appraisal task. The study received approval from the university ethics committee.

3.2.Materials

Personality Assessment Inventory-Borderline (PAI-BOR) (Morey, 2004). The borderline section (PAI-BOR) of the Personal Assessment Inventory It consists of a 24 items self-report scale derived from the Personality Assessment Inventory (Morey, 2004). The scale assesses specific BPD features: affective instability (i.e., "My mood can shift quite suddenly"); identity diffusion (i.e., "My attitude about myself changes a lot"); negative relationships (i.e., "My relationships have been stormy"); and self-harm (i.e., "When I'm upset, I typically do something to hurt myself"). Participants selected the response that best pertained to them on 4-point scales (false/not at all true, slightly true, mainly true, very true). The PAI-BOR has shown strong psychometric properties across different samples (Jackson & Trull, 2001). The measure reached acceptable internal consistency (Cronbach's $\alpha = .69$).

Justice Sensitivity Inventory (JSI) (Schmitt et al., 2010). It assesses four perspectives of justice sensitivity through 10 items each: Victim sensitivity (the readiness to react in situations where the participant is the victim of injustice), perpetrator sensitivity (the readiness to react to situations where the participant actively victimize someone else), observer sensitivity (the readiness to react to situations where the participant is a neutral bystander of injustice), and beneficiary sensitivity (the readiness to react in situations where the participant passively benefit from injustice against someone else). The four scales showed good internal consistency (JS Victim $\alpha = 0.83$; JS Perpetrator $\alpha = 0.88$; JS Observer $\alpha = 0.85$, and JS Beneficiary $\alpha = 0.88$).

Adult-Rejection Sensitivity Questionnaire (Downey & Feldman, 1996; London et al., 2007). For a detailed description of the questionnaire, see the second paragraph of the second chapter.

Trust Appraisal Task. Participants rated on 7-point scales the trustworthiness of each of 16 black and white photographs of unfamiliar Caucasian faces presented in random order from 1 (not all trustworthy) to 7 (very trustworthy). We used 16 different identities (8 male and 8 female) previously pretested to be neutral regarding trustworthiness and attractiveness[4].

3.3.Results

As already mentioned, we tested four main hypotheses. Correlations are reported in Table 1. First, we expected to replicate a negative association between BPD traits and trust appraisal of neutral faces. Surprisingly, BPD features did not show any association with the trust appraisal of neutral faces. BPD features negatively correlated with the appraisal of male faces but not of female faces. Second, according to our expectations, BPD features correlated positively with JS from the Victim, Observer, and Beneficiary point of view but not with JS from a Perpetrator perspective. Third, in line with our hypothesis, BPD features positively correlated with both emotional components of RS (i.e., anger and anxiety for rejection) but not with the cognitive component (i.e., the expectation of rejection).

---

[4] We set two separate pretest for the selection of neutral stimuli on both trustworthiness and attractiveness dimension. The first pretest consisted of 30 undergraduate female students (*M* Age = 23.2, *SD* = 2.2) choosing the stimulus they detected as "mildly happy" along a continuum of 21 pictures created through morphing from neutral expression to happy expression for 18 different identities (9 female, 9 male) from the NimStim database. The second pretest consisted of 17 undergraduate female students (M Age = 24.4, SD = 2.4) rating the average frame chosen for each identity in the previous test on the two dimensions of trustworthiness and attractiveness. From these two pretests, we selected a pool of 16 stimuli (8 female, 8 male) that were not statistically different from the average scale score on the trustworthiness and attractiveness dimensions.

Table 1.

Descriptive statistics and correlations. Study 1, N = 181

	M	*SD*	1	2	3	4	5	6	7	8	9	10	11
1. Justice Sensitivity_Victim	3.48	0.78	1										
2. Justice Sensitivity _Observer	3.21	0.82	.44**	1									
3. Justice Sensitivity _Perpetrator	3.69	0.87	.04	.48**	1								
4. Justice Sensitivity _Beneficiary	2.90	0.96	.17*	.53**	.75**	1							
5. Rejection Sensitivity_Expectation	2.31	0.64	-.12	-.09	-.04	.06	1						
6. Rejection Sensitivity _Anxiety	3.88	0.92	.42**	.30**	.18*	.28**	.05	1					
7. Rejection Sensitivity _Anger	2.74	0.93	.48**	.25**	-.02	.10	-.11	.44**	1				
8. BPD features	55.06	10.39	.50**	.35**	.09	.23**	.12	.36**	.39**	1			
9. Trust Appraisal_Male	3.68	0.94	-.19*	.07	.07	.19*	-.06	-.05	-.08	-.19*	1		
10. Trust Appraisal_Female	4.47	0.92	-.07	.02	.09	.16*	-.10	.04	.01	-.07	.69**	1	
11. Trust Appraisal_Total	4.07	0.86	-.14	.05	.08	.19*	-.09	-.01	-.04	-.14	.92**	.92**	1

Note. $^{*}p < .05$. $^{**}p < .01$.

Finally, we expected JS to mediate the BPD-trust appraisal link. BPD features correlated only with trust appraisal of male faces, and only JS victim and beneficiary correlated with the trust appraisal of male faces.We thus met the preconditions to test the JS victim and beneficiary mediating roles in the association between BPD features and trust appraisal of male faces. We used the PROCESS macro in SPSS (Hayes, 2013). The model for appraisal of trust in male faces, including BPD features as predictor and JS Victim as mediator, did not reveal a significant mediation effect (B = -.14, SE = .10, p = .15, 95% CI [-.35, .05]). The model including BPD features as predictor and JS Beneficiary as mediator was significant, $F(2,177) = 9.14$, $p < .001$, $R^2 = 0.09$, with a significant indirect effect, $B = .24$, $SE = .07$, $p < .001$, 95% CI [.09, .38]. The direct effect of BPD features on trust appraisal of neutral male faces was significant, $B = -.02$, $SE = .01$, $p<.001$, 95% CI [-.04, -.01], indicating a partial mediation. Because the direct effect was negative while the indirect effect was positive, the results indicated a suppressor effect of JS Beneficiary. In other words, JS from the beneficiary perspective decreased the negative effect of BPD features on the trust appraisal of male faces (see Figure 1).

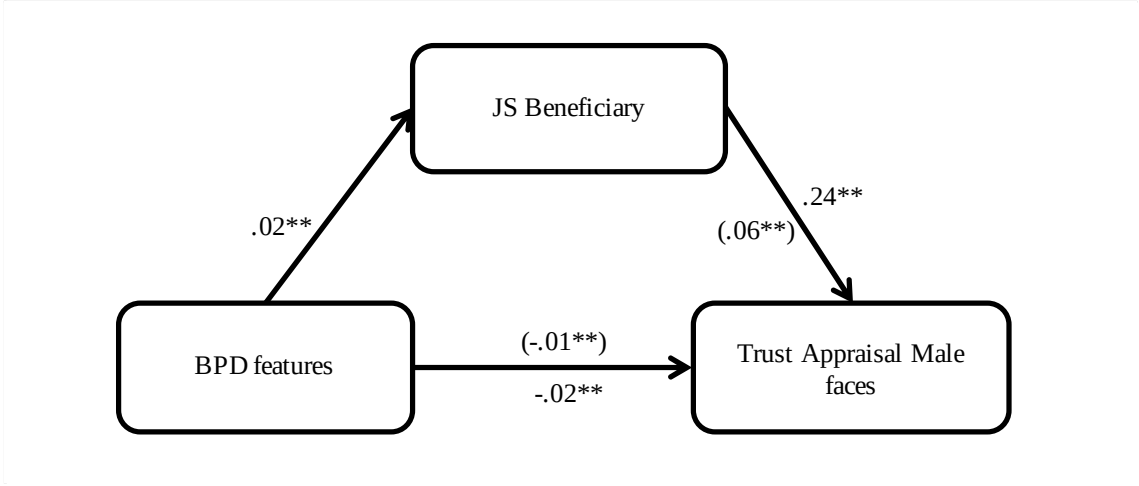


Figure 1. Partial mediation in the total sample, Study 1

Exploratory Analyses

BPD features negatively correlated with trust appraisal of male faces in a sample predominantly composed of female participants (80%). Given our sample asymmetrical gender distribution, we cannot rule out an effect of participants' gender on trust appraisal performances. For this reason, we verified whether a BPD-trust appraisal link occurred in the solely female sample. We thus ran a series of exploratory analyses only on the female sample (N = 143; M age = 21.51, SD = 2.53), considering that the male sample would be too small (N = 38) to obtain reliable results. A sensitivity analysis revealed that the sample would allow us to detect an effect size equivalent to 0.16.

Like in the whole sample, negative correlations among trust appraisal of male faces, BPD features, JS Victim, and JS Beneficiary were present in the female sample (see Table 2). Unlike in the whole sample, JS Observer correlated negatively with trust appraisal of male faces. To probe the distinct and shared contributions of JS from different perspectives we run a parallel mediation model (see Figure 2). The full model was significant, F (4,138) = 6.07, p = .001, R^2 = 0.15. The indirect effects of BPD

features on appraisal of trust in male faces through JS Victim, *B* = -.10, *SE* = .05, 95% CI [-.19, -.01], and JS Observer, *B* = .09, *SE* = .04, 95% CI [.03, .18] were significant whereas the indirect effect of JS Beneficiary was not (*B* = .03, *SE* = .03, 95% CI [-.01, .10]). The direct effect of BPD features was significant (*B* = -.20, *SE* = .09, *p* = .03, 95% CI [-.38, -.02]), suggesting a partial mediation model.

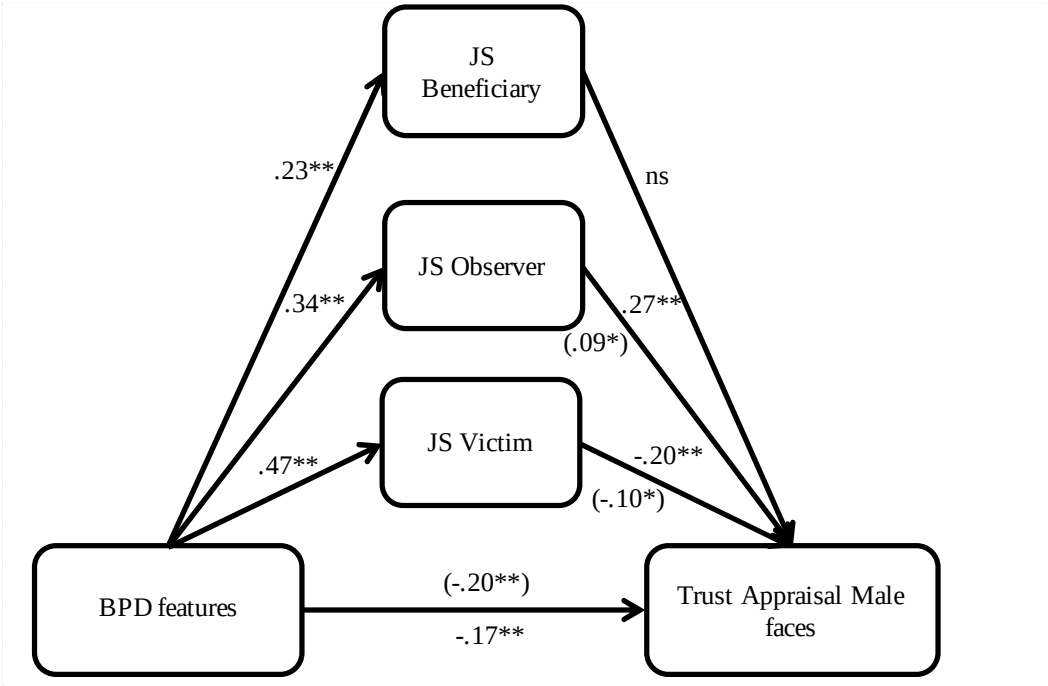


Figure 2. Parallel mediation in the female sample, Study 1

Table 2.

Descriptive statistics and correlations in the female sample. Study 1, N = 143

	M	*SD*	1	2	3	4	5	6	7	8	9	10	11
1. Justice Sensitivity _Victim	3.54	.79	1										
2. Justice Sensitivity _Observer	3.30	.78	.38**	1									
3. Justice Sensitivity _Perpetrator	3.82	.80	-.06	.39**	1								
4. Justice Sensitivity _Beneficiary	2.98	.95	.14	.51**	.75**	1							
5. Rejection Sensitivity _Expectation	2.26	.64	-.07	.02	.11	.13	1						
6. Rejection Sensitivity _Anxiety	3.94	.93	.43**	.29**	.12	.26**	.09	1					
7. Rejection Sensitivity _Anger	2.76	.96	.47**	.24**	-.07	.08	-.03	.43**	1				
8. BPD features	55.70	10.91	.47**	.34**	.06	.23**	.15	.37**	.36**	1			
9. Trust Appraisal_Male	3.66	.91	-.17*	.19*	.15	.20*	-.07	.01	-.07	-.17*	1		
10. Trust Appraisal_Female	4.47	.94	-.06	.09	.16*	.16	-.10	.08	-.01	-.05	.67**	1	
11. Trust Appraisal_Total	4.07	.84	-.13	.16	.17*	.20*	-.10	.05	-.04	-.12	.91**	.92**	1

Note. $^{*}p < .05$. $^{**}p < .01$.

3.4.Discussion

The results of this first study are manifold. First, we found a positive association between BPD features and the emotional RS components, consistent with previous empirical results (De Panfilis et al., 2016; Zimmer-Gembeck & Nesdale, 2013). No correlations occurred with the cognitive RS component (i.e., expectation). These results are in line with previous research and support clinical observations: strong and polarized emotional activations occurring in BPD individuals in interactive situations interfere with good quality interpersonal exchanges even more than the cognitive response (Yeomans et al., 2015).

Second, consistent with Lis and colleagues (2018), we confirmed the positive correlation between BPD features and sensitivity to injustice from the victim's and observer's point of view and the lack of correlation with JS from the perpetrator's perspective. We also found an unprecedented significant association between BPD features and JS from the beneficiary's perspective. The difference between the perpetrator scale and the other subscales could partially explain the lack of correlation. While the victim, observer, and beneficiary scales refer to the readiness to react emotionally to others' unjust acts, the perpetrator scale refers to injustice as a cause of one's active wrongdoing (Schmitt et al., 2010). Our findings suggest that victim sensitivity, observer sensitivity, and beneficiary sensitivity to injustice are altered among individuals with high BPD features. Since individual differences in JS from different perspectives partly account for individuals' behaviors in interpersonal relations, such alterations may partially account for BPD's interpersonal dysfunctions (Lis et al., 2018). For this reason, we invite future researchers to explore further the interplay occurring between alterations in justice sensitivity and BPD features on everyday interpersonal functioning.

Against our expectation, we did not find any link between BPD features and trust appraisal of neutral faces considering male and female stimuli together. However, we found a negative association when considering only male faces. Such a result replicates previous empirical findings. Richetin and colleagues (2018) found a negative association between BPD features and the trust appraisal of male faces but not females in a solely female sample. Unlike Richetin and colleagues (2018), our sample comprised a small portion of male participants (20%). Nonetheless, it is likely that, due to the imbalanced gender distribution in our sample, our results mirrored mainly the female responses. Hence, our findings may further support the hypothesis of an opposite gender effect on the trust appraisal of faces among females. Our findings suggest that female participants with high BPD features tend to appraise little trust when asked to judge male faces, indeed. We found further support for such an opposite gender effect in the correlation between BPD features and trust appraisal after selecting the only female sample. In the only female sub-sample, we found a significant negative association between BPD features and trust appraisal of male faces and not with trust appraisal of female faces.

We then tested the mediating role of JS disposition on the negative association between BPD features and trust appraisal of male faces. We found that Beneficiary JS had a suppressor effect. This partial mediation result may be of high relevance for the understanding of interpersonal impairments among BPD individuals. We shed light on a personal disposition that may act as protective mechanisms against BPD individuals' trust impairments. According to our results, the tendency of individuals with high BPD features to appraise less trust in neutral faces is somehow eased by heightened sensitivity to injustice from the beneficiary point of view.

Finally, our exploratory analyses on the female sub-sample showed that JS Victim negatively mediated (i.e., as JS Victim increased, trust appraisal decreased) and JS Observer positively mediated (i.e., as JS Observer increased the trust appraisal of male faces increased) the link between BPD features and trust appraisal of male faces. JS Beneficiary was not a significant mediator. The opposite mediation effects of JS victim and observer could result from the fear of being exploited and concern for justice for the self that victim sensitivity naturally implies, while observer sensitivity does not (Schmitt et al., 2005). At least among women regarding male targets, we suppose that an increase in one's sensitivity to injustice from a victim perspective leads to a decrease in others' trustworthiness perception because of an underlying concern and fear of being exploited by others. Such concerns and fears do not necessarily occur in individuals with a high sensitivity to injustice from an observer perspective.

4. Study 2

Because of previous accounts for gender differences in trust behaviors, in this second study, we aimed at formally testing the role of gender on the link between BPD features and trust appraisal of neutral faces. To test such a hypothesis, we designed a replication of the first study on a gender-balanced sample. Since this study replicates the first, we expected to replicate its main findings. First, we expected to replicate the lack of association between BPD traits and lower trust appraisal of neutral faces. Second, we expected associations between BPD features and JS Victim, Observer, and Beneficiary. Additionally, we tested the effect of participants' gender on the association between BPD and trust appraisal. We pre-registered the procedure, hypotheses, and analyses plan for the second study on the Open Science Framework portal (OSF Registration DOI: 10.17605/OSF.IO/XJFCY).

4.1.Methods

Sample determination

To establish the sample size for a significant moderation model, we computed a sensitivity analysis with G*Power for a moderation model with three predictors (BPD features, gender, and Gender*BPD features), a power of .80, and alpha of .05. According to such calculation, a sample of 400 participants would detect a small effect (f^2= .019 equivalent to β of the interaction = .275).

Participants, Procedure, and Materials.

Four hundred and eight participants (*M* age = 23.90, *SD* = 4.34; Male = 180; Female = 228) were recruited at University Milano-Bicocca on a voluntary basis. The procedure and materials of this second study were identical to those used in Study 1. The only difference between procedures is the omission of the RSQ questionnaire in the second study. Participants took part in a single laboratory session and received, when applicable, credit courses. After consenting to the study, participants filled in the Personality Assessment Inventory Borderline section (PAI-BOR, Morey, 2004), the Justice Sensitivity Inventory (JSI, Schmitt, et al., 2010), and then completed a trust appraisal task. PAI-BOR and JSI showed good internal consistency (PAI-BOR α = .86; JS Victim α = .83; JS Perpetrator α = .86; JS Observer α = .88 and JS Beneficiary α = .88). The university ethics committee approved the procedure.

4.2.Results

As mentioned in the main aim section, we tested different hypotheses.

There was no association between BPD features and trust appraisal of neutral faces (all *p*s > .05, see Table 3). Regarding JS correlations with BPD features, we found a positive association between BPD features and JS from Victim and Observer

perspectives. Different from Study 1, we found no associations between JS Beneficiary and BPD features (see Table 3).

To test the influence of gender on the association between BPD and trust appraisal, we computed separate moderation models introducing BPD features as the independent variable, gender of participants as moderator, and trust appraisal outcomes (only male faces, only female faces, and both male and female faces) as separate dependent variables. None of the models was significant, indicating that participants' gender did not moderate the relations between BPD features and trust appraisal.

Table 3.

Descriptive statistics and correlations. Study 2, N = 408

	M	*SD*	1	2	3	4	5	6	7	8
1. Justice Sensitivity _Victim	3.87	.94	1							
2. Justice Sensitivity _Observer	3.63	.95	.44**	1						
3. Justice Sensitivity _Perpetrator	4.26	1.01	-.04	.39**	1					
4. Justice Sensitivity _Beneficiary	3.21	1.11	.03	.53**	.67**	1				
5. BPD features	56.40	11.30	.29**	.16**	-.08	.06	1			
6. Trust Appraisal_Male	3.64	.84	-.03	.01	.12*	.14**	-.09	1		
7. Trust Appraisal_Female	4.39	.75	13**	.14**	.21**	.12*	-.07	.53**	1	
8. Trust Appraisal_Total	4.01	.70	.05	.08	.19**	.15**	-.09	.89**	.86**	1

Note. $^{*}p < .05$. $^{**}p < .01$.

4.3.Discussion

In the second study, we did not find any association between trust appraisal of neutral faces (male, female, and both male and female) and BPD features. Although these results partly replicate our findings in the first study, they are not in line with previous studies suggesting that a significant decrease in trust appraisal of neutral faces is associated with BPD (Fertuck et al., 2013, 2019; Nicol et al., 2013). One plausible explanation for such inconsistency relies on the different methodologies to assess BPD features used. We used the PAI-BOR questionnaire to evaluate a non-clinical population while previous studies compared clinically diagnosed BPD individuals (assessed with the SCID-II interview; First et al., 1995) and control groups.

Consistent with Study 1 and with Lis et al.'s findings (2018), we found a positive correlation between BPD features and JS from the victim's and observer's perspectives. These results suggest enhanced sensitivity to injustice from Victims and Observers' point of view among individuals with high BPD features. Given that individual differences in JS partly account for dysfunctional behaviors in interpersonal contexts, we invite future researchers to explore the weight of JS predisposition on interpersonal dysfunctions among individuals with BPD features (Lis et al., 2018).

Finally, we expected to find participants' gender to moderate the association between BPD features and trust appraisal. In line with the socialization theory, we expected to find a greater appraisal of trust by women. In our society, women are socialized to cope with stressful situations by seeking support from others. Consequently, women are often more used to benefit from others' help, even if not expected, compared to men (Sigmon et al., 1995). Being more familiar with receiving support, we expected women to appraise more trust compared to men. Our

expectations found no confirmation. This means that the lack of association between BPD features and trust appraisal occurs regardless of respondents' gender.

5. General discussion

The results of these two studies can be interpreted within the framework of the SeMi model and psychoanalytic Object Relations Theories (ORT). The SeMi model delineates the processes that translate JS into uncooperative behavior (Gollwitzer & Rothmund, 2011). ORT model provided clinicians with a general framework and useful tools for identifying, diagnosing, and conceptualizing patients with Personality Disorders (Kernberg & Caligor, 2005). The ORT model conceptualizes Personality Disorders as the outcome of a lack of integration of positive and negative aspects of self (i.e., identity diffusion) and others (Stern et al., 2018). Such integration is essential for establishing a realistic and stable sense of self and others across time and situations. On the other side, a lack of integration could interfere with an accurate evaluation of self and others, resulting in a misperception of internal and external reality.

In an attempt to integrate the SeMi and ORT models, we hypothesized that the misperception resulting from the failure to integrate negative and positive dimensions of self and object among individuals with BPD features (as suggested by ORT theorists) corresponds to a suspicious mindset (as indicated by SeMi theorists). As a consequence of such a suspicious mindset, individuals with high BPD features could misperceive others as unjust, untrustworthy, or hostile. According to the SeMi model, enhanced JS often translates into uncooperative behaviors (Fetchenhauer & Huang, 2004). Consequently, drawing upon ORT and SeMi, we expected that JS dispositions would weigh on BPD individuals' distorted view of interpersonal relations in terms of lower appraisal of others' trustworthiness.

Our findings show that BPD features are positively associated with JS from the Victim and Observer perspective. Moreover, in our first study JS from the Beneficiary point of view suppressed the negative association between BPD features and trust appraisal. Additionally, the female sample's exploratory analysis also suggests that JS from the victim's perspective strengthens further the link between BPD features and decreased trust appraisal. In contrast, the observer's perspective acts oppositely and suppress the BPD features-trust appraisal link.

From a clinical standpoint, individual differences in JS may account for the rise of detrimental affects in psychotherapy sessions and transference relationships. Previous studies showed that victim JS is positively related to aggression (Bondü, 2018) and uncooperative behaviors (Fetchenhauer & Huang, 2004). In our sample, we found JS from the victim perspective further strengthened the link between BPD features and decreased trust appraisal. Consequently, it is likely for individuals with BPD features who are highly sensitive to injustice from the victim's perspective to react aggressively or uncooperatively to perceived injustice. We believe this could happen in the therapist-patient relationship so that BPD patients may quickly feel judged by therapists and undermine therapeutic quality. In line with our suggestion, several theorists defined BPD patients as "difficult to treat" due to high dropout rates, irregular psychotherapy attendance, and noncompliance (e.g., Gunderson, 2009; Levy et al., 2006; Yeomans et al., 2015). ORT theorists suggest this occurs mainly for the lack of a stable sense of self among BPD individuals.

Both aggression and uncooperative behaviors are hallmarks of BPD pathology and could be considered a potential risk factor for the therapeutic alliance. For these reasons, we believe that JS assessment in clinical work should not be ignored. It could

be a peril for clinicians to ignore the JS victim propensity of their BPD patients. Early detection of increased JS from the victim perspective in patients with BPD features could signal clinicians an enhanced risk for externalizing behaviors in the clinical setting. However, according to our results, enhanced JS from the observer's and beneficiary's perspective may partly suppress the BPD individuals' tendency towards suspiciousness and untrustworthiness. For these reasons, we invite researchers to develop efficient tools for an agile and economic assessment of JS in vis-à-vis clinical settings.

Furthermore, in both studies, no significant association occurred between BPD features and trust appraisal of neutral faces (only in the first study we detected a significant correlation considering only the trust appraisal of male faces). In the second study, we also found no significant gender effect, indicating that the strength of the BPD-trust appraisal link did not change across genders. Clinically, this may suggest that BPD behavioral trust-relevant expressions do not vary across genders. Hence, it may be helpful and useful for clinicians to observe and investigate patients' trust dynamics presenting with BPD features regardless of patients' gender.

Such results challenge previous findings of a strong negative association between BPD features and appraisal of others' trust. As already mentioned, we believe the inconsistencies are mainly due to the different assessment of BPD features used (self-report questionnaire vs. semi-structured interviews). Moreover, we used a non-clinical sample while previous studies compared clinic clinically diagnosed BPD individuals to control groups.

This brings to the discussion of two main limitations of both studies: the recruitment of non-clinical samples and self-report measures of BPD features. The

non-clinical sample selection limits the generalizability of our findings to clinical populations. However, personality disorders can be considered as extremes of normally distributed personality dispositions (Rothschild et al., 2003). In line with such a theoretical approach to personality, personality consists of broad normal traits that are dimensional. At the extreme ends of such traits, maladaptive problems arise (Widiger & Trull, 2007). Consequently, pathological constellations of traits are distributed in the general population at varying levels, and the empirical investigation of pathological manifestations can be carried out within non-clinical samples. However, we invite future researchers to test the replicability of our findings in clinical populations. A second limitation of the present contribution is the use of self-reported measures of BPD features. Future studies on the topic should include other types of pathological features measures such as semi-structured interviews to prevent any biased responses. Future studies should also address the contributions of comorbidities and check the specificity of our findings for BPD features by including clinical control groups.

Despite such limitations, we believe this contribution has clinically meaningful implications. Clinicians often struggle to work with patients with BPD for their interpersonal style. Due to their suspicious and impulsive interpersonal style, patients with BPD features often challenge therapy with disruptive behaviors or active efforts to take control of the therapeutic setting (Gunderson, 2012). Our findings suggest that individual differences in JS may play a significant role in such a suspicious mindset among individuals with BPD features. Thus, we recommend an early assessment of JS to prevent adverse therapeutic outcomes such as dropout or poor quality transference. One strategy to avoid patients with BPD features' suspiciousness to interfere in the

clinical work could be the active participation of patients in the diagnosis phase in the early stages of the treatment. By asking patients with BPD features to participate in their diagnosis, clinicians may prove their trustworthiness, stimulate patients' compliance with treatment, and dismiss their baseline suspiciousness.

CHAPTER 4. SUSPICIOUSNESS AND SELF-CONCEPT IN BORDERLINE PERSONALITY DISORDER

Overview

In chapters 2 and 3 we deepened the understanding of two distinct personal dispositions on the effect of BPD features on trust appraisal: RS and JS. Proceeding with the exploration of the stage of Prior beliefs and Disposition of our trust model, we focus on the contribution of a third disposition: suspiciousness. Rather than exploring the direct effect of suspiciousness on the BPD features-trust appraisal link, we explore the role of suspiciousness on one risk factor for decreased trust appraisal of others, the malevolent self-concept.

Based on psychoanalytic thinking about projection, individuals with a polarized and excessive malevolent self-concept tend to project into others such a negative representation of the self, developing untrustworthy and uncooperative attitudes (Kernberg, 1996, 1967; Klein, 1948). According to Object Relation Theory, projection is one of the core defensive mechanisms of Borderline Personality Organization. Thus, malevolent self-concept can be identified as a potential antecedent of untrustworthiness. Furthermore, theoretical contributions suggest that suspiciousness about others is a common feature among individuals with BPD (Bach, Sellbom & Simonsen, 2016).

Based on such theoretical and empirical backgrounds, we tested whether BPD features predict a malevolent self-concept and the plausible role of suspiciousness on such a link. Moreover, according to previous findings, explicit-as-aware, conscious self-related cognitions, and implicit-as-unaware self-related cognitions can be

consistent or not. For this reason, we also explored the role of implicit malevolent self-concept. To this aim, two hundred thirty-nine participants completed an assessment of BPD features and suspiciousness. Then participants completed a direct appraisal of their self-concept malevolence (i.e., explicit self-concept). Finally, to assess participants' implicit self-concept malevolence, we used an Implicit Association Test procedure (Greenwald, McGhee & Schwartz, 1998). We found evidence that BPD features predict a significant increase in the malevolence of explicit self-concept, and suspiciousness partly mediates such an effect. Furthermore, the implicit self-concept's malevolence moderated such a mediation path. We review the main clinical and research implications in the discussion section of the chapter.

1. Introduction

Borderline Personality Disorder (BPD) is a severe psychiatric condition. Chronic instability in multiple areas such as emotional dysregulation, self-harm, impulsivity, and identity disturbance are core features of BPD (American Psychiatric Association, 2013). According to several empirical studies, individuals with BPD embrace a more negative self-concept than non-clinical individuals (Klein et al., 2001; Roepke et al., 2011; Rüsch et al., 2007). Furthermore, different theoretical frameworks describe self-concept disturbances in BPD, such as Object Relation Theory (ORT; Kernberg, 1975); Biosocial Theory (Linehan, 1993); Mentalization Theory (Fonagy & Target, 2006). Apart from the empirical support and theories on self-concept disturbances among individuals with BPD reported above, there is still a small number of studies investigating the mechanisms underlying dysfunctional self-concept in individuals with BPD.

The self-concept consists of "an organized knowledge structure that contains traits, values and episodic and semantic memories about the self, and that controls the processing of self-relevant information" (Roepke et al., 2011, p. 149). In the social cognition framework, researchers acknowledge self-concept as a multifaceted phenomenon encompassing explicit-as-aware, conscious self-related cognitions, and implicit-as-unaware self-related cognitions (Greenwald et al., 2002). For example, a person that explicitly describes oneself as "benevolent" at the same time experiences a variety of benevolence-related affects and memories of which the person may not be aware. Interestingly, implicit self-concept can be consistent or inconsistent with explicit self-concept (Remue et al., 2014). Indeed, several studies demonstrated dissociations between the explicit self-concept (usually assessed with self-report

measures) and the implicit self-concept (usually assessed via performance-based measures; Schnabel & Asendorpf, 2010). Traditional dual-process models explained such dissociations as due to the presence of two separate modes to process information about the self: the explicit mode responsible for conscious, controlled, and reflective information processing and the implicit mode accountable for unconscious, automatic, and intuitive processes (Bosson et al., 2000; Greenwald & Farnham, 2000; Wilson et al., 2000). Nonetheless, recent contributions called into question such an assumption. It is now argued that correlations between indirect and direct (i.e., self-report) measures tap into whether the respondents consider legitimate the construct underlying the indirect measure (Hahn & Goedderz, 2020). Thus individuals could be aware of them, and low correlations between indirect and direct measures could echo the willingness to reject them as a foundation for answering explicit reports. From a psychodynamic point of view, such dissociations among individuals with BPD might be due to the massive use of splitting as a defense mechanism (Kernberg, 1975). According to ORT, extreme aggressive impulses and negative affects in BPD might lead to splitting the internal representations of self and others to protect the good inner parts from the bad ones (Kernberg, 1967). Consequently, individuals with BPD might show difficulties in integrating positive and negative experiences and developing distorted representations of the self and interpersonal world as either good or bad. The direct clinical manifestation of excessive use of splitting among individuals with BPD is the extreme and repetitive oscillation in the expression of contradictory “all good” and “all bad” self-concepts. Such fluctuations in self-concepts also result in the perception of external objects as “all good” ones and “all bad” ones, and concomitant

abrupt shifts and complete reversals of feelings and conceptualizations about a particular person (Kernberg, 1985).

This tendency to rely on a polarized negative self- and other-concepts leads to dysfunctional evaluations of relationships with others, work, and life in general. Several empirical contributions indicate that through a social comparison process, the evaluations of others are strongly related to evaluations of the self and vice versa (Markus & Wurf, 1987). Thus, how much people rely on polarized negative self-concept can polarize others' social perception and vice versa. Among individuals with BPD, many pathological manifestations (dysfunctional interpersonal relationships, impulsivity, self-destructive behaviors, etc.) root in disturbances in the ability to create, maintain and use benign and integrated images of self and others (Bender & Skodol, 2007). For example, individuals with BPD have been found to have a polarized negative representation of others as untrustworthy, which could connect to a polarized negative representation of self, too (Yeomans et al., 2015). Despite the relevance of polarized self-representation for individuals with BPD, the processes through which BPD features connect to polarized self-concept are still under-researched.

In the present contribution, we focus on investigating "malevolent" self and others-concept in a non-clinical sample of participants with varying levels of BPD features. The choice to explore the broad conceptualization of "malevolence" (vs. "benevolence") of self and others representations is theoretically driven. From a psychoanalytic perspective, one's representations of self and others are affectively colored, and such affective quality ranges from "malevolent" to "benevolent" (Kernberg, 1975; Klein, 1948). Furthermore, previous studies showed that

suspiciousness about others is a reliable marker for BPD pathology (Bach et al., 2016). Suspiciousness is defined as "expectations of and sensitivity to signs of interpersonal ill-intent or harm; having doubts about others' loyalty and fidelity; feelings of persecution" (Skodol et al., 2011, p. 38). Associations between negative self-concept (conceptualized as low self-esteem) and suspiciousness have been demonstrated in both clinical (Barrowclough et al., 2003) and non-clinical samples (Berry et al., 2006; Rotenberg et al., 2005). However, their interplay among individuals with BPD is still unexplored. To our knowledge, there is no investigation of the influence of suspiciousness attitudes on the "deflated" concept of self in individuals with BPD.

2. Main Aims & Hypotheses

We aim to explore the relationships between BPD features, explicit malevolent self-concept, implicit malevolent self-concept, and suspiciousness in a sample of undergraduate students. Based on previous findings, we have two main hypotheses: (1) BPD features would be associated negatively with explicit malevolent self-concept, and (2) suspiciousness would be related to explicit malevolent self-concept. We also tested whether suspiciousness mediates the negative association between BPD features and explicit malevolent self-concept with exploratory aims.

Finally, we were also interested in investigating the possible role of implicit malevolent self-concept on such a mediation model. We investigated whether implicit malevolent self-concept would correlate or not with explicit malevolent self-concept. We explored the role of implicit malevolent self-concept. First, we tested whether the implicit malevolent self-concept was a moderator of the link between suspiciousness and malevolent explicit self-concept (see Figure 1). Second, we examined whether it

moderated the indirect effect of suspiciousness between BPD features and the malevolent explicit self-concept (moderated mediation).

3. Methods

Participants and Procedure

Two hundred thirty-nine individuals recruited at the University of Milano-Bicocca (N = 239) participated in the study. All the participants were 18 to 50 years of age, and the sample was gender-balanced (121 males). We discarded data from 12 participants because of low accuracy in the task (accuracy below 25%). Our final sample consisted of 227 participants (114 males; M_{age} = 24, SD = 4.72) with a varying range of BPD features. Participants took part in a single laboratory session. The testing session lasted approximately 30 minutes. After welcoming the participants, the researchers rapidly explained the procedure and collected their written informed consent. On a computer, participants filled out the Personality Assessment Inventory-Borderline (PAI-BOR; Jackson & Trull, 2001) and the Suspiciousness facet of the PID-5 (Krueger et al., 2011). Then, participants completed an indirect measure of malevolent self-concept with an Implicit Association Test (IAT, Greenwald et al., 1998), a direct measure of malevolent self-concept, and demographic information questions. The procedure received approval from the university ethics committee.

Materials

Personality Assessment Inventory–Borderline Feature Scale (PAI-BOR; Jackson & Trull, 2001; Italian version, Pignolo et al., 2018). For a detailed description of the questionnaire, see paragraph 3.2. in the third chapter. Cronbach's α of the PAI-BOR total scores was good (α = .86).

Suspiciousness Facet-Questionnaire (SFQ, Krueger et al., 2011); Italian version, Fossati et al., 2017). It consists of a six items self-report scale derived from the Personality Inventory for DSM-5 (PID-5). The PID-5 is a 220-item self-report inventory developed to index the five DSM-5 Section III personality domains (Negative Affectivity, Detachment, Antagonism, Disinhibition, and Psychoticism). Suspiciousness is a facet of the Detachment dimension. It consists of "expectations of and sensitivity to signs of interpersonal ill-intent or harm; doubts about loyalty and fidelity of others; feelings of being mistreated, used and/or persecuted by others" (Anderson et al., 2013). The respondent rates each item on a scoring scale ranging from 0 (*very false or often false of me*) to 3 (*very true or often true of me*). Cronbach's α of the SFQ was good (α = .74).

Indirect measure of benevolent self-concept: Benevolent-IAT. The IAT is a measure of the relative strength of association between concept-attribute pairs. We used "Self" and "Others" as target categories, whereas the attribute categories were "Malevolent" and "Benevolent". We selected five stimuli for each category: "Self", "Others", "Malevolent", and "Benevolent" (for further details on the pretest procedure, see Appendix I). Participants categorized the stimuli appearing one at a time on the computer screen using the letter "D" and the letter "K" on the computer keyboard. The categories labels were present in the left and right upper corners of the screen throughout each task. In the first practice block, participants categorized the stimuli to the target categories (i.e., "self" and "others"). In the second practice block, participants classified words (i.e., "malevolent" and "benevolent"). In the first critical block, participants categorized the four categories of words using the same key for "Self" and "Malevolent" and the other key for "Others" and "Benevolent". The fourth

block was again practice and consisted of classifying the words related to benevolent and malevolent but with the opposite key assignment than the one used in the second block. In the second critical block, participants categorized the four categories of words but this time using the same key for "Self" and "Benevolent" and the other key for "Others" and "Malevolent". The order of the critical blocks was counterbalanced between participants. The practice blocks consisted of 20 trials, and the critical blocks consisted of 60 trials. We recorded reaction-times and error responses for all trials. We computed an IAT score following standard procedure (D score, Greenwald et al., 2003). A more positive IAT score indicated that individuals evaluated themselves more malevolent than benevolent compared to others.

Direct measure of malevolent self-concept. Respondents indicated how much ten adjectives were generally describing them on 7-point scales (from 1, *extremely untrue of you* to 7, *extremely true of you*). The adjectives selected from a preliminary study (for further details, see appendix 1) were 5 words describing a "benevolent person" (α =.83) and 5 terms representing a "malevolent person" (α = .81). We calculated a relative measure of malevolent self-concept (compared to benevolent) by subtracting the mean scores of benevolent items from the mean scores of malevolent items. The score computed is "explicit malevolent self-concept" so that larger scores represent a stronger explicit perception of oneself as malevolent.

4. Results

In our sample, there was a positive association between BPD features and explicit malevolent self-concept. The association did not occur between BPD features and implicit malevolent self-concept. Consistent with our predictions, we also found a positive association between suspiciousness and malevolent explicit self-concept.

Finally, there was no correlation between implicit and explicit malevolent self-concepts (see Table 1).

Table 1

Correlations and Descriptive Statistics (N = 227)

	M	*SD*	1	2	3	4
1. BPD Features	52.31	10.50	1			
2. Suspiciousness	0.99	0.49	.40***	1		
3. Explicit malevolent self-concept	-3.10	1.32	.30***	.13*	1	
4. Implicit malevolent self-concept	0.19	0.20	-.01	-.07	.02	1

Note. $^{*}p < 0.05$. $^{***}p < 0.001$.

To investigate whether suspiciousness mediates the effect of BPD features on malevolent self-concept, we carried out a mediation analysis. We ran a mediation analysis using PROCESS for SPSS (Hayes, 2013; model #4) using the Bootstrap estimation method with 5000 samples. We introduced explicit malevolent self-concept as the dependent variable, BPD features as the independent variable, and suspiciousness as the mediator. The indirect effect of BPD features on malevolent self-concept via suspiciousness was significant, $B = .13$, SE = .04, Bootstrap 95% CI [.04, .22]. The direct effect of the BPD features on malevolent self-concept remained significant, $B = .17$, SE = .01, $p = .02$. Thus, suspiciousness partially accounted for the relationship between BPD features and explicit malevolent self-concept (see Figure 1).

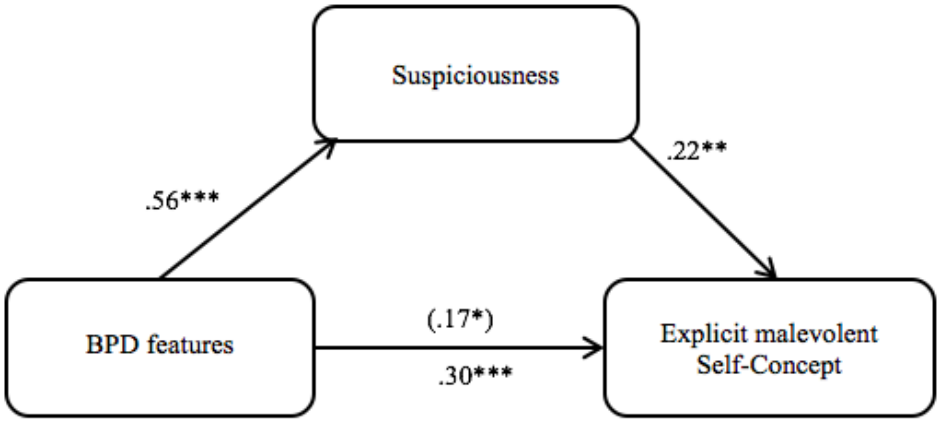


Figure 1. Mediated effect of BPD features on explicit malevolent self-concept via suspiciousness. * p < .05; ** p < .01. *** p < .001.

We then investigated the role of implicit malevolent self-concept as a moderator of the link between suspiciousness and explicit malevolent self-concept using PROCESS for SPSS (Hayes, 2013; model #1; see Figure 2) centreing data to reduce multicollinearity. The model explained 4% of variance and was significant, $F(3,223) = 3.05$, $p = .03$. The main effects of suspiciousness and implicit malevolent self-concept were not significant, $B = .11$; $SE = .06$, 95%CI [-.02; .24] and $B = .05$, $SE = .07$; 95%CI [-.08; .18], respectively. The interaction of implicit self-concept and suspiciousness was significant, $B = .13$, $SE = .06$, 95%CI [.02; .25]. A simple slope analysis indicated that for individuals with a high implicit perception of themselves as malevolent compared to others (+1*SD*), suspiciousness predicted the explicit malevolent self-concept, $B = .25$, $SE = .08$, 95%CI [.08; .41], whereas it did not for individuals with a low (-1*SD*) or medium implicit perception of themselves as malevolent, $B = -.02$, $SE = .10$, 95%CI [-.21; .17] and $B = .11$, $SE = .07$, 95%CI [-.02; .24], respectively. For individuals with a highly malevolent implicit self-concept, the more suspicious they were, the more they explicitly perceived themselves as malevolent (see Figure 2).

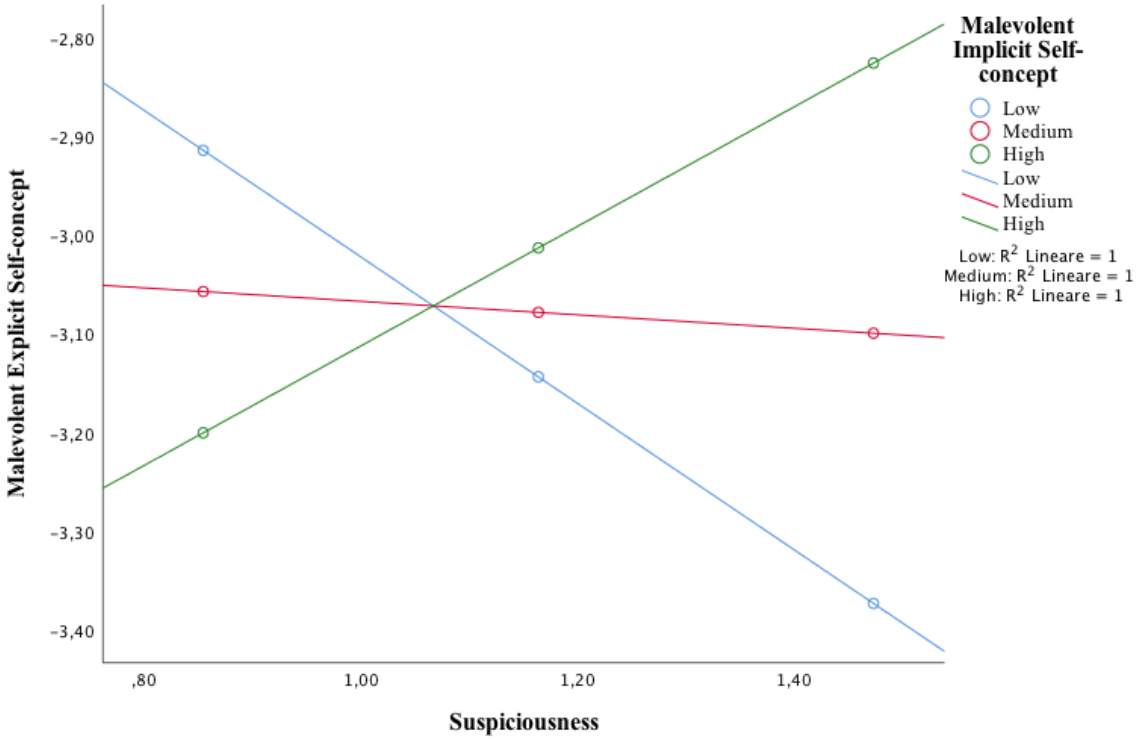


Figure 2. Interaction between suspiciousness and level of implicit self-concept for predicting malevolent explicit self-concept.

We finally evaluated a moderated mediation model (Hayes, 2013; model #14) using the Bootstrap estimation method with 5000 samples to test whether the partially mediated effect of BPD features on explicit malevolent self-concept through suspiciousness was moderated by the implicit malevolent self-concept. The model explains 12% of variance and was significant, $F(4,222) = 7.40$, $p < .001$. The moderated mediation index was significant, $B = .16$, $SE = .06$, 95%CI [.04; .27] indicating that the implicit malevolent self-concept significantly moderated the mediation. The decomposition of the effect revealed that for individuals with a highly malevolent implicit self-concept, suspiciousness mediated the relation between BPD features and explicit malevolent self-concept, $B = .06$, $SE = .04$, 95%CI [.01; .13]. The mediation was not significant for individuals with low and medium malevolent

implicit self-concept, *B* = -.07, *SE* = .05, 95%CI [-.16; .02] and *B* = -.01; *SE* = .03, 95%CI [-.06; .06], respectively.

5. Discussion

This study assessed BPD features, suspiciousness, malevolent explicit self-concept, and malevolent implicit self-concept in a non-clinical sample. The results supported our first hypothesis that the higher participants scored on BPD features, the more they explicitly reported a malevolent self-concept. This finding is consistent with previous research suggesting negative self-concept among individuals with BPD compared to clinical and non-clinical samples (Klein et al., 2001; Roepke et al., 2011; Rüsch et al., 2007, 2011).

We found no association between the implicit and explicit malevolent self-concepts. One interpretation of this result is that participants in our study might have rejected their implicit self-concept evaluations as malevolent in favor of a more positive representation of self-concept in the explicit report. In other words, our participants might have been aware of the self-concept reflected by the indirect measure we used (i.e., IAT) and calibrated their responses at the explicit assessment of the same construct to be socially desirable (i.e., positive self-concept). Our interpretation is in line with contributions suggesting that indirect measures of self-concept echo information people might not access when asked explicitly (Cunningham et al., 2007). For this reason, we suggest explicit and implicit self-concept are worth being separately evaluated. We stress the need for empirical approaches that, besides the explicit aspect of self-representations, also assess self-concept's implicit aspects in future investigations.

Consistent with our second hypothesis, we found an association between suspiciousness and increased malevolence of the explicit self-concept. The explicit self-concept measure we used assessed a specific expression of low self-esteem, namely a representation of the self characterized by malevolent affects and intentions. Thus, the association between suspiciousness and malevolent explicit self-concept is in line with literature indicating associations between deflated self-esteem and suspiciousness in a non-clinical population (Warman et al., 2010). We also found a positive association between suspiciousness and BPD features. This is consistent with previous contributions suggesting biases in evaluating social stimuli among individuals with BPD features (Nicol et al., 2013). Individuals with BPD tend to attribute untrustworthy, hurtful, and neglectful intentions to significant others (for a review, see Poggi et al., 2019). Our findings confirm further such indication. In our sample, higher BPD features were associated with an increase in suspiciousness.

To sum up, our findings are in line with the clinical characterization of individuals with BPD as individuals with a tendency to have very critical views of themselves (i.e., malevolent self-concept) and untrustworthy opinions of others (i.e., suspiciousness). Our contribution also supports the idea that implicit and explicit malevolent self-representation is not necessarily associated (i.e., lack of association between implicit and explicit self-concept). One plausible explanation for the lack of such association is that participants calibrated their explicit report of self-concept. This would imply that participants could have been aware of self cognitions that our implicit assessment of self-concept measure (i.e., IAT) reflected. Consequently, when stimulated to pay specific attention to this cognition through an explicit measure of the

same construct, they calibrated responses to convey a positive representation of themselves.

Our results showed that the participants' suspiciousness partially accounts for the positive relationship between BPD features and explicit malevolent self-concept. This finding supports the presence of a "suspiciousness" route linking BPD features and explicit malevolent self-concept. As BPD severity increases, the individuals' suspicious mindset increases too, and both relate to an increase in the explicit malevolent self-perception. The mechanism is consistent with the modern object-relations conceptualizations of BPD as a pathology of "internal object relations" (Kernberg & Caligor, 2005). According to such conceptualization, BPD's symptomatic expressions root in a lack of integration of positive and negative representations of self and others. To face such a lack of integration, individuals with BPD operate intrapsychically and interpersonally primitive defenses such as splitting, idealization, and devaluation. Such defenses allow individuals with BPD to "split" the full experience of the self. The fragments of self that are perceived as disturbing are divided from desirable fragments and projected into others. As a result of this process, both self and others are perceived as discontinuous over time and shifting abruptly from extremely positive to extremely negative representation. According to ORT, such defenses leave individuals with BPD with a caricatured and distorted representation of others and a corresponding caricatured and distorted representation of the self (Stern et al., 2018). Consistently, we found that BPD features predict an increased distorted and suspicious representation of others that, in turn, predicts a distorted and malevolent representation of the self.

Moreover, we found that BPD features predicted a decrease in the explicit malevolent self-concept through suspiciousness only in individuals with extremely malevolent implicit self-concept. From a psychoanalytic perspective, an extremely malevolent implicit self-concept corresponds to a "split" and negatively polarized representation of the self. To preserve "all good" representations from destruction by a negatively polarized representation of the self, individuals with BPD engage in projection and projective identification. "All bad" part of self-representations are projected into others leading to BPD individuals' typical paranoid tendencies (i.e., suspiciousness; Summers, 1994). Our results show how polarized internal representations of self (i.e., implicit malevolent self-concept) might partly explain this interplay. Individuals with high BPD features and highly malevolent implicit self-concept show suspicious tendencies due to projection into others of their polarized negative self-representations; however, given that the negative internal representation is linked to self, due to identification with and introjection of the negative parts projected into others, they explicitly report a malevolent self-concept.

Our findings may have significant implications for clinical interventions focused on individuals with BPD features' negative self-concept. Specifically, individuals with BPD features, high suspiciousness, and highly malevolent implicit self-concept may be at higher risk for maintaining malevolent explicit and implicit self-concept through projective identification cycles. Since projective identification can be considered a pathway from the intrapsychic to the interpersonal world, polarized, and primitive representations of self and objects often emerge in the transference during therapy with BPD patients (Ogden, 1979). Consequently, therapists often struggle with BPD patients due to their tendency to repeat unstable

interpersonal relations with the therapist and drop-out (Yeomans et al., 1994). BPD patients, for example, usually show a combination of suspiciousness and extreme idealization of the therapist in the clinical setting (Yeomans et al., 2015). According to our results, if individuals with BPD features own a negative implicit self-concept and high suspiciousness, they could easily manifest negative explicit representations of the self in the treatment relationship. Within the clinical setting, a possible manifestation of such a negative self-concept is hostility and expressed aggression against the self (i.e., severe neglect, self-injury, suicidality) or enacted against others, including the therapist. According to our results, BPD patients may benefit from a treatment focused on establishing a coherent and stable sense of self. Therapy sessions should focus on the polarized negative or positive representations of the self, leading to negative self-concept manifestations into therapeutic relationships. In other words, our results support the need to recognize patients' tendency to evacuate intolerable internal conflicts through actions and aggression in the therapeutic relationship. Our conclusions look consistent with the Transference Focused Therapy (TFP) approach to BPD patients. TFP is a manualized treatment for BPD derived from the ORT framework. TFP invite therapists to recognize patients' impulse to actions in therapy settings through transference interpretations. Transference interpretations allow a precise formulation of individuals' primitive self and object representations acting both in the treatment and outside, in everyday situations (Levy et al., 2006).

Our contribution, like any, has limitations. The current study used a sample of non-clinical participants; hence, it is unclear whether these findings can be generalized to a clinical population. We invite future researchers to test the replicability of such results in clinical populations too. Besides, the measures of implicit and explicit

malevolent self-concept we used had radically different structures. According to Payne and colleagues (2008), correlations between explicit and implicit measures of the same construct with substantial structural differences weakly correlate due to such different structures. We cannot rule out that such a lack of structural fit between the implicit and explicit measure of malevolent self-concept might have influenced the lack of correlation between implicit and explicit malevolent self-concept measures. For this reason, there is a need for replication studies using different indirect and direct measures of malevolent self-concept. Furthermore, it has been recently questioned whether IAT and implicit measures of self-concept in general tap actual self-concept (i.e., evaluation of the current self) or ideal self-concept (i.e., representation of the qualities a person would like to have) (Remue et al., 2014). According to this, the IAT scores for some participants in our study may have reflected the extent to which they believed to be good (i.e., actual self-concept). In contrast, it may have reflected the self they would like to be (i.e., ideal self-concept) for other participants. Future studies should use implicit measures of self-concept that distinguish the ideal self-concept from the actual.

Despite such limitations, our study is the first empirical test of the influence of suspiciousness attitudes on the "deflated" concept of self among individuals with BPD. We provide insight into the suspiciousness pathway through BPD features translate into a malevolent explicit representation of the self. To summarize, we found that individuals with high BPD features and highly malevolent implicit self-concept display explicit malevolent self-concept partly through a suspiciousness path. We also showed how implicit self-concept impacts the way BPD individuals' affectively connotate their explicit self-representation. Such findings add to the previous

knowledge about self-representation among individuals with varying BPD features. Based on the evidence, we encourage both researchers and clinicians to disentangle the unique contribution of implicit and explicit self-representation in individuals' functioning. A more comprehensive and accurate picture of the unique contribution of explicit and implicit self-representation on individuals' behavior may help get a better picture of individual subjective experiences.

CHAPTER 5. THE INFLUENCE OF DIRECT AND INDIRECT CUES OF TRUSTWORTHINESS ON BORDERLINE PERSONALITY DISORDER'S BIASED APPRAISAL OF TRUST.

Overview

While Chapter 2, 3, and 4 deepened the investigation of our model's prior beliefs and disposition stage, here we shift our attention towards the core section of the model: trust appraisal. Specifically, we explored the effect of different trust cues on trust appraisal performances across individuals with varying levels of BPD features.

In the social cognition framework, the empirical investigation of trust impairments among individuals with BPD used facial trustworthiness cues. For example, Fertuck and colleagues (2013) explored the effect of emotional facial cues on the explicit appraisal of trust of BPD participants compared to controls. Nonetheless, empirical literature suggests that individuals base their trust judgments on direct (Van't Wouth & Sanfey, 2008) and indirect trustworthiness cues (Klapper, Dotsch, van Rooij & Wigboldus, 2016). We explore the weight of direct and indirect cues of others' trustworthiness on trust appraisal on two separate non-clinical samples. To this purpose, we developed two novel procedures for manipulating others' trustworthiness, relying either on direct or indirect cues of trustworthiness.

In the first study, we present a direct manipulation of trustworthiness consisting of associating pictures of neutral faces with trustworthy or untrustworthy behavior information (i.e., "She practices what she preaches" vs. "He spreads negative gossip about his friends"). In the second study, we employed Virtual Reality (VR) technology to perform an indirect manipulation of trustworthiness. We decided to implement the

task in VR because it offers an opportunity to replicate everyday situations under easily to control conditions. We exploited the high ecological validity provided by VR to convey implicit cues of avatars' trustworthiness. We implemented a modified version of the traditional Trust Game procedure manipulating avatars' reciprocity rates. Participants played with one trustworthy avatar (i.e., returned a profit for participants) and one untrustworthy avatar (i.e., produced a loss for participants). In both studies, participants (N = 166; N = 100), after completing the assessment of BPD features and the trust manipulation, appraised manipulated identities' trust.

According to the results, both the manipulation procedures we implement are effective in manipulating trust. We, thus, provide researchers with two novel trust manipulation procedures to investigate trust processes. Surprisingly, despite empirical literature suggests BPD features predict a decrease in trust appraisal, no association between untrustworthiness bias and BPD features occurred in both our studies. In the discussion, we comment on the implications for future research of our lack of significant results.

1. Introduction

Borderline Personality Disorder (BPD) is a severe clinical condition with a prevalence ranging from 0.5 % to 3.54 % in the general population (ten Have et al., 2016), 9%–14% among psychiatric inpatients, and 12%–18% among outpatients (Doering, 2019). BPD is characterized by marked impulsivity and a pattern of instability that influences self-image, interpersonal relationships, and affects (American Psychiatric Association, 2013). Impairments in interpersonal functioning persist even after remission of other BPD typical symptoms (Gunderson, 2011).

Among these impairments, the importance of a pervasive mistrust of others in the BPD psychopathology has been addressed within several frameworks (Fonagy et al., 2015; Kernberg, 1967). The people with BPD's inclinations in perceiving others as untrustworthy have been conceptualized as an untrustworthiness bias, characterized by the expectation that "others will reject, be dishonest with, negatively judge, or otherwise emotionally hurt" them (Fertuck et al., 2013, p. 196). According to the Object Relation Theory (ORT), individuals with BPD develop impairments in the ability to trust because of a massive "splitting" of interpersonal and emotional experiences into polarized "good" and "bad" representations of objects (Yeomans et al., 2015). In the ORT framework, untrust of others derives from a defensive projection into others of the dichotomous "bad internal object" resulting from splitting (Stern et al., 2018). Within a mentalization framework, individuals with BPD develop a pervasive expectation of others' untrustworthiness because of early aversive experiences with primary caregivers, and such expectation compromises their social functioning (Fonagy, Luyten, et al., 2017).

Many empirical contributions on trust impairments focused on identifying the antecedents (proximal and distal) or the consequences of dysfunctional trust appraisal in people with BPD (Poggi et al., 2019). For example, according to Botsford and colleagues (2019), a significant antecedent consists of a lower disposition to interpersonal trust in individuals with BPD, as compared with clinical and non-clinical controls. Regarding the consequences of dysfunctional trust appraisal, King-Casas and colleagues (2008) observed cooperative behavior in individuals with BPD and controls during a 10-round Trust Game (TG, the most used experimental paradigm to examine trust behaviors). TG consists of an interactive game between two persons: the trustor (usually the participant) and the trustee. The trustor is usually allocated an endowment and asked to choose an amount to send to the trustee; the amount invested is multiplied by three before the trustee receives it; finally, the counterpart will send back the trustor a sum to choose from the tripled amount received. The authors found significantly more frequent cooperation ruptures and less frequent coaxing behavior during the TG procedure in individuals with BPD than controls (King-Casas et al., 2008).

Trust can be seen as a dynamic process in the sense that it is the result of developmental factors, biological components, social circumstances, and individual differences interacting at every moment of one's lifespan (Lewicki et al., 1998). Besides contributions regarding antecedents and consequences of trust impairments among individuals with BPD, trust appraisal's empirical investigation is still limited. In general, trust appraisal is sensible to others' behaviors (Abramov et al., 2020), but little is known about its triggers among individuals with BPD (Abramov et al., 2020; Liebke et al., 2018). Most of the studies investigated emotional facial cues' effect on the explicit appraisal of individuals' trust with BPD (e.g., Fertuck et al., 2013).

Moreover, research has mainly focused on investigating trustworthiness attributions in static contexts and through self-reports. Trust impairments are central among individuals with BPD (for a review, see Poggi et al., 2019). However, we still know very little about which cues influence trust appraisal among individuals with BPD. Individuals rely on different trustworthiness cues to form others' trustworthiness judgments quickly (for a review of trustworthiness factors, see Mayer et al., 1995). Trustworthiness cues can be static (i.e., semantic information about an agent's trustworthy or untrustworthy behaviors) and dynamic (i.e., behavioral signals of agents' trustworthy or untrustworthy stance across situations). Previous contributions found that trustworthiness cues travel via different channels (i.e., vocal and facial), and individuals can be more or less aware of such cues (Mayer et al., 1995; Tsankova et al., 2015). Individuals base their trust judgments on direct (such as knowledge about others' moral status; (van 't Wout & Sanfey, 2008) and indirect trustworthiness' cues (such as others' facial appearance or behaviors; Klapper et al., 2016). Individuals can deliberately process direct cues of others' trustworthiness. However, automatic processing of indirect social information (such as the unaware association between trust and specific facial appearances; Burns & Conchie, 2015) could influence others' trust appraisal. Individuals with BPD are characterized by biased trust evaluations (Ebert et al., 2013; Fertuck et al., 2013). Aware, intentional, and cognitive evaluations of trustworthiness cues could activate untrustworthy response biases among BPD individuals. Still, automatic and out of awareness associations between others-related concepts and untrust-related aspects can easily culminate in untrustworthy response biases among BPD individuals. To our knowledge, there is no prior empirical evidence

on whether indirect cues are more (or less) relevant than direct cues for biased appraisal of others' trustworthiness at different rates of BPD features.

2. Aims of the contribution

Our studies enable us to test the weight of direct and indirect cues of others' trustworthiness on trust appraisal performance across individuals with varying BPD features. We examine the relative role of direct, static, and indirect, dynamic information on trust appraisal by individuals with varying BPD features in the present contribution. To this aim, we performed two separate studies. In the first study, we used a direct manipulation of trustworthiness. We coupled pictures of neutral faces with semantic information about trustworthy (i.e., "She practices what she preaches") or untrustworthy (i.e., "He spreads negative gossip about his friends") actions. In the second study, we conveyed an indirect trustworthiness manipulation using the dynamic cues of a Trust Game procedure in a VR environment. In fact, researchers considered Trust Game procedures as the "ideal vehicle for examining trust dynamically" (Abramov et al., 2020, p. 2). In this sense, cooperativeness or uncooperativeness of others' behavior during the TG procedure vehicle indirectly others' trustworthiness information. In social sciences research, Virtual Reality (VR) technologies have been increasingly used to overcome the limitations of traditional empirical procedures in investigating human interactions. Standard laboratory procedures lack ecological validity, and the experience of immersion in the tasks is significantly reduced compared to VR environments (Blascovich et al., 2002; McCall & Blascovich, 2009). In VR settings, researchers can deliver dynamic, social cues and, at the same time, avoid the effect of confounding factors, such as attractiveness, height, voice, and other features (Fox et al., 2009). Because of these reasons, we believe the

investigation of BPD's interpersonal impairments, such as untrustworthiness bias, may benefit from the opportunities given by VR for the study of social interactions.

After manipulations, we asked participants to appraise trust towards the manipulated identities through an explicit measure in both studies. Thus in study one, the trustworthiness cues are static and direct, whereas, in study two, the cues are dynamic and indirect.

3. Study 1

In the present study, we aimed to test if direct signals of others' trustworthiness influence the association between BPD traits and untrustworthiness bias. By experimentally manipulating others' trustworthiness with direct and static information (i.e., coupling identities with trustworthy or untrustworthy behavior), we expected to find a significant difference in the appraisal of untrustworthy and trustworthy manipulated identities. Furthermore, we expected to replicate previous findings of an association between BPD traits and a decrease in neutral faces' trust appraisal (i.e., Fertuck et al., 2013). We also investigated the interaction between trustworthiness manipulation and BPD traits on trust appraisal. We expected the association between BPD traits and an untrustworthiness bias would be stronger for trustworthy manipulated identities.

3.1.Methods

Sample size determination

We performed a power analysis to determine the minimum sample size for testing our hypothesis. For a repeated-measures ANOVA with a within factor measure (trustworthy vs. untrustworthy faces) with a medium to good effect size (Cohen's d = 0.20), a minimum sample of 134 participants would achieve an actual power of 0.90.

We happened to collect a slightly larger sample (N = 166). We thus conducted a sensitivity analysis with G*Power (Erdfelder, Faul, Buchner, & Lang, 2009) and determined that with our final sample, power set at .80 and alpha equal to .05, we would be able to detect a medium effect of f = 0.155 (corresponding to Cohen's d = 0.310).

Participants and Procedure

One hundred sixty-six participants (M age = 23.6, SD = 4.07, 96 women, 66 men, and 4 did not indicate their gender) participated in the study. In a single laboratory session, participants provided their informed consent and then filled in a questionnaire for the assessment of BPD features (Personality Assessment Inventory-Borderline; PAI-BOR; Morey, 2004). Then, participants participated in the direct trustworthiness manipulation phase, and, finally, they performed the trust appraisal task. The study received approval from the university ethics committee.

3.2.Materials

Personality Assessment Inventory–Borderline Feature Scale (PAI-BOR; Jackson & Trull, 2001). For a detailed description of the questionnaire, see paragraph 3.2. in the third chapter. Cronbach's α of the PAI-BOR total scores in this study was good (α =.85).

Direct trust manipulation. We selected four face stimuli (2 male; 2 female) representing four different identities from the Nimstim database (Tottenham et al., 2009). We chose exclusively Caucasian stimuli to rule out any racial bias effect on trust appraisal and manipulation (Birkás et al., 2014). To select four neutral stimuli on both attractiveness and trustworthiness dimensions, we run a pretest on 52 students (M Age = 22.1, SD = 2.4, all females). Participants in the pretest rated the trustworthiness

and attractiveness of 18 neutral stimuli (9 females; 9 males) on a 7-point scale. Hence, we ruled out extremely trustworthy or untrustworthy stimuli and selected four stimuli perceived as averaged trustworthy in the pretest (for each selected stimulus's statistics, see Appendix II). We selected sentences describing efficiently trustworthy and untrustworthy behaviors with a second pretest on 180 students (M Age = 23.57, SD = 5.58; 39 men and 145 women). For a detailed description of both pretests, see Appendix II and III.

For the manipulation, we assigned each of the four selected identities to three trustworthy (i.e., "He always returns the books I lend him") or untrustworthy (i.e., "She disappears every time I ask for her help") behavioral statements. All participants viewed equal numbers of males and females faces, combined with trustworthy and untrustworthy statements.

Trust Appraisal task. The task consisted of rating the trustworthiness ("How trustworthy do you think this person is?") of the four facial stimuli manipulated in the previous phase on a 7-point scale from 1 ("not at all trustworthy") to 7 ("very/extremely trustworthy"). We presented the facial stimuli in a fixed random order. We computed the trust learning scores as a differential score of the trust appraisal score of trustworthy identities and the trust appraisal score of untrustworthy identities (Δ score).

3.3.Results

We conducted a repeated-measures ANOVA on trust appraisal performances (Trustworthy or Untrustworthy) to check our manipulation's effectiveness. As predicted, the trustworthiness of target face influenced trust appraisal ratings, $F(1, 165) = 312.32$, $p < .001$, $\eta 2 = .65$, indicating that faces associated with trustworthy

behaviors were judged more trustworthy than faces associated with untrustworthy behaviors (see Table 1). The results provide evidence of the effectiveness of the manipulation.

Against our predictions, there was no correlation between BPD features and trust appraisal ratings (see Table 1). To address further the role of BPD features on trust appraisal task performances, we estimated a repeated-measure moderation model with the MEMORE package (Montoya & Hayes, 2017). We introduced the trust appraisal task's performance as the outcome variable, the trustworthiness of the targets (trustworthy vs. untrustworthy) as the predictor, and BPD features as the moderator. The main effect of BPD features is not significant, B = - 0.01, SE = .01; 95%CI [-.03; .01], p = .360. We found a non significant conditional effect of BPD features on the trust appraisal of either trustworthy (B = -0.1; SE = 0.1, 95%CI [-.01; .03], p = .36) and untrustworthy identities (B = .01, SE = 0.1, 95%CI [-.02; .02], p = .900).

Table 1

Descriptive statistics and correlations

	M	*SD*	1	2	3	4
1. PAI_BOR	52.87	10.16	1			
2. Mean Trust Appraisal Trustworthy	5.33	1.30	-.07	1		
3. Mean Trust Appraisal Untrustworthy	2.42	1.23	.01	-.41***	1	
4. Δ Trustworthy – Untrustworthy	2.91	2.12	-.05	.85***	-.83***	1

Note. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

PAI-BOR = Personality Assessment Inventory–Borderline Feature Scale (Jackson & Trull, 2001)

3.4.Discussion

The association of a brief description of trustworthy and untrustworthy behavior with pictorial facial stimuli successfully conveys the information that the identity represented in the picture is trustworthy or untrustworthy. However, against our expectations, BPD features did not correlate with judgments in the trust appraisal task. Regardless of pathological features, we found average trust ratings of manipulated trustworthy identities to be significantly higher than average trust ratings of manipulated untrustworthy identities. The repeated-measures ANOVA's effect size on trust appraisal performances is very large ($\eta^2 = .65$). This value indicates the effect of the manipulation on trust appraisal of trustworthy and untrustworthy identities is very ^{2} = .65 corresponds to Cohen's d = 2.73). Such a result indicates that our manipulation was robust. Due to the manipulation's strength, we questioned whether individual differences could have influenced the trust appraisal performance. We believe that the lack of a negative association between trust appraisal performances and BPD features could be a consequence of such powerful manipulation. The manipulation's strength may have inhibited individual differences (such as BPD features) effect on the outcome measure. If so, one may hypothesize that changing the way to manipulate trustworthiness (i.e., making it hidden and less noticeable) might leave room for the expression of individual differences on the trust appraisal task. This conclusion leads to our second study.

4. Study 2

This study aimed to test the plausible influence of indirect and dynamic signals of others' trustworthiness on the association between BPD traits and trust appraisal.

We experimentally manipulated two male[5] avatars' trustworthiness through their investment rates in a TG performed in a VR environment. Thus, we expected a significant difference in the trust appraisal between the untrustworthy and trustworthy avatars. Furthermore, we expected negative relation between BPD traits and trust appraisal performance according to previous findings on untrustworthiness bias among BPD individuals (i.e., Fertuck et al. 2013). We further explored whether the strength of the link between BPD traits and untrustworthiness bias would vary depending on the target's trustworthiness.

4.1.Methods

Sample size determination

We performed a power analysis to determine the second study's sample size based on Study1's results. Therefore considering a repeated-measures ANOVA with one factor (trustworthy vs. untrustworthy faces), an effect size d = .20, and a correlation among repeated measures r = 0.41 obtained, a sample of 81 participants should be sufficient to achieve a power of 0.90. The sample collected with slightly larger (N = 100). A sensitivity analysis with G*Power (Erdfelder, Faul, Buchner, & Lang, 2009) determined that with our final sample, a power of .80, α = .05, and a correlation among repeated measures r = .41, we would be able to detect a medium effect size of f = 0.154 (Cohen's d = 0.306).

Participants and Procedure

[5] Participants interacted only with two male avatars to rule out any possible effect of trustees' gender, although the literature provides mixed evidence. The general population usually believes males to be less trustworthy than females (Snijders & Keren, 1999; Wright & Sharp, 1979). Yet in economic games, even if women were considered more trustworthy, they were not behaviorally trusted more (Buchan et al., 2008; Eckel & Wilson, 2006; Orbeli et al., 1994).

One hundred participants (*M* Age = 23.1, *SD* = 4.07; 65 women and 34 men) took part in the study. All participants gave written informed consent. In a single laboratory session, participants filled in a questionnaire to assess BPD features (Personality Assessment Inventory-Borderline; PAI-BOR; Morey, 2004). Then, subjects participated in the indirect manipulation phase, and, finally, they performed the trust appraisal task. The study received approval from the University Ethics committee.

4.2.Materials

Personality Assessment Inventory–Borderline Feature Scale (PAI-BOR; Jackson & Trull, 2001). In this sample, Cronbach's α of the PAI-BOR total scores in this study was good (α =.87).

Indirect trust manipulation. We used a Virtual Reality Environment. We replicated Franzen and colleagues' procedure (2011) for the indirect manipulation of trustworthiness but adapted it in a Virtual Reality environment. We used a Trust Game to manipulate trustees' reciprocity rates to convey trustworthy or untrustworthy trustees' behaviors. After wearing the Oculus Rift, participants sat on a chair. They started interacting with avatars in a virtual environment that looked like a room with a table in the middle (for a graphical representation of the VR environment, see appendix IV). The game consisted of two blocks of 9 trials. In each block, the participant played with a different avatar in the role of trustee. The trustworthiness manipulation consisted of changing the percentage of the return amount of the two trustees. One trustee (trustworthy avatar) always returned more than 1/3 of the tripled amount (i.e., a profit for the participant). The other trustee (untrustworthy avatar) always returned less than 1/3 of the tripled amount (i.e., a loss for the participant). We counterbalanced

the assignment of avatars to trustworthy or untrustworthy conditions and the two avatars' presentation order. After introducing the task, one avatar appeared on the opposite side of the table where the participant was seated. On the left side of the table was present a pile of coins ("wallet"). At the start of the game, the wallet contained 10 €. We located the amount of money to share on the table's right side ("amount to share"). Amounts in the wallet and to share (and corresponding piles height) changed according to participant behavior. By pressing the "S" or "K" key on a keyboard, the participant could move coins from left to right and change the amount of money to share. Once the participant confirmed the amount to transfer and the amount to keep in his/her wallet, he/she pressed the space bar. In each of the 18 trials, the participant could choose any amount between 0 € and 10€ to share. The sharing amount was always tripled before getting to the counterpart, the trustee. Finally, the trustee avatar returned a proportion of the amount to the participant. At the end of every trial, participants saw a panel with a trial's summary indicating the amount of money shared, the amount of money returned by the trustee, and the amount of profit or loss made.

Trust Appraisal task. The task was the same as in the previous study. Participants rated the trustworthiness of the picture of the two avatars met in the VR trust game. The trust learning score consisted of the difference between the trustworthy avatars and untrustworthy avatars appraisals (Δ score).

4.3.Results

We performed the same set of statistical analyses as in Study 1. To empirically test whether our manipulation was valid, we again employed a within-subject ANOVA, with target faces (Trustworthy or Untrustworthy) as within-subject factors. As predicted, the ANOVA revealed a main effect, $F(1, 99) = 69.75$, $p < .001$, $\eta^2 = .41$

indicating that trustworthy avatar was judged more trustworthy than untrustworthy (See Table 2). The indirect trustworthiness manipulation in the VR setting was thus effective. Against our predictions, BPD features did not correlate with the trust appraisal ratings (see Table 2), indicating that individuals with high BPD features did not show any untrustworthiness bias.[6]

Table 2

Descriptive statistics and correlations

	M	*SD*	1	2	3	4
1. PAI-BOR	55.31	11.60	1			
2. Mean Trust Appraisal Trustworthy	4.00	1.31	.07	1		
3. Mean Trust Appraisal Untrustworthy	5.69	1.24	-.01	-.26**	1	
4. Δ Trustworthy – Untrustworthy	1.70	2.02	-.05	-.81***	.78***	1

Note. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

PAI-BOR = Personality Assessment Inventory–Borderline Feature Scale (Jackson & Trull, 2001)

4.4.Discussion

Regardless of the effectiveness of the indirect manipulation of trustworthiness, BPD features did not correlate with trust judgments. Again the lack of correlation between BPD features and trust appraisal task performances is not consistent with previous findings suggesting a decrease in trust appraisal increasing BPD features

[6] We calculated correlations between all PAI_BOR subscales and appraisal task performance in the second study too. Consistently to our first study, none significant correlation occurred, indicating the lack of any link between BPD features and performances at the trust appraisal task.

(Fertuck et al., 2013; Franzen et al., 2013). We endorsed a large effect size ($\eta^2 = .413$) using these indirect and dynamic cues in the manipulation. It indicates the effect of the manipulation is strong ($\eta^2 = .413$ corresponds to Cohen's $d = 1.65$). Notwithstanding this, it is weaker than the manipulation we implemented in the first study (Cohen's $d = 2.73$).

Whereas we replicated Franzen and colleagues' (2011) procedure, we used VR technology to approximate real-life economic interactions. Franzen and colleagues' (2011) version of the manipulated TG task lacked the realism and ecological validity we implemented in VR. The close to reality experience offered by the virtual environment-based TG might have stimulated our participants' to pay further attention to fairness and unfairness cues as if they were acting in real human to human interactions. This could be one first reason for inconsistencies between our results and previous.

A second explanation could be that the outcome measure may not have reflected respondents' trust attitudes. The measure may have tapped a deliberate manipulation of responses to regulate participants' impressions of others (Fazio & Olson, 2003). We believe participants may have overestimated their trust to convey that they trust individuals more than they actually do for self-presentation and social desirability concerns. Even if in social science, a common assumption is that, in real-life social-exchanges, being at least minimally suspicious is functional and necessary to accurately judging truth and detecting deception (McCornack & Parks, 1986). At the same time, trustworthy attitudes are more socially accepted than untrustworthy (Simpson, 2013). We argue social desirability and distortions in self-presentation might have widened the gap between actual appraisal of avatars' trustworthiness and

self-reported trust appraisal. In other words, participants might have inflated responses in the trust appraisal of the untrustworthy avatar to perform the task in a socially desirable fashion.

5. General discussion

According to previous contributions, BPD individuals show a lower propensity to trust others, namely untrustworthiness bias (for a review of the evidence, see Poggi et al. 2019). Although others' trustworthiness judgments rely on various cues, there was a lack of a systematic investigation of the role of different kinds of cues on untrustworthiness bias. In the present contribution, we explored the influence of direct and static cues (first study) and indirect dynamic cues (second study) of trustworthiness on trust appraisal by individuals with varying rates of BPD features. To this aim, we implemented two different manipulations of trust in this contribution.

In the first study, we conveyed the association of trustworthiness or untrustworthiness to identities through a direct manipulation procedure in a laboratory setting. In the second study, we used VR technology. Given the complexity of social interaction between humans, implementing both realistic and experimentally controlled procedures to study BPD individuals' interactive behavior is challenging. The use of VR technology might help to overcome this challenge. In our second study, we indirectly manipulated trustworthiness through a VR version of the TG procedures. Both manipulations were successful.

Concerning the association between BPD features and trust appraisal, based on past literature, we expected to find a negative association (Fertuck et al., 2013). Surprisingly, BPD features did not associate with others' trust appraisal in both the studies presented. Such a lack of association occurred regardless of the quality of the

trust manipulation we developed (direct vs. indirect). One possible explanation of the discrepancy with previous findings could be methodological. All previous studies that found diminished trust appraisal performance with explicit measure asked participants to rate more stimuli than 4 or 2, like in our case. Fertuck and colleagues (2013) asked participants to rate the trustworthiness of 144 face stimuli; Miano and colleagues (2013) asked the participants to appraise the trustworthiness (and 16 other dimensions) of 12 different face stimuli. We believe that the request to evaluate just a few stimuli might be more susceptible to participants' intentional inhibition of biased responses than a larger number of stimuli. Whereas short and fast tasks like ours might easily elicit normative responses (e.g., "I trust others"), the exposure to long and effortful tasks seems more appropriate to uncover the untrustworthiness bias (Fertuck et al., 2013; Miano et al., 2013).

It is also possible that we failed to replicate previous findings because of some limitations. First, our samples were not clinical and mainly recruited among psychology university students. Second, we found large prevalence rates of the clinical number of BPD features endorsed by our participants in both studies. According to the PAI scoring rules (Morey, 2004), a total PAI-BOR raw score of 60 or more indicates typical borderline personality functioning. In the first study, 23% (N = 41) of participants showed significant BPD symptomatology. In the second 32% (N = 32). Both rates far exceed the prevalence rate found in other community surveys, and the prevalence of BPD diagnosis in the general population of ∼1.7% (Gunderson et al., 2018). According to a review of community studies using structured clinical diagnostic instruments, rates of personality disorders in community samples range from 3.9% (among university students, Lenzenweger et al., 1997) to 22.3% (Coid et al., 2006).

Third, both our samples were not gender-balanced, and we cannot rule out any participants' gender effect on results. In the future, a more balanced sample in terms of gender and pathological distribution could lead to different results. Finally, we raise concerns regarding the outcome measure we used in both studies. Participants had some degree of control over their responses to this task. Since our outcome measure asked participants to directly rate their appraisal of others' trust, we argue that they could have had some degree of control over their responses at this task. Our participants might have been unwilling to provide accurate reports for social desirability concerns. Hence, participants may have produced controlled responses to provide a representation of themselves as trusting individuals. Our outcome measure may have tapped the trust appraisal of others that participants perceived most desirable rather than the actual one. An indirect measure could reflect one's automatic beliefs about others' trustworthiness and convey information about individuals' implicit trust appraisal of others. For this reason, we invite future researchers to base their exploration of interpersonal trust disposition among individuals with BPD features on implicit measures rather than explicit, as we did.

Despite some limitations, we believe our contribution is relevant for both research and clinical purposes. In the research field, we failed to replicate previous findings of an association between untrustworthiness bias and BPD features. Such a replication failure does not necessarily mean that the original result is incorrect. We believe the failure in replication could be an opportunity for a better understanding of the phenomenon. More precisely, we believe follow-up experiments using implicit trust appraisal measures (rather than explicit) could be worth it. Indirect indicators of untrustworthiness bias (i.e., cooperative behaviors in economic games, reaction times

in appraising trustworthiness) may overcome the influence of social desirability on outcome measures. For this reason, we invite future researchers to widen the empirical investigation of untrustworthiness bias among the BPD clinical population. The present contribution provides researchers with two novel procedures for manipulating unknown others' trustworthiness: one direct (Study 1) and one indirect (Study 2). Importantly, in both studies, we found large effects of manipulation (i.e., trust appraisal of trustworthy ad untrustworthy identities were significantly different after the manipulation). The manipulation procedures developed could be useful tools for researchers investigating trust impairments among individuals with BPD features.

Furthermore, we also showed that the TG's virtual version has several benefits over traditional desktop displayed versions of the game. In particular, the enhanced realism and ecological validity obtained may provide opportunities to measure other relevant psychological phenomena occurring during economic, social encounters. In sum, the measure of interacting behavior during a VR TG could offer an alternative empirical tool for researchers investigating the relative roles of individual differences in the appraisal of others' trustworthiness.

Concerning clinical work, working through trust impairments with patients with high BPD features can be complicated. According to our findings, patients with BPD may tend to deny and devalue their propensity to lower trust at explicit levels of communication. Still, being unaware of untrustworthiness bias subtending BPD patients might be a peril for clinicians. Clinical work in the absence of reciprocal trust between clinicians and patients is usually unproductive and could be frustrating for clinicians (Fonagy et al., 2015). Hence, it may be helpful and practical for clinicians

to consider implicit as well as explicit indicators of untrustworthiness bias presented by patients with BPD features to isolate trust issues even when not noticeable.

Therefore, this contribution implies that researchers and clinicians should consider the relevance of disentangling the two levels of untrustworthiness bias indicators among individuals with BPD: explicit and implicit.

CHAPTER 6. BORDERLINE PERSONALITY DISORDER AND TRUST DISPOSITIONS DURING THE COVID-19 OUTBREAK .

Overview

After exploring the Prior belief and Disposition (in chapters 2, 3, and 4) and the Trust Appraisal (in chapter 5) stages of our model in the previous chapters, we focus on the Situation stage. We present a study that took place during the Coronavirus Disease 2019 (COVID-19) pandemic. Such an unprecedented situation offered us the opportunity to explore the effects of extraordinary circumstances on trust processes among individuals with varying levels of BPD features.

Authorities introduced exceptional interventions such as compulsory social distancing, closure of schools, and telework to deal with the COVID-19 global pandemic. Trust dispositions toward others and institutions might be particularly affected by pandemic situations. We expected that such a situation represents a potential risk factor for individuals with baseline dysfunctional appraisal of others' trustworthiness, such as individuals with BPD features. More precisely, we expected such an extraordinary situation to trigger further untrustworthy dispositions among individuals with high rates of BPD features. To test such a hypothesis, we developed a self-report instrument to assess individuals' interpersonal trust during the COVID-19 outbreak. The first study presented in the next chapter describes the procedure we followed for its development. It consists of four interpersonal trust-relevant scenarios that could occur during confinement. Individuals are asked to indicate how strongly they agree or disagree with the trust behavior described in each of the items. In the second study, a large community sample completed the assessment of BPD features,

suspiciousness, propensity to trust, interpersonal trust, and trust in the COVID-19 context. The results are manifold. According to previous literature, we found a significant difference in the propensity to trust and suspiciousness between high and low BPD features participants. Against our expectations, no significant difference occurred in interpersonal trust in both general and COVID-19 specific scenarios. In the discussion, we comment on such finding considering the role of pandemic situation.

1. Introduction

Since December 2019, the world healthcare community has had to deal with the spread of the Coronavirus Disease 2019 (COVID-19) caused by SARS-CoV-2. On March 11th, 2020 World Health Organization declared a global pandemic. The pandemic has placed an overwhelming burden on authorities to respond effectively and offer appropriate policies in reducing the transmission of the virus. Due to the lack of medical treatment, the WHO encouraged a social distancing policy to reduce infectious contacts. Governments of many western democracies implemented unprecedented interventions, including the closure of schools, compulsory telework, cancellation of public gatherings, mandatory quarantine of uninfected people without known exposure to SARS-CoV-2, and people's movement tracking.

Given human-to-human transmission and the consequent perception of other people as a disease source, the COVID-19 outbreak may have shaped social interactions' quality (not just quantity). Individuals' interpersonal problems are likely to have been substantially impacted by the lack of social contact resulting from confinement and other restrictions. Social-distancing policies restrict individuals' interpersonal connections and the experiences of trustworthy exchanges with others. Research showed a general mistrust and fear in response to the intensity of restrictive public policies and the "virtual invisibility" of COVID-19 (Betsch et al., 2020). Additionally, previous contributions found that individuals' interpersonal trust is a key predictor of cooperativeness, prosocial motivations, and compliance with COVID-19 measures (Jørgensen et al., 2020; Jørgensen et al., 2020b). Consequently, individuals with low interpersonal trust attitudes could be less prone to engage in protective behaviors than individuals with higher interpersonal trust.

Previous studies showed that low interpersonal and institutional trust causes a significant increase in mental health problems such as depression, loneliness, and suicidality (Qualter et al., 2010; Rotenberg et al., 2010). Within the COVID-19 outbreak context, Yao and colleagues documented the risk of suffering additional and increasing problems for people with prior mental disorders (Yao et al., 2020). Furthermore, in a recent review Preti, di Pierro, Fanti, Madeddu, and Calati (2020) highlighted that Personality Disorder patients are significantly affected by pandemic situations. In line with the authors, we expect that the COVID-19 context could be exceptionally arduous for those suffering from Borderline Personality Disorder (BPD) due to their interpersonal difficulties (Preti et al., 2020). BPD is a psychiatric condition characterized by marked interpersonal impairments (American Psychiatric Association, 2013). A core dimension of the unstable patterns of interpersonal behavior among individuals with BPD is the generalized belief that others have malevolent intentions and are untrustworthy (so-called untrustworthiness bias, Fertuck et al., 2013; Miano et al., 2013). Several studies showed an array of trust impairments among individuals with BPD (for a review, see Poggi et al., 2019).

Trust dispositions within the BPD population have been recently investigated in many ways. Empirical contributions mainly focused on a general tendency among individuals with BPD toward a generalized mistrust of others, namely the untrustworthiness bias. Early contributions explored such a tendency through explicit outcome measures such as the trust appraisal of neutral faces. Several studies found individuals with BPD feature appraised less trust than clinical and non-clinical controls (i.e., Fertuck et al., 2013; Richetin et al., 2018). Later, the BPD's untrustworthiness bias was explored by looking at implicit indices such as trust behavior in trust game

procedures. Several works found a significant decrease in cooperativeness, indicating mistrust, among individuals with BPD compared to clinical and non-clinical controls (i.e., King-Casas et al., 2008; Unoka et al., 2009). Nevertheless, trust in everyday life is not limited to trust appraisal of neutral faces or economic transactions. In a recent contribution, Botsford and colleagues aimed at providing an alternative measure of trust covering everyday real-life aspects of interpersonal trust. To this purpose, the authors developed a novel self-report questionnaire to assess interpersonal trust based on scenarios reflecting real-life interpersonal trust situations with a great variety of interacting partners and situational aspects (Interpersonal Trust Scenario Questionnaire, ITSQ; 2019). Furthermore, the authors tested the clinical validity of ITSQ on a sample of individuals with BPD, Major Depressive Disorder, and Seasonal Affective Disorder. Since the untrustworthiness bias is a core feature of BPD, the authors expected to find low scores at the ITSQ by BPD participants. Consistent with their hypothesis, Botsford and colleagues found low ITSQ scores providing evidence of untrustworthiness bias also in the context of real-life interpersonal trust scenarios (2019).

Regarding the COVID-19 situation, recent contributions show that protracted quarantine exacerbates poorer psychological outcomes, increased frustration, and demoralization (Brooks et al., 2020; Rona et al., 2007). Since patients with BPD often experience significant distress due to a perceived lack of agency and control over circumstances (Mortensen et al., 2016) and protracted quarantine is an “out-of-control” situation, the pandemic may exacerbate significant difficulties for individuals with BPD features. To our knowledge, no previous contribution explored the course of untrustworthiness bias among BPD individuals in such a challenging situation.

Based on previous evidence, we expected the COVID-19 outbreak could represent a potentially hazardous context for BPD's baseline untrustworthy dispositions toward others.

2. Aims of the contribution

During infectious disease outbreaks, the spread of an overall climate of fear and mistrust is prevalent (World Health Organization, 2016). The present study explores the influence of COVID-19 circumstance on the relationship between interpersonal trust and BPD features in a large non-clinical sample. We expected such a generalized and worldwide mistrust condition to trigger untrustworthy dispositions among individuals with high rates of BPD features. To test such a hypothesis, we needed to develop a self-report instrument to assess individuals' interpersonal trust during the COVID-19 outbreak. For this purpose, we took inspiration from the ITSQ that measures interpersonal trust in everyday situations. In the preliminary study, we assessed whether a series of scenarios depicting life during confinement involved interpersonal trust. We also tested them for clarity and comprehensiveness. We selected the best fitting scenarios. In the main study, we measured BPD features, individual differences in trust, interpersonal trust during everyday situations and during the COVID-19 outbreak in a large non-clinical community sample (N = 1052). This study aimed at exploring the COVID-19 outbreak's impact on the link between Interpersonal Trust and BPD features. Moreover, we also aimed to test the psychometric properties of the ITSQ Italian version.

More precisely, we first expected to replicate the two-factors structure of the ITSQ suggested by Botsford and colleagues (2019) with an Italian version of the measure. Second, we expected to find a negative association between BPD features

and Interpersonal Trust measures (both ITSQ and Interpersonal Trust in COVID-19 scenarios).

Finally, we expected to find different scores at trust measures comparing participants who endorsed a clinical number of BPD features to non-clinical participants. Besides ITSQ and interpersonal trust in COVID-19 scenarios, for the assessment of trust in our main study, we used two additional self-report measures (i.e., the Propensity to Trust Scale, PTS, Evans & Revelle, 2008; and the Suspiciousness facet of the PID-5, SFQ, Krueger, et al., 2011). Based on previous evidence, we expected participants with a clinical number of BPD features to score higher at the SFQ and lower at the PTS, ITSQ, and COVID-19 scenarios than participants with non-clinical BPD features.

3. Preliminary study

In the preliminary study, we aimed to develop a brief self-report measure of individuals' interpersonal trust during the COVID-19 outbreak. We first generated a large pool of scenarios describing every day trust behaviors during the lockdown circumstance. Then we tested the adequacy and comprehensiveness of these scenarios in a community sample (N = 30).

3.1. Methods

Participants and Procedure

We first asked five volunteers from the general population (three females and two males between 21 and 40 years of age) to write ten scenarios describing every day trust behaviors in the pandemic context. The items represent different COVID-19 interpersonal situations and corresponding trust behaviors. For instance, "You are grocery shopping for your grandparents. You meet a neighbor of theirs who offers to

deliver groceries to your grandparents in your place. He ensures that he will keep safely distant from your grandparents. You have had varied experiences trusting your grandparents' neighbors in the past. How strongly would you agree to the following statement? You thank him and let him deliver groceries to your grandparents". Next, we asked two clinical researchers (one male and one female) to select the ten items they believe best to assess interpersonal trust in the pandemic situation each. This first selection resulted in 16 items. Then, thirty participants (17 female, *M* age = 24, *SD* = 4.5) rated the adequacy of the 16 items for describing trust behaviors and the comprehensiveness of the formulations on 7 point Likert scales from 1 (*not at all*) to 7 (*completely*). The ethics committee of Milano Bicocca University approved the study protocol.

Statistical analyses and Results

We first computed each item's mean score on the adequacy to describe trust behaviors dimension. We preselected items with mean adequacy scores higher than 5, indicating a reasonable distance from the average scale score (i.e., 4). Seven items met such criterion, while nine were discarded. Table 1 includes a brief description of items and their mean score and *SD* on the adequacy dimension.

We then carried out one-sample *t*-tests against the average scale value to determine whether the seven items were statistically different from neutrality on the adequacy dimension. All the seven items with the highest adequacy were statistically different from the average value ($t_{(1,29)}$ ranging from 3.11 to 9.38, all $p < .004$).

Additionally, we run a one-sample *t*-test against the clarity dimension's average value. This analysis allowed the selection of highly comprehensible items. We

excluded three items because not statistically different from the central value indicating average comprehensibility (i.e., S6, S11, and S16).

Table 1

Preselected items statistics

	Adequacy *M*	Adequacy *DS*
S1: Walking neighbor's dog	5.10	1.52
S4: Hosting a friend	6.31	1.34
S6: Allowing cleaners to your home	5.87	1.52
S11: Skipping medical check-up	5.20	1.67
S12: Delegating grocery	5.63	1.43
S13: Keep visiting your sister	5.03	1.56
S16: Discharging grandmother from nursing home	5.10	1.47

In conclusion, from an initial pool of sixteen, we selected four adequate and comprehensible items for the assessment of Interpersonal trust during the COVID-19 outbreak. In table 2, we report detailed descriptive statistics and t-scores of the selected items. For the complete text (both English and Italian version) of the selected scenarios, see Appendix V.

Table 2

Scenarios statistics

	Adequacy				Clarity			
	M	*SD*	*t*	*p*	*M*	*SD*	*t*	*p*
S1: Walking neighbor's dog	5.10	1.52	3.11	.004	3.11	.004	3.69	<.001

S4: Hosting a friend	6.31	1.34	9.38	<.001	9.38	<.001	2.28	.03
S12: Delegating grocery	5.63	1.43	6.27	<.001	6.27	<.001	3.76	.001
S13: Keep visiting your sister	5.03	1.56	3.62	.001	3.62	.001	5.80	<.001

4. Main Study

In our main study, we aim at exploring interpersonal trust in the COVID-19 situation among individuals with varying levels of BPD features. The preliminary study was needed to develop an adequate pool of scenarios to assess interpersonal trust in the COVID-19 context.

To our knowledge, this is the first contribution using an Italian translation of the ITSQ. For this reason, we first expected to replicate the two-factors structure of the ITSQ suggested by Botsford and colleagues (2019) in the Italian version.

To assess BPD features, we use two approaches, a dimensional and a categorical approach. This dual approach for the assessment of BPD features is in line with the current hybrid Alternative Model for diagnosing BPD (i.e., categorical and dimensional) proposed in section III of DSM-5 (American Psychiatric Association, 2013). For the dimensional approach, we used the PAI-BOR (Morey, 2004) total score. Assessing dimensionally BPD, we hypothesized a negative association between BPD features, Interpersonal Trust measures (ITSQ and Interpersonal Trust in COVID-19 scenarios), and other trust measures (Propensity to trust and Suspiciousness) in line with existing research.

For the categorical approach, we defined likely BPD diagnosis based on the PAI-BOR cutoff scores (significant BPD symptoms; Morey, 2004). From this perspective, we expected participants who endorsed a clinical number of BPD features to score lower at trust measures than non-clinical participants.

4.1. Methods

Participants and Procedure

To rule out any effect of prolonged quarantine period or variations in the way participants coped with confinement, we introduced a time boundary to data collection. Survey dissemination started on April 8th, 2020, and ended on April 30th. One thousand one hundred and ninety participants (799 female, 391 male; *Mean* age = 27.65, *SD* = 8.3) completed an online survey voluntarily. All participants gave informed consent. We excluded data from participants who took less than 10 minutes to fill in the survey, the minimum time to complete the survey correctly with attention. Our final sample consisted of 1052 participants (703 female, 66.8%, *Mean* age = 27, *SD* = 7.4). The online survey consisted of a self-report assessment of BPD features and several measures of individual differences in trust-sensitive dimensions such as Propensity to Trust, Suspiciousness, Interpersonal Trust in everyday situations, and four scenarios derived from our preliminary study measuring Interpersonal Trust in the COVID-19. The ethics committee of Milano Bicocca University approved the study protocol.

4.2. Materials

Assessment of BPD features. We used the Personality Assessment Inventory–Borderline Feature Scale (PAI-BOR; Jackson & Trull, 2001). For a detailed description of the questionnaire, see paragraph 3.2. in the third chapter. In our sample, 327 participants obtained a score indicating elevated to clinical BPD features, and 725 participants obtained a non-clinical score. The reliability of the PAI-BOR total score in our sample was good (Cronbach's α =.86).

Assessment of suspiciousness. We used the Suspiciousness Facet Questionnaire (SFQ) derived from the Personality Inventory for DSM-5 (PID-5,

Krueger et al., 2011). For a detailed description of the questionnaire, see the third paragraph of the fourth chapter. The reliability of the SFQ was good (α =.74).

Assessment of the propensity to trust. We used the subscale trust of the Propensity to Trust Survey (PTS; Evans & Revelle, 2008). PTS consists of 7 items (i.e., "I believe that most people would lie to get ahead"). Participants indicated their agreement to each item on a scale ranging from 1 (*not at all*) to 5 (*completely*). The reliability of the PTS was good (α =.73).

Assessment of the Interpersonal Trust. We used the novel Interpersonal Trust Scenario Questionnaire (ITSQ; Botsford et al., 2019). The ITSQ presents respondents with eight interpersonal scenarios in which they can behave trustworthy or untrustworthy towards others. Participants expressed their agreement to the trust behavior described in each item on a scale ranging from 1 1 (*would not agree*) to 5 (*would completely agree*). The instrument showed acceptable psychometric properties (Cronbach's α = 0.72) in a German sample (Botsford et al., 2019). Additionally, the ITSQ distinguishes two separated components of interpersonal trust: entrusting known people and entrusting unknown people. Thus, after obtaining permission from the authors, we translated the items into Italian. A German mother-tongue and Italian speaker underwent a back-translation of the translated items into German. Discrepancies between the original items and the back-translations were analyzed, and adjustments were made to the Italian items accordingly. We verified the translation's adequacy to the original version (for the Italian version of the ITSQ, see appendix VI) through this process.

Assessment of the Interpersonal Trust during the COVID-19 outbreak. We used the four scenarios we pretested for adequacy and comprehensibility in the first

study. We asked participants to rate how strongly they agree or disagree with the trust behavior described in each of the four items COVID-19 specific on a scale from 1 (*would not agree*) to 5 (*would completely agree*). High scores indicate a higher inclination toward trust behavior in the specific context described by the scenario. We aimed to present participants with diverse situations related to trust in the peculiar case of confinement for the COVID-19 outbreak. Given minor inter-items correlations ($r < 0.25$), we did not aggregate scores of the four items in a total score. We will comment on each item separately.

4.3. Results

4.3.1. Test of the two-factor structure of ITSQ in the Italian version

We first examined the psychometric properties of the Italian version of the ITSQ. Botsford and colleagues (2019) suggested a two-factor structure (i.e., *entrusting known others* and *entrusting unknown others*) based on results of a first Exploratory Factor Analyses (EFA) then confirmed with a Confirmatory Factor Analysis (CFA) in an independent sample.

To explore whether our data replicated the two-factor structure, we conducted a Principal Component Analysis (PCA). Following Botsford and colleagues' (2019) procedure, we ran a PCA extracting two factors with oblimin rotation. However, since we found the two factors were uncorrelated ($r = 0.12$), we performed an orthogonal rotation (Varimax). Results indicated that the solution accounted for 49% of the total variance. The first factor explained 28% of the total variance and the second 21%. Factor loadings of each item on the Varimax rotated factors were consistent in pattern and magnitude with Botsford and colleagues' (2019) (see Table 3). All items of the ITSQ showed factor loadings of at least .61 on one of the two factors, replicating a

two-factor structure. According to the factorial form of ITSQ suggested by Botsford and colleagues (2019), items 1, 2, 3, 4, and 5 loaded onto the "entrusting known others" factor that explained 28% of the total variance. In contrast, items 6, 7, and 8 loaded onto the "entrusting unknown others" factor that explained 21% of the total variance. Cronbach's α of the total ITSQ scale was moderate (α = .63) as like as of the Unknown (α = .65) and Known (α = .64) subscales.

Table 3

Standardized factor loadings for the two-factor model and original factor loadings from Botsford et al. (2019)

Items	Entrusting Known		Entrusting Unknown	
	Our study	Botsford et al. (2019)	Our Study	Botsford et al. (2019)
Item 1: Lending a camera to someone	0.642	0.605		
Item 2: Lending money to someone	0.623	0.674		
Item 3: Lending a hard drive to someone	0.654	0.627		
Item 4: Letting someone take an important letter	0.611	0.546		
Item 5: Lending a sound system to someone	0.705	0.725		
Item 6: Going with a stranger			0.752	0.864
Item 7: Driving with a stranger			0.786	0.651
Item 8: Sleeping in a stranger's place			0.768	0.667

Note. Factor Loadings from Botsford and colleagues were obtained through an Exploratory Factor Analysis with Oblimin rotation. For the full procedure, see Botsford and colleagues (2019).

4.3.2. Associations between BPD features and self-report measures of trust.

We tested the associations between BPD features and measures of trust. Regarding ITSQ, BPD features were negatively associated with a general tendency to interpersonal trust. Looking at ITSQ factors separately, BPD features showed negative associations with the entrusting unknown other dimension and no correlation with the entrusting known others (Table 4).

Regarding correlations with the Interpersonal Trust COVID-19 Scenarios (i.e., Walking neighbor's dog; Hosting a friend; Delegating grocery; Keep visiting your sister), no significant correlation with BPD features occurred. The only exception was a weak negative association between BPD features and the "delegating grocery" scenario, indicating a lower propensity to entrusting grandparents' neighbor for delivering groceries to them (for the full text of scenario 12, see appendix V).

Interestingly we found correlations of BPD features with other trust measures that are consistent with previous findings. BPD features were negatively associated with the Propensity to Trust and positively with suspiciousness.

Regardless of BPD features, we found COVID-19 scenarios correlated with all trust self-report measures in the expected directions. We found positive associations between individual Interpersonal Trust, Propensity to Trust, and behavioral trust in COVID-19 scenarios. Reversely, suspiciousness was negatively associated with trustworthy actions in the scenarios (i.e., a lower tendency towards hosting a friend during lockdown).

Finally, ITSQ showed significant correlations with two dimensions related to trust in the expected direction: a general tendency to interpersonal trust was positively associated with PTS and negatively to suspiciousness

Table 4

Descriptive statistics and Correlations

	M	*SD*	1	2	3	4	5	6	7	8	9	10
1. PAI-BOR	53.7	11.31	1									
2. ITSQ	2.97	.64	-.07*	1								
3. ITSQ_S	3.63	.80	-.10***	.86***	1							
4. ITSQ_K	1.87	.88	.02	.63***	.14***	1						
5. SFQ	1.70	.48	.51***	-.22***	-.23***	-.08**	1					
6. PTS	1.23	.68	-.44***	.26***	.24***	.14***	-.63***	1				
7. S1	3.70	1.24	.03	.19***	.17***	.12***	-.12***	.11***	1			
8. S4	2.90	1.40	-.01	.29***	.25***	.18***	-.15***	.13***	.25***	1		
9. S12	2.25	1.43	-.08**	.17***	.16***	.09**	-.05	.09**	.06*	.16***	1	
10.S13	4.01	1.12	-.04	.25***	.24***	.12***	-.09**	.14***	.16***	.23***	.04	1

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

PAI-BOR = Personality Assessment Inventory–Borderline Feature Scale (Jackson & Trull, 2001);

ITSQ = Interpersonal Trust Scenario Questionnaire (Botsford et al., 2019);

ITSQ_S= Interpersonal Trust Scenario Questionnaire Stranger subscale (Botsford et al., 2019);

ITSQ_K= Interpersonal Trust Scenario Questionnaire Known subscale (Botsford et al., 2019);

SFQ = Suspiciousness Facet Questionnaire derived from the Personality Inventory for DSM-5 (Krueger et al., 2011).

PTS = Propensity to Trust Survey (Evans & Revelle, 2008)

4.3.3. Different trust disposition between high and low BPD features samples.

Based upon the evidence of decreased trust among individuals with BPD, we expected participants who endorsed a clinical number of BPD features to score differently at trust measures (ITSQ, PTS, and SFQ) compared to non-clinical (Fertuck et al., 2013; King-Casas et al., 2008; Seres et al., 2009). More precisely, we expected lower scores at the ITSQ and PTS measures and higher SFQ scores by individuals who endorsed a clinical number of BPD features. Such predictions were mainly theoretically driven. To test such a hypothesis, we first computed a dichotomous variable according to PAI-BOR cutoff scores (Morey et al., 2004). We included participants scoring 60 or more at the PAI-BOR in the clinical BPD features group. Those scoring 59 or less represented the non-clinical BPD features group. The high BPD sample consisted of 327 participants (226 female, 69%, *M* age = 26, *SD* = 6.3). The low BPD sample consisted of 725 participants (477 female, 66%, *M* age = 28, *SD* = 7.7).

To compare continuous outcomes variables from two independent samples (high BPD features participants and low), we conducted a multivariate ANOVA. In line with our expectations, high BPD individuals showed a significantly lower propensity to trust than low BPD features participants. Consistently, high BPD individuals showed significantly higher suspiciousness compared to low BPD features participants.

No significant difference at the ITSQ emerged between high BPD features participants and low BPD features.

Table 4

Trust indices of high BPD and low BPD sample

	High BPD sample (*N* = 327)		Low BPD sample (*N* = 725)				
	M	*SD*	*M*	*SD*	*F*	*p*	η^2
ITSQ	2.95	.64	2.97	.64	.21	.64	.00
SFQ	2.00	.47	1.56	.42	206.21	<.001	.16
PTS	.85	.62	1.40	.65	166.15	<.001	.13

5. Discussion

In the present work, we explored the impact and consequences of confinements due to COVID-19 on Interpersonal Trust. Furthermore, we aimed at providing a better understanding of the role of BPD features on trust behaviors in such an unprecedented situation. We used a novel self-report measure of Interpersonal Trust, the ITSQ, and tested the Italian version's factor structure through a Principal Component Factor Analysis. Based on the results, we replicated the two-factor simple model suggested by Botsford and colleagues (2019) in our sample too.

Additionally, we found theoretically meaningful associations of the Italian version of ITSQ with external measures, namely PTS and suspiciousness, establishing convergent and validity of the ITSQ with alternative trust measures. The consistency of these associations to previous conceptualizations of Interpersonal Trust may suggest that the Italian version of the ITSQ is a fruitful assessment tool for interpersonal trust behavior dispositions.

Regarding COVID-19 situations, we can see a significant association between ITSQ scores and interpersonal trust in COVID-19 scenarios, meaning that general interpersonal trust attitudes are associated with the individual disposition to behave trustworthy in confinement situations. Against our prediction, a similar correlation pattern does not occur with BPD features. This brings us to the second phenomenon we explored in the present study: the BPD's untrustworthiness bias trend during the COVID-19 outbreak. Due to the well-established diminished propensity to trust among individuals with BPD, we expected BPD features to predict a decrease in trust behaviors in the COVID-19 context too. More precisely, we expected a negative association between BPD features and interpersonal trust behavior disposition in the pandemic context (measured through scores at the COVID-19 scenarios). Results disconfirmed our expectations showing no associations between BPD features and trust behaviors in confinement circumstances. Additionally, we found no statistical difference between high BPD and low BPD individuals at ITSQ and COVID-19 scenarios scores. According to previous findings, individuals with high BPD features should have scored significantly lower than low BPD features individuals at ITSQ (Botsford et al., 2019; Botsford & Renneberg, 2020). Such a difference did not occur in our sample.

One possibility is that the uniqueness of confinement reduced social and physical contact with others. A sense of isolation from the rest of the world may have stimulated individuals' need for social connections, confounding the interpersonal measures we used. The physical distancing recommendations and orders to stay-at-home disrupted individuals' connectivity in structure, function, and quality. Previous findings suggest that in deprivation of interpersonal exchanges with significant others

circumstances, individuals engage in reparation of connectedness efforts (Provenzi & Tronick, 2020). According to Kaniasty, restoring a feeling of connectedness to others is essential for feeling supported during adversity (Kaniasty, 2012). Based on such suggestions, we argue that the scores at the COVID-19 scenarios and the ITSQ may have tap individuals' motivation to restore connectedness rather than their trust behavior dispositions. Participants in our study filled the survey while in confinement for at least one whole month (the Italian government introduced the lockdown policy on March 8th, and we inaugurated data collection on April 8th). In such a condition, responding to the ITSQ and the COVID-19 scenarios we developed, participants may have reported their urge for restoring connectedness to others rather than their disposition towards trusting others. Consistent with such an inference, we found high BPD features participants scored statistically different at general trust measures (i.e., PTS and SFQ) than low BPD features participants. Researchers should endorse indirect measures of interpersonal attitudes rather than self-report measures in future studies to overcome such limitations. Alternatively, future research may use alternative sources of information such as supervisory or peer ratings. Second, we selected a community sample, and it may have limited generalizability to clinical populations. Third, our sample was not homogeneous in terms of age and gender, and the individual experience of the COVID-19 situation may vary across gender and age. Due to higher mortality rates for men than women and older people compared to young (Jin et al., 2020; Jin et al., 2020b), we could expect the individual perception of COVID-19's threat may have varied.

Despite such limitations, we believe the present contribution has clinical implications. The interpretation of our findings related to behavioral trust dispositions

might be in line with recent contributions suggesting that people engage in behaviors to maximize the emotional and psychological connection with others to cope with the deprivation of interpersonal exchanges and "urgent demand for proximity to someone" (Ventura Wurman et al., 2020, pp. 13). We found similar trust behavior dispositions (ITSQ and COVID-19 scenarios) despite significant differences in general trust (PTS and Suspiciousness) between high and low BPD features participants. Such a general tendency toward increased disposition to connecting with others may also occur in the therapist-patient relationship. Patient's social isolation and increased need for social contact could be particularly challenging for clinicians. Clinicians, often forced to opt for remote therapy rather than face to face settings during COVID-19 outbreak, may feel pressured to provide increased contact and support to patients than in standard settings.

Furthermore, we found that in the COVID-19 situation, the expression of untrustworthiness bias (i.e., disposition towards trust behaviors) was the same across individuals with high and low BPD features despite their significant difference in general trust. Although our data do not directly address this, the imposed isolation situation may have caused variations in individuals' expression of their behavioral trust disposition towards others. Such interpretation of our findings is in line with recent contributions suggesting that, rather than static, personality should be considered a dynamic process whose expressions (i.e., behaviors) may vary over time across different situations (Fleeson & Noftle, 2012). Clinical descriptions of therapeutic work with BPD suggest that such variability in the expression of pathological features occurs in clinical settings too. For example, Gunderson (2007) observed that individuals with BPD fluctuate between extremes of idealization and

devaluation in their therapist's view. Facing such fluctuations could be challenging for clinicians who often navigate critical relational problems with BPD patients. We thus invite researchers to consider the role of situational variables on the expression of BPD's untrustworthiness bias. Besides uncovering empirical models integrating the structure of stable individual differences (i.e., untrustworthiness bias among individuals with BPD) and dynamic processes (i.e., reaction to situations) in the expression of mistrust among individuals with BPD, such a research field may inform clinical practice. We claim that clinicians may benefit from a deeper understanding of the situational factors that could stimulate BPDs' relational difficulties in the clinical setting as well in real life.

To sum up, we showed that the impact of the COVID-19 global pandemic and related factors is complex. Despite the well-established untrustworthiness bias linked to BPD pathology, we found similar interpersonal trust rates between individuals with high BPD features and low BPD features at the ITSQ and COVID-19 scenarios. We interpreted our findings based on the extensive influence that the COVID-19 situation showed in both the community and people with pre-existing mental health disorders (Brooks et al., 2020; Moreno et al., 2020). We suggest that individual expression of untrustworthiness bias may vary across different situations. Consequently, we invite both researchers and clinicians to pay special attention and care, evaluating the clinical presentation of BPD features by each different patient across situations. Individual differences in coping with the challenges posed by the COVID-19 global pandemic must be recognized and incorporated in patients' clinical management in these unprecedented times.

GENERAL DISCUSSION

Here we aimed at providing a better understanding of trust processes (i.e., appraisal and learning regarding others' trustworthiness) among individuals with Borderline Personality Disorder (BPD) features. In the last decades, psychological science has yielded significant advances in the understanding of interpersonal trust. For instance, social psychology theorists found that trust behaviors cannot be dissociated from the situation in which they unfold (Todorov, 2008; Willis & Todorov, 2013). Thus, from a social psychology perspective, it is necessary to consider the influence of situations on the appraisal of others' trustworthiness (Funder et al., 2012). Otherwise, economic theories and game theories inspired the growing field of neuroeconomics (Evans & Krueger, 2009). Theorists from the neuroeconomics field explore the psychological (i.e., motivation, affect, and cognition; Dunn & Schweitzer, 2005) and neural (i.e., brain circuits, hormones/neurotransmitters, and genes; Roberts et al., 2018) underpinnings of trust behaviors using economic exchange games (e.g., trust game; Krueger & Meyer-Lindenberg, 2019).

Yet, a conceptual framework integrating separate research findings into a comprehensive model of trust processes was still lacking. A wide variety of contributions coming from different theoretical frameworks focused on the investigation of interpersonal trust. We leveraged this diversity of conceptualizations to lay a conceptual foundation for subsequent theoretically robust investigations of trust impairments among individuals with varying levels of BPD features. Building upon previous empirically and theoretically driven conceptualizations of trust, we developed a multi-stage, heuristic model of normative trust processes in the first chapter. The proposed model aims to explain how several psychological components

shape trust behavior over time in an iterative, multi-step process. Such a model provided a framework to review how BPD patients differ from the general population concerning trust processing stages. According to the suggested model, the quality of early trust experiences (i.e., distal antecedents), as well as factors occurring moments before trust-relevant exchanges (i.e., proximal antecedents), influence actual and present trust appraisal of individuals. Furthermore, the model identifies the behaviors that trust appraisal predicts (such as cooperation) and how trust dispositions update according to the outcomes of social encounters (i.e., betrayal or, otherwise, validation of trust behaviors by others' actions).

We argue this model offers a relevant and systematized standpoint to study trust processes, providing an enhanced understanding of trust impairments across different diagnoses. Throughout the present dissertation, our specific focus was exploring variations from typical trust processes related specifically to BPD, a clinical condition associated with trust impairments (Herpertz & Bertsch, 2014; Poggi et al., 2019; Roepke et al., 2012). Our review of previous literature suggested dysfunctional processes at each stage of the proposed model among BPD individuals. Based on negative trust exchanges in early developmental stages with caregivers, individuals with BPD develop a rigid set of beliefs about others' untrustworthiness (e.g., Butler et al., 2002; Orme et al., 2019). Individuals with BPD features appraise less trust and behave less trustworthy in social contexts than non-BPD individuals (e.g., Fertuck et al., 2013; King-Casas et al., 2008). Finally, individuals with BPD show impairments updating information about others' trustworthiness based on the real-time quality of interactions (e.g., Abramov et al., 2020). Although reviewing main empirical findings on trust impairments in BPD we found empirical support for each stage of the model,

we believe our knowledge of processes underlying trust impairments among individuals with BPD features is still deficient. By systematizing previous empirical contributions, we disclosed the main limitations, inconsistencies, and uncovered areas on the topic. In the second part of the dissertation, we empirically addressed three under-investigated areas. First, we investigated the influence of trust-related individual differences on the associations between BPD and trust impairments (i.e., Rejection sensitivity in chapter 2, Justice Sensitivity in chapter 3, and suspiciousness in chapter 4). Second, we explored the relative importance of target characteristics in triggering trust impairments among individuals with BPD features (i.e., direct, static cues and indirect, dynamic cues of trustworthiness in chapter 5). Finally, we focused on the role of context and situations on the BPD-trust impairments link (i.e., Covid-19 outbreak in chapter 6).

In the second chapter, we presented an empirical investigation of the role of Rejection Sensitivity (RS) on trust appraisal in a non-clinical sample. In this study, we demonstrated that RS is a process through which BPD features connect to trust appraisal. More precisely, we found that only the emotional components of RS (i.e., anger and anxiety for plausible social rejection) mediate the BPD features-trust appraisal association, while RS's cognitive component (i.e., the expectation of being socially rejected) doesn't.

In the third chapter, we studied a personal disposition with theoretical overlaps to RS: Justice Sensitivity (JS). Based on the mediating role of RS on the BPD features-trust appraisal association found in the second chapter, we explored the role of JS in the context of trust impairments. Our results suggest the JS from the perspective of the victim and observer have opposite actions on the BPD features-trust appraisal

association. On one side, JS from the observer's perspective partly suppresses the negative association between BPD features and trust appraisal. On the other side, JS from the victim's perspective further strengthens the link between BPD features and decreased trust appraisal.

In the fourth chapter, we explored the self-concept ("an organized knowledge structure that contains traits, values and episodic and semantic memories about the self, and that controls the processing of self-relevant information"; Roepke et al., 2011, p. 149). Fluctuations in self-concepts are a core feature of BPD and often lead to dysfunctional evaluations of self, abrupt shifts in relationships with others, work, and life in general. Relying on an incoherent self-concept could be one underpinning of BPDs' tendency to attribute untrustworthy, hurtful, and neglectful intentions to significant others, namely suspiciousness (i.e., "expectations of and sensitivity to signs of interpersonal ill-intent or harm; having doubts about others' loyalty and fidelity; feelings of persecution"; Skodol et al., 2011, p. 38). We found that BPD features predicted an increased distorted and suspicious representation of others that, in turn, predicted a distorted and malevolent representation of the self. This result provided insight into the suspiciousness pathway through which BPD features translate into a malevolent explicit representation of the self.

Taken together, these first three studies shed light on potentially risky and protective personal dispositions underlying interpersonal trust behaviors among individuals with BPD. Describing the unique contribution carried by RS, JS, and Suspiciousness, we add to the previous knowledge of individual differences relevant to BPD's interpersonal trust. Concerning RS, several studies suggested it is a personal disposition with strong associations to BPD features (e.g., Staebler et al., 2011) and

fear of abandonment (Arntz et al., 1999; Arntz, Dreessen, et al., 2004; Ayduk et al., 2008). Our contribution showed that RS is also relevant for dysfunctional trust processing among individuals with BPD features. More precisely, BPD features translate in lower trust appraisal of neutral faces through RS's emotional components (i.e., anger and anxiety for rejection) only and not of the cognitive one (i.e., expectation of rejection).

Regarding JS, the interplay occurring between JS and BPD features was little explored by previous researchers. Just one recent contribution by Lis and colleagues (2018) showed that BPD individuals are prone to develop cognitive, emotional, and behavioral reactions to injustice experiences, which may weigh on interpersonal dysfunctions. More precisely, the researchers investigated the relationship between BPD features and JS, finding evidence of a positive correlation between victim and observer JS and BPD features (Lis et al., 2018). Based on our findings, we suggest that concerns about others' sense of Justice among individuals with BPD could lead to the perception of others' untrustworthiness or uncooperative behaviors, namely untrustworthiness bias.

Furthermore, our findings from the fourth chapter shed light on a maladaptive interplay between a malevolent implicit self-concept and a suspicious representation of others. More precisely, we found that BPD features predict an increased distorted and suspicious representation of others that, in turn, predicts a distorted and malevolent representation of the self. Moreover, the suspiciousness' mediation of the relation between BPD features and explicit malevolent self-concept is significant for participants with an highly malevolent implicit self-concept (and not for participants with an averaged or low malevolent implicit self-concept). This result is in line with

the ORT conceptualization of BPD as a pathology that roots in a lack of integration of positive and negative representations of self and others due to the massive use of the splitting defensive mechanism (Stern et al., 2018).

Based on our findings from the second, third, and fourth chapters, we can conclude that individuals with BPD features display different sensitivities to social dimensions (i.e., RS, JS, and suspiciousness). Furthermore, such sensitivities lead to diverse maladaptive interpersonal outcomes such as untrustworthiness bias or uncooperative behaviors.

In the fifth chapter, we moved our focus to explore what may trigger untrustworthiness bias among individuals with BPD. To this purpose, we developed two separate studies using different trustworthiness cues: direct and static cues in the first study and indirect and dynamic cues in the second study. Surprisingly, we found no association between BPD features and trust appraisal in response to neither direct nor indirect cues of trustworthiness. Our failure in replicating previous findings of an association between BPD features and trust appraisal could be mainly methodologically driven. More precisely, we believe that the outcome measure we used may have weighed on results. Using an explicit self-report measure of trust appraisal after the exposure to direct and indirect cues, participants could have had some degree of control over their responses. Nonetheless, such findings are meaningful and suggest that BPD's untrustworthiness bias expression may vary over time, and the procedure used to measure it.

Finally, since the initial conceptualizing of this dissertation, the world has dramatically changed due to the Covid-19 pandemic. To measure trust behavioral dispositions during confinement due to lockdown measures, we developed a scenario-

based measure. We recruited a non-clinical sample to explore the influence of Covid-19 circumstances on the relationship between interpersonal trust and BPD features. Interestingly, we found similar trust rates at the scenario-based measure of trust dispositions between individuals with high BPD features and low BPD features despite significant differences in trait trust measures (i.e., Propensity to Trust and Suspiciousness). In line with the previous chapter, such findings suggest that individuals' state trust dispositions are not static and highly sensitive to situations. At the same time, BPD's untrustworthiness at the trait level is more stable across situations.

Referring to the proposed model, the results obtained through the empirical contributions in this thesis allow achieving a more comprehensive knowledge of the BPD's impairments at the stages of prior beliefs and dispositions (chapters 2, 3, and 4), trust appraisal (chapter 5), and situation perception (chapter 6). Moreover, we believe that the studies presented in this contribution lead to interesting implications for clinicians and researchers.

1. Implications for research

In the systematic review of literature presented in the first chapter, we showed that impairments in trust processes are theoretically and diagnostically central to BPD. Lower interpersonal trust among individuals with BPD is associated with interpersonal impairments and other adverse clinical outcomes (Jeung & Herpertz, 2014). Further, interpersonal impairments play an important role in the prognosis and course of BPD diagnosis (Links & Heslegrave, 2000). Although all these contributions suggest that trust impairments are extremely relevant in BPD, based on the existing body of research, the processes underpinning such impairments were still quite unclear. In the

present dissertation, we provided some contributions to our knowledge about trust impairments among individuals with BPD features. However, due to the complexity of trust dynamics, further empirical work is still needed. In fact, some limitations in our empirical contributions exist and should be addressed in future research to further advance our understanding of trust impairments.

First, given the cross-sectional nature of our studies, conclusions regarding the course of untrustworthiness bias over time are not possible. For this reason, we invite future researchers to move from cross-sectional and self-report designs to more sophisticated procedures using more ecologically valid methods. Although interpersonal trust impairments have been usually regarded as a stable feature of BPD, our findings from the fifth and sixth chapters suggest that trust disposition may change across situations and over time. Future research aiming to characterize trust impairments in BPD should examine the situations that elicit such dysfunctions. A deeper understanding of situations that trigger trust impairments would provide more accurate and informative indicators of such impairments from a diagnostic perspective. To this aim, the Ecological Momentary Assessment (EMA) approach could be promising. In EMA procedures, participants answer prompts on their smartphones at various points throughout an established period (i.e., a few days, one week, or even a month). This approach allows for studying psychological phenomena in individuals' "natural habitats". Participants provide self-report measures while exposed to many environmental and interpersonal factors that usually occur in their everyday lives, which cannot be resembled in laboratory procedures (Wilhelm & Roth, 1998). Such assessments involving participants in their daily environments confer ecological validity to measures. Furthermore, researchers can eventually capture within-person

fluctuations of individual dispositions such as interpersonal trust disposition through EMA procedures. For instance, EMA could be applied to the investigation of trust course in therapy relationships. The deployment of an EMA approach to the study of trust course while under treatment may facilitate the screening and monitoring of the evolution of trust towards the therapist in BPD patients. In other words, we believe that the use of EMA procedures may enable timely assessment of trust issues and support the collection of ecologically valid and longitudinal data about dysfunctional processing of trust toward the therapist among individuals with BPD features. It would also be interesting to apply EMA to assess typical trust dynamics of individuals with BPD in everyday relationships with significant others. Traditionally, information on the quality of individuals' relationships derives from self-report questionnaires (Joel et al., 2020). Compared to traditional self-report measures, EMA procedures provide increased validity by timely asking for reports on recent specific social interactions rather than remembered mental aggregates as in self-report questionnaires (e.g., reporting social behaviors or emotions over the last week).

Second, although trust processes have been addressed in many domains (psychology, neuroscience, sociology, behavioral economics, etc.), we used the two most used procedures to measure trust, self-report questionnaires and investments in an economic game (Hale et al., 2018). Nonetheless, both methods have some limitations. Self-report measurements can be susceptible to individual interpretations of the question posed, and participants may not always have access to their inner world (Ben-Ner & Halldorsson, 2010; Hahn & Goedderz, 2020). Regarding investment behavior, previous studies suggest it is highly sensitive to stable individual characteristics, such as altruism (Ashraf et al., 2006) and risk-seeking (Schechter,

2007) that could be considered as proxies for trust. For these reasons, it might be that, nor self-report ratings of one's trust dispositions nor investment behaviors provide "pure" measurements of trust dispositions. Thus, we invite future researchers to design novel procedures for measuring trust. To this purpose, we suggest Virtual Reality (VR) technologies could be a promising field. VR can offer the opportunity to closely replicate every day trust situations under easy to control conditions conferring ecological and construct validity to trust measures. Moreover, measuring trust behavior in VR technology would allow researchers to tap both implicit trust responses (such as the time for entrusting counterparts' advice in the virtual space or gaze direction in the virtual space) and explicit ones (i.e., trust behaviors). In this direction, Hale and colleagues recently developed a novel tool for measuring trust, namely the Virtual Maze (Hale et al., 2018). In the Virtual Maze task, participants navigate through a maze and can choose to trust or not trust the virtual characters they meet along the maze about which way to proceed. With this procedure, the researchers provided a more ecological measurement of trust that reflects both implicit (i.e., time occurring before deciding to ask for advice) and explicit (i.e., following or not following advice) aspects of trust processes. We invite future researchers to apply the Virtual Maze or develop new procedures for measuring trust and deceive limitations of traditional measures of trust.

Furthermore, the influence of other mental health conditions on trust processes could be explored. BPD is a clinical condition with particularly high comorbidity rates with other disorders (e.g., post-traumatic stress disorder, substance use disorders, avoidant PD, narcissistic PD, etc.). Due to high comorbidities, it is challenging to analyze the specificity of BPD features of trust impairments. It is plausible that when

certain pathological features coexist with BPD features, trust impairments' expression change compared to clinical presentations among individuals with solely BPD features. It would be beneficial to determine the degree to which trust impairments are BPD-specific and characterize BPD compared to other disorders and other Personality Disorders in particular. Parsing apart BPD-specific trust impairments from difficulties that are more broadly associated with different clinical conditions would be fruitful for differential diagnostic purposes. In a recent review of the literature, we first attempted to address the BPD's specificity of trust impairments compared to other Personality Disorders (Poggi et al., 2019). We compared the main findings on trust impairments among individuals with BPD, Narcissistic, and Paranoid Personality Disorder. Concerning narcissism, results suggest that mistrust toward others might provoke aggressive reactions and outbursts (Kwiatkowska et al., 2019). Regarding paranoid features, findings indicate that trust appraisal decreases as a function of skepticism (Furnham & Crump, 2015). Nonetheless, considering the modest number of empirical studies on untrustworthiness biases in other Personality Disorders, there is a need for further studies exploring trust dispositions in clinical conditions different from BPD to state more firm conclusions.

For this reason, we invite future researchers to extend investigations on trust impairments in other Personality Disorders or to include control groups covering the span of diagnostic overlap with BPD in those investigations focused on specific BPD's impairments. Furthermore, we invite researchers to employ clinician-rated measures (e.g., SCID-5 PD) for accurate BPD features' assessments. One of our studies' main limitations was using self-report scales to assess BPD features in our community samples (i.e., BPDCL; PAI-BOR). Self-report measures of pathological features echo

the individual's unique perspective on his or her typical functioning. Since people with clinical BPD features showed inaccurate evaluations of themselves and others (i.e., negativity bias; Nicol et al., 2013), their self-report assessments might be biased.

Finally, our findings would need to be replicated in clinical samples and larger and socially and ethnically diverse populations. Because we used non-clinical samples in all the empirical studies presented, the generalizability of our findings to clinical populations should be tested in clinical samples. Besides, we recruited mainly Italian participants in our studies. Finally, since interpersonal trust attitudes vary as a function of ethnicity (e.g., Soroka et al., 2003; Stolle et al., 2008), it would be interesting replicating our studies in ethnically diverse populations (e.g., non-Western samples).

2. Clinical implications

In this work, we proposed a novel model to systematize impairments in trust processes in BPD. Then, we presented five separate empirical contributions exemplifying some of the alterations in such processes typical of individuals with BPD features. Dysfunctional judgments of others' trustworthiness have been investigated as a plausible cause of BPD's interpersonal impairments (Duan et al., 2020). According to a longitudinal study, improvements in interpersonal functioning (and, therefore, trust) are more challenging for BPD patients than other clinical conditions and persist even after ten years in 15% to 25% of individuals with BPD (Choi-Kain et al., 2010). We believe that the coherent model of trust we provided in this contribution may help clinicians develop a practice-level model to facilitate trust functioning among patients with BPD features. As already mentioned, the studies presented involved non-clinical participants only, limiting the generalizability to clinical

populations with BPD or patients showing BPD features despite a lack of clinical BPD diagnosis. However, results may cautiously inspire some clinical considerations.

We shape the clinical considerations of the present work resembling the main evolutionary steps of psychodynamic theories. According to contemporary theorists, two important developments have occurred in the evolution of psychodynamic theory and practice. The first is the shift from a monadic theory of mind (one-person psychology) to an interactional relational theory of mind (two-person psychology). The second is the evolution of postmodern approaches that pose our experience of reality as constructed by culture and society (Sandler, 1992). Likewise, we progressively moved from exploring intrapsychic phenomena towards cultural and social factors passing through interpersonal components in the present contribution.

The Freudian orthodox perspective of one-person psychology assumes that the fundamental unit of study is the individual, and interpersonal events are explained through intrapsychic concepts (such as individual drives and defenses; Freud, 1917, 1958). Similarly, in our second and third chapters, we assumed the individual as a focal unit and highlighted the specific contribution of some intrapsychic dispositions on trust impairments. Based on our findings, diverse individuals' dispositions can be associated with trust or mistrust outcomes. For example, in the second chapter, we found that BPD features predict untrustworthiness bias through individual sensitivity to social rejection. In the third chapter, we found that individual differences in sensitivity to injustice may account for diverse trust appraisal outcomes. If taken together, these findings may suggest alternative ways to address interpersonal trust issues in clinical settings rather than working on trust issues directly. Our results seem to indicate that clinicians may secondarily trigger interpersonal trust changes by

focusing clinical work on trust-related personal dispositions and, eventually, reduce the psychological pain provoked by interpersonal trust failures.

Afterward, two-persons theorists conceptualized the behavior resulting from the interactions between internal, intrapsychic events in one's mind and external, relational phenomena (Ainsworth, 1991; Sullivan, 1953). Accordingly, in the fourth and fifth chapters, we moved to a two-person perspective considering BPD's trust impairments from a relational perspective. In the fourth chapter, we exemplified the path through which detrimental and incoherent intrapsychic representations translate to the interpersonal world. We found evidence that individuals with BPD rely on an unintegrated polarized intrapsychic representation of the self, and through projective identification, they develop a suspicious interpersonal disposition. These results align with the Object Relations Theory (ORT) conceptualization of interpersonal difficulties among individuals with a Borderline Personality Organization. According to ORT, such a dysfunctional projective identification process may easily culminate in pathological misperception and misattribution of malicious motives to others, including the therapist (Stern et al., 2018). As a consequence of such misattribution of malevolent motives to the therapists, individuals with BPD may engage in self-destructive actions or indirect attacks to the clinical setting as a means of defeating the malignant therapist (Kernberg, 1984; Yeomans et al., 2002, 2013). In clinical practice, these maladaptive transaction circles usually occur in the form of paranoid transference. According to ORT, implicit attacks to the setting usually happen in the context of the deepening of transference, when the patient begins to experience real bond and dependency toward the therapist. We suggest that the therapist's

interventions should help the patient get aware of the split representations underpinning their drive dispositions to mistrust to defend against others.

From a clinical standpoint, the shift from one-person psychology towards two-person conceptualizations established that the therapist is no longer considered a neutral, objective observer, but rather a participant in the analytic process (Levine, 1994). Such a shift in the view of the therapeutic situation also changed the way of using the most characteristic psychodynamic technique: interpretation (Higa & Gedo, 2012; Lyons-Ruth et al., 1998). Increased interest in both therapists' and patients' contributions to the therapeutic relationship is reflected in interpretations directed to interpersonal processes rather than predominantly on resistances (such as traditional interpretation techniques). Consistently with ORT, we suggest that transference interpretations could be among the most effective techniques to help patients dealing with interpersonal difficulties. Transference interpretations may offer the patients opportunities to integrate inconsistent self and object representations, leading to observable changes in behavior and the quality of interpersonal relationships (Yeomans et al., 2015). When patients enact unintegrated internal representations of the self and objects in the therapeutic situation, clinicians should focus on the intrapsychic mental representations behind their actions through transference interpretations. We believe this therapeutic strategy could break the vicious circle leading individuals with BPD to untrustworthy dispositions towards others. Nonetheless, the validity of such clinical inference needs to be further corroborated in future studies.

Finally, postmodern psychodynamic theories posit that culture, society, and situations inform the configuration of the self (Greenberg, 1983; Singer et al., 1989).

In our last chapter, we recalled such a shift in the study of psychological functioning exploring the influence of social context (i.e., Covid-19 pandemic). The social and economic uncertainty derived from the pandemic allowed us to examine the impact of a generalized and worldwide mistrust condition on BPD's untrustworthy dispositions. Interestingly, we found similar interpersonal trust appraisal rates between individuals with high BPD features and low BPD features despite a significant difference at the trait trust level. From a clinical standpoint, this suggests that social context may influence the clinical presentation of BPD features. Therefore, we stress the importance of considering the context in clinical work and the need for setting interventions aiming to support or change the environment in which patients live (e.g., intervention with relatives of the patient). Based on our findings, we claim that clinicians should pay special attention to assessing the social environment where patients are embedded.

Currently, the traditional view of the self as a coherent, enduring psychological structure is obsolete. In the words of Stern, "The self is not unitary but multiple, not static but in flux, not a separate centre of initiative, but intersubjectively constituted" (Stern, 2002, p. 694). Looking at the present dissertation as a whole, we offered an integrated perspective that combines one-person, two-person, and postmodern angles to explore mistrust clinical expressions among individuals with BPD features.

3. Conclusions

In the first chapter of the present work, we proposed a conceptual framework integrating separate research findings on interpersonal trust into a comprehensive model. Throughout the dissertation, we examined and found support for atypical processing at some stages of the model in individuals with BPD features (i.e., prior

dispositions, the influence of the situation, and trust appraisal). Also, we commented on the main clinical implication of our findings in each contribution. This achievement may suggest that the model is quite complete and useful both from an empirical and clinical perspective.

To conclude, we added to the previous knowledge of interpersonal trust processes among individuals with BPD features and provided a practical, comprehensive trust process model. Nonetheless, we believe there is still a need for more work to integrate such a model. We claim that the proposed model might guide future investigations to advance our understanding of interpersonal trust and encourage clinicians to develop efficient strategies to restore interpersonal trust among patients with BPD features.

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. https://doi.org/10.1371/journal.pone.0236170

Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Sciences*, *3*(12), 469–479. https://doi.org/10.1016/S1364-6613(99)01399-6

Ainsworth, M. D. (1989). Attachments beyond infancy. *The American Psychologist*, *44*(4), 709–716. https://doi.org/10.1037//0003-066x.44.4.709

Ainsworth, M. D. (1991). An Ethological Approach to Personality Development. *American Psychologist*, *46*(4), 333–341. https://doi.org/10.1037/0003-066X.46.4.333

Ainsworth, M. D., & Bowlby, J. (1991). An ethological approach to personality development. *American Psychologist*, *46*(4), 333–341. https://doi.org/10.1037/0003-066X.46.4.333

Allen, J. G., Fonagy, P., & Bateman, A. W. (2008). *Mentalizing in clinical practice*. American Psychiatric Pub.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub. https://doi.org/10.1176/appi.books.9780890425596

Anderson, J. L., Sellbom, M., Bagby, R. M., Quilty, L. C., Veltri, C. O. C., Markon, K. E., & Krueger, R. F. (2013). On the Convergence Between PSY-5 Domains and PID-5 Domains and Facets: Implications for Assessment of DSM-5 Personality Traits. *Assessment*, *20*(3), 286–294. https://doi.org/10.1177/1073191112471141

Arntz, A., Dietzel, R., & Dreessen, L. (1999). Assumptions in borderline personality disorder: specificity, stability and relationship with etiological factors. *Behaviour Research and Therapy*, *37*(6), 545–557. https://doi.org/10.1016/S0005-7967(98)00152-1

Arntz, A., Dreessen, L., Schouten, E., & Weertman, A. (2004). Beliefs in personality disorders: a test with the Personality Disorder Belief Questionnaire. *Behaviour*

Research and Therapy, *42*(10), 1215–1225. https://doi.org/https://doi.org/10.1016/j.brat.2003.08.004

Arntz, A., & Veen, G. (2001). Evaluations of others by borderline patients. *The Journal of Nervous and Mental Disease*, *189*(8), 513–521.

Ashraf, N., Bohnet, I., & Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, *9*(3), 193–208. https://doi.org/10.1007/s10683-006-9122-4

Axelrod, R., & Dion, D. (1988). The Further Evolution of Cooperation. In *Science* (Vol. 242, Issue 4884).

Ayduk, O., May, D., Downey, G., & Higgins, E. T. (2003). Tactical Differences in Coping With Rejection Sensitivity: The Role of Prevention Pride. *Personality and Social Psychology Bulletin*, *29*(4), 435–448. https://doi.org/10.1177/0146167202250911

Ayduk, O., Zayas, V., Downey, G., Cole, A. B., Shoda, Y., & Mischel, W. (2008). Rejection sensitivity and executive control: Joint predictors of borderline personality features. *Journal of Research in Personality*, *42*(1), 151–168. https://doi.org/10.1016/j.jrp.2007.04.002

Bach, B., Sellbom, M., Bo, S., & Simonsen, E. (2016). Utility of DSM-5 section III personality traits in differentiating borderline personality disorder from comparison groups. *European Psychiatry*, *37*(September 2018), 22–27. https://doi.org/10.1016/j.eurpsy.2016.04.006

Barefoot, J. C., Maynard, K. E., Beckham, J. C., Brummett, B. H., Hooker, K., & Siegler, I. C. (1998). *Trust , Health , and Longevity*. *21*(6), 517–526. https://doi.org/10.1023/A:1018792528008

Barrowclough, C., Tarrier, N., Humphreys, L., Ward, J., Gregg, L., & Andrews, B. (2003). Self-esteem in schizophrenia: Relationships between self-evaluation, family attitudes, and symptomatology. *Journal of Abnormal Psychology*, *112*(1), 92–99. https://doi.org/10.1037/0021-843X.112.1.92

Bartz, J., 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

Bateman, A., Campbell, C., Luyten, P., & Fonagy, P. (2018). A mentalization-based approach to common factors in the treatment of borderline personality disorder. *Current Opinion in Psychology*, *21*, 44–49. https://doi.org/10.1016/j.copsyc.2017.09.005

Bateman, A., & Fonagy, P. (2010). Mentalization based treatment for borderline personality disorder. *World Psychiatry*, *9*(1), 11–15. https://doi.org/10.1002/j.2051-5545.2010.tb00255.x

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. In *Psychological Bulletin* (Vol. 117, Issue 3, pp. 497–529). American Psychological Association. https://doi.org/10.1037/0033-2909.117.3.497

Beck, A. T., Butler, A. C., Brown, G. K., Dahlsgaard, K. K., Newman, C. F., & Beck, J. S. (2001). Dysfunctional beliefs discriminate personality disorders. *Behaviour Research and Therapy*, *39*(10), 1213–1225. https://doi.org/10.1016/S0005-7967(00)00099-1

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

Bender, D. S., & Skodol, A. E. (2007). Borderline personality as a self-other representational disturbance. *Journal of Personality Disorders*, *21*(5), 500–517. https://doi.org/10.1521/pedi.2007.21.5.500

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

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142.

Berry, K., Wearden, A., Barrowclough, C., & Liversidge, T. (2006). Attachment styles, interpersonal relationships and psychotic phenomena in a non-clinical student sample. *Personality and Individual Differences*, *41*(4), 707–718. https://doi.org/10.1016/j.paid.2006.03.009

Betsch, C., Wieler, L. H., & Habersaat, K. (2020). Monitoring behavioural insights

related to COVID-19. *The Lancet, 395*(10232), 1255–1256. https://doi.org/10.1016/S0140-6736(20)30729-7

Birkás, B., Dzhelyova, M., Lábadi, B., Bereczkei, T., & Perrett, D. I. (2014). Cross-cultural perception of trustworthiness: The effect of ethnicity features on evaluation of faces' observed trustworthiness across four samples. *Personality and Individual Differences*, *69*, 56–61. https://doi.org/10.1016/j.paid.2014.05.012

Black, D. W., Blum, N., Pfohl, B., & Hale, N. (2004). Suicidal behavior in borderline personality disorder: Prevalence, risk factors, prediction, and prevention. *Journal of Personality Disorders*, *18*(3), 226–239. https://doi.org/10.1521/pedi.18.3.226.35445

Blais, M. A., Hilsenroth, M. J., & Fowler, J. C. (1999). Diagnostic efficiency and hierarchical functioning of the DSM-IV borderline personality disorder criteria. *The Journal of Nervous and Mental Disease*, *187*(3), 167–173.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*(2), 103–124. https://doi.org/10.1207/S15327965PLI1302_01

Bohus, M., Kleindienst, N., Limberger, M. F., Stieglitz, R. D., Domsalla, M., Chapman, A. L., Steil, R., Philipsen, A., & Wolf, M. (2009). The short version of the Borderline Symptom List (BSL-23): Development and initial data on psychometric properties. *Psychopathology*, *42*(1), 32–39. https://doi.org/10.1159/000173701

Bondü, R. (2018). Is bad intent negligible? Linking victim justice sensitivity, hostile attribution bias, and aggression. *Aggressive Behavior*, *44*(5), 442–450. https://doi.org/10.1002/ab.21764

Bosson, J. K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology*, *79*(4), 631–643. https://doi.org/10.1037/0022-3514.79.4.631

Botsford, J., & Renneberg, B. (2020). Autobiographical memories of interpersonal trust in borderline personality disorder. *Borderline Personality Disorder and*

Emotion Dysregulation, *7*(1), 1–10. https://doi.org/10.1186/s40479-020-00130-w

Botsford, J., Schulze, L., Bohländer, J., & Renneberg, B. (2019). Interpersonal Trust: Development and Validation of a Self-Report Inventory and Clinical Application in Patients With Borderline Personality Disorder. *Journal of Personality Disorders*, *182*, 1–22. https://doi.org/10.1521/pedi_2019_33_462

Bowlby, J. (1973). Attachment and loss: Volume II: Separation, anxiety and anger. In *Attachment and Loss: Volume II: Separation, Anxiety and Anger* (pp. 1–429). London: The Hogarth press and the institute of psycho-analysis.

Bowlby, J. (1988). *A secure base: Clinical applications of attachment theory* (Vol. 393). Taylor & Francis.

Brennan, K. A., Clark, C. L., & Shaver, P. R. (1998). *Self-report measurement of adult attachment: An integrative overview.*

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 and Life-Threatening Behavior*, *36*(3), 313–322. https://doi.org/10.1521/suli.2006.36.3.313

Bromiley, P., & Cummings, L. L. (1995). Transactions Costs in Organizations with Trust. *Research on Negotiations in Organizations*, *5*(January 1995), 219–247.

Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., & Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it: rapid review of the evidence. *The Lancet*, *395*(10227), 912–920. https://doi.org/10.1016/S0140-6736(20)30460-8

Buchan, N. R., Croson, R. T. A., & Solnick, S. (2008). Trust and gender : An examination of behavior and beliefs in the Investment Game. *Journal of Economic Behavior & Organization*, *68*, 466–476. https://doi.org/10.1016/j.jebo.2007.10.006

Burns, C., & Conchie, S. M. (2015). Measuring implicit trust and automatic attitude activation. In *Handbook of research methods on trust*. Edward Elgar Publishing.

Butler, A. C., Brown, G. K., Beck, A. T., & Grisham, J. R. (2002). Assessment of dysfuntional beliefs in borderline personality disorder. *Behaviour Research & Therapy*, *40*, 1231–1240.

Cameron, C. (2003). *Behavioral game theory: experiments in strategic interaction*. 2018.

Chanen, A., Jovev, M., McCutcheon, L., Jackson, H., & McGorry, P. (2008). Borderline Personality Disorder in Young People and the Prospects for Prevention and Early Intervention. *Current Psychiatry Reviews*, *4*(1), 48–57. https://doi.org/10.2174/157340008783743820

Chang, B., Sharp, C., & Ha, C. (2011). The criterion validity of the borderline personality features scale for children in an adolescent inpatient setting. *Journal of Personality Disorders*, *25*(4), 492–503. https://doi.org/10.1521/pedi.2011.25.4.492

Choi-Kain, L. W., Zanarini, M. C., Frankenburg, F. R., Fitzmaurice, G. M., & Reich, D. B. (2010). A longitudinal study of the 10-year course of interpersonal features in borderline personality disorder. *Journal of Personality Disorders*, *24*(3), 365–376.

Clark, L. A. (2005). Temperament as a unifying basis for personality and psychopathology. *Journal of Abnormal Psychology*, *114*(4), 505–521. https://doi.org/10.1037/0021-843X.114.4.505

Clarkin, J. F., Hull, J. W., & Hurt, S. W. (1993). Factor structure of borderline personality disorder criteria. *Journal of Personality Disorders*, *7*(2), 137–143. https://doi.org/10.1521/pedi.1993.7.2.137

Coid, J., Yang, M., Tyrer, P., Roberts, A., & Ullrich, S. (2006). Prevalence and correlates of personality disorder in Great Britain. *British Journal of Psychiatry*, *188*(MAY), 423–431. https://doi.org/10.1192/bjp.188.5.423

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*(December 2018), 609–617. https://doi.org/10.1016/j.psychres.2018.12.102

Costa Jr., P. T., & McCrae, R. R. (2008). The Revised NEO Personality Inventory (NEO-PI-R). In *The SAGE handbook of personality theory and assessment, Vol 2: Personality measurement and testing.* (pp. 179–198). Sage Publications, Inc. https://doi.org/10.4135/9781849200479.n9

Crick, N. R., Murray-Close, D., & Woods, K. (2005). Borderline personality features

in childhood: A short-term longitudinal study. *Development and Psychopathology*, *17*(4), 1051–1070. https://doi.org/10.1017/S0954579405050492

Cross, S. E., Bacon, P. L., & Morris, M. L. (2000). The relational-interdependent self-construal and relationships. *Journal of Personality and Social Psychology*, *78*(4), 791–808. https://doi.org/10.1037/0022-3514.78.4.791

Cross, S. E., & Madson, L. (1997). Models of the self: Self-construals and gender. *Psychological Bulletin*, *122*(1), 5–37. https://doi.org/10.1037/0033-2909.122.1.5

Cunningham, W. A., Zelazo, P. D., & Packer, D. J. (2007). *The iterative reprocessing model: a multilevel framework for attitudes and evaluation*. *25*(5), 736–760.

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 Panfilis, C., Meehan, K. B., Cain, N. M., & Clarkin, J. F. (2016). Effortful Control, Rejection Sensitivity, and Borderline Personality Disorder Features in Adulthood. *Journal of Personality Disorders*. https://doi.org/10.1521/pedi_2015_29_226

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

DeBruine, L. M. (2002). Facial resemblance enhances trust. *Proceedings of the Royal Society B: Biological Sciences*, *269*(1498), 1307–1312. https://doi.org/10.1098/rspb.2002.2034

DePaulo, B. M., Kirkendol, S. E., Kashy, D. A., Wyer, M. M., & Epstein, J. A. (1996). Lying in Everyday Life. *Journal of Personality and Social Psychology*, *70*(5), 979–995. https://doi.org/10.1037/0022-3514.70.5.979

Doering, S. (2019). Borderline Personality Disorder in Patients With Medical Illness: A Review of Assessment, Prevalence, and Treatment Options. *Psychosomatic Medicine*, *81*(7).

https://journals.lww.com/psychosomaticmedicine/Fulltext/2019/09000/Borderline_Personality_Disorder_in_Patients_With.4.aspx

Domes, G., Schulze, L., & Herpertz, S. C. (2009). Emotion recognition in borderline personality disorder-a review of the literature. *Journal of Personality Disorders*, *23*(1), 6–19. https://doi.org/10.1521/pedi.2009.23.1.6

Downey, G., & Feldman, S. I. (1996). Implications of Rejection Sensitivity for Intimate Relationships. In *Journal of Personality and Social Psychology* (Vol. 70, Issue 6).

Duan, Z., Ye, T., Poggi, A., & Ding, X. (2020). Gaze towards my choice: Noneconomic social interaction changes interpersonal trust only with positive feedback. *Psychonomic Bulletin and Review*, *27*, 1362–1373. https://doi.org/10.3758/s13423-020-01785-w

Dunn, J. R., & Schweitzer, M. E. (2005). *Feeling and Believing : The Influence of Emotion on Trust*. *88*(5), 736–748. https://doi.org/10.1037/0022-3514.88.5.736

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

Eckel, C. C., & Wilson, R. K. (2006). *Detecting Trustworthiness: Does Beauty Confound Intuition?*

Erdfelder, E., Faul, F., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Erikson, E. H. (1993). *Childhood and society*. WW Norton & Company.

Evans, A. M., & Krueger, J. I. (2009). The Psychology (and Economics) of Trust. *Social and Personality Psychology Compass*, *3*(6), 1003–1017. https://doi.org/10.1111/j.1751-9004.2009.00232.x

Evans, A. M., & Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality*. https://doi.org/10.1016/j.jrp.2008.07.011

Fazio, R. H., & Olson, M. A. (2003). Implicit Measures in Social Cognition Research: Their Meaning and Use. *Annual Review of Psychology*, *54*, 297–327.

https://doi.org/10.1146/annurev.psych.54.101601.145225

Fertuck, E. A., Fischer, S., & Beeney, J. (2018). Social Cognition and Borderline Personality Disorder: Splitting and Trust Impairment Findings. In *Psychiatric Clinics of North America* (Vol. 41, Issue 4, pp. 613–632). W.B. Saunders. https://doi.org/10.1016/j.psc.2018.07.003

Fertuck, E. A., Grinband, J., Mann, J. J., Hirsch, J., Ochsner, K., Pilkonis, P., Erbe, J., & Stanley, B. (2019). Trustworthiness appraisal deficits in borderline personality disorder are associated with prefrontal cortex, not amygdala, impairment. *NeuroImage: Clinical*, *21*. https://doi.org/10.1016/j.nicl.2018.101616

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

Fertuck, E. A., Keilp, J., Song, I., Morris, M. C., Wilson, S. T., Brodsky, B. S., & Stanley, B. (2012). Higher executive control and visual memory performance predict treatment completion in borderline personality disorder. *Psychotherapy and Psychosomatics*, *81*(1), 38–43. https://doi.org/10.1159/000329700

Fetchenhauer, D., & Huang, X. (2004). Justice sensitivity and distributive decisions in experimental games. *Personality and Individual Differences*, *36*(5), 1015–1029. https://doi.org/https://doi.org/10.1016/S0191-8869(03)00197-1

Fineberg, S. K., Leavitt, J., Stahl, D. S., Kronemer, S., Landry, C. D., Alexander-Bloch, A., Hunt, L. T., & Corlett, P. R. (2018). Differential Valuation and Learning From Social and Nonsocial Cues in Borderline Personality Disorder. *Biological Psychiatry*, *84*(11), 838–845. https://doi.org/10.1016/j.biopsych.2018.05.020

First, M. B., Spitzer, R. L., Gibbon, M., Williams, J. B. W., Davies, M., Borus, J., Howes, M. J., Kane, J., Pope, H. G., & Rounsaville, B. (1995). The structured clinical interview for DSM-III-R personality disorders (SCID-II). Part II: Multi-site test-retest reliablity study. *Journal of Personality Disorders*, *9*(2), 92–104. https://doi.org/10.1521/pedi.1995.9.2.92

First, Michael B, Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *User's guide for the Structured clinical interview for DSM-IV axis I disorders SCID-I:*

clinician version. American Psychiatric Pub.

Fleeson, W., & Noftle, E. E. (2012). *Personality research.*

Fonagy, P., & Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy*, *51*(3), 372.

Fonagy, P., Campbell, C., & Bateman, A. (2017). Mentalizing, attachment, and epistemic trust in group therapy. *International Journal of Group Psychotherapy*, *67*(2), 176–201. https://doi.org/10.1080/00207284.2016.1263156

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., Luyten, P., Allison, E., & Campbell, C. (2017). What we have changed our minds about: Part 2. Borderline personality disorder, epistemic trust and the developmental significance of social communication. *Borderline Personality Disorder and Emotion Dysregulation*, *4*(1), 9. https://doi.org/10.1186/s40479-017-0062-8

Fonagy, P., & Target, M. (2006). The mentalization-focused approach to self pathology. *Journal of Personality Disorders*, *20*(6), 544–576. https://doi.org/10.1521/pedi.2006.20.6.544

Fossati, A., Somma, A., & Borroni, S. (2017). The multidimensionality of pathological narcissism from the perspective of social ostracism: A study in a sample of Italian University students. *Personality and Individual Differences*. https://doi.org/10.1016/j.paid.2017.05.009

Fox, J., Arena, D., & Bailenson, J. N. (2009). Virtual Reality: A Survival Guide for the Social Scientist. *Journal of Media Psychology*, *21*(3), 95–113. https://doi.org/10.1027/1864-1105.21.3.95

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

Franzen, N., Hagenhoff, M., Baer, N., Schmidt, A., Mier, D., Sammer, G., Gallhofer,

B., Kirsch, P., Lis, S., King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., Montague, P. R., Coifman, K. G., Berenson, K. R., Rafaeli, E., Downey, G., … de Graaf, R. (2013). From negative to positive and back again: Polarized affective and relational experience in borderline personality disorder. *Journal of Personality Disorders*, *6*(3), 668–679. https://doi.org/10.1037/a0028502

Freud, S. (1917). *Introductory lectures on psychoanalysis*. WW Norton & Company.

Freud, S. (1958). The dynamics of transference. In *The standard edition of the complete psychological works of Sigmund Freud, Volume XII (1911-1913): The case of Schreber, papers on technique and other works* (pp. 97–108).

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2–18. https://doi.org/10.1037/a0024338

Funder, D. C., Guillaume, E., Kumagai, S., Kawamoto, S., & Sato, T. (2012). The Person-situation Debate and the Assessment of Situations. *The Japanese Journal of Personality*, *21*(1), 1–11. https://doi.org/10.2132/personality.21.1

Furnham, A., & Crump, J. (2015). A big five facet analysis of a paranoid personality disorder. *Journal of Individual Differences*.

Giesen-Bloo, J., Arntz, A., & Schouten, E. (2006). The borderline personality disorder checklist: psychometric evaluation and factorial structure in clinical and nonclinical samples. *Crossing Borders: Theory, Assessment and Treatment in Borderline Personality Disorder*, 85–102.

Gmytrasiewicz, P. J., & Doshi, P. (2004). *Interactive POMDPs: Properties and Preliminary Results*.

Golembiewski, R. T., & McConkie, M. (1975). The centrality of interpersonal trust in group processes. *Theories of Group Processes, 131*, 185.

Gollwitzer, M., & Rothmund, T. (2011). What exactly are victim-sensitive persons sensitive to? *Journal of Research in Personality*, *45*(5), 448–455. https://doi.org/10.1016/j.jrp.2011.05.003

Gollwitzer, M., Rothmund, T., Pfeiffer, A., & Ensenbach, C. (2009). Why and when justice sensitivity leads to pro- and antisocial behavior. *Journal of Research in Personality*, *43*(6), 999–1005. https://doi.org/10.1016/j.jrp.2009.07.003

Gollwitzer, M., Rothmund, T., & Süssenbach, P. (2013). The Sensitivity to Mean Intentions (SeMI) model: Basic assumptions, recent findings, and potential avenues for future research. *Social and Personality Psychology Compass*, *7*(7), 415–426. https://doi.org/10.1111/spc3.12041

Granovetter, M. (1985). *Economic Action and Social Structure: The Problem of Embeddedness'*. https://about.jstor.org/terms

Greenberg, J. (1983). *Object relations in psychoanalytic theory*. Harvard University Press.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., Rudman, L. A., Nosek, B. A., Banaji, M. R., Farnham, S. D., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*(1), 3–25. https://doi.org/10.1037/0033-295X.109.1.3

Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*(6), 1022.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. In *Journal of Personality and Social Psychology* (Vol. 85, Issue 2, pp. 197–216). American Psychological Association. https://doi.org/10.1037/0022-3514.85.2.197

Grilo, C. M., Sanislow, C. A., Skodol, A. E., Gunderson, J. G., Stout, R. L., Bender, D. S., Yen, S., Shea, M. T., Morey, L. C., Zanarini, M. C., & McGlashan, T. H. (2007). Longitudinal diagnostic efficiency of DSM-IV criteria for borderline personality disorder: A 2-year prospective study. *Canadian Journal of Psychiatry*, *52*(6), 357–362. https://doi.org/10.1177/070674370705200604

Gunderson, J. G. (2012). Borderline personality disorder: Borderline personality disorder in adolescents. *Psychiatric Times*, *29*(5), 22–24. https://doi.org/10.1056/NEJMcp1007358

Gunderson, J. G. (2009). *Borderline personality disorder: A clinical guide.*

American Psychiatric Publisher.

Gunderson, J. G. (2007). Disturbed Relationships as a Phenotype for Borderline Personality Disorder. *American Journal of Psychiatry*. https://doi.org/10.1176/appi.ajp.2007.07071125

Gunderson, J. G., & Lyons-Ruth, K. (2008). BPD's Interpersonal Hypersensitivity Phenotype: A Gene-Environment-Developmental Model. *Journal of Personality Disorders*. 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., 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

Gunderson, J. G., (2011). Borderline Personality Disorder. *The New England Journal of Medicine*, *364*, 2037–2079. https://doi.org/10.1056/NEJMcp1007358

Gunderson, J. G., Herpertz, S. C., Skodol, A. E., Torgersen, S., & Zanarini, M. C. (2018). Borderline personality disorder. *Nature Reviews Disease Primers*, *4*(1), 1–20.

Hahn, A., & Goedderz, A. (2020). Trait-Unconsciousness, State-Unconsciousness, Preconsciousness, and Social Miscalibration in the Context of Implicit Evaluation. *Social Cognition*, *38*(Supplement), 115–134. https://doi.org/10.1521/soco.2020.38.supp.s115

Hale, J., Payne, M. E., Taylor, K. M., Paoletti, D., & De C Hamilton, A. F. (2018). The virtual maze: A behavioural tool for measuring trust. *Quarterly Journal of Experimental Psychology (2006)*. https://doi.org/10.1080/17470218.2017.1307865

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis; a regression-based approach.* Guilford Press.

Hazan, C., & Shaver, P. (1987). Romantic Love Conceptualized as an Attachment Process. *Journal of Personality and Social Psychology*, *52*(3), 511–524. https://doi.org/10.1037/0022-3514.52.3.511

Herpertz, S. C., & Bertsch, K. (2014). The social-cognitive basis of personality

disorders. *Current Opinion in Psychiatry*, *27*(1), 73–77. https://doi.org/10.1097/YCO.0000000000000026

Higa, J. K., & Gedo, P. M. (2012). Transference interpretation in the treatment of borderline personality disorder patients. *Bulletin of the Menninger Clinic*, *76*(3), 195–210. https://doi.org/10.1521/bumc.2012.76.3.195

Hill, J., Pilkonis, P., Morse, J., Feske, U., Reynolds, S., Hope, H., Charest, C., & Broyden, N. (2008). *Social domain dysfunction and disorganization in borderline personality disorder*. 135–146. https://doi.org/10.1017/S0033291707001626

Hill, R. M., Penner, F., Vanwoerden, S., Mellick, W., Kazimi, I., & Sharp, C. (2019). Interpersonal Trust and Suicide Ideation Among Adolescent Psychiatric Inpatients: An Indirect Effect via Perceived Burdensomeness. *Suicide and Life-Threatening Behavior*, *49*(1), 240–252. https://doi.org/10.1111/sltb.12433

Hoc, J.M. (2001). Towards ecological validity of research in cognitive ergonomics. *Theoretical Issues in Ergonomics Science*, *2*(3), 278–288.

Hopwood, C. J., Wright, A. G. C., Ansell, E. B., & Pincus, A. L. (2013). The interpersonal core of personality pathology. *Journal of Personality Disorders*, *27*(3), 270–295. https://doi.org/10.1521/pedi.2013.27.3.270

Hosmer, L. T. (2011). Trust: the connecting link between organizational theory and philosphical ethics. *Academy of Management Review*, *20*(2), 379–403. https://doi.org/10.5465/amr.1995.9507312923

Hula, A., Vilares, I., Dayan, P., & Montague, P. R. (2017). *A Model of Risk and Mental State Shifts during Social Interaction*. 1–20. http://arxiv.org/abs/1704.03508

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage.

Jackson, H. J., & Burgess, P. M. (2004). Personality disorders in the community: Results from the Australian National Survey of Mental Health and Well-Being Part III - Relationships between specific type of personality disorder, Axis 1 mental disorders and physical conditions with disability and. *Social Psychiatry and Psychiatric Epidemiology*, *39*(10), 765–776. https://doi.org/10.1007/s00127-004-0821-x

Jackson, K. M., & Trull, T. J. (2001). The factor structure of the Personality Assessment Inventory-Borderline Features (PAI-BOR) Scale in a nonclinical sample. *Journal of Personality Disorders*, *15*(6), 536–545. https://doi.org/10.1521/pedi.15.6.536.19187

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

Jin, J. M., Bai, P., He, W., Wu, F., Liu, X. F., Han, D. M., Liu, S., & Yang, J. K. (2020). Gender Differences in Patients With COVID-19: Focus on Severity and Mortality. *Frontiers in Public Health*, *8*(April), 1–6. https://doi.org/10.3389/fpubh.2020.00152

Jin, Z., Zhao, K., Xia, Y., Chen, R., & YU H, T. T. T. (2020). Psychological responses to the coronavirus disease (COVID-19) outbreak. *China*, *14*, 1.

Joel, S., Eastwick, P. W., Allison, C. J., Arriaga, X. B., Baker, Z. G., Bar-Kalifa, E., Bergeron, S., Birnbaum, G. E., Brock, R. L., Brumbaugh, C. C., Carmichael, C. L., Chen, S., Clarke, J., Cobb, R. J., Coolsen, M. K., Davis, J., de Jong, D. C., Debrot, A., DeHaas, E. C., … Wolf, S. (2020). Machine learning uncovers the most robust self-report predictors of relationship quality across 43 longitudinal couples studies. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(32), 19061–19071. https://doi.org/10.1073/pnas.1917036117

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

Jones, G. R., & George, J. M. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review*, *23*(3), 531–546. https://doi.org/10.5465/AMR.1998.926625

Jørgensen, F., Bor, A., Lindholt, M. F., & Petersen, M. B. (2020). Lockdown Evaluations During the First Wave of the COVID-19 Pandemic. *(Preprint)*, 1–32.

Jørgensen, F., Bor, A., & Petersen, M. B. (2020). *Compliance Without Fear: Predictors of Protective Behavior During the First Wave of the COVID-19*

Pandemic. https://doi.org/10.31234/osf.io/uzwgf

Kaniasty, K. (2012). Predicting social psychological well-being following trauma: The role of postdisaster social support. *Psychological Trauma: Theory, Research, Practice, and Policy*, *4*(1), 22–33. https://doi.org/10.1037/a0021412

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*, *3*, 1–27. https://doi.org/10.1146/annurev.clinpsy.3.022806.091432

Kernberg, O. F. (1967). Borderline personality organization. *Journal of the American Psychoanalytic Association*, *15*(3), 641–685.

Kernberg, O. F. (1975a). A systems approach to priority setting of interventions in groups. *International Journal of Group Psychotherapy*, *25*(3), 251–275.

Kernberg, O. F. (1975b). Transference and countertransference in the treatment of borderline patients. *Journal of the National Association of Private Psychiatric Hospitals*, *7*(2), 14–24.

Kernberg, O. F. (1984). *Severe Personality Disorders - Psychotherapeutic Strategies*. Yale University Press.

Kernberg, O. F. (1985). *Borderline Conditions and Pathological Narcissism*. Rowrnan & Littlefield Publishers, Inc. A.

Kernberg, O. F. (1996). A psychoanalytic theory of personality disorders. In *Major theories of personality disorder.* (pp. 106–140). Guilford Press.

Kernberg, O. F., & Caligor, E. (2005). A psychoanalytic theory of personality disorders. In G. Press (Ed.), *Major theories of personality disorder* (Vol. 2, pp. 114–156).

Kernberg, O. F. (1967). Borderline Personality Organization. In *Journal of the American Psychoanalytic Association* (Vol. 15, Issue 3). https://doi.org/10.1177/000306516701500309

Kim, D., Subramanian, S. V, & Kawachi, I. (2008). Social capital and physical health. In *Social capital and health* (pp. 139–190). Springer. https://doi.org/10.1007/978-0-387-71311-3_8

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., & Montague, P. R. (2008). The Rupture and Repair of Cooperation in Borderline Personality Disorder. *Science*, *321*(5890), 806–810.

https://doi.org/10.1126/science.1156902

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., & Montague, P. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science (New York, N.Y.)*, *321*(5890), 806–810. https://doi.org/10.1126/science.1156902

Klapper, A., Dotsch, R., van Rooij, I., & Wigboldus, D. H. J. (2016). Do we spontaneously form stable trustworthiness impressions from facial appearance? *Journal of Personality and Social Psychology*, *111*(5), 655–664. https://doi.org/10.1037/pspa0000062

Klein, M. (1948). Contributions to psychoanalysis, 1921-1945. In *Contributions to psychoanalysis, 1921-1945*. Hogarth Press.

Klein, M., Wonderlich, S. A., & Crosby, R. (2001). Self-concept correlates of the personality disorders. *Journal of Personality Disorders*, *15*(2), 150–156. https://doi.org/10.1521/pedi.15.2.150.19214

Kollock, P. (1998). Social Dilemmas: The Anatomy of Cooperation. *Annual Review of Sociology*, *24*(1), 183–214. https://doi.org/10.1146/annurev.soc.24.1.183

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, *435*(7042), 673–677. https://doi.org/10.1038/nature03701

Koslow, S. (2000). Can the truth hurt? How honest and persuasive advertising can unintentionally lead to increased consumer skepticism. *Journal of Consumer Affairs*, *34*(2), 245–267. https://doi.org/10.1111/j.1745-6606.2000.tb00093.x

Krueger, F., & Meyer-Lindenberg, A. (2019). Toward a Model of Interpersonal Trust Drawn from Neuroscience, Psychology, and Economics. *Trends in Neurosciences*, *42*(2), 92–101. https://doi.org/10.1016/j.tins.2018.10.004

Krueger, F., Eaton, N. R., Clark, L. A., Watson, D., Markon, K. E., Derringer, J., Skodol, A., & Livesley, W. J. (2011). Deriving an empirical structure of personality pathology for DSM-5. *Journal of Personality Disorders*, *25*(2), 170–191. https://doi.org/10.1521/pedi.2011.25.2.170

Kurzban, R., & Houser, D. (2001). Individual differences in cooperation in a circular public goods game. *European Journal of Personality*, *15*(1 SUPPL.). https://doi.org/10.1002/per.420

Kwiatkowska, M. M., Jułkowski, T., Rogoza, R., Żemojtel-Piotrowska, M., & Fatfouta, R. (2019). Narcissism and trust: Differential impact of agentic, antagonistic, and communal narcissism. *Personality and Individual Differences*, *137*(2018), 139–143. https://doi.org/10.1016/j.paid.2018.08.027

Lambert, M. J. (1992). Psychotherapy outcome research: Implications for integrative and eclectical therapists. In *Handbook of psychotherapy integration.* (pp. 94–129). Basic Books.

Lane, C., & Bachmann, R. (1998). *Trust within and between Organisations* (Oxford University Press (ed.)).

Langley, G. C., & Klopper, H. (2005). Trust as a foundation for the therapeutic intervention for patients with borderline personality disorder. *Journal of Psychiatric & Mental Health Nursing*, *12*(1), 23-32 10p. https://doi.org/10.1111/j.1365-2850.2004.00774.x

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

Lenzenweger, M. F., Loranger, A. W., Korfine, L., & Neff, C. (1997). Detecting personality disorders in a nonclinical population: Application of a 2-stage procedure for case identification. *Archives of General Psychiatry*, *54*(4), 345–351.

Levine B, H. (1994). The analyst's participation in the analytic process. *International Journal of Psycho-Analysis*, *75*, 665–676.

Levy, K. N., Clarkin, J. F., Yeomans, F. E., Scott, L. N., Wasserman, R. H., & Kernberg, O. F. (2006). The mechanisms of change in the treatment of borderline personality disorder with transference focused psychotherapy. *Journal of Clinical Psychology*, *62*(4), 481–501.

Lewicki, R. J., McAllister, D. J., & Bies, R. I. (1998). Trust and distrust: New relationships and realities. *Academy of Management Review*, *23*(3), 438–458. https://doi.org/10.5465/AMR.1998.926620

Lewis, J. D., & Weigert, A. J. (1985). Trust as a Social Reality. *Social Forces*, *63*(4), 967–985. https://doi.org/10.1093/sf/63.4.967

Liebke, L., Thome, J., Hauschild, S., Bungert, M., Defiebre, N., Lis, S., Koppe, G., Bohus, M., Schmahl, C., & Izurieta Hidalgo, N. A. (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*. The Guilford Press.

Linehan, M. M. (2018). *Cognitive-behavioral treatment of borderline personality disorder*. Guilford Publications.

Links, P. S., & Heslegrave, R. J. (2000). Prospective studies of outcome: Understanding mechanisms of change in patients with borderline personality disorder. *Psychiatric Clinics of North America*, *23*(1), 137–150.

Lis, S., Schaedler, A., Liebke, L., Hauschild, S., Thome, J., Schmahl, C., Stahlberg, D., Kleindienst, N., & Bohus, M. (2018). Borderline personality disorder features and sensitivity to injustice. *Journal of Personality Disorders*, *32*(2), 192–206. https://doi.org/10.1521/pedi_2017_31_292

London, B., Downey, G., Bonica, C., & Paltin, I. (2007). Social causes and consequences of rejection sensitivity. *Journal of Research on Adolescence*, *17*(3), 481–506. https://doi.org/10.1111/j.1532-7795.2007.00531.x

Loranger, A. W. (1997). International personality disorder examination (IPDE). In A. Janca, A. W. Loranger, & N. Sartorius (Eds.), *Assessment and Diagnosis of Personality Disorders: The ICD-10 International Personality Disorder Examination (IPDE)* (pp. 43–51). Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511663215.005

Lyons-Ruth, K., Bruschweiler-Stern, N., Harrison, A. M., Morgan, A. C., Nahum, J. P., Sander, L., Stern, D. N., & Tronick, E. Z. (1998). Implicit relational knowing: Its role in development and psychoanalytic treatment. *Infant Mental Health Journal*, *19*(3), 282–289. https://doi.org/10.1002/(sici)1097-0355(199823)19:3<282::aid-imhj3>3.0.co;2-o

Macko, A., Malawski, M., & Tyszka, T. (2014). Belief in others' trustworthiness and trusting behaviour. *Polish Psychological Bulletin*, *45*(1), 43–51. https://doi.org/10.2478/ppb-2014-0007

Maltese, S., Baumert, A., Schmitt, M. J., & MacLeod, C. (2016). How victim

sensitivity leads to uncooperative behavior via expectancies of injustice. *Frontiers in Psychology*, *6*(JAN), 1–11. https://doi.org/10.3389/fpsyg.2015.02059

Markus, H., & Wurf, E. (1987). The Dynamic Self-Concept: A Social Psychological Perspective. *Annual Review of Psychology*, *38*(1), 299–337. https://doi.org/10.1146/annurev.psych.38.1.299

Masland, S. R., & Hooley, J. M. (2019). When Trust Does Not Come Easily: Negative Emotional Information Unduly Influences Trustworthiness Appraisals for Individuals With Borderline Personality Features. *Journal of Personality Disorders*, 1–16. https://doi.org/10.1521/pedi_2019_33_404

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734.

McCall, C., & Blascovich, J. (2009). How, When, and Why to Use Digital Experimental Virtual Environments to Study Social Behavior. *Social and Personality Psychology Compass*, *3*(5), 744–758. https://doi.org/10.1111/j.1751-9004.2009.00195.x

McCornack, S. A., & Parks, M. R. (1986). Deception detection and relationship development: The other side of trust. *Annals of the International Communication Association*, *9*(1), 377–389.

Metalsky, G. I., & Abramson, L. Y. (1981). Attributional Styles: Toward a Framework for Conceptualization and Assessment. In *Assessment Strategies for Cognitive–Behavioral Interventions*. Academic Press. https://doi.org/10.1016/b978-0-12-404460-9.50008-2

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*, 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

Miano, A., Grosselli, L., Roepke, S., & Dziobek, I. (2017). Behaviour Research and

Therapy Emotional dysregulation in borderline personality disorder and its in fl uence on communication behavior and feelings in romantic relationships. *Behaviour Research and Therapy*, *95*, 148–157. https://doi.org/10.1016/j.brat.2017.06.002

Minzenberg, M. J., Fan, J., New, A. S., Tang, C. Y., & Siever, L. J. (2007). Fronto-limbic dysfunction in response to facial emotion in borderline personality disorder: An event-related fMRI study. *Psychiatry Research - Neuroimaging*, *155*(3), 231–243. https://doi.org/10.1016/j.pscychresns.2007.03.006

Miskewicz, K., Fleeson, W., Arnold, E. M., Law, M. K., Mneimne, M., & Michael Furr, R. (2015). A contingency-oriented approach to understanding borderline personality disorder: Situational triggers and symptoms. *Journal of Personality Disorders*, *29*(4), 486–502. https://doi.org/10.1521/pedi.2015.29.4.486

Mohseni, M., & Lindstrom, M. (2007). Social capital, trust in the health-care system and self-rated health: The role of access to health care in a population-based study. *Social Science and Medicine*, *64*(7), 1373–1383. https://doi.org/10.1016/j.socscimed.2006.11.023

Montoya, A. K., & Hayes, A. F. (2017). Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods*, *22*(1), 6–27. https://doi.org/10.1037/met0000086

Moreno, C., Wykes, T., Galderisi, S., Nordentoft, M., Crossley, N., Jones, N., Cannon, M., Correll, C. U., Byrne, L., Carr, S., Chen, E. Y. H., Gorwood, P., Johnson, S., Kärkkäinen, H., Krystal, J. H., Lee, J., Lieberman, J., López-Jaramillo, C., Männikkö, M., … Arango, C. (2020). How mental health care should change as a consequence of the COVID-19 pandemic. *The Lancet Psychiatry*, *7*(9), 813–824. https://doi.org/10.1016/S2215-0366(20)30307-2

Morey, L. C. (2004). Personality Assessment Inventory (PAI). *The Encyclopedia of Clinical Psychology*, 1–10. https://doi.org/10.1002/9781118625392.wbecp284

Mortensen, J. A., Evensmoen, H. R., Klensmeden, G., & Håberg, A. K. (2016). Outcome Uncertainty and Brain Activity Aberrance in the Insula and Anterior Cingulate Cortex Are Associated with Dysfunctional Impulsivity in Borderline Personality Disorder. *Frontiers in Human Neuroscience*, *10*, 207. https://doi.org/10.3389/fnhum.2016.00207

Nicol, K., Pope, M., Sprengelmeyer, R., Young, A. W., & Hall, J. (2013). Social judgement in borderline personality disorder. *PLoS ONE*, *8*(11), 1–6. https://doi.org/10.1371/journal.pone.0073440

Ogden, T. (1979). On projective identification. *The International Journal of Psycho-Analysis*, *60 Pt 3*, 357–373.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Oosterhof, N. N., & Todorov, A. (2009). Shared Perceptual Basis of Emotional Expressions and Trustworthiness Impressions From Faces. *Emotion*, *9*(1), 128–133. https://doi.org/10.1037/a0014520

Orbeli, J., Dawes, R., & Sehwartz-Shea, P. (1994). Trust, Social Categories, and Individuals: The Case of Gender. In *Motivation and Emotion* (Vol. 18, Issue 2). https://doi.org/10.1007%2FBF02249396.pdf

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), 17–19. https://doi.org/10.1186/s40479-019-0110-7

Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why Do Implicit and Explicit Attitude Tests Diverge? The Role of Structural Fit. *Journal of Personality and Social Psychology*, *94*(1), 16–31. https://doi.org/10.1037/0022-3514.94.1.16

Pignolo, C., Di Nuovo, S., Fulcheri, M., Lis, A., Mazzeschi, C., & Zennaro, A. (2018). Psychometric properties of the Italian version of the Personality Assessment Inventory (PAI). In *Psychological Assessment* (Vol. 30, Issue 9, pp. 1226–1236). American Psychological Association. https://doi.org/10.1037/pas0000560

Pincus, A. L., & Ansell, E. B. (2003). Interpersonal Theory of Personality. In *Handbook of Psychology* (Vol. 5). https://doi.org/10.1002/0471264385.wei0513

Pincus, A. L., & Hopwood, C. J. (2012). A Contemporary Interpersonal Model of Personality Pathology and Personality Disorder. *The Oxford Handbook of Personality Disorders*, 371–398. https://doi.org/10.1093/oxfordhb/9780199735013.013.0018

Ping Li, P. (2012). When trust matters the most: The imperatives for contextualising trust research. *Journal of Trust Research*. https://doi.org/10.1080/21515581.2012.708494

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

Preti, E., Casini, E., Richetin, J., De Panfilis, C., & Fontana, A. (2018). Cognitive and Emotional Components of Rejection Sensitivity: Independent Contributions to Adolescent Self- and Interpersonal Functioning. *Assessment*. https://doi.org/10.1177/1073191118817866

Preti, E., Di Pierro, R., Fanti, E., Madeddu, F., & Calati, R. (2020). Personality Disorders in Time of Pandemic. *Current Psychiatry Reports*, *22*(12). https://doi.org/10.1007/s11920-020-01204-w

Preuss, N., Brändle, L. S., Hager, O. M., Haynes, M., Fischbacher, U., & Hasler, G. (2016). Inconsistency and social decision making in patients with Borderline Personality Disorder. *Psychiatry Research*, *243*, 115–122. https://doi.org/10.1016/j.psychres.2016.06.017

Provenzi, L., & Tronick, E. (2020). The Power of Disconnection During the COVID-19 Emergency: From Isolation to Reparation. *Psychological Trauma: Theory, Research, Practice, and Policy*, *12*, 2016–2018. https://doi.org/10.1037/tra0000619

Qualter, P., Brown, S. L., Munn, P., & Rotenberg, K. J. (2010). Childhood loneliness as a predictor of adolescent depressive symptoms: An 8-year longitudinal study. *European Child and Adolescent Psychiatry*, *19*(6), 493–501. https://doi.org/10.1007/s00787-009-0059-y

Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 888–896. https://doi.org/10.1037//0278-7393.17.5.888

Rekosh, J. H., & Feigenbaum, K. D. (1966). The Necessity of Mutual Trust for Cooperative Behavior in a Two-Person Game. *The Journal of Social Psychology*, *69*(1), 149–154. https://doi.org/10.1080/00224545.1966.9919713

Remue, J., Hughes, S., De Houwer, J., & De Raedt, R. (2014). To be or want to be: Disentangling the role of actual versus ideal self in implicit self-esteem. *PLoS ONE*, *9*(9). https://doi.org/10.1371/journal.pone.0108837

Richardson-Vejlgaard, R., Broudy, C., Brodsky, B., Fertuck, E., & Stanley, B. (2013). Predictors of psychotherapy alliance in Borderline Personality Disorder. *Psychotherapy Research*, *23*(5), 539–546. https://doi.org/10.1080/10503307.2013.801001

Richetin, J., Poggi, A., Ricciardelli, P., Fertuck, E., & 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*(2).

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron*, *35*(2), 395–405. https://doi.org/10.1016/S0896-6273(02)00755-9

Roberts, I. D., Krajbich, 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., Schröder-Abé, M., Schütz, A., Jacob, G., Dams, A., Vater, A., Rüter, A., Merkl, A., Heuser, I., & Lammers, C. H. (2011). Dialectic behavioural therapy has an impact on self-concept clarity and facets of self-esteem in women with borderline personality disorder. *Clinical Psychology and Psychotherapy*, *18*(2), 148–158. https://doi.org/10.1002/cpp.684

Roepke, S., Vater, A., Preißler, S., Heekeren, H. R., & Dziobek, I. (2012). Social cognition in borderline personality disorder. *Frontiers in Neuroscience*, *6*(JAN), 1–12. https://doi.org/10.3389/fnins.2012.00195

Rona, R. J., Fear, N. T., Hull, L., Greenberg, N., Earnshaw, M., Hotopf, M., & Wessely, S. (2007). Mental health consequences of overstretch in the UK armed forces: First phase of a cohort study. *British Medical Journal*, *335*(7620), 603–607. https://doi.org/10.1136/bmj.39274.585752.BE

Rotenberg, K. J., Addis, N., Betts, L. R., Corrigan, A., Fox, C., Hobson, Z., Rennison, S., Trueman, M., & Boulton, M. J. (2010). The relation between trust

beliefs and loneliness during early childhood, middle childhood, and adulthood. *Personality and Social Psychology Bulletin*, *36*(8), 1086–1100. https://doi.org/10.1177/0146167210374957

Rotenberg, K. J., Boulton, M. J., & Fox, C. L. (2005). Cross-sectional and longitudinal relations among children's trust beliefs, psychological maladjustment and social relationships: Are very high as well as very low trusting children at risk? *Journal of Abnormal Child Psychology*, *33*(5), 595–610. https://doi.org/10.1007/s10802-005-6740-9

Rothschild, L., Cleland, C., Haslam, N., & Zimmerman, M. (2003). A taxometric study of borderline personality disorder. *Journal of Abnormal Psychology*, *112*(4), 657–666. https://doi.org/10.1037/0021-843X.112.4.657

Rotter, J. B. (1980). Interpersonal trust, trustworthiness, and gullibility. *American Psychologist*, *35*(1), 1–7. https://doi.org/10.1037/0003-066X.35.1.1

Rüsch, N., Lieb, K., Göttler, I., Hermann, C., Schramm, E., Richter, H., Jacob, G. A., Corrigan, P. W., & Martin Bohus, P. (2007). Shame and Implicit Self-Concept in Women With Borderline Personality Disorder. *Am J Psychiatry*, *164*. https://doi.org/10.1176/ajp.2007.164.3.500

Rüsch, N., Schulz, D., Valerius, G., Steil, R., Bohus, M., & Schmahl, C. (2011). Disgust and implicit self-concept in women with borderline personality disorder and posttraumatic stress disorder. *European Archives of Psychiatry and Clinical Neuroscience*. https://doi.org/10.1007/s00406-010-0174-2

Russell, J. J., Moskowitz, D. S., Zuroff, D. C., Sookman, D., & Paris, J. (2007). Stability and Variability of Affective Experience and Interpersonal Behavior in Borderline Personality Disorder. *Journal of Abnormal Psychology*, *116*(3), 578–588. https://doi.org/10.1037/0021-843X.116.3.578

Sandler, J. (1992). Reflections on developments in the theory of psychoanalytic technique. *International Journal of Psycho-Analysis*, *73*, 189–198.

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. *American Journal of Psychiatry*, *159*(2), 284–290.

https://doi.org/10.1176/appi.ajp.159.2.284

Sansone, R. A., & Sansone, L. A. (2013). Responses of mental health clinicians to patients with borderline personality disorder. *Innovations in Clinical Neuroscience*, *10*(5–6), 39. https://doi.org/PMC3719460

Saunders, K. E. A., Goodwin, G. M., & Rogers, R. D. (2016). Borderline personality disorder, but not euthymic bipolar i disorder, is associated with prolonged post-error slowing in sensorimotor performance. *Journal of Affective Disorders*, *198*, 163–170. https://doi.org/10.1016/j.jad.2016.03.027

Schechter, L. (2007). Traditional trust measurement and the risk confound: An experiment in rural Paraguay. *Journal of Economic Behavior & Organization*, *62*(2), 272–292. https://doi.org/https://doi.org/10.1016/j.jebo.2005.03.006

Schmeck, K., Schlüter-Müller, S., Foelsch, P. A., & Doering, S. (2013). The role of identity in the DSM-5 classification of personality disorders. *Child and Adolescent Psychiatry and Mental Health*, *7*(1), 1–11. https://doi.org/10.1186/1753-2000-7-27

Schmitt, M., Gollwitzer, M., Maes, J., & Arbach, D. (2005). Justice sensitivity. *European Journal of Psychological Assessment*, *21*(3), 202–211.

Schmitt, M., Neumann, R., & Montada, L. (1995). Dispositional Sensitivity to Befallen Injustice. *Social Justice Research*, *8*(4), 385–407. https://doi.org/10.1007/BF02334713

Schmitt, M., Baumert, A., Gollwitzer, M., & Maes, J. (2010). The Justice Sensitivity Inventory: Factorial Validity, Location in the Personality Facet Space, Demographic Pattern, and Normative Data. *Social Justice Research*, *23*(2), 211–238. https://doi.org/10.1007/s11211-010-0115-2

Schnabel, K., & Asendorpf, J. B. (2010). The self-concept: New insights from implicit measurement procedures. In *Handbook of implicit social cognition: Measurement, theory, and applications.* (pp. 408–425). The Guilford Press.

Schneider, I. K., Konijn, E. A., Righetti, F., & Rusbult, C. E. (2011). A healthy dose of trust: The relationship between interpersonal trust and health. *Personal Relationships*, *18*(4), 668–676. https://doi.org/10.1111/j.1475-6811.2010.01338.x

Seres, I., Unoka, Z., & Kéri, S. (2009). The broken trust and cooperation in

borderline personality disorder. *NeuroReport*, *20*(4), 388–392. https://doi.org/10.1097/WNR.0b013e328324eb4d

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, *132*(3), 617–627. https://doi.org/10.1093/brain/awn279

Shapiro, S. P. (1987). The social control of impersonal trust. *The American Journal of Sociology*, *93*(3), 623–658. https://doi.org/10.1086/228791

Shoda, Y., Wilson, N. L., Chen, J., Gilmore, A. K., & Smith, R. E. (2013). Cognitive-Affective Processing System Analysis of Intra-Individual Dynamics in Collaborative Therapeutic Assessment: Translating Basic Theory and Research Into Clinical Applications. *Journal of Personality*, *81*(6), 554–568. https://doi.org/10.1111/jopy.12015

Sigmon, S. T., Stanton, A. L., & Snyder, C. R. (1995). Gender differences in coping: A further test of socialization and role constraint theories. *Sex Roles*, *33*(9–10), 565–587. https://doi.org/10.1007/BF01547718

Simms, L. J., & Clark, L. A. (2006). *The schedule for nonadaptive and adaptive personality (SNAP): A dimensional measure of traits relevant to personality and personality pathology* (2nd ed.). Springer Publishing Co.

Simpson, T. W. (2013). Trustworthiness and Moral Character. *Ethical Theory and Moral Practice*, *16*(3), 543–557. https://doi.org/10.1007/s10677-012-9373-4

Singer, J. L., Sincoff, J. B., & Kolligian Jr, J. (1989). Countertransference and cognition: Studying the psychotherapist's distortions as consequences of normal information processing. *Psychotherapy: Theory, Research, Practice, Training*, *26*(3), 344.

Sinnaeve, R., van den Bosch, L. M. C., & van Steenbergen-Weijenburg, K. M. (2015). Change in interpersonal functioning during psychological interventions for borderline personality disorder—a systematic review of measures and efficacy. *Personality and Mental Health*, *9*(3), 173–194. https://doi.org/10.1002/pmh.1296

Skodol, A. E., Bender, D. S., Oldham, J. M., Clark, L. A., Morey, L. C., Verheul, R., Krueger, R. F., & Siever, L. J. (2011). Proposed changes in personality and

personality disorder assessment and diagnosis for DSM-5 part II: Clinical application. *Personality Disorders: Theory, Research, and Treatment*, *2*(1), 23–40. https://doi.org/10.1037/a0021892

Snijders, C., & Keren, G. (1999). *Determinants of trust.*

Spitzer, R. L., Williams, J. B. W., Gibbon, M., & First, M. B. (1990). User's guide for the structured clinical interview for DSM-III-R: SCID. In *User's guide for the structured clinical interview for DSM-III-R: SCID.* American Psychiatric Association.

Staebler, K., Helbing, E., Rosenbach, C., & Renneberg, B. (2011). Rejection sensitivity and borderline personality disorder. *Clinical Psychology and Psychotherapy*. https://doi.org/10.1002/cpp.705

Stanley, B., & Siever, L. J. (2010). The Interpersonal Dimension of Borderline Personality Disorder: Toward a Neuropeptide Model. *American Journal of Psychiatry*, *167*(1), 24–39. https://doi.org/10.1176/appi.ajp.2009.09050744

Stern, A. (1938). Psychoanalytic investigation of and therapy in the border line group of neuroses. *The Psychoanalytic Quarterly*, *7*(4), 467–489.

Stern, B. L., Caligor, E., Hörz-Sagstetter, S., & Clarkin, J. F. (2018). An Object-Relations Based Model for the Assessment of Borderline Psychopathology. *Psychiatric Clinics of North America*, *41*(4), 595–611. https://doi.org/10.1016/j.psc.2018.07.007

Stern, S. (2002). The self as a relational structure: A dialogue with multiple-self theory. *Psychoanalytic Dialogues*, *12*(5), 693–714. https://doi.org/10.1080/10481881209348701

Storebø, O. J., Stoffers-Winterling, J. M., Völlm, B. A., Kongerslev, M. T., Mattivi, J. T., Jørgensen, M. S., Faltinsen, E., Todorovac, A., Sales, C. P., Callesen, H. E., Lieb, K., & Simonsen, E. (2020). Psychological therapies for people with borderline personality disorder. *Cochrane Database of Systematic Reviews*, *2020*(5). https://doi.org/10.1002/14651858.CD012955.pub2

Sullivan, H. S. (1953). *The interpersonal theory of psychiatry*. Routledge.

Summers, F. (1994). *Object relations theories and psychopathology: A comprehensive text*. Routledge.

ten Have, M., Verheul, R., Kaasenbrood, A., van Dorsselaer, S., Tuithof, M.,

Kleinjan, M., & de Graaf, R. (2016). Prevalence rates of borderline personality disorder symptoms: A study based on the Netherlands Mental Health Survey and Incidence Study-2. *BMC Psychiatry*, *16*(1). https://doi.org/10.1186/s12888-016-0939-x

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*(1), 128–132. https://doi.org/10.1016/j.yhbeh.2009.03.019

Thielmann, I., Hilbig, B. E., & Niedtfeld, I. (2014). Willing to Give but Not to Forgive: Borderline Personality Features and Cooperative Behavior. *Journal of Personality Disorders*, *28*(6), 778–795. https://doi.org/10.1521/pedi_2014_28_135

Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences*, *1124*, 208–224. https://doi.org/10.1196/annals.1440.012

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Trull, T. J. (1995). *Borderline Personality Disorder Features in Nonclinical Young Adults : 1 . Identification and Validation*. *7*(1), 33–41.

Tsankova, E., Krumhuber, E., Aubrey, A. J., Kappas, A., Möllering, G., Marshall, D., & Rosin, P. L. (2015). *The Multi-modal Nature of Trustworthiness Perception*. http://www.isca-speech.org/archive

Unoka, Z., Seres, I., Aspán, 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 ‘t Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, *108*(3), 796–

803. https://doi.org/10.1016/j.cognition.2008.07.002

Ventura Wurman, T., Lee, T., Bateman, A., Fonagy, P., & Nolte, T. (2020). Clinical management of common presentations of patients diagnosed with BPD during the COVID-19 pandemic: the contribution of the MBT framework. *Counselling Psychology Quarterly*, *00*(00), 1–27. https://doi.org/10.1080/09515070.2020.1814694

Verba, S., & Almond, G. A. (1963). *The civic culture: Political attitudes and democracy in five nations*. https://doi.org/10.1515/9781400874569-010

Wagner, A. W., & Linehan, M. M. (1999). Facial expression recognition ability among women with borderline personality disorder: Implications for emotion regulation? *Journal of Personality Disorders*, *13*(4), 329–344. https://doi.org/10.1521/pedi.1999.13.4.329

Warman, D. M., Lysaker, P. H., Luedtke, B., & Martin, J. M. (2010). Self-esteem and delusion proneness. *Journal of Nervous and Mental Disease*, *198*(6), 455–457. https://doi.org/10.1097/NMD.0b013e3181e086c5

Whitener, E. M., Brodt, S. E., Korsgaard, M. A., & Werner, J. M. (1998). Managers as Initiators of Trust: An Exchange Relationship Framework for Understanding. In *Source: The Academy of Management Review* (Vol. 23, Issue 3).

Widiger, T. A., & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder: shifting to a dimensional model. *The American Psychologist*, *62*(2), 71–83. https://doi.org/10.1037/0003-066X.62.2.71

Wilhelm, F. H., & Roth, W. T. (1998). Taking the laboratory to the skies: Ambulatory assessment of self- report, autonomic, and respiratory responses in flying phobia. *Psychophysiology*, *35*(5), 596–606. https://doi.org/10.1017/S0048577298970196

Willis, J., & Todorov, A. (2006). *First Impressions Making Up Your Mind After a 100-Ms Exposure to a Face*.

Willis, J., & Todorov, A. (2013). Making up your mind after 100-ms exposure to face. *Psychological Science*, *17*(7), 592–598.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A Model of Dual Attitudes. *Psychological Review*, *107*(1), 101–126. https://doi.org/10.1037/0033-295X.107.1.101

Wittchen, H.-U., Zaudig, M., & Fydrich, T. (1997). Skid. Strukturiertes klinisches Interview für DSM-IV. Achse I und II. Handanweisung. *Zeitschrift Fur Klinische Psychologie-Forschung Und Praxis*, *28*(1), 18–27. https://doi.org/10.1026/0084-5345.28.1.68

World Health Organization. (2020). *Guidance for managing ethical issues in infectious disease outbreaks*. World Health Organization. https://apps.who.int/iris/handle/10665/250580

Wright, T. L., & Sharp, E. G. (1979). Content and grammatical sex bias on the Interpersonal Trust Scale and differential trust toward women and men. *Journal of Consulting and Clinical Psychology*, *47*(1), 72–85. https://doi.org/10.1037/0022-006X.47.1.72

Yao, H., Chen, J. H., & Xu, Y. F. (2020). Patients with mental health disorders in the COVID-19 epidemic. *The Lancet Psychiatry*, *7*(4), e21. https://doi.org/10.1016/S2215-0366(20)30090-0

Yeomans, F. E., Clarkin, J. F., & Kernberg, O. F. (2002). *A primer of transference-focused psychotherapy for the borderline patient*. Jason Aronson.

Yeomans, F. E., Clarkin, J. F., & Kernberg, O. F. (2015). *Transference-focused psychotherapy for borderline personality disorder: A clinical guide.* American Psychiatric Publishing, Inc.

Yeomans, F. E., Gutfreund, J., Selzer, M. A., Clarkin, J. F., & Al, E. (1994). Factors related to drop-outs by borderline patients: Treatment contract and therapeutic alliance. In *Journal of Psychotherapy Practice & Research* (Vol. 3, Issue 1, pp. 16–24). American Psychiatric Press.

Yeomans, F. E., Levy, K. N., & Caligor, E. (2013). Transference-focused psychotherapy. *Psychotherapy*, *50*(3), 449.

Young, J. E., Klosko, J. S., & Weishaar, M. E. (2006). *Schema therapy: A practitioner's guide*. Guilford Press.

Zanarini, M. C. (2003). Zanarini Rating Scale for Borderline Personality Disorder (ZAN-BPD): A continuous measure of DSM-IV borderline psychopathology. *Journal of Personality Disorders*, *17*(3), 233–242. https://doi.org/10.1521/pedi.17.3.233.22147

Zanarini, M. C., Frankenburg, F. R., Hennen, J., Reich, D. B., & Silk, K. R. (2005).

(MSAD): Overview and Implications of the First Six Years of Prospective Follow-up. *19*(5), 505–523.

Zimmer-Gembeck, M. J., & Nesdale, D. (2013). Anxious and Angry Rejection Sensitivity, Social Withdrawal, and Retribution in High and Low Ambiguous Situations. *Journal of Personality*, *81*(1), 29–38. https://doi.org/10.1111/j.1467-6494.2012.00792.x

Zimmer-Gembeck, M. J., Nesdale, D., Mcgregor, L., Mastro, S., Goodwin, B., & Downey, G. (2013). Comparing reports of peer rejection : Associations with rejection sensitivity , victimization , aggression , and friendship. *Journal of Adolescence*, *36*(6), 1237–1246. https://doi.org/10.1016/j.adolescence.2013.10.002

Žitný, P., & Halama, P. (2011). Self-esteem, locus of control and personality traits as predictors ofsensitivity to in justice. *Studia Psychologica*, *53*(1), 27–40.

Zucker, L. G. (1986). Production of trust: Institutional sources of economic structure, 1840-1920. *Research in Organizational Behavior*, *8*, 53–111.

APPENDIX I

IAT Stimuli

Main aim: the selection of terms representing trustworthy and untrustworthy dispositions.

Procedure: we presented 6 items for the trustworthy category (i.e. fiducioso, fiducia, fidarsi, sicuro, affidare, credere) and 6 items for the untrustworthy category (i.e., diffidente, sfiducia, sospettoso, dubitare, cauto). Participants rated each of the items on a 7-point scoring scales: "How much do you think this term describe a "trustworthy" person?"

Results: 52 undergraduate female students (*Mean* Age = 22.1, *SD* = 2.4) took part to the pretest. We selected the four terms that showed extreme scores toward the "trustworthy" and "untrustworthy" poles of the continuum. We report the selected stimuli in bold font.

	N=52		$t_{(1,51)}$
	M	*SD*	
affidare	1.23	.61	-26.64***
confidare	1.63	.81	-16.46***
credere	1.92	.92	-12.28***
sicuro	2.27	1.01	-8.77***
fidarsi	1.21	.53	-30.76***
fiducia	1.25	.62	-26.07***
fiducioso	1.27	.81	-19.63***

Note. *** $p < 0.001$

	N=52		$t_{(1,51)}$
	M	*SD*	
diffidente	5.92	.26	64.93***
dubitare	4.98	.82	12.89***
sfiducia	5.35	.88	15.07***
sospettoso	4.94	.66	15.54***
scettico	5.10	.69	16.60***
diffidare	5.85	.36	46.43***
cauto	4.62	1.03	7.79***

Note. *** $p < 0.001$

APPENDIX II

Face Stimuli Pretest

Main aim: the selection of pictures representing faces neutral on both trustworthiness and attractiveness dimension.

Procedure: we presented the pictures of 18 faces (9 female, 9 male) from the NimStim database in a counterbalanced order. All the figures showed a Caucasian identity and in greyscale.

Participants rated each of the pictures on two 7-point scoring scales: “How attractive do you think this person is?” and “How trustworthy do you think this person is?”.

Results: 52 undergraduate female students (*Mean* Age = 22.1, *SD* = 2.4) took part to the pretest. We selected the four identities (2 females, 2 males) that showed scores close to the mid-point on attractiveness and trustworthiness. We report the selected stimuli in bold font (20, 36 for male stimuli and 1, 10 for female stimuli).

	N=52		$t_{(1,51)}$		N=52		$t_{(1,51)}$
	M	*SD*			*M*	*SD*	
Attractiveness_26M	1.58	.93	-18.66***	Trustworthiness_26M	2,50	1,39	-7,76***
Attractiveness _20M	**1.56**	**.77**	**-22.65*** **	**Trustworthiness _20M**	**3,46**	**1,44**	**-2,68****
Attractiveness _21M	1.52	.85	-21.00***	Trustworthiness _21M	2,75	1,31	-6,87***
Attractiveness _22M	2.10	1.19	-11.51***	Trustworthiness _22M	3,17	1,16	-5,11***
Attractiveness _25M	2.65	1.29	-7.48***	Trustworthiness _25M	3,17	1,23	-4,83***
Attractiveness _30M	1.87	1.06	-14.42***	Trustworthiness _30M	3,54	1,36	-2,44**
Attractiveness _36M	**3.13**	**1.37**	**-4.54*** **	**Trustworthiness _36M**	**3,35**	**1,23**	**-3,81*****
Attractiveness _32M	1,71	1,07	-15,38	Trustworthiness _32M	3,04	1,29	-5,34***
Attractiveness _34M	2,58	1,24	-8,26	Trustworthiness _34M	3,02	1,42	-4,97***

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

Bold font for selected stimuli

	N=52		$t_{(1,51)}$		N=52		$t_{(1,51)}$
	M	*SD*			*M*	*SD*	
Attractiveness_1F	**2.71**	**1.46**	**-6.36*****	**Trustwortiness_1F**	**3.02**	**1.16**	**-6.08*****
Attractiveness_2F	1.90	1.03	-14.62***	Trustwortiness_2F	3.65	1.29	-1.92***
Attractiveness_3F	1.90	.91	-16.55***	Trustwortiness_3F	3.83	1.29	-.96
Attractiveness_10F	**3.37**	**1.35**	**-3.37*****	**Trustwortiness_10F**	**3.48**	**1.21**	**-3.08***
Attractiveness_5F	1.87	1.06	-14.42***	Trustwortiness_5F	4.06	1.36	.30
Attractiveness_9F	2.67	1.26	-7.57***	Trustwortiness_9F	4.46	1.27	2.61*
Attractiveness_6F	2.54	1.33	-7.89***	Trustwortiness_6F	4.15	1.52	.72
Attractiveness_7F	2.17	1.11	-11.81***	Trustwortiness_7F	4.42	1.48	2.05*
Attractiveness_8F	1.67	.90	-18.61***	Trustwortiness_8F	3.65	1.38	-1.80

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Bold font for selected stimuli

APPENDIX III

Sentences Stimuli Pretest

The main aim is to select six sentences describing a trustworthy person and six sentences describing an untrustworthy person controlling for any likability effect.

Procedure: In a preliminary brainstorming session, five separate researchers ideated sentences describing trustworthy behaviors toward the self (*N* =48), untrustworthy behaviors toward the person (*N*=39), and unlikeable (but not untrustworthy) behaviors toward the self (*N*=34). The total pool of sentences consisted of 121 items. Subsequently, the number of items researchers reduced by consensus around the best proper sentences to fifteen for each category. To sum up, the final pool of sentences consisted of 45 items.

Participants rated each of the 45 pre-selected sentences on two 7-point scoring scales: "How likable do you think this person is?" and "How trustworthy do you think this person is?"

Results: 184 participants (*Mean* Age = 23.57; *SD* = 5.58; male *N* = 39; female *N* = 145) took part to the study on voluntary basis.

We run principal components factorial analyses with varimax rotations. We determined to select a two-factor solution based on our needs to choose a pool of sentences describing trustworthy behaviors and sentences describing untrustworthy actions. The extracted factors had eigenvalues (7.97 and 3.56) higher than those generated in a random data set. To further interpret the two-factor solution, we examined the items that had salient pattern coefficients. The first factor included sentences solely from the untrustworthy category, while items in the second factor came exclusively from the trustworthy.

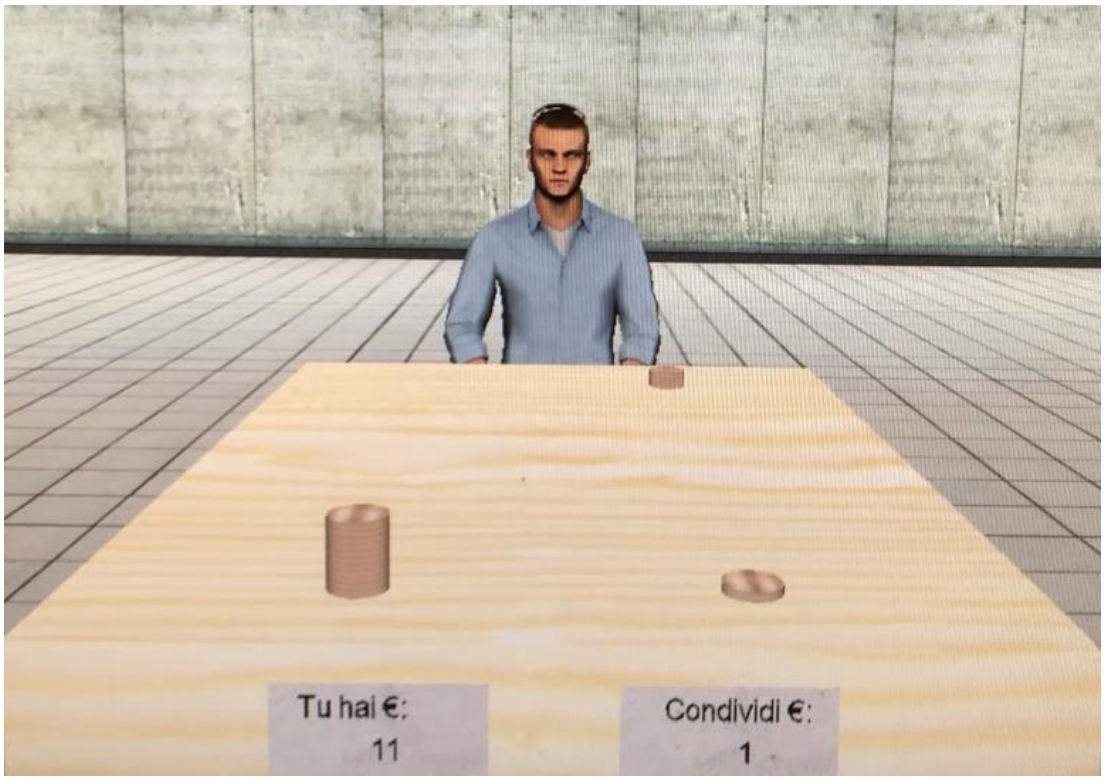
The principal component factor analysis suggests that the first factor represented untrustworthiness and the second trustworthiness.

In the table, we report the items we selected according to higher saturation scores.

Trustworthy factor	Untrustworthy factor
X never lies to me	X betrayed me
X fulfilled his commitment to me	X likes to fool me
X always returns the books that I lend to him/her	X made other people aware of my confidences
X reached me after finding my ID on the ground	X forgot to feed my dog even if I asked her/him
X keeps the secrets I share with him/her	X used sensitive information on my account to dishonor me
X always keeps his/her word	X disappears every time I ask for help

APPENDIX IV

VR Environment



APPENDIX V

Interpersonal Trust in Covid-19 outbreak items English version.

S1. Your neighbor must attend a two-week quarantine period and he/she cannot leave home. He/She has a dog that he/she cannot walk. Your neighbor ensures that he/she frequently toilet the dog. You have had varied experiences trusting your friend in the past.

How strongly would you agree to the following statement?

Get the dog on his door and take him to the park for a walk.

S4. Your best friend's roommate must attend after contact with a COVID-19 positive person. The friend asks you to spend two weeks at yours. He/She ensures not having contact with his/her roommate for at least two weeks. You have had varied experiences trusting your friend in the past.

How strongly would you agree to the following statement?

You host your friend.

S12 You are grocery shopping for your grandparents. You meet a neighbor of theirs who offers to deliver groceries to your grandparents in your place. He ensures that he will keep safely far from your grandparents. You have had varied experiences trusting your grandparents' neighbor in the past.

How strongly would you agree to the following statement?

You thank him and let him deliver groceries to your grandparents

S13. You go and visit your sister after she broke her arm. She is unable to attend to the household on her own. While you are at your sister's, she sneezes. She ensures it is just cold. You have had varied experiences trusting your sister in the past.

How strongly would you agree to the following statement?

You keep visiting her to help with the household.

Interpersonal Trust in Covid-19 outbreak items Italian version.

S1. Il tuo vicino di casa deve osservare un periodo di due settimane di quarantena e non può uscire di casa. Possiede un cane che non potrebbe portare fuori a passeggiare durante la quarantena. Il tuo vicino ti assicura che lava frequentemente il cane. Finora hai avuto esperienze di vario tipo rispetto all'affidabilità del tuo vicino.

Quanto saresti d'accordo con questa opzione?

Vai a prendere il cane sul suo uscio e lo accompagni al parco per una passeggiata.

S4. Il coinquilino del/la tuo/a migliore amico/a deve passare due settimane in quarantena perché ha avuto contatti ravvicinati con una persona che ha contratto COVID-19. Il/la tuo/a amico/a ti chiede di trascorrere le due settimane a casa tua e ti assicura che non ha avuto contatti con il suo coinquilino da almeno due settimane. Finora hai avuto esperienze di vario tipo rispetto all'affidabilità del/la tuo/a amico/a.

Quanto saresti d'accordo con questa opzione?

Ospiti il/la tuo/a amico/a.

S12. Stai facendo la spesa per i tuoi nonni che abitano in un paese vicino a quello in cui risiedi. Incontri un loro vicino di casa che si offre di portare la spesa ai tuoi nonni al posto tuo. Ti assicura che manterrà le distanze di sicurezza nel consegnare la spesa ai tuoi nonni.
Finora hai avuto esperienze di vario tipo rispetto all'affidabilità del vicino dei tuoi nonni.

Quanto saresti d’accordo con questa opzione?

Ringrazi e fai consegnare la spesa ai tuoi nonni.

S13. Vai a trovare tua sorella perché si è rotta un braccio e non è in grado di svolgere le faccende domestiche autonomamente. Mentre sei da tua sorella, lei starnutisce. Ti assicura che il medico le ha diagnosticato un semplice raffreddore. Finora hai avuto esperienze di vario tipo rispetto all’affidabilità di tua sorella.

Quanto saresti d’accordo con questa opzione?

Continui a farle visita per aiutarla con i lavori domestici.

APPENDIX VI

ITSQ - Italian Version

Sotto troverai alcune tipiche situazioni di vita quotidiana. Ad ognuna di queste situazioni corrisponde una potenziale reazione. Prova a immedesimarti in ciascuna situazione e considera brevemente se e quanto reagiresti nel modo indicato.

4. Un/a amico/a ti chiede di prestargli/le una macchina fotografica molto costosa per una vacanza. Finora hai avuto esperienze varie rispetto all'affidabilità del/la tuo/a amico/a.

Quanto saresti d'accordo con questa opzione: Presti al tuo/a amico/a la macchina fotografica.

2. Stai pianificando un viaggio con il tuo partner. Dovete prenotare il volo ma il tuo partner, momentaneamente a corto di soldi, ti chiede di pagare tu il volo piuttosto caro, sebbene anche tu debba stare attento/a alle tue finanze. Il tuo partner ti promette che ti restituirà i soldi nel più breve tempo possibile, appena li avrà. Finora hai avuto esperienze varie rispetto all'affidabilità del/la tuo/a partner.

Quanto saresti d'accordo con questa opzione: Dai al tuo partner i soldi per pagarsi il biglietto.

3. Il tuo partner ti chiede se può prendere in prestito il tuo hard disk esterno per una presentazione. L'hard disk contiene dati molto importanti e privati. Finora hai avuto esperienze varie rispetto all'affidabilità del/la tuo/a partner.

Quanto saresti d'accordo con questa opzione: Dai in prestito al tuo partner l'hard disk.

4. Devi spedire in tempo dei documenti molto importanti tramite posta. Un/a amico/a che si trova da te per una visita si offre di spedire la lettera dato che sulla

via di casa passa davanti ad una cassetta della posta. Finora hai avuto esperienze varie rispetto all'affidabilità del/la tuo/a amico/a.

Quanto saresti d'accordo con questa opzione: Dai al/la tuo/a amico/a la posta da spedire.

5. Un amico/a ti chiede se puoi prestargli un'attrezzatura musicale costosa per una festa e ti assicura che in caso di danno la farà riparare. Finora hai avuto esperienze varie rispetto all'affidabilità del/la tuo/a amico/a.

Quanto saresti d'accordo con questa opzione: Dai al/la tuo/a amico/a l'attrezzatura.

6. Sei a una fermata del bus, lontano da casa tua, e hai perso l'ultimo bus. Un automobilista si ferma e ti offre di portarti nel posto dove abiti. L'automobilista ha tratti facciali rudi, ma anche un'espressione amichevole.

Quanto saresti d'accordo con questa opzione: Ringrazi, sali in macchina e accetti il passaggio.

7. Vuoi sorprendere tua madre con una visita per il suo compleanno. Le tue finanze sono limitate e su Internet trovi solo un'offerta per un viaggio condiviso. L'autista che offre la corsa non ha ricevuto alcuna recensione da parte dai passati passeggeri e quindi non è chiaro se sia una persona simpatica e con uno stile di guida sicuro. Inoltre, non ci sono informazioni sull'aspetto della persona.

Quanto saresti d'accordo con questa opzione: Prenoti comunque la corsa offerta.

8. Stai pianificando un tour di una città e hai un budget limitato. Come sistemazione per il pernottamento sono disponibili una camera d'albergo relativamente economica e una camera gratuita tramite Couchsurfing. Il Couchsurfing è una piattaforma Internet molto conosciuta, attraverso la quale le persone possono mettere a disposizione le loro camere gratuitamente. A giudicare dalla foto su internet,

il giovane che offre la camera non sembra né particolarmente amichevole né poco amichevole.

Quanto saresti d’accordo con questa opzione: Opti per la camera gratuita su Couchsurfing.