Personality Dysfunction Manifest in Words: Understanding Personality Pathology Using Computational Language Analysis

Charlotte Entwistle, BSc (Hons), MRes

Department of Psychology

Lancaster University

This thesis is submitted for the degree of

Doctor of Philosophy

October 2023

Personality Dysfunction Manifest in Words: Understanding Personality Pathology Using Computational Language Analysis

Charlotte Entwistle, BSc (Hons), MRes.

Abstract

Personality disorders (PDs) are some of the most prevalent and high-risk mental health conditions, and yet remain poorly understood. Today, the development of new technologies means that there are advanced tools that can be used to improve our understanding and treatment of PD. One promising tool – indeed, the focus of this thesis – is computational language analysis. By looking at patterns in how people with personality pathology use words, it is possible to gain access into their constellation of thinking, feelings, and behaviours. To date, however, there has been little research at the intersection of verbal behaviour and personality pathology. Accordingly, the central goal of this thesis is to demonstrate how PD can be better understood through the analysis of natural language. This thesis presents three research articles, comprising four empirical studies, that each leverage computational language analysis to better understand personality pathology. Each paper focuses on a distinct core feature of PD, while incorporating language analysis methods: Paper 1 (Study 1) focuses on interpersonal dysfunction; Paper 2 (Studies 2 and 3) focuses on emotion dysregulation; and Paper 3 (Study 4) focuses on behavioural dysregulation (i.e., engagement in suicidality and deliberate self-harm). Findings from this research have generated better understanding of fundamental features of PD, including insight into characterising dimensions of social dysfunction (Paper 1), maladaptive emotion processes that may contribute to emotion dysregulation (Paper 2), and psychosocial dynamics relating to suicidality and deliberate self-harm (Paper 3) in PD. Such theoretical knowledge subsequently has important implications for clinical practice, particularly regarding the potential to inform psychological therapy. More broadly, this research highlights how language can provide implicit and unobtrusive insight into the personality and psychological processes that underlie personality pathology at a large-scale, using an individualised, naturalistic approach.

Contents

List of Tables

List of Figures

List of Abbreviations

ANCOVA	Analysis of Covariance
APA	American Psychiatric Association
ASPD	Antisocial Personality Disorder
BDI	Beck Depression Inventory
BPD	Borderline Personality Disorder
BSL	Borderline Symptom List
CASSIM	Conversation-Level Syntax Similarity Metric
CCRT	Core Conflictual Relationship Theme
CI	Confidence Interval
DSH	Deliberate Self-Harm
DSM	Diagnostic and Statistical Manual of Mental Disorders
DV	Dependant Variable
EV	Emotion Vocabulary
GLMM	Generalised Linear Mixed Model
ICD	International Classification of Diseases
IV	Independent Variable
KMO	Kaiser-Meyer-Olkin
LIWC	Linguistic Inquiry and Word Count
LMM	Linear Mixed Model
LSM	Language Style Matching
M	Mean
NLP	Natural Language Processing
PAI-BOR	Personality Assessment Inventory-Borderline Scale
PCA	Principal Component Analysis
PD	Personality Disorder
RCT	Randomised Controlled Trial
RQ	Research Question
SD	Standard Deviation
SE	Standard Error
ST	Schema Therapy

Acknowledgements

Over the past 4 years, I have been supported by numerous people who have helped me through my PhD journey, of whom I would like to sincerely thank.

First and foremost, I wish to profoundly thank my supervisory team for providing invaluable knowledge and guidance throughout my PhD: Dr Ryan Boyd, Dr Sophie Nightingale, and Prof Paul Taylor. Specifically, I would like to express my deepest gratitude to Ryan for teaching me *everything* I know, while being a continuous source of much needed support and encouragement, extending beyond supervisory obligations; I am eternally grateful. I am also extremely grateful to Sophie for stepping in as my primary supervisor towards the end of my PhD, I could not have asked for a better person to have filled this role; thank you for all of your supportive guidance that got me through the final year of my PhD. Finally, huge thanks to Paul for providing much needed practical guidance, while helping me to see the big picture of my work.

In addition to my supervisory team, I would like to extend my sincere thanks to all collaborators who have passed on their wisdom and helped make our collaborative research a success. I would especially like to thank Dr Katie Hoemann for all of your unobligated input, guidance, and generally insightful thoughts throughout my PhD endeavour. Furthermore, thanks should also go to all of the Psychology Department at Lancaster University for making my PhD an enjoyable, intellectually challenging, and rewarding experience. Lastly, big thank you to everybody who participated in this research, and of course to the Engineering and Physical Sciences Research Council for funding this studentship; this research would not have been possible otherwise.

My final round of thanks goes to all of my family and friends who have provided much needed support, both practical and moral, and encouragement over the past 4 years. In particular, thank you to my partner, Will, for your continuous love and support. Additionally, special thanks to everyone who has dedicated their time to help with childcare – Mum, Grandma, Nana, and Richard – I would not have gotten this far without you. I would of course also like to thank my cat for all of the restoration, joy, and entertainment that you bring each day. Finally, I fully dedicate this thesis to Ava, my inspiration, my light at the end of the tunnel.

Declaration

I declare that the work presented in this thesis is, to the best of my knowledge and belief, original and my own work. The material has not been submitted, either in whole or in part, for a degree at this, or any other university. This thesis does not exceed the maximum permitted word length of 80,000 words including appendices and footnotes, but excluding the bibliography. I have submitted the following chapters of this thesis for peer-reviewed publication:

- Chapter 1.1: “Personality Disorder and Verbal Behaviour” – published in The Handbook of Language Analysis in Psychology 2022.
- Chapter 3: “Uncovering the Social-Cognitive Contributors to Social Dysfunction in Borderline Personality Disorder Through Language Analysis” – published in Journal of Personality Disorders 2023.
- Chapter 4: “Natural Emotion Vocabularies and Borderline Personality Disorder” – published in Journal of Affective Disorders Reports 2023.
- Chapter 5: “Suicidality and Deliberate Self-Harm in Borderline Personality Disorder: A Digital Linguistic Perspective” – to be submitted to Clinical Psychological Science.

Charlotte Entwistle
31/10/2023

Publications

The following publications have been produced while developing this thesis, of which formed several of the thesis chapters:

Entwistle, C., & Boyd, R. L. (2023). Uncovering the social-cognitive contributors to social dysfunction in borderline personality disorder through language analysis. *Journal of Personality Disorders, 37*(4), 444–455. https://doi.org/10.1521/pedi.2023.37.4.444 *(Appears in Chapter 3).*

Entwistle, C., Horn, A. B., Meier, T., Hoemann, K., Miano, A., & Boyd, R. L. (2023). Natural emotion vocabularies and borderline personality disorder. *Journal of Affective Disorders Reports, 14,* 100647. https://doi.org/10.1016/j.jadr.2023.100647 *(Appears in Chapter 4).*

Entwistle, C., Marceau, E., & Boyd, R. L. (2022). Personality disorder and verbal behavior. In M. Dehghani & R. L. Boyd (Eds.), *The Handbook of Language Analysis in Psychology* (pp. 335–356). The Guilford Press. *(Appears in Chapter 1.1)*

Contribution Statements

The following authorship contribution statements outline the contributions of all co-authors to each multi-authored chapter (i.e., the manuscripts and book chapter). The PhD candidate, Charlotte Entwistle (CE), was the principal author of each chapter. Dr Ryan Boyd (RLB) was the project's primary supervisor for the majority of the PhD, followed by Dr Sophie Nightingale (SJN) who took over as the primary supervisor in the final year of the PhD program. Dr Katie Hoemann (KH), Dr Andrea Horn (ABH), Dr Tabea Meier (TM), Dr Annemarie Miano (AM), and Dr Ely Marceau (EMM) contributed to research in this thesis and are each a co-author of one or more of the chapters. The following details the author contributions to each chapter:

Chapter 1.1: Personality Disorder and Verbal Behaviour by CE, EMM, & RLB

CE and RLB conceptualised and designed the book chapter. CE conducted literature searches and wrote the first draft of the book chapter. RLB provided critical feedback on numerous drafts of the book chapter. All authors provided feedback on the final draft and contributed to the final version of the book chapter.

Chapter 3: Uncovering the Social-Cognitive Contributors to Social Dysfunction in Borderline Personality Disorder Through Language Analysis by CE & RLB

CE and RLB conceptualised and designed the study. CE collected all data for the study. CE cleaned the data and prepared it for analysis with expert guidance from RLB. CE analysed the data with expert input from RLB. Both CE and RLB interpretated the results at various stages through numerous discussions. CE wrote the first draft of the manuscript and RLB provided critical feedback on various drafts, including the final version of the manuscript.

Chapter 4: Natural Emotion Vocabularies and Borderline Personality Disorder by CE, ABH, TM, KH, AM, & RLB

CE and RLB conceptualised and designed Study 1 and AM conceptualised and designed Study 2. CE collected data for Study 1 and AM collected data for Study 2. ABH and TM organised and supervised the transcription of spoken conversations from video recordings in Study 2 and assisted with preparing Study 2 data for analysis. CE and RLB analysed and interpretated the data for both studies; all authors provided critical feedback. CE wrote the first draft of the manuscript. All authors contributed to the writing up of the manuscript at various stages and provided critical revisions to the final submitted manuscript. All authors approved the final submitted version of the manuscript.

Chapter 5: Suicidality and Deliberate Self-Harm in Borderline Personality Disorder: A Digital Linguistic Perspective by CE, KH, SJN, RLB

CE and RLB originally conceptualised the study; authors, CE, RLB, and KH all had substantial contribution to the conception and design of the study. The principal author (CE) primarily undertook the tasks involved in conducting this study, including applying for ethical approval, data collection and pre-processing, designing, organising, and overseeing the manual coding of the dataset (carried out by undergraduate research assistants), implementation and application of computational language analysis methods, analysis, and writing up the manuscript. RLB extracted the data from Reddit and provided expert guidance on data pre-processing procedures. All authors provided critical input regarding the analysis of data at various stages, and all authors were involved in interpretating the results. CE wrote the first draft of the manuscript; all authors provided critical feedback and revisions to the manuscript.

Co-Author Certifications

The signatures below provide certification from all co-authors that the stated contributions to the thesis are accurate, and they grant permission for the candidate to include these manuscripts/chapters in this thesis.

Co-author: Dr Ryan Boyd
Date: 29/08/2023

Co-author: Dr Sophie Nightingale

Date: 29/08/2023

Co-author: Dr Katie Hoemann

Date: 29/08/2023

Co-author: Dr Andrea Horn

Date: 29/08/2023

Co-author: Dr Tabea Meier

Date: 29/08/2023

Co-author: Dr Annemarie Miano

Date: 30/08/2023

Co-author: Dr Ely Marceau

Date: 31/08/2023

CHAPTER 1: Introduction

1.1 Personality Disorder and Verbal Behaviour

Charlotte Entwistle, Ely Marceau, Ryan L. Boyd[1]

Text as it appears in:
Entwistle, C., Marceau, E., & Boyd, R. L. (2022). Personality disorder and verbal behavior. In M. Dehghani & R. L. Boyd (Eds.), *The Handbook of Language Analysis in Psychology* (pp. 335–356). The Guilford Press.[2]

The central goal of psychological science, broadly defined, is to discover and understand universal rules that govern our mental worlds — namely, our thoughts, feelings, and behaviors. Social psychology, for example, explores how different social forces influence a person's emotional, cognitive, and behavioral processes. Educational psychology is typically concerned with how people absorb, process, recall, and deploy new ideas and information. The study of personality and individual differences, on the other hand, is geared towards identifying, describing, and explaining the ways in which people *differ* from one another. Why are some people highly motivated to learn new skills whereas others are not? What causes some people to get anxious more easily than others? Why do some people run into a burning building to save a life while everyone else runs away?

In most cases, the study of personality centres around normative differences that typify important, but relatively "neutral", variations between individuals. Whether a

[1] The authors would like to thank Morteza Dehghani for his insightful feedback and comments on an earlier draft of this manuscript.
[2] This text is written in American English, as required for publication in *The Handbook of Language Analysis in Psychology*.

person prefers reading books or going to parties, what we generally care about is understanding the potential ways in which people differ, and how all of these possible "ways of differing" contribute to each person's unique psychological composition. When we talk about personality dimensions like extraversion, for example, we implicitly acknowledge that most people fall somewhere in the "meaty" part of the bell curve; very few people are *extremely* extraverted or *extremely* introverted. Despite the fact that we often talk about personality as "either/or" types, psychologists quietly acknowledge that studying the relatively small number of extreme cases in either direction (e.g., those statistically rare cases of *extraordinarily* extraverted or introverted people) does not usually tell us much about how *most* people function, psychologically.

Some areas of personality research, however, are focused precisely on these more extreme variations between people; the people at the tail ends of the bell curve. When an individual's personality deviates from social norms to the point of causing personal and interpersonal complications, we move into the territory of talking about pathological personalities, or personality *disorders*. Personality disorders are typically defined as pervasive patterns of maladaptive traits and behaviors, beginning in early adult life, which lead to substantial personal distress or social dysfunction, or both, and disruption to others (American Psychiatric Association, 2013). An adult who bursts into tears at the slightest inconvenience, a person who desperately and excessively latches on to others, and a person who sabotages all of their friendships due to envy all exhibit non-normative or extreme behaviors. Importantly, these behaviors are likely to have seriously negative consequences in their day-to-day lives: the emotionally fragile individual may unintentionally drive others away, the socially desperate individual may end up with feelings of abandonment and loneliness when rejected, and the saboteur may be subjected to serious social blowback.

Today, the development of new technologies means that there are many advanced tools that can be used to improve our understanding of personality disorder, and, in turn, the treatment of personality disorder. One particularly promising tool — indeed, the focus of this chapter — is computerized language analysis. Through the exploration and analysis of verbal behavior, it is possible to empirically develop new insights into personality disorder, broadly defined. That is, by looking at patterns in the way that people with personality disorder use language — the words that they use and the way in which they use them — we can gain access into their broad constellation of thinking, feelings, and behaviors, as well as how *precisely* each of these features contributes to their pathology.

To date, however, there has been very little research at the intersection of verbal behavior and personality pathology. Accordingly, the goal of this chapter is to describe and discuss how personality disorder may become better understood through the application of natural language analysis, providing a rough roadmap for the development of personality disorder studies using modern methods. Specifically, in this chapter we will provide:

1. A brief background and overview of personality disorder;
2. An overview of how natural language processing (NLP) methods have advanced understanding within the wider field of psychology, focusing on personality psychology and psychopathology specifically;
3. Examples that demonstrate how NLP methods can help to resolve some of the fundamental, unanswered questions and debates in the personality disorder literature.

1.1.1 A (Very) Brief Overview of Personality Pathology

The idea of personality pathology has a long history, tracing back at least as far as 192 AD. The ancient Roman physician and philosopher Galen conceptualised four “temperaments,” or personality types, on the basis of four bodily fluids known as the Hippocratic humours. The four primary humours — blood, yellow bile, black bile, and phlegm — were understood according to general cosmological theory, whereby they were thought to be manifestations of four primary elements: air, fire, earth, and water. For example, black bile (i.e., "melanchole") was thought to be a manifestation of earth, characterised by coldness and dryness. An excess of “cold and dry” qualities were thought to characterise depression, both metaphorically and literally, suggesting some form of association between the melanchole humour itself and depression (Stelmack & Stalikas, 1991).

In recent decades, personality disorders have received considerable scientific, clinical, and societal attention (Tyrer et al., 2015), and are now among some of the most commonly diagnosed psychological disorders. In a recent systematic review and meta-analysis examining the global prevalence of personality disorder, a worldwide prevalence of 7.8% was reported for personality disorder in the general population (Winsper et al., 2020). To put this into context, the worldwide prevalence of anxiety disorders has been estimated to be 6.7% (Steel et al., 2014) and schizophrenia less than 1% (Charlson et al., 2018). Prevalence rates of personality disorders are even higher in clinical populations: around a quarter of all patients in primary care, half of all patients in psychiatric outpatient settings, and two-thirds of prisoners meet the diagnostic criteria for at least one personality disorder (Tyrer et al., 2015), illustrating the high social and economic costs associated with personality disorder.

Valuably, greater empirical attention has led to improvements in our knowledge of personality disorders. For instance, psychologists have begun to find that various types of personality disorder all share some fundamental commonalities — for example, it is now widely agreed that interpersonal and affective dysfunction are right at the core of personality disorder (Wright & Simms, 2016). Specifically, people with personality disorder tend to experience some combination of social difficulties (e.g., social withdrawal; fear of abandonment), issues around their identity (e.g., being unsure of who they are as a person), and emotional problems (e.g., extreme emotional fluctuations; shallow emotions).

Relatedly, like most forms of psychopathology, personality disorders are almost universally typified by problematic behavior. We all regularly engage in behaviors to cope with or regulate our thoughts and feelings, such as exercising or listening to music. However, we sometimes adopt problematic self-regulatory behaviors that are harmful to ourselves and/or others — we may pick a fight with our spouse or overeat when feeling overwhelmed by stress from our job, for example. Such behaviors are known as maladaptive regulatory behaviors and are seen at elevated rates (and in more extreme forms) in people with personality disorder. For instance, self-injurious behavior (such as intentionally cutting oneself) is an example of a maladaptive regulatory behavior particularly common among people with personality disorder, in which this behavior is often undertaken in an attempt to deal with or relieve feelings of intense negative emotion (e.g., Buckholdt et al., 2015). One explanation as to why people with personality disorder engage in maladaptive regulatory behaviors at an elevated rate is that these behaviors could be an attempt to manage the emotional dysregulation that they experience (Carpenter & Trull, 2013). Further, it is likely that people suffering from personality pathology have exhausted other options for relief from their emotions.

Personality disorder is also associated with greatly elevated threats to well-being, such as increased rates of aggression, physical ailments, and death by suicide (Frankenburg & Zanarini, 2004; Gilbert et al., 2013; Schneider et al., 2008). Concerns over the well-being and life outcomes of people with personality disorder are further amplified by the fact that personality disorders have historically been notoriously difficult to treat — medications are generally ineffective for managing social and identity problems, and individuals with personality disorder are sometimes resistant to therapy (Gabbard, 2012), with more than one third of people with personality disorder dropping out of treatment prematurely (McMurran et al., 2010). Thus, individuals with personality disorder are at a particularly high risk for negative outcomes.

Despite advances in characterizing the etiology of personality disorder development and manifestation over time (e.g., Winsper, 2018), there remains much to be uncovered regarding the underlying structure and manifestation of the disorder and the provision of effective treatment. Given the high risk and high prevalence, more empirical research driven towards developing a greater understanding of personality disorder is essential. Such advances in knowledge would crucially inform clinical practice and, in turn, would benefit those with lived experience of personality disorder, their family and carers, and wider society (Barr et al., 2020). Valuably, natural language analysis is one technique that has the potential to improve our understanding of personality disorder.

1.1.2 Psychology and Language

Verbal behavior analysis has a long history in psychology, particularly in understanding personality and psychopathology. Given that personality disorder rests at the intersection of personality and psychopathology, it is instructive to consider how NLP is often applied to each respective area individually. If NLP methods can help to improve our understanding of both personality and psychopathology individually, we are optimistic that these methods will be critical tools in helping us to better understand personality disorder as well.

1.1.2.1 Language Analysis and Personality

Personality psychology has strong roots in the study of language. Indeed, much of our current knowledge surrounding personality dimensions descends directly from the “lexical approach” to individual differences. Briefly described, the lexical approach to understanding personality and personality structure is based on the assumption that meaningful individual differences will naturally become encoded in the ways in which we describe ourselves and others — our words. Put another way, the lexical approach to personality generally assumes that humans naturally evaluate what makes each person different from one another, and that we logically use words to express, understand, and convey those interpersonal differences that are psychosocially important. The lexical hypothesis has been elaborated on by several personality researchers, including the reduction of trait descriptors down to the most “central” dimensions of personality (Allport, 1937; Cattell, 1943; John et al., 1988).

Whereas the lexical hypothesis is often used to describe how patterns in language can inform our understanding of personality in the broadest terms, a sizeable body of research has demonstrated that *individual* patterns of language use can also be psychologically revealing. Rather than mapping out the structure of personality from Webster’s dictionary, the idiosyncratic ways in which a given person speaks, writes, and types have been shown to reveal what a person pays attention to in the world around them or, put simply, their “attentional habits” (see Boyd & Schwartz, 2021). For example, we expect that — by definition — extraverts will attend more to their social environments than introverts; indeed, there is considerable evidence to date that extraverts use relatively high rates of social words (e.g., “friend,” “family,” and “people”) when compared to introverts (Mairesse et al., 2007). Similarly, people with insecure attachment styles have been found to attend more to themselves as individuals and attend less to themselves in connection with others, as evidenced by higher rates of 1st person singular pronouns, or “*I*-words”, and lower rates of 1st person plural pronouns, or “*we*-words” (Dunlop et al., 2020).

Imagine two people who go out to dinner with a group of mutual acquaintances. Both individuals go through similar behaviors: they each take a shower, get dressed, drive to the restaurant, order a meal, eat, socialize, and return home. However, when asked “What did you do last night?”, each person answers the question differently. The first person, Nathan, says “I went to a restaurant and got myself some dinner.” The second person, Colin, says “All of us met up at a restaurant and enjoyed a lovely meal with friends.” There is a world of difference, psychologically speaking, between Nathan and Colin; the two sentences not only have different meaning in the literal, linguistic sense, but they also provide a logical route to each person’s subjective thoughts and experience of the event. For instance, Nathan’s statement is self-focused and relatively neutral, telling us that he is likely to be far less socially connected than Colin. These differences are both subtle and striking at the same time. Colin’s use of we-words (“us”) and social words (“friends”) can be easily detected by a computer program, despite the fact that the program will have no idea what either sentence actually *means*. In turn, this means that even the simplest computer programs can be used to take a person’s language and convert it into measures of their attentional patterns and, consequently, their psychological traits.

Importantly, the use of language analysis has provided unique insights into personality theories that would otherwise have been difficult to capture through traditional assessment methods. To illustrate how NLP methods have helped to improve understanding of personality, two examples of important lessons we have learned using language analysis include: 1) core dimensions of personality can be traced in language, and 2) how the core personality components fit together to “form” one’s personality and how these components operate in the real-world.

1.1.2.1.1 Lesson 1: Dimensions of Personality can be Traced in Language

The use of natural language analysis has revealed new and interesting dimensions of personality that have not been possible to uncover from traditional methods. For example, insightful early research was conducted by Pennebaker and King (1999), which involved conducting factor analysis on linguistic features derived from natural language data; namely, from student essays. From the language factors generated, core dimensions of personality, or “thinking-styles”, were revealed to be reflected in language. Valuably, this uncovered the possibility of construing personality at an individual level in terms of the language a person uses, demonstrating how language can be used to gain insight into the underlying structure of personality.

Building on this further, a recent study used language from social media posts to develop a new structural model of personality (Kulkarni et al., 2018). In this study, factor analysis was used to derive a trait model based on everyday language; analysing people's words to infer their psychological traits. From this, it was made clear that, perhaps unsurprisingly, different people tend to talk in different ways. These various "dimensions of language" can be thought of as different "dimensions of thinking", which predicted important outcomes, such as intelligence and socioeconomic status. Interestingly, the trait model generated from language differed considerably from the traditional Big 5 personality model. The language-derived trait model therefore allowed for previously unknown insights into the underlying structure of personality, in that it helped to uncover personality dimensions from a new angle. Moreover, this trait model was able to compete with the Big 5 model in terms of generalisability and stability of factors, and was found to have test-retest reliability, predictive validity, and face validity. Thus, this indicates the potential of using NLP methods to learn about the core components and structure of personality in a way that supersedes traditional psychometric approaches, allowing for new contributions to existing personality theories.

1.1.2.1.2 Lesson 2: Language Analysis Reveals How Personality Components Operate in the Real-World

To date, there has been an impressively large body of research working to map out the underlying structure of personality (Digman, 1990; Eysenck, 1991). However, much of the goal of personality research is to understand *how* personality operates in the real world and influences a person's actual behavior — that is, we are often interested in not just the "form" of personality, but the "function" of personality for the individual. The analysis of natural language can provide insight in this respect. For example, extraversion has been associated with greater words spoken and more social language, and this pattern of verbal behavior was also associated with nonverbal social behavior, such as spending more time on the phone and around other people (Tackman et al., 2020). The use of language as a behavioral measure of a person's psychology demonstrates the possibility of gaining new insights into what personality looks like and how it impacts on a person's actual behavior from a new perspective, revealing interesting interactions between personality and real-life situations not seen before.

In addition, NLP methods have the potential to detect individual differences in the real-world with greater accuracy than traditional self-report methods. That is, research incorporating NLP methods can overcome some of the systematic biases associated with self-reports — particularly self-enhancement biases — to uncover meaningful individual differences in psychological well-being (Wojcik et al., 2015). In fact, the conclusion from studies with findings based on self-reports — that political conservatives have greater happiness and psychological well-being than political liberals (Onraet et al., 2013) — was directly contradicted by compelling findings derived from behavioral measures. Contrary to questionnaire-based findings, the analysis of verbal and nonverbal behavior revealed that liberals in fact *experience* and *express* greater happiness than conservatives, evidenced by behavioral indicators such as more intense and genuine smiles and higher rates of positive emotion language (Wojcik et al., 2015). Conservatives report being happier on a questionnaire, but their actual behavior does not support this, suggesting that the questionnaire findings were at least partially driven by self-enhancement motives, highlighting the limitations of relying on self-report measures alone to study individual differences.

1.1.2.2 Language Analysis and Psychopathology

Research on psychopathology is, in many ways, historically interwoven with the idea that our words reflect some of our deepest thoughts, feelings, and behaviors, often unconsciously. For instance, Freud viewed language as a pathway to studying the unconscious forces at work in our minds, and he focused a considerable amount of his life on understanding verbal behavior. In his early work, Freud (1891) proposed a theory of language, whereby he discussed its nature in relation to thought and consciousness and its origins as an instrument of social communication. He also specifically associated language with psychosis, suggesting that dysfunction in word-presentation association processes was the underlying cause of incoherent speech in people with psychosis (Freud, 1915).

In more recent years, psychodynamic thinking remained closely tied to the study of language as a way to understand and explain psychological disorders. Colin Martindale (1975a), for example, proposed that cognition occurs along a continuum, ranging from regressed (unconscious, primary-process) to conscious (secondary-process) thought. Martindale (1975a), like many others of his time, believed that psychopathology (including personality pathology) was a consequence of being in a state of regressed thought and language. From this perspective, people experiencing psychopathology were thought to be "stuck" at an unconscious level of thought, focused solely on primary drives (e.g., sexual drives) and lacking higher level cognitive processes, such as insight and self-awareness, and this cognitive state was believed to be directly visible in a person's language.

Relatedly, in the field of psychotherapy process-outcome research, early pioneering work considered language analysis as a potent methodological tool. For example, Mergenthaler and Kächele (1988) reported on establishing a "computerized databank" to store, organise, and analyze a large volume of verbatim transcripts of psychotherapy sessions. More recently, at the turn of the new millennium, the development of the computerized Gottschalk-Gleser content analysis method (Gottschalk, 2000) facilitated measurement of a magnitude of psychobiological states and traits, such as anxiety and hostility, from the content analysis of verbal behavior. Importantly, computerized methods for quantifying verbal behavior were able to overcome the high demands of manual application, such as significant training requirements and time-intensive hand scoring of transcripts. Such early research incorporating language analysis methods generated a burgeoning interest in the use of modern NLP methods to better manage the complexities of psychotherapeutic processes and better understand treatment outcomes (Pace et al., 2016).

With the rise of personal computing, social media, and smart technology, there has been a recent surge of empirical research incorporating NLP methods to study and understand psychopathology. Much like research on personality, language analysis can help us to understand psychopathology by providing implicit and unobtrusive insight into the core underlying psychology, motivations, and behaviors of people with psychological disorders, allowing for greater understanding of the true nature of such disorders.

Parallel to our examples above, we will briefly illustrate how NLP methods can help to grow our understanding of psychopathology, broadly defined. Namely, we again highlight lessons from NLP research that 1) have helped to pinpoint the nature and structure of psychological disorders, and 2) demonstrate the ability of natural language data to unobtrusively measure and track the progression of psychopathology over time.

1.1.2.2.1 Lesson 1: Language Analysis Allows Insight into the Nature and Structure of Psychological Disorders

Perhaps the most consistent and exemplary finding in clinical NLP research to date is that people with depression tend to use language differently than those without depression, reflecting a generally different social and attentional orientation. Across dozens of studies, individuals with depression are consistently found to use 1st person singular pronouns — that is, self-referential words such as *I*, *me*, and *my* — at relatively high rates, indicating something of an excessive self-focus or an inability to “get out of their own heads” (e.g., Edwards & Holtzman, 2017; Sonnenschein et al., 2018; Zimmermann et al., 2017).

Additional work has helped to extend the nomological network surrounding depression, allowing for better and more accurate typification of the disorder. For example, research exploring the language of individuals suffering from depression finds that they are more prone to “all-or-nothing” thinking, as evidenced by relatively high use of “absolutist” language, such as *“always”* and *“never”* (Al-Mosaiwi & Johnstone, 2018), and use language indicative of greater cognitive load (e.g., *“think”*, *“ought”*; Eichstaedt et al., 2018). Thus, the analysis of language has allowed for valuable insights into the underlying nature of depression, in that it has revealed that self-focus is in fact a consistent, trait-like characteristic of depression, rather than simply a small feature of depression that is only sometimes present. Ideally, this knowledge will improve treatment through providing a target for clinical interventions.

In the domain of psychotic disorders, language analysis has also helped to improve our understanding of the nature of schizophrenia. Schizophrenia is primarily characterized by psychotic symptoms (e.g., hallucinations, delusions), including externalizing biases and paranoid thinking, as well as interpersonal dysfunction and disorganized speech and behavior (APA, 2013). Interestingly, these clinical characteristics can generally be found in language. In particular, people with schizophrenia will often use considerably more external references (i.e., 3rd person plural pronouns, such as *“they”*) in their language compared to the general population (Coppersmith et al., 2015; Fineberg et al., 2015; Lyons et al., 2018). This aligns with the core clinical features of schizophrenia — specifically, the interpersonal dysfunction, externalising biases, and paranoid thinking components — suggesting that elevated use of 3rd person pronouns might be a useful indicator of the disorder.

Furthermore, NLP research has also uncovered markers of mental distress in people with schizophrenia, with associations found between schizophrenia and greater use of health-related words, negative emotion words, and 1st person singular pronouns (Zomick et al., 2019). The finding of the relatively high use of health-related words among people with schizophrenia is particularly interesting as this provides new insights into the nature of the disorder, in that excessive focus on health may be a central component of schizophrenia, a notion that has not yet been theoretically established.

1.1.2.2.2 Lesson 2: Language can Assist with Measuring and Tracking Psychopathology Over Time

Accurate assessment and monitoring of psychological disorders is necessary for informing appropriate diagnosis and treatment. Inaccurate measurement of a disorder can have profoundly negative consequences and has the potential to result in life changing outcomes for the people affected. For example, misdiagnosis of bipolar disorder has resulted in delays in the provision of appropriate treatment, subsequently leading to increased risk and negative outcomes for those affected, such as increased suicide risk, length of hospitalization, and social impairment (Altamura et al., 2015). The ability to measure and track psychopathology also allows for the evaluation of clinical treatments, by monitoring individual responses to treatment in real-time, and so it plays a vital role in the development of effective, individualized treatments. Moreover, it is essential that psychological disorders can be accurately and unobtrusively monitored so that it is possible to observe how a given disorder manifests over time, which would provide insight into the developmental trajectory of a psychological disorder. Vitally, NLP methods have the potential to make important contributions to both the measurement and tracking of psychopathology over time.

To date, numerous studies have used linguistic markers to detect and track changes in psychopathology. For instance, through measuring linguistic markers of mental distress (e.g., 1st person singular pronouns, negative emotion words) and observing changes in these patterns over time, studies have been able to measure and observe changes in general psychological well-being and mental distress at a large-scale (e.g., Bagroy et al., 2017; Guntuku et al., 2020). Specifically, in one study, a machine-learning model built on social media data could detect mental health expressions (i.e., words and phrases related to mental health) with 97% accuracy, which resulted in the development of a "Mental Well-being Index" (Bagroy et al., 2017). Vitally, this index was able to predict the prevalence of mental health issues across different universities.

Importantly, the ability to detect the presence of psychopathology through language means that linguistic features could be used alongside other, more traditional measures in the assessment of psychological disorders, providing an unobtrusive and implicit contribution to the measurement of psychopathology. For example, reliable linguistic markers of mental distress (e.g., 1st person singular pronouns) could be incorporated as additional outcome measures, alongside other clinical outcome measures (e.g., self-report mental health measures, clinician-rated measures), in the assessment of general mental distress in people undergoing a psychological therapy, to evaluate the effectiveness of such therapy.

Moreover, language has been used to monitor specific psychological disorders. This has consistently been demonstrated with depression (e.g., Dean & Boyd, 2020; Schwartz et al., 2014; Park & Conway, 2017), whereby changes in depressive states have been successfully measured through language. Relatedly, changes in suicidal ideation can also be precisely detected and measured through language (e.g., De Choudhury et al., 2016; Ma-Kellams et al., 2016). Specifically, in one study, through measuring changes in language on social media, it was possible to predict with high accuracy whether a given person would make a post on an online suicide help forum, which is a strong indicator of suicidal ideation (De Choudhury et al., 2016). The ability to detect depression and suicidal ideation through language has obvious clinical implications: when detected, clinical interventions can be provided to try to address such depressive symptoms and suicidal thoughts before they worsen.

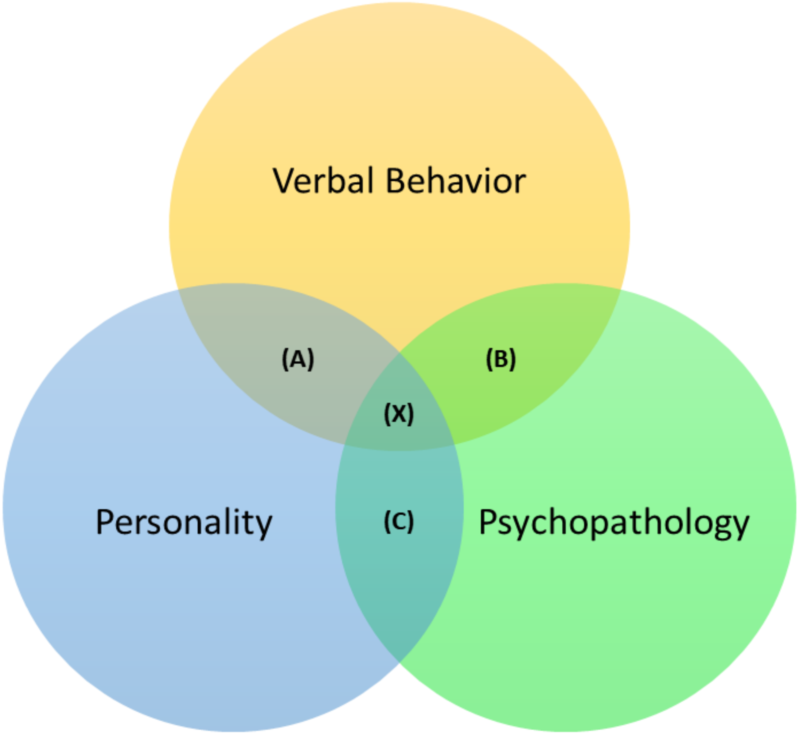
Similarly, language has been used to measure changes in psychotic symptoms. Through the examination of changes in language used in social media posts, research has been able to measure and predict changes in psychotic symptoms, such as the occurrence of delusions and hallucinations (Birnbaum et al., 2019). Most importantly, this research also revealed the possibility of identifying early warning signs of psychotic relapse through linguistic and behavioral markers. Specifically, increased use of negative emotion words, swear words, death words, and 1st and 2nd person pronouns were strong predictors of psychotic relapse. A machine learning model developed from linguistic and behavioral features identified in the study was able to predict psychotic relapse with 71% accuracy. Importantly, this ability to accurately detect and monitor psychopathology means that individualized clinical interventions can be provided in a timely manner.

1.1.2.3 Personality Disorder and Language Analysis

Today’s automated language analysis methods have made it possible, and very accessible, to conduct large-scale, objective linguistic analyses to gain insight into people’s underlying psychological and personality processes. However, what is peculiar is that, despite the success of NLP methods in the fields of psychopathology and personality psychology individually, these areas of research have not been brought together in a general, formalized way. This lack of unification across disciplines is particularly strange given the relatedness and intertwining nature of psychopathology and personality psychology. Right at the core of the intersection of personality and psychopathology are personality disorders (see Figure 1.1).

Figure 1.1

The Intersection of Verbal Behavior, Personality, and Psychopathology



Note. To date, there has been considerable research at the pairwise intersections of Verbal Behavior, Personality, and Psychopathology, but almost no research that integrates all three domains (X).

Compared to psychopathology and personality research, the use of computational language analysis in the study of personality disorder is much more rare. If we can conduct insightful personality and psychopathology research using NLP methods, we should be able to conduct insightful personality disorder research using the same approaches, given the high degree of interconnectedness across each area. Accordingly, the remainder of the chapter will integrate core ideas at the personality–language intersection with those at the psychopathology–language intersection, providing views and recommendations for how we can begin to fill the major gaps in knowledge in this area.

1.1.3 Using Language Analysis to Understand Personality Disorder

As in personality and psychopathology research, language analysis methods have the potential to help to improve our understanding of personality disorder, which, in turn, would have positive implications on the treatment of personality disorder. Specifically, we can use NLP methods to help to shed light on the answers to some of the major open questions and debates in the personality disorder literature. In particular, two large, frequently debated topics in the personality disorder literature surround the assessment of personality disorder and the developmental trajectory of personality disorder across the lifespan. Accordingly, we will discuss and provide examples describing how language analysis can insightfully contribute to these debates.

1.1.3.1 How Should We Assess Personality Disorder?

How personality disorders should be assessed is heavily debated (see Kim & Tyrer, 2010), particularly in terms of the measurement, assessment, and classification procedures. This topic is especially important given that the assessment and classification of personality disorder directly impacts on treatment decisions: an individual can only be provided with appropriate treatment for an affliction when it has been accurately determined *what* affliction treatment needs to be provided for. If the assessment and diagnosis procedure fail to accurately identify a person as having a personality disorder, the opportunity to provide appropriate and essential treatment in a timely manner may be missed.

1.1.3.1.1 How is Personality Disorder Currently Assessed?

Personality disorder diagnosis typically involves multiple clinical assessments through structured interviews, in which a clinician asks about a person’s life, feelings, thoughts, and behaviors, along with self-report and observer-report (i.e., reports from other people who are not being assessed, such as family or friends) measures to assess the quantity and severity of personality disorder features. Clinical observations of how a person behaves and examinations of psychiatric history and medical records (e.g., previous hospitalizations) are also sometimes carried out as part of the assessment procedure. Such assessment methods are currently used to measure and classify distinct personality disorders (e.g., borderline personality disorder) based on clinical features specific to a given personality disorder.

To illustrate the personality disorder diagnostic procedure, imagine a woman named Lucy. Lucy has recently become an inpatient in a psychiatric hospital due to intense suicidal ideation. After having discussions with Lucy about her mental health and past experiences, and after examining her clinical history and observations from clinical staff, the psychiatrist responsible for Lucy's care suspected that she might have a personality disorder. To investigate the possibility that Lucy does indeed have a personality disorder, the responsible clinician decides to carry out a clinical interview using a personality disorder assessment measure, which assesses Lucy's personality disorder symptoms (e.g., impulsivity, self-harm) based on diagnostic criteria. From this assessment, and taking into account Lucy's clinical notes and clinical history, the clinician concluded that Lucy does in fact meet the criteria for personality disorder — specifically, borderline personality disorder — resulting in a formal diagnosis being provided.

Diagnosis of personality disorders are undertaken on the basis of specified diagnostic criteria outlined within diagnostic classification manuals of psychological disorders. Currently, there are two dominant systems: The *Diagnostic and Statistical Manual of Mental Disorders* (DSM) and the *International Classification of Diseases* (ICD). In the latest version of the DSM — the DSM-5 — ten distinct personality disorders were outlined and categorized into three clusters, based on shared characteristics (e.g., anxiety, dependency on others, fear of abandonment). Notably, however, the typological approach to personality disorder classification and the underlying structure of personality disorder presented in the DSM is strongly debated. Such debates are primarily a result of most people now coming to understand that individuals generally do not fit into clear categories or types (e.g., Wilmot et al., 2019), as presented in the DSM. In recent years, many have argued for a major change in the entire DSM classification system (e.g., Clark et al., 2017; Newson et al., 2020).

Consequently, in the latest version of the ICD — the ICD-11 — major changes have been outlined regarding personality disorder classification, which are set to come into effect in 2022. The ICD-11 has completely shifted from the traditional typological approach to personality disorder classification (i.e., classification based on the quantity of disorder-related symptoms and behaviors that reach a particular threshold) and instead adopted a dimensional, trait-based approach (i.e., along a continuum of normal-abnormal personality). In this approach, distinct personality disorders are not outlined — meaning that a person would not be classified with a discrete disorder, such as "antisocial personality disorder". Instead, the focus is on core personality dysfunction and global level of severity, with the ability to classify personality disorders across three levels of severity (i.e., "mild", "moderate", "severe"). The diagnostic criteria for personality disorder diagnosis is based on a global evaluation of personality functioning, in comparison to arbitrary symptom thresholds. Personality disorder severity assessment is dependent on the overall degree to which personality dysfunction causes disturbances in relation to aspects of the self, interpersonal relationships, affect, cognition, and behavior. An individual with a "severe" personality disorder, for example, might have strong suicidal tendencies and regularly act aggressively towards themselves and others, whereas an individual with a "mild" personality disorder may regularly experience intense mood swings and have a fear of abandonment, but may not act on their negative thoughts.

Discrepancies between the two major psychological disorder diagnostic systems in how personality disorders should be classified exemplifies how little we really know about personality disorder. Moreover, although the ICD approach to personality disorder classification is a move in the right direction, there are still concerns regarding the way in which personality disorders are typically assessed; that is, with a heavy reliance on self-report measures. One important flaw in this regard surrounds the biases that accompany self-report measures, such as social desirability bias and other self-serving biases. Most individuals are motivated to view and present themselves in a positive light, which results in skewed responding to self-reports about problematic and clinically meaningful (but socially undesirable) thoughts, feelings, and behaviors. People may also lack insight into their own thoughts and feelings to allow accurate reporting on a psychological questionnaire or in a clinical interview. Self-report biases are of particular importance in clinical research, as if people are not reporting their symptoms accurately, it could lead to misdiagnosis and inappropriate treatment.

In addition to the issues that typically accompany self-report measures as a result of the questionnaire-*takers,* there are also issues surrounding the design of self-report measures as a result of the questionnaire-*makers*. A self-report questionnaire can only capture what we ask it to measure. It would be difficult to discern that a person is engaging in self-harm behavior if we only ask them about their exercise regimen, for example. Thus, it is important that we are not only aware of what it is that we *want* to learn from a given measure, but also that we create and deploy measures that allow participants to tell us information that we might *need* but may never have thought to ask.

1.1.3.1.2 How Can Language Analysis Assist with Personality Disorder Assessment?

The analysis of natural language has the potential to improve the personality disorder assessment procedure; predominantly resulting from the ability to measure various psychological and personality constructs through language. By looking at people's word use and sentence structure, as illustrated with the examples of the lessons learned from NLP research discussed earlier, we can measure broadly defined emotions, thoughts, and motivations. For example, NLP methods can describe a person's current emotional state (Park & Conway, 2017), core values (Boyd et al., 2015), and cognitive processes (Khawaja et al., 2014).

There is a vast amount of evidence showing the potential of making precise psychological measurements by analyzing natural language (e.g., Golbeck, 2016; Hall & Caton, 2017; Yarkoni, 2010). Consequently, it is possible to concurrently measure a wide range of psychological and personality constructs using NLP methods. The ability to measure various psychological and personality constructs all at once allows for a greater understanding of how such constructs are related to one another, and so can provide detailed insight into the structure of one's underlying psychology. Therefore, analysis of the language of people with personality disorder should help to provide insight into the composition and structure of personality disorder. For example, if we find that people with personality disorder who are particularly high in impulsivity always express a substantial amount of emotion in their language, this would provide insight into how two main components of personality disorder — impulsivity and affective dysfunction — interact with one another and influence a person's coping behaviors within and across situations.

1.1.3.1.3 Understanding the Structure of Personality Disorder

If language analysis can facilitate a more refined knowledge of the structure of personality disorder, it should be possible to improve the assessment of personality disorder as well. Through language analysis, it is possible to uncover central components of personality disorder. For example, if we find that people with personality disorder consistently use language indicative of disconnection from others, such as relatively few "*we*-words" and relatively high rates of "*I*-words" and "*they*-words", this would reveal that a critical feature of the interpersonal dysfunction underlying personality disorder is a fundamental view of oneself as disconnected from others.

Greater understanding of the core components of personality disorder and how they operate would have implications for treatment, as such detailed understanding of core personality disorder components would allow for these components to be a central focus in clinical practice. For example, if the analysis of language provides insight that identity problems arise from an interaction between the formulation of self-concept and aspects of interpersonal dysfunction, this would allow for the development of targeted interventions in clinical practice. Through the computational analysis of the words that people use (and do not use) in close proximity to self-references, then, we may be able to find important, precise aspects of an individual's self-schema that would benefit from further development or restructuring.

Moreover, improved understanding of the underlying structure of personality disorder could feed into the future development of self-report measures in assessing personality disorder, thereby improving the validity of such measures by ensuring that they measure what they are supposed to measure (i.e., core personality disorder features). For example, if the analysis of autobiographical narrative text finds that "early romantic frustrations" are a prevalent theme among individuals with personality disorder, this information can be adapted into a more straightforward self-report format, addressing the above-mentioned issues surrounding whether we are "asking the right questions". Additionally, insight into the structure of personality disorders should help to improve the classification procedure by helping to understand whether there are clearly distinct personality disorders with separate underlying processes (i.e., the typological approach), or whether they all have the same underlying processes with slight variations in traits (i.e., the dimensional approach), or potentially some combination of the two.

Definitions of the underlying structure and composition of personality disorder are frequently debated, which is primarily a consequence of highly inconsistent findings across research programs (Wright & Zimmermann, 2015). For instance, there are still no definitive answers as to what, and how many, core factors comprise personality disorders, or how core factors (e.g., negative affect, interpersonal dysfunction) vary across different types of personality disorder. Stemming from these unanswered questions surrounding the structure of personality disorders, there remains major debates around whether personality disorders are typological or dimensional by nature. And, although many are now moving in favor of the dimensional model, it is not yet known whether the ICD-11 captures this sufficiently.

At present, research has attempted to answer questions around the structure of personality disorders by conducting factor analyses to evaluate models of the factors, or processes, underlying personality disorders (Wright, 2017). This typically involves applying factor analytic techniques (i.e., looking at patterns of covariation in measurable behaviors to identify factors) to personality disorder diagnosis criteria (e.g., Wright et al., 2015) and specific personality disorder symptoms and features (e.g., Trull et al., 2012) to determine the underlying factorial structure of personality disorders. However, despite the many valuable insights that factor analysis methods have provided regarding the structure of personality disorder, there is still no consensus on how exactly personality disorders are structured.

Accordingly, through the use of NLP methods, there are several ways in which we can improve understanding of the underlying structure and composition of personality disorder. For instance, because words reflect attention (Boyd et al., 2019), it is possible to implicitly see what a person is attending to every time they speak or write. If we can use language to measure many different things that a person is paying attention to, this would allow access to a substantial amount of insightful information about the person, which would take a considerably long time to obtain from a self-report questionnaire and would be sorely lacking in objectivity. In the context of personality disorder, if a person is found (through language analysis) to be paying considerable attention to goal attainment, for example, and negative affect also becomes extremely salient almost always simultaneous to when attending to such goals, this would highlight a critical interaction between the motivational and affective dysfunction components of personality disorder.

The ability to extract a large amount of meaningful information regarding people's attentional patterns and personality traits from a small amount of language data, such as a social media post, means that language data could be incorporated into factor analyses and complement self-report data in attempting to determine the central factors and processes underlying personality disorder. Complementing self-report data with natural language data will also help to overcome some of the issues associated with self-report measures, as the person will generally not be aware that they are being assessed through their language, reducing the likelihood of biased data.

The potential of language analysis methods to provide insight into the underlying structure and composition of personality disorder can be seen from the examples of the lessons learned from language analysis in general personality research (e.g., Kulkarni et al., 2018; Pennebaker & King, 1999; Yarkoni, 2010). The success of language analysis methods in helping to explore and understand the structure of personality demonstrates that it is very possible to do the same with personality disorder, given that the approach would be no different. Additionally, there has in fact been some research that has touched upon the idea of using language analysis to try to understand the structure of personality disorders. Such research has primarily been conducted on what are known to be "dark" personalities, such as psychopathic and narcissistic personalities. For example, language analysis studies have been conducted that have revealed linguistic markers of features, or subfactors, of psychopathy (e.g., callousness; Hancock et al., 2018). Importantly, this demonstrates how personality disorder features can be uniquely and implicitly measured through language and how it is possible to use language to differentiate between specific factors and processes that underlie personality disorders.

It is therefore clear that there is potential to use language analysis methods to gain insight into the underlying structure and composition of personality disorders. Consequently, this would help to improve the accuracy of the assessment and classification of personality disorders, given that it needs to be known what exactly comprises a personality disorder before it can be accurately assessed and classified. Importantly, accurate assessment means accurate diagnosis, which means more appropriate and potentially more effective treatment. Additionally, once it is possible to diagnose someone accurately — not just with *what* personality disorder they have, but what *specific* nuances of problems across domains are present (e.g., social, affective, cognitive, etc.) — this would allow intervention to be better targeted, in a more individualized and personalized way.

1.1.3.1.4 Improving the Treatment of Personality Disorder

We have discussed how language analysis can *indirectly* impact the treatment of personality disorder — through providing insight into the structure of personality disorder and improving assessment and classification procedures — but it is also important to discuss how language analysis may *directly* inform improved personality disorder treatment. Improvements in the treatment of personality disorder is essential as further research is needed to support the efficacy of personality disorder treatments currently available (Bateman et al., 2015). While several studies suggest that psychotherapy is an effective treatment option for borderline personality disorder (Cristea et al., 2017), there is relatively sparse evidence for others, such as antisocial personality disorder (Gibbon et al., 2020). Thus, there in a vital need for improvements in research studying treatments for personality disorder.

Traditionally, treatments for personality disorder (or any clinical condition), whether pharmaceutical or psychotherapeutic, are evaluated through randomized controlled trials (RCTs), otherwise known as clinical trials. Generally speaking, clinical trials involve randomly assigning people to one of two or more conditions, one of which will be a control condition (i.e., not the treatment intervention) and at least one other will be the treatment condition. To evaluate the effectiveness of treatment, baseline outcome measures (e.g., general psychological functioning; symptoms) are assessed before the intervention and again after the intervention, usually with multiple follow-up assessments. These assessments are then compared before and after the intervention and between conditions, allowing for empirical evaluation of the treatment. The RCT approach is seen as the “gold standard” for evaluating the effectiveness of a clinical treatment.

Despite the fact that clinical trials should be the gold standard for evaluating clinical treatments, they are often not conducted according to the standard necessary — that is, in accordance with the guidance on things like appropriate sample size and follow-up length. Regarding personality disorder treatment specifically, one particular issue surrounds the outcome measures used to assess personality disorder treatment, as they are surprisingly inconsistent, varied, and often do not measure the same constructs

(Bateman et al., 2015), making comparisons between findings difficult. The issue surrounding personality disorder treatment measurement again illustrates our current shortcomings in knowledge around exactly which aspects of personality disorder *should* improve with treatment and indicates the importance of incorporating new approaches to improve the research evaluating treatments of personality disorder.

With advances in technology (e.g., smartphone sensors), there has been a major uptake in using innovative approaches to monitor psychopathology in the real world (e.g., Ben-Zeev et al., 2015; Seppala et al., 2019; Shatte et al., 2019). Put another way, much attention in clinical research is now being directed towards using technology to observe people's responses to treatment in real-time, outside of therapeutic settings (i.e., in people's natural environments). Accordingly, this is where natural language data fits in. It is possible to monitor the verbal behavior of people with personality disorder throughout the course of treatment, measuring changes in their underlying psychology, emotions, and motivations in response to treatment in real-time. Importantly, this can be done alongside clinical trials, using linguistic features as complementary outcome measures, thereby strengthening the outcome measures currently used in personality disorder research.

The possibility of using natural language data to monitor changes in psychology and mental health over time among people receiving clinical treatment has been well evidenced (e.g., Arevian et al., 2020), and so could be fruitfully applied to clinical personality disorder research. For instance, natural language data could be used to measure psychotherapeutic change in people with personality disorder in response to treatment, which has in fact been explored already (Arntz et al., 2012). Valuably, this exemplifies how linguistic outcome measures could complement other traditional outcome measures of treatment effectiveness for personality disorders, providing an implicit measure of current mental state and psychological progress.

As well as utilising natural language measures to investigate outcomes over the course of psychological treatment, a closely linked opportunity for future NLP personality disorder research concerns prediction of treatment outcomes. To illustrate an example of how language analysis could help with the prediction of treatment outcomes, some research has utilized manual transcript scoring methods, such as the Core Conflictual Relationship Theme (CCRT; Luborksy, 1998), to investigate factors related to treatment response (e.g., Hegarty et al., 2020). However, manual scoring of transcripts is very time consuming and effortful. In future, one potential opportunity afforded by NLP methods could be the semi-automation of manual scoring methods, such as the CCRT, through computerization of such methods. Although this approach would require careful consideration to ensure clinical validity, a significant benefit would be the relative ease via which computerized scoring would permit analysis of large-scale data.

Furthermore, the client-therapist relationship represents a particularly salient area to explore given its strong association in predicting psychotherapy outcomes (Flückiger et al., 2018; Wampold et al., 2015). Specifically, language can be used to explore and describe therapeutic alliance through various NLP techniques. One possible technique involves observing similarities in the language style (the way in which words are used and sentences are structured) between the client (e.g., person with personality disorder) and therapist. A language style matching (LSM) score can be quantified to empirically measure this (Gonzales et al., 2010), as can several other NLP techniques, such as conversation-level syntax similarity metric (CASSIM; Boghrati et al., 2018). Importantly, language style similarities are thought to map on to the interpersonal coordination of psychological states (Ireland & Pennebaker, 2010). Thus, measuring the matching of language styles between a client and their therapist should be insightful regarding their psychological connectedness and rapport.

To illustrate the potential to use language to measure client-therapist alliance, one study found that higher LSM between the therapist and client at the start of therapeutic treatment predicted greater therapeutic rapport (Borelli et al., 2019). Additionally, machine learning models built on linguistic features from psychotherapy sessions have been found to have modest accuracy in predicting therapeutic alliance (Goldberg et al., 2020). The ability to implicitly predict the likelihood of a therapeutic alliance forming should help to guide future directions in therapy practice. For example, one could imagine the potential for using language to elucidate client-therapist dyadic interactions at a very early stage in therapy and use this insight to make therapeutic recommendations that will enhance treatment outcomes.

The examples presented illustrate how the analysis of language can be used to guide and improve the treatment of personality disorder. In relation to this, a recent article by Goldberg and colleagues (2020) highlights a number of useful practical suggestions and future recommendations for using machine learning and NLP methods in psychotherapy research, which are informative for future study of personality disorder. Examples of these recommendations include using large datasets, having reasonable expectations, and developing interdisciplinary collaborations across the fields of clinical psychology and computer and data science in particular. The development of such interdisciplinary collaborations should improve the likelihood of the implementation of new technologies into clinical practice, by helping to reduce the barriers between science and practice. Already there are a small number of studies utilizing machine learning in psychotherapy research more broadly (e.g., Aafjes-van Doorn et al., 2020), and so the opportunity for studies in personality disorder populations is a promising area of future research.

1.1.3.2 What is the Developmental Trajectory of Personality Disorders?

In addition to debates around how personality disorders should be assessed, classified, and treated, the developmental trajectory of personality disorders across the lifespan is another area of complexity in personality disorder literature. Put simply, there is much to learn about how personality disorders develop and manifest over time. Specifically, understanding the complexity of gene-environment interactions in both vulnerability and resilience factors is an important area for further research (Amad et al., 2014; Bulbena-Cabre et al., 2018; Marceau et al., 2018; Witt et al., 2017). Questions surrounding how personality disorders manifest over the course of the lifespan have also been debated, primarily due to mixed findings. For instance, in past decades, the main consensus was that personality disorders remain relatively stable over time, whereas recent research suggests that personality disorders have both stable and dynamic aspects (Hopwood & Bleidorn, 2018). Moreover, there is little knowledge on how personality disorders manifest in later life (Oltmanns & Balsis, 2011).

Understanding how personality disorders develop and manifest over time is of major importance, as knowing where personality disorders originate from and at what point they are likely to cause serious problems would allow for risk factors to be identified that could inform preventative early interventions. Additionally, knowing how personality disorders are likely to progress throughout life is of value to determine factors, such as particular stages or events in life, that may exacerbate or reduce symptoms. This is useful information regarding the treatment of personality disorders as demographic specific treatment could be provided based on this knowledge, ensuring that treatment is better tailored to the individual. In line with this, if it is known when symptoms are likely to be at their worst, this makes it possible to predict “flare ups”, such as extreme emotional outbursts, which could then be appropriately targeted in a timely manner.

Research into the origins and developmental trajectory of personality disorders is somewhat limited at present (Bulbena-Cabre et al., 2018; Hopwood & Bleidorn, 2018), which at least partially explains the lack of clarity on these topics. Another contributing factor to the current lack of consensus on how personality disorders develop stems from the fact that there are considerable variations in how personality disorders manifest over time between individuals, suggesting a need for an individualized, person-centred approach (Hicks et al., 2017). Indeed, an individualized approach is exactly what language analysis methods can provide. Through exploring people’s language, there is potential to observe the development and manifestation of personality pathology at an individual level and at a large scale.

1.1.3.2.1 How Can Language Analysis Improve Understanding of the Developmental Trajectory of Personality Disorders?

Analysis of natural language can help to provide insight into the development and manifestation of personality disorders given that it is possible to identify markers of psychopathology from language. Research discussed earlier which shown evidence of reliable linguistic markers of psychological disorders, such as depression (e.g., Edwards & Holtzman, 2017) and schizophrenia (e.g., Coppersmith et al., 2015), illustrates the possibility of using language to detect the presence of psychopathology. Accordingly, linguistic markers of personality pathology could be used, in conjunction with other measures, to help triangulate the onset and course of personality disorders, and thus what has led up to this onset.

To illustrate an example of how such methods may be used to better understand the development of personality disorder, imagine a study which retrospectively measures longitudinal changes in language patterns of individuals diagnosed with personality disorder in the years or even decades leading up to their diagnosis, through easily accessible natural language data, such as social media posts dating back years. If drastic changes in language are uncovered, this may point to factors related to the onset of personality disorder symptoms. Valuably, this would help with the identification of important precursors and risk factors for personality disorder development and could complement the increasing focus on prevention and early intervention in the treatment of personality disorder (e.g., Chanen et al., 2017).

Furthermore, the ease of collecting and analyzing large-scale language data makes it possible to take an individualized approach to observing changes in personality pathology over the lifespan, which can be done in real-time. Through precise language measurement, it is possible to observe how underlying psychological and personality processes of personality disorders change over time using a longitudinal, repeated measures design. The ability to measure and track the progress of psychological disorders, such as personality disorder, through language is strongly evidenced by the studies discussed earlier demonstrating the possibility of using language to measure changes in depression (e.g., Park & Conway, 2017), suicidal ideation (e.g., De Choudhury et al., 2016), and psychotic symptoms (e.g., Birnbaum et al., 2019). This indicates the potential of closely studying language patterns of people with personality disorder across the lifespan, through regular language measurements (e.g., from diaries), to identify markers of personality disorder symptoms and fluctuations in severity. Importantly, the ability to measure manifestations of personality pathology at a large-scale should help to guide treatment on an individual basis and provide an avenue for interventions aimed at prevention and early intervention.

1.1.4 Challenges and Limitations

In this chapter, we have outlined and provided numerous examples to show how natural language analysis methods can be used to better understand personality disorder. However, despite the many exciting possibilities NLP methods bring to the study of personality disorder, it is important to acknowledge the potential challenges and limitations of the application of such methodology in the clinical field. First, it seems necessary to highlight that language analyses conducted on large, naturalistic datasets often result in small to moderate effect sizes (Matz et al., 2017). Likewise, many of the NLP research findings discussed in this chapter consisted of small to moderate effect sizes, which raises questions about the potential impact and practical relevance of such findings.

Nonetheless, the notion that large-scale linguistic analyses often result in small effect sizes can be at least partially be explained by the fact that language data analyzed will typically be naturalistic, meaning that the effects are taking place in the real-world and outside of carefully-designed experimental settings that are largely under meticulous researcher control. The real world is messy and full of confounds, but we can be confident that even when we find a small effect size in such settings, the effect is real and not due to chance. Moreover, as emphasized by Matz and colleagues (2017), modern big-data analyses provide psychologists with new opportunities to identify "weak but reliable signals in a complicated world". Thus, small effect sizes resulting from big-data analyses can still be highly meaningful and impactful and provide the opportunity to understand the underlying psychology and behavior of billions of people from around the world.

Second, many (perhaps most) language-based studies to date are limited by the fact that they were almost exclusively conducted in the English language. It is therefore unclear how such findings would translate to other languages. For example, it is not yet known whether the findings that people with depression use language indicative of more absolutist, all-or-nothing thinking (Al-Mosaiwi & Johnstone, 2018) and of a higher cognitive load (Eichstaedt et al., 2018) would be replicated in non-English languages, such as the Russian language. It would be highly valuable for future research to investigate whether associations between language and mental health are universally present or language-specific, as this would have important clinical implications. If such associations are universally present, this would mean that there are universal markers of psychopathology that can be incorporated into clinical practice worldwide. However, if the associations found between language and mental health in the English language, such as the association between depression and greater self-focused language, cannot be replicated in other languages, this would contextualise the findings and ensure more accurate interpretation.

Third, we note the importance of ensuring clinical validity of all psychometrics — computationally-derived and language-based measures notwithstanding. Given the potential impact that NLP and other big-data computational methods could have on clinical practice, it is critical that these methods allow clinical validity to be maintained. Naturally, the resourceful, easy to access, large-scale nature of NLP methods will greatly appeal to clinical researchers, particularly when compared to more traditional, resource intensive psychological research methods. However, it is important that such modern methods are used ethically and with care by researchers and are informed and guided by clinical expertise and theory. Computational methods will often be best used in *combination* traditional methods, rather than being used alone, such as the use of linguistic features as complementary outcome measures to assess responses to clinical treatment alongside other traditional outcome measures (e.g., self-reported well-being). Triangulation of research methods in the study of personality disorder will allow for the strengths of such methods to be combined, and thus would generate better and more reliable understanding of personality pathology.

Relatedly, and finally, possibly the most important challenge to using NLP methods to improve understanding and the care and treatment of personality disorder, and psychopathology more generally, surrounds overcoming the barriers between science and clinical practice. This is a particular challenge given the need to consider the clinical validity of computational research methods, as just discussed, and also given that there has traditionally been resistance among clinicians to adopt such methods (Goldberg et al., 2020). Thus, greater communication and collaboration between scientists and clinicians is essential for moving forward, in that scientists and clinicians alike would benefit from the sharing of knowledge and advice and developing mutual understanding. The development of such collaborations would undoubtably improve the likelihood of the implementation of new technologies, such as NLP methods, into clinical practice, by helping to reduce the barriers between science and practice.

1.1.5 Conclusions

Personality disorders are presently some of the most prevalent and high-risk psychological disorders, yet remain poorly understood. Despite extensive study, there is still a lack of clarity on some of the most fundamental aspects of personality disorder, such as the underlying structure and dynamic manifestation over time. Importantly, as evidenced in the wider fields of personality and psychopathology research, natural language processing methods have the potential to improve our understanding of personality disorder. At present, however, there is limited research incorporating language analysis methods in the study personality disorder. Given that personality pathology lies right at the intersection of personality psychology and psychopathology, we believe that the study of personality disorder is abundantly ready for language-based exploration.

In this chapter, we provided illustrative examples of how language analysis can be used to enhance understanding of personality disorder and address some of the fundamental, unanswered questions in the personality disorder literature, including the assessment and classification of personality disorder and developmental trajectory across the lifespan. Such examples demonstrate how language can provide implicit and unobtrusive insight into personality and psychological processes underlying personality pathology at a large-scale, using an individualized approach.

The growth in sophisticated language analysis methods and powerful statistical techniques allow rich sources of data in the thriving digital world to be analyzed in ways that promote new understanding of psychopathologies, including personality pathology. Crucially, this research direction represents an opportunity to ensure that both empirical research and clinical practice can reciprocally inform and enhance each other. In order to move forward, interdisciplinary collaborations across clinical and computational research, as well as communication and collaboration between empirical research and clinical practice, are essential. For such promising methods to reach their full potential and have a real-world impact, it is important that they lead to insights that can directly inform clinical interventions and approaches. Taken in all, we hope that this chapter will inspire researchers and clinicians alike to come together and take advantage of the many benefits that the application of natural language processing methods can bring to personality disorder research and practice.

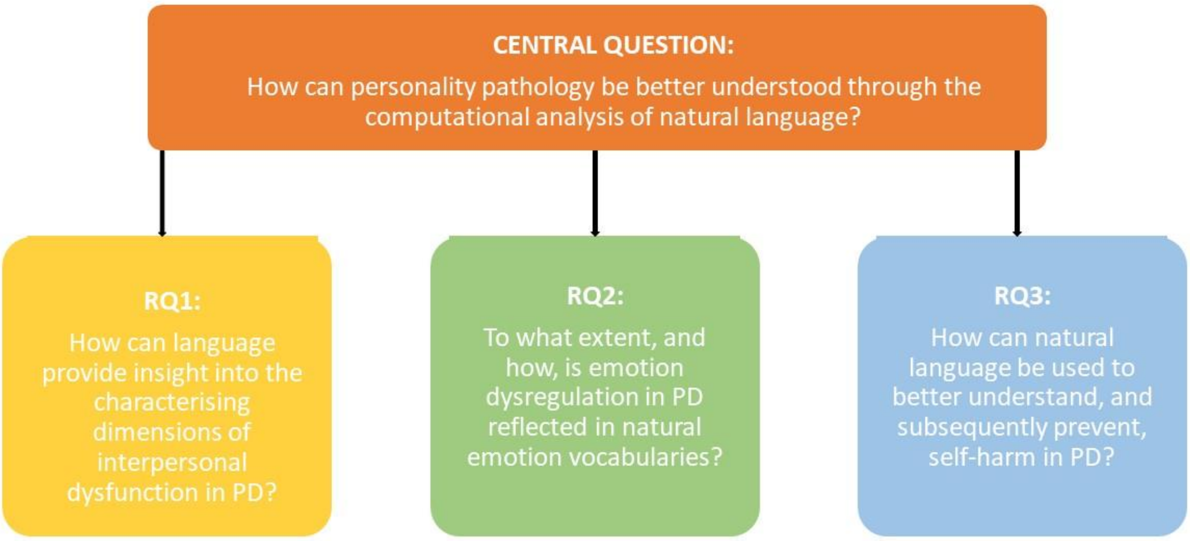
1.2 The Current Research

The previous section (Chapter 1.1) highlighted how, despite growing empirical attention, there remains unanswered questions with regard to fundamental aspects of personality pathology. Moreover, the previous section illustrated the promising potential of computational language analysis methods to address some of these unanswered questions, albeit with some caveats and challenges. Accordingly, the primary goal of the present research is to begin to fill this gap in the literature by employing a range of natural language analysis methods to provide new insights into personality pathology. Specifically, the central, overarching research question guiding this thesis is: *How can personality pathology be better understood through the computational analysis of natural language?*

To address the central research question, this thesis presents three research articles, comprising four empirical studies, that each leverage computational language analysis methods to better understand personality pathology. Each of the papers focuses on a different core feature of PD, mapping on to specific, distinct research questions (see Figure 1.2 for mapping of thesis research questions), while incorporating language analysis methods. Namely, the core PD features investigated in the present research are as follows: interpersonal dysfunction (Paper 1 – Study 1; RQ1), emotion dysregulation (Paper 2 – Studies 2 and 3; RQ2), and behavioural dysregulation (Paper 3 – Study 4; RQ3). Such features were the main focal points of the present research given that social, emotional, and behavioural dysfunctions lie at the heart of personality pathology, and span across all variations of PD (e.g., APA, 2013; Videler et al., 2019). More specifically, problems with regulating emotions (e.g., Crowell et al., 2009), relationships (e.g., Hill et al., 2008), and behaviour (e.g., Reichl & Kaess, 2021) are theoretically proposed and empirically evidenced as the predominant facets of PD, and therefore arguably, in their interactions with one another (see, e.g., Linehan, 1993), form the broad construct of PD. Thus, detailed investigation of these core PD features will allow for greater knowledge of the main facets of personality pathology, which, in turn, permits better understanding of the nature of PD.

Figure 1.2

Central Thesis Research Questions


CENTRAL QUESTION:
How can personality pathology be better understood through the computational analysis of natural language?
RQ1:
How can language provide insight into the characterising dimensions of interpersonal dysfunction in PD?
RQ2:
To what extent, and how, is emotion dysregulation in PD reflected in natural emotion vocabularies?
RQ3:
How can natural language be used to better understand, and subsequently prevent, self-harm in PD?

Note. This figure illustrates the core research questions addressed in the present thesis. The central, overarching question surrounds how we can use computational language analysis methods to better understand personality pathology, in a general sense. The specific research questions, under the central research question, each relate to a distinct core feature of PD. RQ1 focuses on social dysfunction and is examined in Study 1 (Paper 1; Chapter 3). RQ2 focuses on emotional dysregulation and is examined in Studies 2 and 3 (Paper 2; Chapter 4). RQ3 focuses on behavioural dysregulation and is examined in Study 4 (Paper 3; Chapter 5).

To provide a brief overview of the research conducted, Study 1 (*N* = 530; Paper 1) addressed RQ1 by analysing people's language from written essays about their interpersonal relationships, while also assessing borderline personality disorder (BPD) features, to uncover core social-cognitive dimensions that characterise interpersonal dysfunction in BPD. Studies 2 and 3 (Paper 2) addressed RQ2 by analysing natural language from written essays (Study 2; *N* = 530) and spoken conversations between women diagnosed with BPD and their romantic partners (Study 3; *N* = 64 couples) to describe the natural emotion vocabularies (i.e., the variety of emotion words actively used) associated with BPD, to provide insight into the (maladaptive) emotion processes that comprise emotion dysregulation in BPD. Study 4 (*N* = 992; Paper 3) addressed RQ3 by examining the natural language of individuals with self-identified BPD on online BPD discussion forums (i.e., BPD Reddit forums); these data were subsequently analysed to better understand the psychosocial dynamics (evident in language) of self-harm (i.e., suicidality and deliberate self-harm [DSH]) in this population, and in this particular context (i.e., online support platforms).

It seems noteworthy to point out that all of the present research has been conducted in the context of BPD specifically (relative to other types of PD), which was primarily due to BPD being the most commonly recognised, established, and diagnosed PD (see, e.g., Chad et al., 2018), thus making it a more accessible and translatable form of PD. Moreover, given its heterogeneous nature, combined with high rates of co-morbidity with other PD types (e.g., Shah & Zanarini, 2018), BPD has been conceptualised by some as reflective of "general personality pathology" (e.g., Wright et al., 2016). In terms of supporting empirical evidence for this conceptualisation, when adopting a dimensional approach to PD, BPD has been found to strongly map on to a "general factor" of personality pathology (Sharp et al., 2015; Wright et al., 2016). Consequently, this theoretical position and supporting empirical evidence imply that findings in the context of BPD should largely be generalisable to personality pathology more broadly.

Moreover, it is also necessary to make apparent that the same sample and dataset was used in both Study 1 (Paper 1) and Study 2 (Paper 2), as these data were originally collected as part of a broad investigation into BPD, verbal behaviour, and various psychosocial processes. This broad dataset thus served as a valuable resource for investigating both the social (Study 1) and emotional (Study 2) dysfunctional facets of personality pathology through language analysis. Nevertheless, although the same sample of participants were included in both studies (*N* = 530), the two studies utilised different measures from the broader dataset, and a different analytic approach was adopted for each of the studies, primarily due to differing study aims and scope.

To provide more detail regarding the differing measures utilised between Study 1 and Study 2, and the rationale for incorporating different measures between these studies, a measure of the Dark Triad traits (i.e., the Dirty Dozen, Jonason & Webster, 2010) was included in Study 1 but not in Study 2. The reasoning behind including the

Dark Triad measure in the Study 1 analysis but not Study 2 surrounds the relevance of the construct of the Dark Triad – which includes the dark personality traits psychopathy, narcissism, and Machiavellianism – to the scope and goals of the particular study. That is, given strong relationship between Dark Triad traits and social (dys)function (e.g., Lyons, 2019), it was meaningful to include a measure of the Dark Triad in Study 1 to differentiate and disambiguate characterising features of social dysfunction in BPD from social dysfunction more broadly. In contrast, it would not have been meaningful to include the Dark Triad measure as a general comparison to BPD in Study 2 given that the focus of this study is on emotion processes and dysregulation, of which is not typically – or consistently – associated with construct of the Dark Triad. Put simply, the Dark Triad measure was included as a comparison to BPD in Study 1 given its' relevance to social functioning (the focus of Study 1), but was not included as a comparison to BPD in Study 2 given its' lack of clear and consistent relationship to emotion functioning (the focus of Study 2).

Furthermore, writing samples from which participants were prompted to write about their everyday behaviours were included in the Study 2 analysis but were not included in Study 1, which was again due to differing study aims. Specifically, one of the main goals of Study 2 was to explore whether associations between BPD and emotion language vary depending on the context, as an indicator of context-sensitivity. The inclusion of an additional writing sample to the writing sample in the context of interpersonal relationships of a different, more general topic (i.e., everyday behaviour) permitted a general comparison sample of natural language data to address the goal of exploring the context sensitivity of emotion language in BPD. On the other hand, it was not necessary to include the comparison everyday behaviour writing sample in the Study 1 analysis given that the sole aim of this study was to investigate and characterise social processes and dysfunction in BPD, meaning that only the interpersonal relationship writing samples were relevant to analyse. Similarly, the way in which the computational language analyses of the language data were carried out and the specific linguistic features included in the analyses naturally differed between the two studies, in accordance with the particular aims of each study (e.g., a broad range of psychosocial linguistic features were included in the Study 1 analysis but only emotion language variables were included in the Study 2 analysis).

Finally, a small number of differences in the measures used and analytic procedure between the two studies emerged as a direct result of the journal publication and peer-review process for Paper 1 and Paper 2. In particular, in comparison to Study 1 that examined associations with BPD at the level of total BPD feature scores (due to simplicity and the big-picture nature of this study), the Study 2 analysis additionally included a breakdown of the results by particular BPD features (e.g., affective dysfunction; social impairments) as a result of the peer-review process, in which this breakdown of results was requested by reviewers (which we agreed was valuable to include). Likewise, an analysis of missing data was included in Study 1 due to this being agreed upon during the peer-review process, whereas this analysis was not deemed necessary or meaningful for Study 2.

1.3 Rationale for Alternative Format Thesis

There are several justifications behind the decision to write and submit an "Alternative Format" thesis. Most importantly, the primary rationale for writing this as an Alternative Format thesis surrounds the fact that writing-up empirical articles from research findings has been a central focus since the beginning of my doctoral studies, with the eventual aim of publishing such articles in academic journals as a way of disseminating research findings and widening the audience – and subsequently the impact – of my research. Specifically, the focus on writing-up research results throughout my PhD stimulated the production of distinct empirical articles at various stages, with most of which now published in academic journals. The Alternative Format thesis route thus allows me to better portray the work carried out during my PhD.

1.4 Construction of Alternative Format Thesis

In this section, I provide a general account of how the present Alternative Format thesis was constructed, step-by-step.

1) A detailed literature review was carried out at the start of my PhD studies, immediately prior to writing-up my book chapter (i.e., Chapter 1.1), which formed Chapter 2 (i.e., the literature review chapter). However, the literature

review chapter was edited following the write-up and inclusion of the book chapter and the three research articles to ensure that only relevant research that was not discussed in detail in the book chapter or in any of the three research papers was included in the literature review, to avoid repetition.

2) The book chapter of which I was the lead author, entitled “Personality Disorder and Verbal Behavior”, was written during the first half of my PhD (following the initial literature review) – whilst simultaneously carrying out research studies on the topic – and was published in *The Handbook of Language Analysis in Psychology*. This (published) book chapter subsequently formed Chapter 1.1 (i.e., the introductory chapter) of the present thesis, as it provides a detailed account of the broad rationale driving the current work.
3) The three empirical articles presented in Chapters 3, 4, and 5 were written prior to the remainder of this thesis. The order in which these papers are presented in this thesis (i.e., Paper 1, Chapter 3; Paper 2, Chapter 4; Paper 3, Chapter 5) is consistent with the order in which the studies were conducted and written-up.
4) The supplemental material for the three research papers formed the Appendices for this thesis, with a separate Appendix for each paper (e.g., supplemental material for Paper 1 formed Appendix A).
5) Following the inclusion of the literature review (step 1), book chapter (step 2), and the three research papers and their associated supplemental material (steps 3 and 4), all materials preceding the main body of the thesis were then constructed (e.g., title page, abstract, contents page).
6) Next, the remainder of the other sections of the introduction chapter (Chapter 1) were constructed (e.g., “The Current Research”), in the order in which they are presented in this thesis.
7) The general discussion chapter (Chapter 6) was then constructed, which brings together all findings in addressing the central RQs and highlights the contributions and implications of this thesis, while discussing some of the key strengths and limitations with suggestions for future research.
8) Finally, the consolidated bibliography was created, which involved merging all of references contained within the thesis into a single bibliography section.

1.5 Contribution to Research Undertaken

As evidenced in the “Contribution Statements” section (pg. 11-13), I (the student) primarily undertook the main tasks involved in conducting all research projects outlined in this thesis. My contributions – with expert input and guidance from supervisors (and other co-authors) at various stages – included playing a major role in the conceptualisation and design of all research projects, applying for ethical approval for all studies, collecting data for all studies, data pre-processing and annotation (where relevant), implementation and application of computational language analysis methods, statistical analysis of all data, and writing up each of the chapters. I am the principal author of each of the chapters presented in this thesis, and thus wrote the first draft of each of the chapters, of which subsequently received feedback and input from various collaborators. Refer to the “Contribution Statements” section for a detailed breakdown of each co-author’s role in each of the studies and chapters in this thesis, signed by all co-authors.

CHAPTER 2:

Literature Review

This literature review chapter provides a review of relevant literature in the realm of personality pathology and computational language analysis that has not already been introduced or sufficiently described in other thesis chapters, including the book chapter (Chapter 1.1) and the three research papers (Chapters 3, 4, & 5). To clarify, the reasoning behind the inclusion of this literature review in the present thesis is that the papers (and book chapter) do not provide a full review of all relevant research that has utilised computational language analysis to study personality dysfunction, thus necessitating a dedicated literature review chapter. I first give an overview of the current status of the problematic and pathological personality language literature (Section 2.1), while highlighting the need to take into account more generic (language-based) psychopathology findings when interpretating personality pathology research. I then provide a review of both non-clinical problematic personality (i.e., unfavourable personality traits) work (Section 2.2) and clinical PD (Section 2.3) research that has incorporated computational language analysis methods, before summarising the literature review (Section 2.4).

2.1 Current Status of the Personality Dysfunction and Language Literature

As highlighted throughout the introduction chapter, research incorporating computational language analysis methods in the study of personality pathology is scarce. Much of the work that has been done in this domain so far has primarily focused on the detection of PDs and unfavourable personality traits through the identification of linguistic markers (e.g., Sumner et al., 2012). Such linguistic markers have largely been measured using automated word counting programs, such as Linguistic Inquiry and Word Count (LIWC; Boyd et al., 2022), which typically rely on internal dictionaries to map words on to various psychologically meaningful categories. Moreover, the existing research incorporating language analysis methods has predominantly focused on (non-clinical) problematic personality traits – often referred to as "dark traits"; in particular – narcissism, psychopathy, and Machiavellianism – otherwise known as the "Dark Triad" of personality (Paulhus & Wiliams, 2002). There has been considerably less empirical attention directed towards clinical PDs.

Before going ahead with a review of the existing problematic personality and personality pathology literature, it is useful to draw attention to key findings generated from more generic mental health language analysis work. In particular, it is informative to describe the central linguistic features typically associated with depression, given that substantial work has been carried out in this domain, and it is possible – and highly likely – that these associated features are in fact transdiagnostic (i.e., associated with various mental health conditions). Thus, knowledge of the linguistic features typically associated with depression will help to contextualise the personality pathology language literature.

2.1.1 Depression and (Transdiagnostic) Language

One of the most consistent and robust findings to date in clinical language analysis research surrounds the relationship between depression and self-focused language (i.e., first-person singular pronouns). As touched upon in the introduction chapter (Chapter 1.1), research has consistently found greater use of self-focused language to be a linguistic marker of depression (e.g., Sonnenschein et al., 2018; Tackman et al., 2019; Zimmermann et al., 2017). Most notably, in a comprehensive meta-analysis of correlations between depression and first-person singular pronoun use (Edwards & Holtzman, 2017), a significant correlation was evidenced between them, which was not found to be moderated by demographic factors, such as age and gender. Importantly, no evidence of publication bias was found in this literature, demonstrating the validity and reliability of first-person singular pronouns as a linguistic marker of depression.

Interestingly, frequent use of self-focused language has also been revealed to be associated with various other mental health conditions, including anxiety (e.g., Anderson et al., 2008), schizophrenia (e.g., Zomick et al., 2019), autistic-spectrum disorder (e.g., Nguyen et al., 2013), and PDs (e.g., Molendijk et al., 2010). The associations between frequent use of self-focused language and various mental health conditions suggests that self-focused language may not be specific to depression, but rather a transdiagnostic marker of general mental distress. Indeed, the notion that self-focused language is a marker of general mental distress was directly tested and confirmed in a study by Lyons and colleagues (2018), in which they investigated five different mental health conditions (including BPD) and found that people who experienced any form of mental distress used relatively more first-person singular pronouns.

Relatedly, in addition to self-focused language, research has also directed attention towards uncovering other linguistic markers of depression, including use of more negative and less positive emotive language (e.g., Tolboll, 2019), more conjunctions (e.g., *'and', 'but'*), cognitive processes words, causal words, health-related words, and tentative language (e.g., *'might', 'could'*), and less first-person plural pronouns (e.g., Coppersmith et al., 2015). Such language markers are consistent with the core areas of dysfunction in depression, including maladaptive emotion processes, cognitive dysfunction, and social impairments. As with first-person singular pronouns, it is likely that these linguistic markers are indicative of general mental distress rather than specific to depression, as the broad areas of dysfunction that they reflect are generally transdiagnostic. Indeed, many of the linguistic markers of depression have also been shown to be associated with numerous other mental health conditions (e.g., Coppersmith et al., 2015; Lyons et al., 2018). One possible reason behind this overlap in associations with linguistic features surrounds the fact that depression is highly comorbid with most, if not all, mental health conditions (e.g., Steffen et al., 2020). Relatedly, if adopting a transdiagnostic approach to mental health (e.g., Dalgleish et al., 2020), the linguistic features associated with depression may in fact simply be transdiagnostic. Either way, given that PDs are typically associated with high levels of mental distress – and are highly comorbid with depressive disorders (e.g., Friborg et al., 2014) – it is most probable that PDs will also reflect depression-consistent language patterns; this is important to acknowledge and take into consideration when interpretating findings from the PD literature, as well as when interpretating the present findings.

I will now move on to discussing the existent problematic personality and personality pathology language literature. I will start by describing the more extensive research base that has applied computational language analysis to the study of non-clinical problematic personalities (Section 2.2) – or unfavourable personality traits (i.e., the Dark Triad of personality) – and then move on to detailing the more limited research that has applied these methods to clinical PDs (Section 2.3).

2.2 Non-Clinical Problematic Personality Language Literature

In this section, I focus on the Dark Triad personality traits (i.e., narcissism, psychopathy, and Machiavellianism), as considerable empirical research has utilised language analysis methods to study and better understand the Dark Triad, in terms of the construct as a whole as well as the individual "dark" traits. In particular, although I am not aware of any published research to date that has applied computational language analysis to the study of Machiavellianism specifically (other than as part of the overarching Dark Triad construct), there has been significant work that has applied these methods to subclinical narcissism and psychopathy as individual constructs (i.e., not in the context of the broader Dark Triad). Thus, I will discuss subclinical narcissism and psychopathy research that has incorporated computational linguistic methods in more detail – as individual constructs – following a review of literature on the Dark Triad as a whole construct. Importantly, the inclusion of research on non-clinical problematic personality in the present literature review is in alignment with the now widely supported dimensional approach to personality pathology (and psychopathology more broadly; e.g., Wright et al., 2016).

2.2.1 The Dark Triad

As briefly touched upon earlier, the Dark Triad is composed of three "dark" personality constructs – namely, narcissism, psychopathy, and Machiavellianism (Paulhus & Wiliams, 2002). To broadly define such constructs, generally speaking, subclinical narcissism is conceptualised as the most "self-absorbed" of the Dark Triad traits. Narcissistic personality characteristics generally include sustained efforts to maintain a grandiose self-view and unrealistic positive self-beliefs (see, e.g., Zajenkowski & Szymaniak, 2021). Psychopathy, on the other hand, is generally considered the most malevolent and dangerous. Subclinical psychopathy is broadly characterised by antisocial behaviour, impulsivity, shallow affect, manipulation, deception, and callousness (e.g., Crego & Widiger, 2022). Similar to subclinical psychopathy, Machiavellianism is also characterised by a tendency to deceive, manipulate, and exploit other people, usually for personal gain. Further, Machiavellianism reflects a cynical and selfish orientation accompanied by a willingness to use whatever means necessary to achieve one's goal (Marcus & Zeigler-Hill, 2016).

With respect to the Dark Triad as a whole, although the Dark Triad traits are considered distinct constructs, they also share many characteristics; as Paulhus and Williams (2002) stated, they are "overlapping but distinct constructs". Characteristics common among all Dark Triad traits include disagreeableness, deceitfulness, egocentrism, coldness, and manipulative and exploitative behaviours (e.g., Lyons, 2019). Moreover, the Dark Triad traits are all generally associated, to varying degrees, with social malevolence, emotional and/or empathetic deficiencies, and self-promotion (Lyons, 2019). Thus, it will not come as a surprise to know that the Dark Triad personality is strongly associated with antisocial and criminal behaviour (Muris et al., 2017).

Yet, in comparison to PDs, Dark Triad traits are considered to have both maladaptive *and adaptive* features (see, e.g., Zeigler-Hill & Marcus, 2016), which differentiates these subclinical dark personality constructs from clinical PDs. To provide an illustrative example of such differentiation, if an individual is highly narcissistic, but this narcissism is dominated by adaptive narcissistic features that typically result in positive outcomes for the individual (e.g., leadership qualities, high self-esteem), it is unlikely that this would cause significant persistent distress or impairment that would certify a narcissistic PD diagnosis; meaning that this person would be "narcissistic" but not have narcissistic PD. In contrast, if a narcissistic individual is dominated by the maladaptive (or vulnerable) features of narcissism (e.g., impulsivity, interpersonal dysfunction), and this was causing them significant persistent distress or impairment, a clinical narcissistic PD diagnosis would be applicable (see Lyons, 2019, for further discussion relating to differentiating features of subclinical versus clinical dark personality traits [i.e., the “vulnerable Dark Triad”]). Put simply, it is important to acknowledge that, in comparison to PD traits, the mere presence of Dark Triad traits does not always have negative consequences, and in some cases can in fact have positive consequences.

2.2.1.1 The Dark Triad and Verbal Behaviour

If one were to make a prediction at what patterns would be present in the language of people with dark personalities, one would probably guess that their hostility, negativity, self-serving nature, and disconnectedness from others would be what you would find. Indeed, such associations have generally been indicated in empirical research. For example, in one of the earliest studies to use Twitter data (analysed via the automated word counting program LIWC) to predict dark personality traits, Sumner and colleagues (2012) found that people scoring higher on psychopathy and Machiavellianism used more swear words, negative emotion words, and anger-related words, and fewer positive emotion words and first-person plural pronouns. Additionally, psychopathy was also associated with greater use of death words and filler words (e.g., *‘so’, ‘like’*) and fewer first-person singular pronouns. As for narcissism, individuals scoring higher on this measure tended to use more sexual words.

Likewise, also leveraging Twitter data, a study conducted by Preotiuc-Pietro and colleagues (2016) identified linguistic markers of the Dark Triad personality that were generally consistent with the notion that people with dark personality traits use darker, more hostile language. That is, it was revealed that people higher on the Dark Triad personality used more language relating to alcohol, negative emotion words – particularly anger and disgust words – swear words, sexual words, filler words, and language orientated to the present tense. As for specific Dark Triad traits, individuals that were more psychopathic were revealed to have the most distinctive language patterns. Not surprisingly, psychopathy was associated with considerably more hostile and aggressive language, with greater violent and angry content and greater expression of negative emotion. Notably, these language patterns were predictive of psychopathy even when controlling for narcissism and Machiavellianism. Such findings thus portray how hostility and “darkness” in personality can be traced in natural language.

In addition to Twitter data, meaningful research has also leveraged Facebook data to study the Dark Triad personality. As could be expected, findings from research conducted on Facebook data have largely overlapped with those utilising Twitter data, and have also reflected the core characteristics of the dark traits. For example, one study found that individuals who were more psychopathic and narcissistic expressed more negatively valenced language and had more "odd" semantic representations in their Facebook posts (Garcia & Sikström, 2014). Furthermore, the semantic content of Facebook posts was found to significantly predict psychopathy and narcissism. Machiavellianism, however, was not predicted by the semantic content of posts. The results from this study therefore indicate that particular (problematic) personality traits appear to manifest themselves in language more than others, thereby highlighting the importance of examining language in relation to specific dimensions of personality (dysfunction).

Valuably, also utilising Facebook data, one study generated insight into the generalisability of the linguistic findings relating to the Dark Triad conducted in English (as in the studies described above) to non-English languages by investigating this in the Russian language (Bogolyubova et al., 2018). Promisingly, these findings were broadly consistent with those of studies conducted in the English language. With regard to specific associations found, people higher on Machiavellianism used less first-person plural pronouns, third-person plural (e.g., *'them'*) and singular (e.g., *'she', 'he'*) pronouns, and second-person plural pronouns (e.g., *'yours'*), indicative of less social orientation and connectedness. As for the semantic features, these were generally consistent with characteristics of the dark personality traits. That is, people that were more narcissistic tended to use language relating to social interaction, self-image, social status, and reasoning, and people that were more psychopathic tended to use language relating to basic needs (e.g., money, food), politics, and authority related issues. In comparison, only negative associations were uncovered between Machiavellianism and semantic features, including less content relating to social relationships, positive affect, negative events, religion, mental processes, and individual characteristics (e.g., nationality), indicating more reserved and less revealing and open (and thus potentially more manipulative) language. Critically, overlap in the findings conducted in the Russian language with the findings discussed in the English language provides an initial indication that linguistic markers of problematic traits may be reliable across different languages, which has implications regarding knowledge of how such traits may (or may not) manifest differently in different languages and cultures. However, considerably more empirical work is needed to truly understand such language and cultural variations.

Although the research discussed so far in the realm of the Dark Triad and language paints a fairly clear picture, in a more recent study examining relationships between language features, Dark Triad traits, and the need for power via Facebook data (Yuan et al., 2020), the associations found between linguistic features in Facebook posts and Dark Triad traits were somewhat contrary to findings of past work. In particular, although the finding that Machiavellianism is associated with greater use of negative emotion words and clout is aligned with previous findings and/or the conceptualisation of Machiavellianism (i.e., clout portraying higher perceived social status), this study also revealed individuals higher on Machiavellianism to use more first-person singular pronouns; an association that has not been evidenced in previous studies that have examined this. Rather than reflecting psychological distress – as in the language-based clinical literature – the association between Machiavellianism and self-focused language evidenced in this study may simply be reflective of the selfish and self-orientated nature of individuals with highly Machiavellian traits. In alignment with this interpretation, self-focused language was also found to be positively associated with narcissism in Yuan et al. (2020). Finally, and perhaps surprisingly, psychopathy was revealed to be associated with the use of less analytic language, yet more authentic language. Notably, the association found between psychopathy and authenticity stands out from findings of other work in this domain, as well as with respect to the conceptualisation of psychopathy as marked by deceitful and manipulative behaviour. Discrepancies between the linguistic markers of Dark Triad traits evidenced here and those of related studies highlight the need for more thorough, precise, and large-scale investigations into the verbal behaviour of individuals with problematic (or pathological) personalities, using a broader range of computational linguistic techniques and more diverse samples.

Taken in all, the existent literature investigating the construct of the Dark Triad suggests that there are some general, overarching linguistic markers of the Dark Triad personality, such as greater use of negative emotion words (particularly anger words), filler words, swear words, and fewer first-person plural pronouns. Generally speaking, the linguistic markers identified appear intuitive given that they reflect hostility, negativity, manipulative tendencies, and a greater disconnection from others among people with dark personalities, which is aligned with the characterising features of such dark traits. In terms of particular Dark Triad traits, with the exception of one study conducted in the Russian language (Bogolyubova et al., 2018), research has consistently revealed language patterns associated with psychopathy to be the most distinct and extreme, with relatively few *distinctive* language patterns associated with Machiavellianism. Rather, the linguistic patterns associated with Machiavellianism tended to overlap with those of psychopathy. Such findings reflect the highly overlapping characteristics of Machiavellianism and subclinical psychopathy and suggest similarity in the underlying nature of these traits. Moreover, findings from the language analysis research discussed provide some support to the whole concept of the Dark Triad – in that the three traits are conceptually related – as they reveal shared linguistic features between the three traits (e.g., hostile, negative, and socially disconnected language), implying underlying psychological and personality processes among these traits. On a broader level, the research discussed in this section, and the interpretations generated, demonstrates how natural language can be analysed to better understand the nature of problematic (or pathological) personality traits.

I now move on to reviewing the literature that has incorporated computational language analysis methods to the study of (subclinical) narcissism and psychopathy specifically, as individual constructs.

2.2.1.2 The Language of a (Subclinical) Narcissist

Although there does not currently appear to be any published research that has leveraged language analysis methods to study narcissism as a clinical PD (i.e., narcissistic PD), there is a considerable amount of research that has applied these methods to the study of narcissism as a non-clinical personality construct (i.e., in the field of normative personality psychology). Yet, in adopting the dimensional approach to personality pathology, it could be presumed that there would be considerable overlap in the verbal behaviour of individuals with narcissistic PD and people in the general population with highly narcissistic traits, particularly those high on the narcissistic vulnerability dimension.

In accordance with the notion that excessive self-focus is a central feature of narcissism, research that has investigated linguistic patterns associated with narcissism has naturally devoted attention to the relationship between narcissism and self-focused language (i.e., first-person singular pronouns). Specifically, it is intuitively presumed that narcissistic people talk about themselves more. The clinical characteristics of narcissistic PD also add to this presumption, given that central characteristics include grandiose self-perception and self-belief. Accordingly, the first study to have used language analysis methods to empirically study narcissism was conducted by Raskin and Shaw (1988), of whom explored the relationship between narcissism and self-focused language. The results of this early study revealed that – in line with intuitive expectations – people who had more narcissistic traits used more first-person singular and fewer first-person plural pronouns; no relationships were found between narcissism and second-person or third-person pronouns. Such findings therefore provided empirical support to the presumption that narcissistic people talk about themselves more, and also indicated that individuals high on narcissistic traits associate themselves with others less and are more socially disconnected (reflected by fewer we-words, or first-person plural pronouns). Interestingly, traces of social disconnectedness in the verbal behaviour of narcissists are also in line with the clinical characteristics of narcissistic PD, particularly in relation to deficits in empathy and social functioning.

Notably, Raskin and Shaw's (1988) early findings sparked future researchers to utilise first-person singular pronouns as a linguistic marker of narcissism (e.g., Aktas et al., 2016; DeWall et al., 2011). However, the relationship between narcissism and self-focused language has been failed to be replicated in almost all of the more recent empirical research investigating this. The most compelling evidence for this can be seen from findings of a study carried out by Carey and colleagues (2015), whereby they investigated this relationship in a large sample of over 4,800 people, using five different measures of narcissism, among two languages (English and German), and across various communication contexts. Overall, this study revealed no significant relationship between narcissism and first-person singular pronoun use. Moreover, several other more recent studies also evidenced no association between narcissism and self-focused language (e.g., Rathner et al., 2018; Underberg et al., 2019). Thus, recent research has demonstrated strong empirical evidence to contradict the common misconception (and initial findings from the early Raskin & Shaw, 1988, study) that narcissists talk about themselves more than others do.

Yet, there has been one recent study evidencing a significant relationship between narcissism and first-person singular pronoun use (Dorough, 2018), although this study examined the language of individuals high on vulnerable narcissism, rather than narcissism in general. Specifically, this study revealed that those who scored higher on vulnerable narcissism used more first-person singular pronouns. However, it is important to acknowledge that vulnerable narcissism did not predict self-focused language use any more than the common predictors – depression and negative emotionality – suggesting that the association found between vulnerable narcissism and self-focused language use was likely due to the negative emotionality and mental distress experienced by individuals with vulnerable narcissism, rather than a consequence of narcissism specifically.

Importantly, findings from Dorough's (2018) study are informative as they imply key differences in the underlying psychology (reflected in language) between individuals high on the narcissistic vulnerability dimension and those high on the narcissistic grandiosity dimension of narcissism. Likewise, such differences are also consistent with the differences in the clinical conceptualisation of these two dimensions, with narcissistic vulnerability comprising more maladaptive features and narcissistic grandiosity comprising more adaptive features. Critically, findings from Dorough's (2018) study also indicate that the linguistic markers of vulnerable narcissism overlap with the linguistic markers of depression and general mental distress, which is consistent with clinical literature showing vulnerable narcissism to be strongly associated with depression (e.g., Erkoreka & Navarro, 2017). Given that the characteristics of vulnerable narcissism overlap with narcissistic PD characteristics, it seems plausible to hypothesise that people with narcissistic PD would also use relatively greater self-focused language, which is additionally probable due to the mental distress that is associated with narcissistic PD. In support of such interpretations and the findings from Dorough (2018), a recent case study examining the personal letters of a well-known serial killer, Jack Unterweger – of whom was classified as having malignant narcissism (an extreme and particularly dark form of narcissism, closely related to narcissistic PD) – was revealed to use high frequencies of first-person singular pronouns in his letters (Marko & Leibetseder, 2023). Interpretating these findings together, while narcissism in everyday life (particularly grandiose narcissism) does not appear to be associated with self-focused language, darker and more vulnerable forms of narcissism (as in narcissistic PD) do appear to be reflected by the use of more self-focused language, indicative of the psychological distress that accompanies such maladaptive narcissistic traits.

In a similar line of work, other research has expanded the focus from personal pronouns to uncover other linguistic markers of narcissism. In particular, one study conducted by Holtzman and colleagues (2019) investigated linguistic markers of narcissism in a large-scale meta-analysis comprising 15 distinct samples. From this, the main linguistic markers of narcissism identified include: more words relating to sports, more second-person pronouns, more swear words, and more sexual words, as well as fewer anxiety words, fewer perception-related words, and less tentative language. Although these linguistic markers of narcissism are highly insightful, they mostly relate to narcissistic grandiosity (which was the dimension investigated), and so are less applicable to pathological narcissism, in which narcissistic vulnerability dominates. For example, narcissistic PD is associated with hypersensitivity to fear and anxiety (Ronningstam & Baskin-Sommers, 2013), and so less frequent use of anxiety and fear related words is unlikely to be a linguistic marker of narcissistic PD; in fact, it is more plausible that *greater* use of anxiety words would be a marker of narcissistic PD. Nevertheless, it is probable that some of the linguistic markers identified in Holtzman et al. (2019) will also be generalisable to pathological narcissism, particularly those that relate to impulsivity and hostility, such as the greater use of swear words and sexual words.

In contrast to the Holtzman et al. (2019) findings highlighting the verbal behaviour of narcissistic individuals to be more disagreeable and hostile (e.g., greater use of swear words), more recent research has in fact revealed grandiose narcissism to be associated with more agreeable language, including a more open-minded language style (Cutler et al., 2021). Such discrepancies in findings regarding the language patterns associated with narcissism may be a reflection of social desirability motivations and impression management tactics when participants have been explicitly prompted by researchers to write/talk, relative to naturally and unobtrusively collected spontaneous language data of which individuals' language are analysed unknowingly, or without direct prompting from a researcher (i.e., naturally occurring texts, such as social media posts).

Relatedly, a very recent study that investigated linguistic markers of narcissism in older adults (aged 65-89) also generated some discrepant findings relative to past work (Zhang et al., 2023). In particular, this study revealed the most predictive linguistic features of narcissism to be greater use of first-person plural pronouns, achievement words (e.g., *'win', 'success'*), work words (e.g., *'office', 'meeting'*), sexual words, and language that signalled a desired state (e.g., *'need', 'want'*). Although the findings highlighting individuals with more narcissistic traits to be more self-centred (i.e., more focused on their own needs), work and achievement orientated, and impulsive (i.e., more frequent talk of sexual activities) are broadly consistent with previous findings (and the conceptualisation of narcissism), the positive association evidenced between narcissism and first-person plural pronouns directly contradicts past findings (e.g., Raskin & Shaw, 1988). Moreover, such association is also inconsistent with the theoretical conceptualisation that narcissism is characterised by social disconnectedness and disagreeableness, as this linguistic marker (e.g., we-words) traditionally implies greater social connectedness. These discrepancies may reflect differences the social aspects of narcissism – and the way in which they manifest in language – in older adults compared to younger individuals (which is the more typical research population). Alternatively, the authors (Zhang et al., 2023) proposed that the association evidenced between narcissism and first-person plural pronouns could have resulted from greater use of the Royal We, in which "we" is used to reflect commands (e.g., Schimpff, 2019), thus indicating superiority and authority status. However, such interpretation is speculative, as this possibility was not directly tested in the study (and also contradicts previous findings), and thus requires empirical investigation. Moreover, this study again did not focus specifically on vulnerable narcissism, meaning that these findings may not generalise to individuals with pathological forms of narcissism.

To summarise, considerable empirical attention has been directed towards exploring the verbal behaviour of narcissism in everyday life. Narcissistic PD, on the other hand, is yet to receive such empirical attention. Critically, the findings from the (subclinical) narcissism literature that are most generalisable to narcissistic PD are those that specifically relate to vulnerable narcissism. Naturally, the primary focus of research so far has been on examining the relationship between narcissism and self-focused language. Although the findings in this regard are somewhat inconsistent, the empirical literature available predominantly suggests that there is not a direct association between self-focused language and narcissism, contradicting the common assumption that narcissistic people talk about themselves more. However, when investigating vulnerable (or darker forms of) narcissism specifically (as in narcissistic PD), this has in fact been found to be associated with more self-focused language, reflecting mental distress and negative emotionality. Other linguistic markers of pathological narcissism still need exploration, as findings from the current literature base mostly relate to grandiose narcissism.

2.2.1.3 The Language of a (Subclinical) Psychopath

As with subclinical narcissism, there is a considerable amount of research that has leveraged computational linguistic methods to study psychopathy. It is important to acknowledge that although psychopathy is not a diagnosable mental health condition, there is a fine line between psychopathy and antisocial personality disorder (ASPD) – a diagnosable PD (described in detail in Section 2.3.1.2) – as they are conceptually related constructs (e.g., Werner et al., 2015). Of the central characterising features of subclinical psychopathy, it is the antisocial behaviour aspect that is most overlapping with ASPD. Conversely, the main distinctions between psychopathy and ASPD are that with ASPD, the primary focus is on the (antisocial) behavioural characteristics, whereas with psychopathy, the focus is directed towards (dysfunctional) interpersonal, affective, and psychological processes (e.g., shallow affect), in addition to antisocial behaviour. Moreover, some have simply argued that psychopathy is an extreme form of ASPD (e.g., Black, 2019; Coid & Ullrich, 2010). Accordingly, it could be presumed that psychopathy and ASPD would share overlapping linguistic features, especially those that relate to the antisocial behavioural aspects. Nonetheless, it could also be expected that there would be distinct linguistic patterns specific to psychopathy, given that it is perceived as being distinguishably more extreme than ASPD. I review the limited literature on the application of computational language analysis methods to the study of ASPD specifically in the subsequent clinical personality pathology section (Section 2.3), whereas the focus in this section is on the more extensive (subclinical) psychopathy language-based research.

Notably, one vital facet of psychopathy that has received considerable attention by language researchers is the affective dysfunction dimension; likely a consequence of the characterisation put forward in Cleckley's (1976) *"Mask of Sanity"*, in which psychopathy is conceptualised as a semantic disorder whereby the affective components of language are not well integrated. Indeed, indicators of affective dysfunction have been empirically evidenced in the language of psychopathic individuals. For example, language-based research carried out in the context of affect revealed the verbal behaviour of psychopathic offenders to be relatively inconsistent and incoherent, with frequent contradictory statements and topic changes (e.g., Brinkley et al., 1999; Hancock et al., 2013; Marko & Leibetseder, 2023). Critically, such findings show how computational linguistic researchers can integrate psychological theory in their work to generate clinically meaningful findings, of which can subsequently inform psychological theory.

Yet, in contrary to what one might expect, and in contrast to linguistic findings generated from research on various clinical conditions (including PDs), greater negative emotion has not been reflected in the language of psychopaths (Hancock et al., 2013; Hancock et al., 2018; Marko & Leibetseder, 2023), with the exception of anger (Hancock et al., 2018; Le et al., 2017). More intriguingly, one study in fact found psychopathic offenders to use fewer positive *and* negative emotion words than non-psychopathic offenders, again with the exception of anger (Le et al., 2017). In addition, Le et al. (2017) also revealed that the most significant predictors of psychopathy were lower frequencies of anxiety words and higher frequencies of personal pronouns. The explanatory power of fewer anxiety words on psychopathy variance is intelligible given that (primary) psychopathy is associated with low anxiety and fearlessness (e.g., Neumann et al., 2013). These findings therefore further differentiate psychopathy from clinical mental health conditions, of which are generally associated with greater references to negative emotion (and anxiety in particular) in natural language (e.g., Coppersmith et al., 2015).

Moreover, adding to this differentiation between psychopathy and clinical mental health conditions, and in contrast to findings from Le et al. (2017), a case study of a psychopathic serial killer discussed in the previous narcissism section (Section 2.2.1.2) in fact revealed a strong presence of positive emotion words in the personal letters of this psychopathic individual (Marko & Leibetseder, 2023). Although intriguing, given that this finding has only been evidenced in this single case study, it is possible that it could be specific to this particular individual, or to the population of serial killers. Nevertheless, the literature described here provides support to the notion that (subclinical) psychopathy is conceptually distinct from related diagnosable mental health conditions (i.e., clinical PDs), illuminating the potential of language analysis approaches to contribute to psycho(patho)logy knowledge and theory.

In other insightful work, numerous studies have also uncovered the language of psychopathic individuals to reflect psychological distancing. For example, one study, conducted by Hancock and colleagues (2013), compared the verbal behaviour of psychopaths convicted of murder with non-psychopaths convicted of murder describing their crime narratives. From this analysis, it was found that psychopathic murderers used more past tense and fewer present tense words, and more articles and causal words, than non-psychopathic murderers; psychopathic murderers also used more words related to basic material needs (e.g., food, money, shelter). Such language patterns suggest that psychopathic individuals were more psychologically distanced from the murders (Woodworth & Porter, 2002), and that they viewed the murders as more instrumental than non-psychopathic individuals (i.e., using more cause-and-effect descriptors). Although these findings are specific to the context of murder, the overall findings from the computational linguistic research discussed support the conceptualisation that psychopathy is associated with shallow affect and emotional and cognitive detachment.

Moving on to the antisocial aspect of psychopathy, given that antisocial behaviour is strongly associated with psychopathy (and ASPD), one would expect to see this antisocial nature reflected in the verbal behaviour of individuals with highly psychopathic traits. Indeed, research has consistently revealed psychopathic individuals to use more hostile language, with a greater frequency of swear words and anger words, for example (e.g., Le et al., 2017). Not surprisingly, research has also shown psychopathy to be associated with language patterns reflective of social dysfunction. For example, relative to non-psychopathic people, psychopathic individuals have been found to make fewer references to other people, reflecting less social orientation (Hancock et al., 2013; Le et al., 2017). From these findings, it is clear that the antisocial nature of psychopathy is, at least to some extent, reflected in the verbal behaviour of psychopathic individuals, which could prove useful in assisting with the monitoring of antisocial conditions in the general population.

Although the language-based research findings discussed here are insightful regarding the underlying nature of psychopathy, one major limitation of the research discussed so far is that it has all been conducted in forensic populations. That is, all of the studies discussed have comprised offender samples. The limited focus on forensic samples is problematic as it means that the findings may not be generalisable to the general population, given that the language analysed in these studies may only be reflective of more extreme psychopathic individuals (or those of greater danger to others), and not people with highly psychopathic traits that have not been incarcerated.

Importantly, one recent empirical study indeed investigated the linguistic markers of psychopathy in a non-forensic population (Hancock et al., 2018), by exploring associations between psychopathy (assessed via a self-report scale) and linguistic features present in everyday online communication, including emails, SMS messages, and Facebook messages. Interestingly, findings from Hancock et al. (2018) were mostly consistent with findings from the other psychopathy research conducted on forensic populations. For instance, participants higher in psychopathy were found to refer less frequently to other people, show more psychological distancing, produce less comprehensible language, and use more hostile language (e.g., anger and swear words). However, participants higher on psychopathy did not focus more on basic needs, as evidenced in forensic populations, suggesting that there may be some linguistic differences between more extreme psychopathic offenders and individuals with psychopathic traits in the general population. What is also particularly interesting from the Hancock et al. (2018) study is that they found significant differences in linguistic patterns between subfactors of psychopathy. For example, psychological distancing language was associated with Callous Affect and Erratic Lifestyle, but not Interpersonal Manipulation or Criminal Tendency subfactors. Promisingly, such differences imply that precise language patterns may relate to and differentiate specific facets of psychopathy, and thus very specific problematic personality components more broadly. Critically, this ability to precisely identify particular problematic or pathological personality components from natural language is likely to become increasingly important with the adoption of the dimensional approach to personality pathology classification.

Overall, studies that have leveraged computational language analysis methods to better understand psychopathy have generally evidenced linguistic markers of psychopathy consistent with its conceptualisation, including less emotional expression in language, greater hostility and aggression, incoherent language, and language patterns indicative of psychological distancing. Although the literature discussed in this section provides useful insight into the underlying nature of psychopathy – via a naturalistic behavioural approach – almost all of the studies discussed (with the exception of Hancock et al., 2018) comprised forensic populations, and so the findings may not be generalisable to the general population. Nevertheless, the research discussed in this section has further demonstrated the promising potential of computational language analysis methods to contribute to psychological theory, particularly regarding informing the conceptualisation of psychopathological constructs.

2.3 Clinical Personality Disorder Language Literature

In this section, I review the limited literature that has applied computational language analysis methods to the study of clinical PDs. To date, the majority of the work in this realm has focused on BPD, with a handful studies also conducted on ASPD. Further, there have also been a small number of studies that have applied computational language analysis methods to study personality pathology in a more general sense (i.e., not focusing on any particular type of PD). Accordingly, I will discuss the literature available that has applied computational language analysis methods to the study of personality pathology in general, followed by BPD and ASPD specifically, in turn, in the following sections.

2.3.1 Personality Pathology and Language

Regarding the handful of studies that have leveraged language analysis to better understand personality pathology in a general sense, one early study revealed some initial insight into the way in which personality pathology manifests in natural language (Molendijk et al., 2010). In alignment with the notion emphasised earlier that the linguistic markers of depression (e.g., self-focused language) may in fact be transdiagnostic, Molendijk and colleagues (2010) revealed that, irrespective of whether individuals with PD had a concurrent or historical depression diagnosis or not, those with PD were found to use more first-person singular pronouns, negative emotion words, and fewer positive emotion words compared to individuals without PD. Such findings thus provide further support to the notion that many of the linguistic markers of depression may in fact be reflective of psychological distress more broadly (see, e.g., Lyons et al., 2018).

Building on this work, what is more impactful than simply identifying language patterns associated with (or reflective of) personality pathology is the ability to utilise such linguistic features to track, in real time, psychotherapeutic change, as this has direct implications for clinical practice. Indeed, in a study conducted by Arntz and colleagues (2012), language was utilised as a measure of psychotherapeutic change in response to long-term treatment in individuals with PD. Promisingly, the results revealed a significant decline in linguistic indicators of poor mental health – including the use of first-person singular pronouns, negative emotion words, causation words, negation words, and past and future tense verbs – over the course of treatment, with increases linguistic indicators of positive mental health (i.e., present-tense verbs and positive emotion words) coinciding. Moreover, such changes resulted in the verbal behaviour of individuals with PD becoming similar to that of the general population, and they also predicted better treatment outcome. In particular, reductions in the use of negative emotion words and negations were the strongest predictors of outcome, which is intuitive given that less negative language reflects less attention to (and, presumably, rumination around) negative emotion. Vitally, findings from Arntz et al. (2012) demonstrate how tracking changes in word use can be used to assess psychological treatment effectiveness and measure therapeutic change in individuals with PD, or mental health problems more broadly.

In a similar vein, other interesting and creative applications of computational linguistic techniques to the study of personality pathology have been demonstrated in recent empirical work. For instance, one impactful study utilised sophisticated computational language analysis methods to generate linguistic semantic vectors (i.e., numerical representations of language that take the context of words into account) of PDs – indicating the extent to which forms of personality pathology are prevalent in language – of which shown predictive validity in their associations with the Big-5 personality traits (Neuman & Cohen, 2014). In addition to this, a recent study leveraged NLP and machine learning methods to detect possible causes of PDs based on language in Twitter posts, specifically through topic identification and mood detection techniques (Ellouze et al., 2021). Valuably, such naturalistic, technologically advanced research exemplifies the endless possibilities of applying computational language-based methods to the domain of personality pathology. Yet, at present, research utilising language analysis has only touched the surface with respect to the limitless potential of such approaches to better understand, and subsequently help to treat, personality pathology. I now move on to reviewing literature that has applied computational language analysis to the study of BPD and ASPD specifically.

2.3.1.1 Borderline Personality Disorder

BPD is the most common PD in clinical populations (Leichsenring et al., 2011). Although a problematically heterogenous construct (e.g., Cavelti et al., 2021), BPD is generally characterised by longstanding patterns of intense emotional fluctuations (and emotion dysregulation more broadly), stormy interpersonal relationships, self-harming behaviours, high impulsivity, and a diminished sense of identity (APA, 2013). Notably, it is arguable that BPD is one of the most high-risk, problematic mental health conditions, given that it is associated with a high risk of suicide, intensive use of treatment, and high costs to society (Leichsenring et al., 2011). Thus, BPD represents a serious public health problem.

Valuably, continuing the message emphasised throughout this thesis, exploring the natural language of individuals with BPD makes it possible to gain implicit insight into the psychological processes underlying the disorder. For instance, it could be expected that the impairments associated with BPD would be reflected in the verbal behaviour of individuals with BPD. In line with this expectation, expressive language disturbances have been shown to be typical of those with BPD (Carter & Grenyer, 2012; Rosenbach & Renneberg, 2015). Specifically, research has revealed that BPD is associated with expressive language impairment in the form of lower levels of lexical complexity (complexity of words used) and syntactic complexity (complexity of construction of sentences; Carter & Grenyer, 2012). Given that the language data analysed in these studies were in response to emotional stimuli, this aligns with the notion that people with BPD experience elevated cognitive complexity deficits when emotions are high (e.g., Roepke et al., 2013).

Relatedly, since BPD is strongly accompanied by mental distress, it could be presumed that this would be prevalent in the natural language of individuals with BPD. Indeed, numerous linguistic markers of mental distress have been uncovered to be prolific in the language of those with BPD. For instance, the use of first-person singular pronouns, negative emotion words, health words, conjunctions, causal words, tentative language, past-tense orientated language, and death words have all been found to be significantly greater in the language of people with BPD compared to the general population (Carter & Grenyer, 2012; Coppersmith et al., 2015; Leavitt, 2019; Lyons et al., 2018; Rosenbach & Renneberg, 2015). The relatively elevated use of death words may also be reflective of the suicidal tendencies associated with BPD. Taken together, these findings highlight the overlap in the linguistic markers of BPD and depression, or psychological distress more broadly, thus implying that many (if not all) of these linguistic traces are likely reflective of general distress or psychopathology, rather than necessarily specific to borderline pathology.

Accordingly, what is more insightful than findings demonstrating similarities between the language patterns associated with both BPD and general mental distress surround findings uncovering language patterns distinct to BPD. Indeed, one linguistic feature revealed to be more strongly associated with BPD – relative to other mental health conditions – is greater reference to anger (Coppersmith et al., 2015; Leavitt, 2019; Rosenbach & Renneberg, 2015). Informatively, this elevated reference to anger in the language of people with BPD likely illuminates the central role of anger in the emotion dysregulation experienced by these individuals. Furthermore, research has also uncovered language patterns indicative of impulsivity to be associated with BPD, such as greater use of swear words, and more so than with other mental health conditions (Carter & Grenyer, 2012; Coppersmith et al., 2015). The excessive use of anger and swear words likely reflects the impulsive, disinhibited, and emotionally dysregulated nature of individuals with BPD. Interestingly, given that these language patterns are largely distinctive to BPD, they therefore do not simply reflect general mental distress (as with the other linguistic markers), and thus are of value in differentiating the nature of BPD from other mental health conditions, which could help to inform tailored BPD treatments.

In applying such investigations to other core areas of dysfunction in BPD, empirical examinations of pronoun use and socially-relevant language have revealed traces of interpersonal dysfunction to be prevalent in the natural language of individuals with BPD. In particular, studies have shown BPD to be associated with the use of lower rates of first-person plural pronouns and higher rates of third-person pronouns (Carter & Grenyer, 2012; Coppersmith et al., 2015; Leavitt, 2019; Lyons et al., 2018). Notably, the less frequent use of words such as *"we"* and more frequent use of words such as *"she/he"* appears to reflect typical social impairments associated with BPD, in that they are talking *about* other people more but affiliating themselves *with* others less. Moreover, the use of third-person singular pronouns has even been found to be greater among people with BPD when compared to those with other mental health conditions (Lyons et al., 2018). Additionally, with respect to social language, one study evidenced people with BPD to use more social words – particularly in relation to family – when compared to people with depression and the general population (Rosenbach & Renneberg, 2015). However, this finding is not surprising given that the participants had the task of retrieving and describing memories of rejection. Such findings therefore indicate that people with BPD primarily associated their memories of rejection with family, which is in accordance with the notion that the development of BPD is commonly associated with the (adverse) childhood familial environment (e.g., Carr & Francis, 2009).

Valuably, the research discussed so far has allowed for potential linguistic markers of BPD to be identified. Building on this research, one study built a classifying model, based on linguistic markers, to predict the likelihood of a particular person having a particular mental health condition, of which was found to have a precision of 58% in detecting BPD from language in social media posts (Coppersmith et al., 2015). Advancing this work further, other studies have also used NLP and machine learning techniques to attempt to automatically detect BPD from natural language, of which have generated promising classification accuracy rates (Khazbak et al., 2021; Wang et al., 2020). Such research again demonstrates the endless possibilities and potential of the clinical implications of work in this realm. That said, due to the potential black-box (i.e., obscure or ambiguous) nature of this type of advanced computational linguistic work, it is critical that such research is intertwined with psychological theory and perspective, to ensure psychologically meaningful and practicable findings.

While limited in quantity and scope, the existing research in the realm of BPD and verbal behaviour has made a significant contribution to the field and has helped to develop a deeper understanding of the psychosocial processes underlying BPD. For instance, markers of mental distress were consistently shown to be prevalent in the verbal behaviour of individuals with BPD (i.e., more first-person singular pronouns, negative emotive language, health words, death words, causal words, and tentative language, and fewer first-person plural pronouns). Not surprisingly, such patterns subsequently illuminate mental distress to be strongly connected with the construct of

BPD. More interestingly, several linguistic markers have been consistently evidenced to be more distinctive to BPD (relative to other mental health conditions); namely, the excessive use of anger words, swear words, and third-person pronouns. Promisingly, the linguistic markers identified from this work are in accordance with the clinical and behavioural characteristics of BPD, reflecting the (negatively) emotionally intense, impulsive, socially impaired nature of individuals suffering with BPD.

Nonetheless, the language-based research that has been conducted on BPD so far is constrained by the fact that it has consisted of fairly small sample sizes, limiting the generalisability of the findings. Although this is an issue in clinical research more broadly (due to difficulties in recruitment, etc.), it appears to be a particularly common issue within PD research specifically. Importantly, the use of large-scale computational linguistic analysis methods allows researchers to overcome such recruitment issues by utilising data that is widely available and easy to access (Boyd et al., 2020). Moreover, a large-scale linguistic analysis approach can provide in-depth, generalisable insight into the underlying psychology of individuals with pathological personalities.

2.3.1.2 *Antisocial Personality Disorder*

Unlike psychopathy (a conceptually closely related construct, as discussed in the "Language of a (Subclinical) Psychopath" section [Section 2.2.1.3]), ASPD is a diagnosable mental health condition, and is defined as a pervasive pattern of disregard for and violation of the rights and considerations of others, starting in childhood or adolescence, and lacking remorse (APA, 2013). It is clear from this definition (and label) that ASPD is predominantly characterised by antisocial behaviour; cognitive and affective deficits are given some attention, but to much a lesser extent. Individuals with ASPD will regularly engage in behaviours that are considered antisocial or deviant within a given society, such as displaying physical aggression towards others or destroying property. In addition to frequent displays of antisocial behaviour, people with ASPD will typically experience diminished sense of identity combined with problems with emotion regulation and social functioning. Not surprisingly, as with psychopathy (and the other Dark Triad traits), ASPD is associated with criminal behaviour, with high prevalence rates of ASPD in the forensic population (Black et al., 2010). Moreover, ASPD is also associated with a high risk of suicide, substance misuse, and comorbidity with other PDs (Black et al., 2010; Compton et al., 2005), reflecting the particularly high-risk nature of ASPD.

Regarding the limited work conducted on ASPD using computational language analysis methods, such research has revealed individuals with ASPD to express greater emotion, particularly negative emotion, in their language compared to people without ASPD (Eichler, 1965; Gawda, 2010a, 2010b, 2013), which contrasts with the psychopathy language findings. Such differences in emotion language patterns could reflect a key discrepancy in emotion processes between ASPD and psychopathy, with emotional deficits (i.e., shallow affect) primarily characteristic of psychopathy and not ASPD. Instead, the findings imply that ASPD is characterised more by affective dysregulation (as in other types of personality pathology, such as BPD) than shallow affect, which is generally in line with the clinical literature showing ASPD to be associated with maladaptive emotion regulation (e.g., Garofalo et al., 2018). Discrepancies in the psychological processes underpinning ASPD and psychopathy are additionally supported by research revealing ASPD to be associated with greater self-focused language (Eichler, 1965; Gawda, 2010a, 2010b, 2013) – indicating psychological distress – whereas research on psychopathy evidenced no such association (Hancock et al., 2013; Hancock et al., 2018; Le et al., 2017). Taken together, these findings suggest that ASPD is characterised by mental distress and negative emotionality, whereas this does not appear to (necessarily) be the case for psychopathy. On a broader level, such work highlights how language analysis methods can assist with distinguishing between closely related psycho(patho)logical constructs, of which is of vital importance with respect to personality and psychopathology assessment and classification.

In the same way as differences in the linguistic profiles between ASPD and psychopathy are revealing of their distinctive psychological features, similarities in language patterns indicate shared underlying psychological processes. Interestingly, one linguistic pattern that has been consistently found to be associated with both ASPD and psychopathy is incoherent and incomprehensive language. Specifically, the rise in advanced computerised text analysis tools has allowed for incoherent language to be explored in-depth through capturing subtle aspects of language, such as filler words and disfluencies (e.g., *'err', 'um'*). Through these methods, numerous studies have uncovered the verbal behaviour of individuals with ASPD or psychopathy to be less coherent, with a greater frequency of filler words and disfluencies in their language (e.g., Eichler, 1965; Hancock et al., 2013). Accordingly, these linguistic patterns likely reflect the cognitive dysfunctions associated with both ASPD and psychopathy. Alternatively, given that the use of disfluencies and filler words allows the speaker to stall their response, these linguistic features may also be reflecting the manipulative tendencies associated with both ASPD and psychopathy.

In related work, findings from a more recent study investigating language in written texts (researcher prompted) of prisoners diagnosed with ASPD (compared with prisoners not diagnosed with ASPD and individuals in the general population) provided further confirmation of the linguistic features typically associated with ASPD (Gawda, 2022). Interestingly, rather than focusing on individual linguistic features, this work uncovered two narrative styles, named *demonstrative-digressive-egocentric-emotional-dogmatic* and *reserved-focused on the topic-repetitive*, of which were both evidenced to be positively predictive of ASPD (or psychopathic deviate traits). Importantly, these two narrative styles reflected divergent language patterns among individuals with ASPD (and high psychopathic deviance), indicating varied styles of communication in this population. The *demonstrative-digressive-egocentric-emotional-dogmatic* narrative style broadly reflects the prototypical characteristics of ASPD – and is consistent with the ASPD language literature – with this style comprising longer texts, greater reference to emotion, manipulative and persuasive language, greater negativity and disagreeableness in language, more incoherent language, greater use of absolutist (i.e., dichotomous) language (indicative of all-or-nothing thinking), and greater use of self-focused language. In comparison, the *reserved-focused on the topic-repetitive* narrative style reflected an alternative linguistic style associated with ASPD, comprising shorter texts, less manipulative and persuasive language, repetitive language, more incoherent language, and greater negativity in language. Such varied linguistic styles associated with ASPD indicates the highly heterogeneous nature of this construct (and PDs more broadly), with the distinct language styles likely reflecting distinct forms of ASPD. Moreover, the varied narrative styles also illuminate the complex nature of the relationship between language and personality (pathology), with this relationship presumably influenced by a broad range of mediating factors (e.g., situational factors, topic discussed, demographic characteristics).

In general, research that has utilised computational language analysis methods to study both ASPD and psychopathy has provided valuable insight into the psychological processes underlying these antisocial constructs. Consistent findings revealing similarities in the verbal behaviour associated with ASPD and psychopathy – such as incoherent language, more hostile and negative language, and less social language – indicate shared psychological processes common among both antisocial disorders, mostly reflecting their antisocial nature and associated cognitive dysfunction. Conversely, the main discrepancies in the linguistic profiles of ASPD and psychopathy surround language patterns relating to emotion processes. Broadly speaking, research has revealed individuals with ASPD to express more emotion – particularly negative emotion – and mental distress in their language, and psychopathic individuals to express less emotion in their language (with the exception of anger), with no associations with linguistic indicators of mental distress. Such divergent findings are intelligible given that they are generally in line with the clinical characteristics of ASPD and psychopathy. However, interpretations from the research discussed here should be made with caution, given that there is very limited research that has explored the verbal behaviour of people with ASPD, meaning that comparisons between the linguistic markers (and the reflected psychological processes) of ASPD and psychopathy are not very reliable.

2.4 Summary of the Personality Dysfunction Language Literature

Importantly, empirical research that has studied the verbal behaviour associated with problematic personalities (i.e., the Dark Triad of personality) and clinical PDs has provided useful insight into the underlying psychology and true nature of these problematic or pathological personality constructs, using an individualised, behavioural approach. Moreover, all of the research discussed has illuminated the promising potential of applying computational language analysis methods to the domain of problematic/pathological personalities. However, at present, considerably more empirical attention has been directed towards subclinical problematic personality traits (i.e., the Dark Triad) than clinical PDs. Further, the personality dysfunction and language literature base is constrained by the fact that the predominant focus of such research has been on identifying linguistic markers of problematic personality traits or clinical PDs through associating frequencies of linguistic features with such constructs. Undoubtably, this is important validation work that helps to “set the stage” for promising new horizons. Nonetheless, as illustrated in the Personality Disorder and Verbal Behaviour chapter (Chapter 1.1), there are uncountable other, more sophisticated and psychologically insightful ways in which computational language analysis techniques can be leveraged to better understand problematic and pathological personality constructs.

Given that personality pathology is still very poorly understood, significantly more empirical research regarding all aspects of personality pathology is essential for generating the level of understanding necessary to improve the assessment process and improve the quality and effectiveness of treatment. Only once we have a better understanding of personality pathology will we be better equipped to serve those affected, including individuals suffering with PD and those close to them. It is therefore crucial that future research maximises the potential of the limitless possible applications of computational language analysis methods in improving our understanding of personality pathology.

CHAPTER 3:

Uncovering the Social-Cognitive Contributors to Social Dysfunction in Borderline Personality Disorder Through Language Analysis

Charlotte Entwistle, Ryan L. Boyd

Text as it appears in:

Entwistle, C., & Boyd, R. L. (2023). Uncovering the social-cognitive contributors to social dysfunction in borderline personality disorder through language analysis. *Journal of Personality Disorders, 37*(4), 444–455. https://doi.org/10.1521/pedi.2023.37.4.444 [3]

[3] This text is written in American English, as a result of publication in the *Journal of Personality Disorders*.

Abstract

Borderline personality disorder (BPD) is characterized by severe interpersonal dysfunction, yet the underlying nature of such dysfunction remains poorly understood. The present study adopted a behavioral approach to more objectively describe the social-cognitive contributors to interpersonal dysfunction in BPD. Participants (*N* = 530) completed an online survey comprising validated measures of BPD features and other problematic interpersonal traits (e.g., narcissism), as well as a writing prompt where they were asked to share their personal thoughts about relationships. Computerized language analysis methods were used to quantify various psychosocial dimensions of participants' writing, which were incorporated into a principal component analysis. Analyses revealed four core social dimensions of thought: 1) *Connectedness/Intimacy*; 2) *Immediacy;* 3) *Social Rumination;* 4) *Negative Affect*. All four dimensions correlated with BPD features in intuitive ways, some of which were specific to BPD. This study highlights the value of natural language analysis to explore fundamental dimensions of personality disorder.

Keywords: borderline personality disorder, language analysis, interpersonal dysfunction, factor analysis, social-cognitive dimensions

3.1 Introduction

Borderline personality disorder (BPD) is a severe mental health condition marked by long-term patterns of emotion dysregulation and distorted self-perceptions, affecting an estimated 1.6% of the general population and around 20% of the psychiatric inpatient population (Ellison et al., 2018). BPD is especially characterized by interpersonal dysfunction (e.g., Hill et al., 2008; Miano et al., 2020), which typically manifests as problematic dependent and/or avoidant attachment patterns (Levy, 2005) and patterns of intense and stormy relationships (APA, 2013). The severity of such social dysfunction is underscored by its association with extremely negative outcomes among people with BPD, with BPD-driven social problems often triggering psychological distress and engagement in maladaptive behaviour, such as self-harm and suicide attempts (e.g., Berenson et al., 2016). Despite widespread awareness of the severity of consequences associated with interpersonal dysfunction in BPD, little is known about the root psychological features that characterize these problematic interpersonal patterns.

A large body of evidence suggests that impairments in social cognition play a major role in the social dysfunction characteristic of BPD (see Lazarus et al., 2014, for a review). For instance, a general disorganisation of social processes within and across social domains (e.g., friendships, colleagues) has been proposed as underpinning interpersonal dysfunction in BPD (Hill et al., 2008). Under this theory, individuals with BPD are believed to have diminished awareness of social boundaries and have difficulty their adapting their behavior according to the social context (e.g., one may frequently discuss unwarranted personal and intimate topics among work colleagues), causing interpersonal problems. Additionally, social dysfunction in BPD may be driven, in part, by deficits in one's ability to recognise others' emotions and understand the perspectives of others (e.g., Domes et al., 2009). However, findings on the specific nature of social-cognitive impairments in BPD are inconsistent; for example, some evidence suggests relatively greater empathetic accuracy among individuals with BPD, particularly when faced with relationship threat (Miano et al., 2017a).

In addition to social-cognitive impairments, affective dysregulation has been strongly argued to be a fundamental component underpinning interpersonal dysfunction, and potentially all impairments, in BPD (e.g., Euler et al., 2019; Lazarus et al., 2014). In essence, affective dysregulation can result in rapid mood swings and intense emotional outbursts, leading to interpersonal conflict and social rejection. In particular, lower thresholds for feelings of anger as well as greater alexithymia (i.e., the inability to identify and describe one's emotions) have been found to be associated with greater interpersonal difficulties in BPD (Berenson et al., 2018). As well as instigating social problems in BPD, emotion dysregulation has been found to mediate the relationship between BPD status and social dysfunction, specifically in relation to social rejection sensitivity (Dixon-Gordon et al., 2013) and mentalization ability (Sharp et al., 2011), among others. Emotion dysregulation therefore also plays a major role in propagating social impairments in BPD. Further, disturbances in intimacy are particularly prevalent in individuals with BPD and have also been identified as a characterizing feature of their interpersonal dysfunction (Jeung & Herpertz, 2014).

Despite several propositions, clear consensus on the specific, core dimensions characterizing interpersonal dysfunction in BPD, and how they inform our understanding of the disorder, is yet to be established. Recent methodological advances may provide an avenue to improve understanding of social dysfunction in BPD; in particular, by looking at the ways in which people conceptualize and talk about their social connections. A substantial body of research has shown that it is possible to analyze language patterns to unobtrusively reveal the substance and style of thought (see Pennebaker, 2011), which can overcome limitations inherent to traditional assessment methods in personality disorder, such as self-report questionnaires (Entwistle et al., 2022). It could be expected, then, that directly quantifying *how* people think about relationships should be revealing of key social-cognitive dimensions, which may help to characterize interpersonal dysfunction in BPD. Indeed, scholars have emphasized the notion that one's language use acts as a behavioral indicator of their mental representations of interpersonal relationships (Horn & Meier, 2022), suggesting that the analysis of natural language can allow insight into social-cognitive processes.

In the broader personality literature, numerous studies have highlighted how specific dimensions of personality can be traced in language (e.g., Kulkarni et al., 2018; Pennebaker & King, 1999; Yarkoni et al., 2010). More relevantly, research has also revealed how relationships can be reliably characterized by fundamental social(-cognitive) dimensions (e.g., *power; trust*) at large scale through analyzing language from conversations (Choi et al., 2020). Such research highlights the potential of using computational language analysis methods to gain novel insights into core psychosocial dimensions in a way that goes beyond traditional psychometric approaches. Despite the promising potential of language analytic methods to provide insight into the social-cognitive impairments of individuals with BPD, to date, no such studies have been conducted to our knowledge.

Accordingly, in the present study, we aim to address the following central research question: *What are the core social-cognitive dimensions that characterize interpersonal dysfunction in BPD?* To address this, we analyze the language that people use when describing their relationships to infer core social-cognitive dimensions, then use these dimensions to inform our understanding of the nature of interpersonal dysfunction in BPD. We also measure how such social-cognitive dimensions relate to other constructs associated with problematic social functioning and behavior, to explore the extent to which they are specific to BPD or are reflective of interpersonal dysfunction more generally. Specifically, we assess the “Dark Triad” of personality traits (i.e., psychopathy, narcissism, and Machiavellianism; Jonason & Webster, 2010) in order to capture, in a generalized way, a sphere of interpersonal dysfunction that can be differentiated from BPD.

3.2 Methods

Data were collected as part of a larger investigation on the associations between natural language and various sociopsychological processes, including BPD features.

3.2.1 Participants and Procedure

Participants were recruited via targeted sampling from online forums. The study was advertised by distributing an anonymous link to a Qualtrics questionnaire across various forums. Alongside targeting some general discussion forums, recruitment was particularly targeted toward mental health forums, including a large forum dedicated towards BPD, with the aim of enhancing sample diversity in mental health status. Participants were excluded if they reported that they could not speak or write in fluent English or if they were under the age of 18. Participants were not offered any incentives for participating in the study.

After providing informed consent and demographic information, participants responded to a series of psychological questionnaires and prompts that were presented in a randomized fashion. Participants who did not provide sufficient data – i.e., those who did not provide any responses to the problematic personality measures (*N* = 70) or did not write a minimum of 50 words in the relationship writing task (*N* = 67) – were omitted (total *N* excluded = 137); refer to Appendix A.1 for an analysis of missing/excluded data (i.e., comparing those included in subsequent analyses with those excluded due to minimum word count criteria on key outcome variables). This process resulted in a final sample of 530 participants (*M* age = 26.22, *SD* = 8.41; 72.88% female; 75.62% White; see Table A.1 in Appendix A.2 for full sociodemographic characteristics of the sample).

3.2.2 Materials

3.2.2.1 Measures

Personality Assessment Inventory-Borderline Scale (PAI-BOR). The PAI-BOR (Morey, 1991) was used to assess BPD features, specifically assessing four core features: affective instability, identity problems, interpersonal dysfunction, and self-harm. Each of these features are assessed through 6 items (24 items in total) on a 4-point response scale ranging from 0 (false) to 3 (very true). The mean total PAI-BOR score in the present study was 37.24 (*SD* = 13.03; range = 4 – 72; skewness = .01), which falls just below the established cut-off score of 38, whereby BPD features are considered to be significantly present and therefore worthy of further diagnostic investigation (Morey, 1991; see Table A.1 for descriptive statistics for BPD features).

Dirty Dozen. In order to disambiguate dimensions characterizing interpersonal dysfunction in BPD from other constructs associated with problematic social functioning, participants were asked to complete the Dirty Dozen questionnaire – a well-validated tool for assessing the “Dark Triad” of personality traits: psychopathy, narcissism, and Machiavellianism (Jonason & Webster, 2010). The Dirty Dozen is a 12-item measure, with four items assessing each of the three Dark Triad traits on a 5-point response scale, ranging from 1 (strongly disagree) to 5 (strongly agree), including items such as “I tend to manipulate others to get my way” (Machiavellianism), “I tend to be callous or insensitive” (psychopathy), and “I tend to want others to admire me” (narcissism). The mean total score from the Dirty Dozen measure in the current sample was 31.80 (*SD* = 10.15; range = 12 – 60; skewness = .61; see Table A.1 for descriptive statistics for specific Dark Triad traits).

3.2.2.2 Writing Task

To collect natural language data reflecting participants’ social cognitions, a prompt was included which asked participants to write about their relationships, broadly defined:

> *When you think about your relationships with other people, what comes to mind? For the next 7 minutes (or more), we would like for you to write about how you get along with people. This can include your relationships with coworkers, family, friends, and romantic partners. Try to say as much as you can about both the good and the bad. Do not worry about spelling or grammar.*

> *Simply write everything that comes to mind, giving as much detail as possible. Once you begin writing, try to write continuously until you have finished. If you run out of things to say, re-tell what you have previously said in other words.*

Participants' essays were corrected for common misspellings (e.g., "boyfreind" instead of "boyfriend") and elongations (e.g., "sooo unhappy"). All written responses containing fewer than 50 words were removed from the dataset, to ensure validity of measurement and reliable scores (see, e.g., Boyd, 2017; Cutler et al., 2021; Pennebaker & Ireland, 2011). On average (after removing texts with < 50 words), participants wrote 211.60 words (*SD* = 186.22).

3.2.3 Language Analysis

Language data were analyzed using the automated word-counting, text analysis program Linguistic Inquiry and Word Count (LIWC2015; Pennebaker et al., 2015). Briefly described, the LIWC software calculates the percentage of words belonging to psychologically meaningful dimensions in each text using an internal dictionary that maps words onto meaningful categories. Categories measured by LIWC include the extent to which people are thinking about themselves, other people, emotions, leisure activities, work, and so on. The use of LIWC has been extensively validated across diverse disciplines, spanning fields such as psychology, health and medicine, and computer science, and has been particularly prominent in mental health research (Tausczik & Pennebaker, 2010).

3.3 Results

Following the approach of Pennebaker and King (1999), LIWC scores were incorporated into a principal component analysis (PCA) to reduce the linguistic features into core, language-based, social-cognitive dimensions. Specifically, the PCA was conducted using LIWC scores (derived from the relationships essays) as factors, using the Maximum Likelihood method of extraction with a varimax rotation applied. All LIWC variables from the 2015 built-in dictionary were included in the PCA, with the exception of summary categories (e.g., "analytic", "clout"), filler words, non-fluencies, and punctuation, in order to minimize redundancies and the inclusion of measures comprised entirely of other LIWC measures. In total, 70 LIWC variables were included in the PCA (see Table A.2 in Appendix A.3 for a full list of included LIWC variables and their descriptive statistics). Although the Kaiser-Meyer-Olkin (KMO) statistic was on the lower side (KMO = 0.41), Bartlett's test of sphericity was significant ($\chi^2(2415) = 28371.11$, $p < .001$), indicating the appropriateness of the factor analytic model for this dataset.

The PCA resulted in the extraction of four social-cognitive components, comprising 18 LIWC variables (see Table 3.1 for factor loadings for each component). The 4-component solution was generated on the basis of eigenvalues (all components >= 3), inspection of the scree plot, and interpretability of components in the context of social-cognitive functioning. Only LIWC variables with an absolute factor loading greater than 0.5 were retained.

Table 3.1

Principal Component Analysis, Rotated Matrix, and Factor Loadings for LIWC Variables

LIWC Variable	**Mean (*SD*)**	**Component 1** *(Connectedness/ Intimacy)*	**Component 2** *(Immediacy)*	**Component 3** *(Social Rumination)*	**Component 4** *(Negative Affect)*
Affiliation	5.31 (2.60)	0.75			
Family	1.30 (1.34)	0.68			
Social	12.61 (3.31)	0.66			
Drives	10.50 (3.31)	0.66			
Cognitive processes	15.67 (3.84)	-0.50			
Impersonal pronouns	5.00 (2.21)	-0.52			
Personal pronouns	15.39 (3.49)		0.62		
Verb	18.23 (3.06)		0.60		
Auxiliary verb	10.29 (2.43)		0.57		
1st person singular pronouns	12.58 (3.10)		0.56		
Focus present	14.96 (3.43)		0.54		
Relativity	11.29 (3.29)			0.60	
Focus past	2.33 (2.02)			0.56	
Time	4.69 (2.18)			0.54	
Positive emotion	4.55 (2.30)			-0.53	
Negative emotion	3.11 (1.90)				0.65
Affect	7.84 (2.84)			-0.56	0.62
Anger	0.77 (0.89)				0.54

Note. Mean values represent the mean percentage of total words used. Only LIWC measures with an absolute factor loading > 0.5 are presented.

The four components extracted are described (including snippets of quotes from participants' relationship essays as examples) as follows:

1) *Connectedness/Intimacy* – high socially connected and affiliated language (i.e., social, family, and affiliation words), and low cognitive processing language and impersonal pronouns (e.g., "My best friend and I are very alike, and we get along quite well. She's like my sister"; "I get on very well with my partner. We are best friends and lovers. He is my soul mate.").

2) *Immediacy* – high self-focused, present-tense (in the "here-and-now"), action-oriented (i.e., verbs) language, and personal pronouns (e.g., "I'm probably too attached to her if I'm being honest. I need to move out"; "I've got it so they [family members] won't call me unless it is something important for me, otherwise I will only call them when I need something or in the mood to talk").

3) *Social Rumination* – high relativity and time-oriented language, past-tense reflective language, and low positive emotion (e.g., "I used to try to make everyone laugh when I was little, which eventually put me in a position where I was treated as a joke by all my friends"; "My mother and father divorced when I was 7. My mom held me personally responsible for it and hated me for it.").

4) *Negative Affect* – high affective language, general negative emotion, and anger (e.g., "I do not love any other family members. I rather hate them"; "I constantly lose people in my life, they realise how awful I am as a person and leave. I've come to hate people.").

The model accounted for 25.60% of the total variance. Although this percentage may seem somewhat small with respect to traditional factor analyses, it is in alignment with typical factor analyses conducted using natural language data (see, e.g., Chung & Pennebaker, 2008).

Scores for the four social-cognitive components were generated for each participant and correlated with BPD features and Dark Triad scores using partial Pearson's correlations (two-tailed). Age and gender were controlled for in all correlation analyses due to their differential associations with language use and BPD features, as well as due to significant associations with the social-cognitive components (note that the overall findings remained the same when not controlling for age or gender

– see Appendix A.4). Although another potential confound, education level was not included as a control variable as it was not significantly associated with any of the four social-cognitive components. Analyses revealed that BPD features significantly correlated with all four social-cognitive components. Specifically, BPD features correlated negatively with *Connectedness/Intimacy* and positively with *Immediacy*, *Social Rumination*, and *Negative Affect* (see Table 3.2 for statistics).

Further, several social-cognitive components and Dark Triad traits were correlated at statistically significant levels. *Connectedness/Intimacy* correlated negatively with overall Dark Triad traits, Machiavellianism, and psychopathy, but not with narcissism. *Immediacy* correlated positively with psychopathy only. *Social Rumination* did not correlate with any of the Dark Triad traits. *Negative Affect* correlated positively with overall Dark Triad traits, Machiavellianism, and narcissism, but did not correlate with psychopathy. All correlation analysis results can be seen in Table 3.2.

Table 3.2

Correlations between Social-Cognitive Components and BPD Features and Dark Triad Traits

	Connectedness/ Intimacy	Immediacy	Social Rumination	Negative Affect
BPD Features (*n* = 483)	-.12**	.10*	.15**	.20***
Overall Dark Triad (*n* = 497)	-.10*	.08	.00	.11*
Machiavellianism (*n* = 497)	-.11*	.08	-.02	.13**
Narcissism (*n* = 497)	-.02	.00	-.03	.11*
Psychopathy (*n* = 497)	-.10*	.10*	.04	.03

***p < .001, **p < .01, *p < .05.
Note. All tests are two-tailed and include age and gender as control variables.

3.4 Discussion

The goal of the present study was to generate new insights into social-cognitive dimensions that characterize and contribute to interpersonal dysfunction in BPD, through analyzing natural language. The analysis of a large sample writing about their relationships with others revealed four core social-cognitive components that were related to BPD in intuitive ways: 1) *Connectedness/Intimacy*; 2) *Immediacy*; 3) *Social Rumination*; 4) *Negative Affect*. Several components were associated with both BPD and Dark Triad traits, while others appeared specific to BPD.

Of methodological interest, several dimensions found in the present study strongly overlap with those found in the classic Pennebaker and King (1999) study. Specifically, the *Immediacy* and *Social Rumination* dimensions show distinctive similarities to the "Immediacy" and "The Social Past" dimensions found by Pennebaker and King, including the language variables that comprise them. Such similarities highlight how language-based dimensions of thought can be somewhat reliably replicated across samples. Moreover, similarities between our findings and findings from general personality research suggest that such social-cognitive dimensions may characterize social (dys)function in personality disorder and in normative personality more broadly, thereby supporting the current consensus that personality disorders are dimensional in nature (e.g., Wilmot et al., 2019).

Findings that BPD features were associated with lower levels of intimacy (*Connectedness/Intimacy* component) and more negative affect, and anger in particular (*Negative Affect* component), in the discussion of social connections highlight how individuals manifesting BPD conceptualize relationships in a highly negative and disconnected way; in turn, such maladaptive mental representations of relationships likely contribute to their social dysfunction. These findings support the notion that affective dysregulation is a fundamental feature characterizing, and likely contributing to, interpersonal dysfunction in BPD (e.g., Lazarus et al., 2014). Yet, the same associations were also found with Dark Triad traits, implying that problems with intimacy and affect are important components characterizing social dysfunction in general, and not necessarily specific to BPD.

Interestingly, *Immediacy* (present-tense, action-orientated language) was solely associated with BPD and psychopathy, in that those with higher levels of BPD features and psychopathy scored higher on this dimension. Notably, the positive association with *Immediacy* appears to reflect the notion that relationships and social processes may be characterized by immediacy/urgency in individuals manifesting BPD (and psychopathy), such as seeking instantaneous gratification from one's social connections. The notion that social(-cognitive) processes in BPD are shaped by immediacy may help to explain the problems with intimacy associated with BPD, as the highly instantaneous social-interactive nature of individuals with BPD would likely make it difficult for them to form longstanding relationships characterized by intimacy and connection. Such immediacy is also intuitively linked to the high impulsivity associated with BPD (and psychopathy), thereby providing further indication that impulsivity may be a characterizing dimension of interpersonal dysfunction in BPD (Euler et al., 2019), as well as other severe problematic interpersonal constructs. From a clinical perspective, severe interpersonal dysfunction could, then, be potentially addressed by targeting the immediacy/impulsivity that characterizes maladaptive social-cognitive processes through therapeutic intervention.

Importantly, the *Social Rumination* dimension (time-orientated, past-tense, non-positive language) was found to be associated with BPD exclusively, with people with higher levels of BPD features scoring higher on this dimension; a novel finding regarding the central themes characterizing interpersonal dysfunction in BPD. It is likely that this dimension reflects the (negative) past-orientated nature of individuals with BPD (Miano et al., 2020), as well as how such individuals may have difficulty developing and maintaining healthy new relationships due to being stuck processing past relationships, events, and trauma. Vitally, as *Social Rumination* was revealed to be exclusively related to BPD (when compared to Dark Triad traits), it may be that this component distinguishes interpersonal dysfunction in BPD from other problematic interpersonal traits. Yet, it is highly likely that this notion of being 'stuck' processing past (negative) events/relationships (i.e., social rumination) may also be characteristic of other personality disorder types characterized by rumination, that were not assessed in the present study, such as obsessive-compulsive PD and avoidant PD; further research is needed to clarify distinctions across types of personality pathology.

Nevertheless, in individuals who experience longstanding, persistent patterns of severe interpersonal dysfunction, this notion of being ‘stuck in the past’ could potentially be the predominant mechanism *prolonging* such interpersonal dysfunction across time and contexts. Thus, the treatment of longstanding patterns of social dysfunction may benefit from focusing on individuals’ past relationships and experiences, with the eventual aim of shifting their focus from past traumatic relationships to developing new healthy relationships. Indeed, Schema Therapy (ST) – an effective therapeutic treatment for BPD (Sempértegui et al., 2013) – somewhat incorporates this notion of identifying and shifting the focus from past traumatic experiences to improve functioning. However, ST primarily focuses on improving general cognitive functioning, with less emphasis placed on past relationships and interpersonal functioning specifically. Adapting the focus of ST to place more emphasis on improving social functioning may prove effective in treating longstanding patterns of social dysfunction.

Altogether, the present study has allowed for the discovery of social-cognitive dimensions that help to better understand social dysfunction in BPD through the analysis of natural language. Nonetheless, the study is not without limitations. One potential limitation of the present research surrounds the fact that the study did not comprise a clinical sample, and so we cannot be certain as to whether our findings would replicate in those with clinical BPD diagnoses. A further limitation surrounds the use of self-report measures of problematic personality constructs, which are subject to various biases. In particular, there are established shortcomings surrounding assessment methods of the Dark Triad, as well as the entire construct itself (see Miller et al., 2019). Although the purpose of the Dark Triad measure used in the present study was solely to capture social dysfunction that can be differentiated from BPD, we acknowledge that other constructs could have been assessed that are more closely related to BPD (e.g., other personality disorder traits). Finally, the correlational nature of the data means that causality cannot be inferred from the present findings.

It would be valuable for future research to attempt to replicate our findings in a clinical sample to see whether the social-cognitive components extend to describing interpersonal dysfunction in individuals with a BPD diagnosis, and those with other mental health conditions. Moreover, the linguistic factor analytic approach adopted in the present study is just one of many possible approaches for better understanding social dysfunction in BPD; there are numerous techniques for analyzing qualitative data that have the potential to shed further light into the nature of BPD, attenuating the need for self-report methods (e.g., methods that assess the conceptual level of individuals' verbal behavior).

To conclude, through the analysis of natural language, the present study uncovered four social-cognitive components all related to BPD: *Connectedness/Intimacy*; *Immediacy; Social Rumination;* and *Negative Affect*. Problems with intimacy and affect appear to characterize interpersonal dysfunction across a range of constructs, whereas immediacy and social rumination are more specific to BPD. Notably, our findings suggest that social rumination may distinguish interpersonal dysfunction in BPD from other problematic interpersonal constructs. Computational analysis of natural language has therefore allowed for the identification of fundamental social-cognitive components that provide novel insights into the nature of interpersonal dysfunction in BPD. Our findings provide paths to new research questions surrounding the origins, trajectory, and treatment options for BPD.

CHAPTER 4:

Natural Emotion Vocabularies and Borderline Personality Disorder

Charlotte Entwistle, Andrea B. Horn, Tabea Meier, Katie Hoemann, Annemarie Miano, Ryan L. Boyd [4]

Text as it appears in:

Entwistle, C., Horn, A. B., Meier, T., Hoemann, K., Miano, A., & Boyd, R. L. (2023). Natural emotion vocabularies and borderline personality disorder. *Journal of Affective Disorders Reports*, *14*, 100647. https://doi.org/10.1016/j.jadr.2023.100647 [5]

Abstract

Background

[4] The authors would like to thank Zilla M. Huber, Olenka Dworakowski, and Rebecca Oertel for their valuable work in transcribing the couples' conversations from the video recordings.

[5] Note that the terms "Study 1" and "Study 2" used throughout this manuscript refer to Studies 2 and 3, respectively, in this thesis.

Emotion dysregulation is a characteristic central to borderline personality disorder (BPD). Valuably, verbal behaviour can provide a unique perspective for studying emotion dysregulation in BPD, with recent research suggesting that the varieties of emotion words one actively uses (i.e., active emotion vocabularies [EVs]) reflect habitual experience and potential dysregulation therein. Accordingly, the present research examined associations between BPD and active EVs across two studies.

Methods

Study 1 (*N* = 530) comprised a large non-clinical sample recruited from online forums, whereby BPD traits were measured via self-report. Study 2 (*N* = 64 couples) consisted of mixed-gender romantic couples in which the woman had a BPD diagnosis, as well as a control group of couples. In both studies, participants' verbal behaviours were analysed to calculate their active EVs.

Results

Results from both studies revealed BPD to be associated with larger negative EV (i.e., using a broad variation of unique negative emotion words), which remained robust when controlling for general vocabulary size and negative affect word frequency in Study 2. The association between BPD and negative EV was insensitive to context.

Limitations

Limitations of this research include: 1) the absence of a clinical control group; 2) typical constraints surrounding word-counting approaches; and 3) the cross-sectional design (causality cannot be inferred).

Conclusions

Our findings contribute to BPD theory as well as the broader language and emotion literature. Importantly, these findings provide new insight into how individuals manifesting BPD attend to and represent their emotional experiences, which could be used to inform clinical practice.

Keywords: borderline personality disorder, personality pathology, emotion dysregulation, emotion processes, language analysis, emotion vocabularies

Highlights

- BPD is associated with actively using a diverse range of words for negative emotion.
- The association between BPD and negative emotion word use is insensitive to context.
- These findings likely reflect extensive experience with negative emotion.
- Extensive but inflexible attention to negative emotion may drive dysfunction in BPD.

4.1 Introduction

Borderline personality disorder (BPD) is a severe mental health condition generally characterised by longstanding patterns of dysregulated emotional functioning, problematic interpersonal relationships, and disturbed identity (APA, 2013). Further, BPD is a problematically heterogeneous construct (e.g., Cavelti et al., 2021) and is highly comorbid with various other mental health conditions (e.g., Shah & Zanarini, 2018), prompting some scholars to conceptualise BPD as reflective of "general psychopathology" (for empirical evidence, see, e.g., Gluschkoff et al., 2021; Sharp et al., 2015; Wright et al., 2016). However BPD is conceptualised, emotion dysregulation remains a defining and critical feature (e.g., Crowell et al., 2009).

Emotion dysregulation in BPD is thought to consist of four components: 1) emotion sensitivity, 2) heightened and variable negative affect, 3) a deficit of appropriate emotion regulation strategies, and 4) a reliance on maladaptive regulation strategies (e.g., self-harm; Carpenter & Trull, 2013). Further, prominent clinical theories suggest that the way in which individuals manifesting BPD understand and attend to their negative emotions moderates the impact of these emotions, through either disrupting (positive impact) or driving (negative impact) patterns of emotional cascades or negative rumination (e.g., Beck et al., 1979; Linehan, 1993).

Despite a large literature discussing emotion dysregulation in BPD, there remain foundational gaps in the understanding of such dysfunction, particularly surrounding the emotion processes and experiences themselves. Valuably, advances in affective theory and methods provide researchers with a unique opportunity to address these gaps in knowledge by utilising verbal behaviour to access patterns of attention to, and representation of, emotion (Vine et al., 2020). Examining the words individuals use is an unobtrusive way of quantifying how people think about and formulate their emotional experiences (e.g., Pennebaker, 2011). Moreover, the analysis of emotion words has also been used to assess the extent to which individuals refer to specific emotions, and is sometimes employed as a measure of emotion differentiation (i.e., the ability to experience distinct emotions; e.g., Williams & Uliaszek, 2022). When combined with natural language data, automated analyses of verbal behaviour can provide insight into emotion (dys)regulation in the real world and in context.

In the following sections, we briefly review the relationship between BPD and emotion differentiation, highlight the value of language in studying emotion, and present a recently developed method for quantifying emotion words. We then introduce how this method is applied in the present research to better understand emotion dysregulation in BPD.

4.1.1 BPD and Low Emotion Differentiation

Emotion differentiation refers to one's ability to experience distinct and nuanced emotions, and is thought to reflect accrued knowledge or concepts for emotion (Barrett et al., 2001; Hoemann et al., 2023). In theory, higher emotion differentiation should be associated with better emotion regulation because experiencing specific emotions facilitates a person's ability to enact emotion-specific regulation strategies in a context-specific manner (e.g., Southward et al., 2019). Considerable research supports this assumption (e.g., Kalokerinos et al., 2019), with lower emotion differentiation likewise associated with emotion regulation difficulties (for a review, see Seah & Coifman, 2021).

The central role of emotion dysregulation in BPD suggests that low emotion differentiation would be associated with the construct. Again, previous research provides ample support for this association (e.g., Derks et al., 2017; Fitzpatrick et al., 2019; Suvak et al., 2011). Moreover, lower emotion differentiation predicts greater engagement in maladaptive behaviours in individuals with BPD (Dixon-Gordon et al., 2014), including non-suicidal self-injury (Zaki et al., 2013).

4.1.2 Emotion (Dys)regulation and Natural Language

Despite such compelling findings, the current emotion differentiation literature may only tell part of the story. Emotion differentiation is typically assessed by asking participants to repeatedly rate their momentary experience on a set of pre-specified emotion terms (for review, see Thompson et al., 2021). This approach generates a measure of how pre-selected emotion concepts are implemented but may not reflect how people spontaneously (or typically) represent affective experiences in everyday life. The terms pre-specified by researchers may fail to capture certain types of emotions and, in any case, explicitly prompt participants to attend to their experiences in ways that they might not otherwise do (for discussion, see Li et al., 2020; Vine et al., 2020). For these reasons, researchers have explored natural language as an ecologically valid and minimally intrusive means of capturing the experience and expression of emotion *in-situ* (e.g., Williams & Uliaszek, 2022).

Research using verbal behaviour to better understand emotion dysregulation is somewhat scarce. One prominent approach has involved the examination of emotion word frequencies. In this work, researchers employ word-counting software (e.g., Linguistic Inquiry and Word Count; Pennebaker et al., 2015) to identify the proportion of affectively-laden words in a given text, reflecting general attention to emotion (see Boyd & Schwartz, 2021). In line with the conceptualisation of emotion dysregulation in BPD, research examining emotion word frequencies has consistently demonstrated that people manifesting BPD typically use a high frequency of negative affect words, and anger words in particular (e.g., Coppersmith et al., 2015; Lyons et al., 2018), as well as relatively fewer positive affect words (e.g., Rosenbach & Renneberg, 2015). Although such research provides a useful starting point for exploring emotion processes in BPD using naturalistic language-based methods, there remains considerable room for more in-depth investigation that goes beyond simply exploring general emotion word frequencies.

4.1.3 Active Emotion Vocabularies

Recently, Vine and colleagues (2020) introduced a new method for quantifying emotion language; namely, the measurement of active emotion vocabularies (EVs). This approach captures the *variety* of emotion words used spontaneously (i.e., without prompting from researchers); a greater variety of words used to refer to a particular emotion would indicate a larger active EV for that emotion concept. Based on linguistic theory (see, e.g., Pennebaker, 2011; Zipf, 1949), the emotion words one spontaneously uses should generally correspond with one's typical or frequent, salient experiences. Smaller positive EV (reflecting less experience with positive emotion) and larger negative EV (reflecting more experience with negative emotion) should, then, be associated with poorer emotional functioning. To test this hypothesis, Vine and colleagues (2020) conducted two studies comprising different general population samples. Findings revealed that, in general, larger negative EVs were associated with indicators of poorer physical and psychosocial health, whereas larger positive EVs were associated with indicators of better physical and psychosocial health.

The findings from Vine et al. (2020) provide initial empirical support for linguistic theory suggesting that active EVs should correspond with individuals' typical emotional experiences. Extending the analysis of active EVs to BPD (and any other clinical group) has the potential to provide new insight into the inner emotional world of individuals with BPD and the emotion dysregulation that is central to the disorder.

4.1.4 Current Research

In the current investigation, we conducted two studies – comprising different samples and types of language modality – to examine active EVs expressed in natural language and investigate how they relate to BPD. Study 1 comprised written essays from participants recruited from online forums, enabling us to investigate the relationship between emotion-relevant verbal behaviour and BPD traits in a large non-clinical sample. Study 2 consisted of spoken interactions between romantic partners in a clinical BPD sample, as well as a comparison group of control (i.e., non-clinical) couples.[6] Data from both studies enabled us to examine emotion vocabularies in BPD in the context of close relationships, where emotion dysregulation is especially consequential and likely to be more prominent (e.g., Hill et al., 2008).

[6] See https://osf.io/3j7mk/?view_only=ea2d77ad7244420bbabfbc8eec4f710e for data for both studies.

The current research was driven by two main goals: 1) investigate the relationship between BPD and active EVs, and 2) explore whether associations between BPD and active EVs vary depending on the context, as an indicator of their context-sensitivity. Based on linguistic theory (e.g., Zipf, 1949) and findings from Vine et al. (2020), we hypothesised that BPD would be associated with relatively larger negative EV and smaller positive EV. Further, we hypothesised that the associations with negative EV would be stronger and more robust than the associations with positive EV due to research evidencing that emotion dysregulation in BPD is most prominently distinguished by heightened negative emotion (e.g., Chu et al., 2016). Given the lack of research on the stability of active EVs across time and context, our second research aim was exploratory in nature.

4.2 Study 1

4.2.1 Method

Data for Study 1 were collected as part of a larger investigation on the associations between natural language and various psychological and personality processes, including BPD traits. This study was approved by the Faculty of Science and Technology Research Ethics Committee (FSTREC) at Lancaster University.

4.2.1.1 Participants and Procedure

Participants were recruited for the study via targeted sampling from various online forums (all approved by forum moderators). In particular, advertisement of the study involved online distribution of anonymous links to the Qualtrics study information sheet. Both general discussion forums and mental health forums were targeted for recruitment, with the aim of enhancing sample diversity in mental health status. Following consent procedures, participants were presented with several questions that measured their sociodemographic characteristics (e.g., age, gender). The remaining questions in the study assessed various psychological, social, and personality processes, with the order of all questions, questionnaire items, and writing prompts randomised between participants.

The only inclusion criteria for the study included the ability to write or speak in fluent English and being a minimum of 18 years of age. There were no exclusion criteria relating to mental health conditions; we wanted to allow the sample to be as (psychologically) diverse and inclusive as possible. No incentives were offered for participation in the study. Participants who did not provide sufficient data for key measures – that is, those who did not provide any responses to the BPD measure or whose relationship essays did not meet the minimum word count criteria (see below) – were removed from the dataset (*N* = 137), resulting in a total of 530 participants (see Table 4.1 for sociodemographic characteristics).

Table 4.1

Sociodemographic Characteristics of Participants in Study 1 (N = 530)

Characteristic	Mean	*SD*
Age (*n* = 528)	26.22	8.41
	n	%
Gender (*n* = 520)		
Female	379	72.88
Male	126	24.23
Non-binary	15	2.88
Ethnicity (*n* = 521)		
Asian	40	7.68
Black	10	1.92
Hispanic or Latino	37	7.10
Mixed	34	6.53
White	394	75.62
Other	6	1.15
Marital Status (*n* = 521)		
Single	268	51.44
Married/partnered	237	45.49
Divorced/separated	16	3.07
Education Level (*n* = 522)		
Less than high school	17	3.26

High school/some college	260	49.81
College	74	14.18
University/postgraduate degree	171	32.76
Employment Status (*n* = 524)		
Unemployed	117	22.33
Student	172	32.82
Employed	213	40.65
Self-employed	16	3.05
Retired	6	1.15

Note. Differences in *n*s between the various demographic measures reflect the data provided by participants, as all participants (i.e., *n* = 530) did not provide responses to all questionnaire measures, hence the differing *n*s (e.g., more participants provided data for age than for gender).

4.2.1.2 Measures

Borderline Pathology Features. BPD features were assessed using the Personality Assessment Inventory-Borderline Scale (PAI-BOR; Morey, 1991). The PAI-BOR is a 24-item questionnaire that assesses 4 core features of BPD: affective instability (6 items; α = .78), identity problems (6 items; α = .70), social dysfunction (6 items; α = .62), and self-harm (6 items; α = .80). Responses are measured through a 4-point response scale ranging from 0 (false) to 3 (very true); total PAI-BOR scores can range from 0 to 84. The average total PAI-BOR score for our sample was 37.24 (*SD* = 13.03; α = .88). A total PAI-BOR score of 38 or more is proposed to indicate the presence of "significant BPD features", whereas a score of 60 or more indicates typical borderline pathology (i.e., clinically significant levels; Morey, 1991). While our sample centres around PAI-BOR scores indicative of the presence of BPD traits in the general population (i.e., a score of 38), only a very small portion of the sample (less than 4%) reach clinically significant levels of BPD (i.e., scores of 60+).

Relationship Essays. Participants were asked to write in a spontaneous, 'stream of consciousness' fashion about their relationships with other people, to capture rich data on participants' psychology around the topic of interpersonal relationships. The prompt read as follows:

> *When you think about your relationships with other people, what comes to mind? For the next 7 minutes (or more), we would like for you to write about how you get along with people. This can include your relationships with co-workers, family, friends, and romantic partners. Try to say as much as you can about both the good and the bad. Do not worry about spelling or grammar. Simply write everything that comes to mind, giving as much detail as possible. Once you begin writing, try to write continuously until you have finished. If you run out of things to say, re-tell what you have previously said in other words.*

Everyday Behaviour Essays. To collect natural language data as a more general comparison, participants were also prompted to write about their daily behaviours over the past seven days. The prompt presented was slightly modified from that used in a previous study relating to everyday behaviour and values (Boyd et al., 2015). Specifically, the prompt read:

> *"For the next 7 minutes (or more), write about everything that you have done in the past 7 days. For example, your activities might be simple, day-to-day types of behaviors (such as eating dinner with your family, making your bed, writing an e-mail, and going to work). Your activities in the past week might also include things that you do regularly, but not necessarily every day (such as going to church, playing a sport, writing a paper, having a romantic evening) or even rare activities (such as skydiving, taking a trip to a new place). Try to recall each activity that you have engaged in, starting a week ago and moving to the present moment. Be specific. Once you begin writing, try to write continuously until you have finished."*

4.2.1.3 Pre-Processing and Language Analysis

Participants' written essays were corrected for common misspellings and idiosyncrasies prior to analysis, and all texts containing fewer than 50 words were excluded from subsequent analysis to ensure reliability of language analysis and validity of measurement (see, e.g., Boyd, 2017; Cutler et al., 2021). Participants wrote an average of 211.60 words (*SD* = 186.22) for the relationship essays and 185.79 words (*SD* = 107.05) for the everyday behaviour essays.

Following pre-processing procedures, participants' active EVs were computed from their language using the same methodology as in Vine et al. (2020). Specifically, we used BUTTER (Boyd, 2020) – a text analysis software for the social sciences – to calculate active EVs. For an overview of the methodology behind the automated program, active EVs are quantified by counting the frequency of unique emotion words (e.g., "sad", "depressed") to describe an emotion concept (e.g., sadness) in a given text. As an example, the sentence "I'm so *disappointed* – he made me feel very *sad* and *upset*" illustrates a larger negative EV than the sentence "I'm so *upset* – I can't believe this, it was so *upsetting*, he made me feel so *upset*", despite displaying the same negative emotion word frequency. Active EV scores generated therefore reflect the percentage of unique emotion words relative to the total word count, rather than simply reflecting general emotion word frequencies. Using pre-determined word-mappings (see Vine et al., 2020), the software calculates active EV scores for positive and negative emotion as well as for specific negative emotions nested under the overall negative emotion category, namely: sadness, anxiety/fear, anger, and undifferentiated negative emotion (reflecting stress). Additionally, individuals' general vocabulary size (i.e., the diversity of unique words used in general) is also computed and was included as a control variable in the subsequent EV analyses, as described in Vine et al. (2020).

In EV analyses, it is also informative to control for overall emotion word frequencies to determine whether the results generated directly reflect the *diversity* of one's emotion word use, as opposed to simply being a function of the overall frequency. Accordingly, to generate emotion word frequency scores to be controlled for, we used the word-counting program Linguistic Inquiry and Word Count (LIWC2015; Pennebaker et al., 2015). LIWC is an extensively validated word counting program that uses an internal dictionary to calculate the percentage of words belonging to psychologically meaningful dimensions (e.g., affective processes; social processes) in a given text. Specifically, we used LIWC to measure the frequency of overall positive and negative affect words, computed as percentages of the total word count.

4.2.2 Results

4.2.2.1 Data Analysis

To test our hypotheses that BPD would be associated with larger negative EV and smaller positive EV, we first examined associations between BPD features and positive and negative EVs via two-tailed, bivariate Pearson's correlations. We also correlated BPD features with the negative EV subtypes (i.e., anxiety/fear, anger, sadness, and undifferentiated negative EV) as follow-up specificity tests. To determine the robustness of associations found between BPD and active EVs, we then conducted linear regressions, where we included key control variables as covariates. Specifically, in all regression models, total BPD feature scores were entered as the outcome variable and active EVs, general vocabulary size, and the corresponding emotion word frequency (derived from LIWC) were added as predictors. Although another potential confound, education level was not included as a control variable as it was not significantly associated with positive or negative EV (or any of the negative EV subtypes). Separate regression models were conducted for positive and negative EV. Analyses were performed in the same way on data from both the relationship and everyday behaviour essays.

4.2.2.2 Descriptive Analyses

In terms of the overall emotion word frequencies in participants' relationship essays, using LIWC2015, it was found that an average of 7.84% (*SD* = 2.84) of the words used were of emotional content, including both negative (*M* = 3.11, *SD* = 1.90) and positive emotion (*M* = 4.55, *SD* = 2.30). With regard to unique emotion words used, or active emotion vocabulary scores, the average positive EV was 0.59 (*SD* = 0.56) and average negative EV was 0.80 (*SD* = 0.80). Each EV was significantly associated with the corresponding emotion word frequency (all *p*'s < .001). General vocabulary size positively correlated with negative EV (*r* = .10, *p* = .020), but was not significantly associated with positive EV.

As for the everyday behaviour essays, an average of 3.44% (*SD* = 2.27) of the words used were of emotional content, composed of negative (*M* = 1.47, *SD* = 1.60) and positive emotion (*M* = 1.92, *SD* = 1.37). With regard to active EVs, the average positive EV was 0.17 (*SD* = 0.35) and average negative EV was 0.32 (*SD* = 0.52). Each EV was again significantly associated with the corresponding emotion word frequency (all *p*'s <

.001). General vocabulary size was not found to correlate significantly with negative or positive EVs in the behaviour essays.

4.2.2.3 Research Aim 1: Examining the Relationship Between BPD and Active EVs

Table 4.2 illustrates the correlation coefficients between BPD features and active EVs derived from the relationship essays. In general, results showed that BPD features were positively associated with negative EV and negatively associated with positive EV. Follow-up analyses revealed that the association with overall negative EV was primarily driven by anxiety/fear and anger EV. That is, anxiety/fear EV was marginally associated with total BPD feature scores (*r* = .09, *p* = .055) and significantly associated with BPD social feature scores (*r* = .12, *p* = .009); Anger EV was significantly associated with total BPD feature scores (*r* = .09, *p* = .048), and BPD affect (*r* = .12, *p* = .008) and identity (*r* = .10, *p* = .030) feature scores.

Table 4.2

Pearson Correlations between BPD Features and Emotion Vocabularies in Relationship Essays

	Positive EV	Negative EV
BPD Total (*n* = 498)	-.10*	.12*
Affect (*n* = 515)	-.10*	.11*
Identity (*n* = 510)	-.08†	.13**
Social (*n* = 513)	-.07	.13**
Self-harm (*n* = 516)	-.06	-.01

***p* < .01, **p* < .05, †*p* < .10.

Note. All tests are two-tailed. The *n*s reflect the PAI-BOR data provided by participants (from a total of *n* = 530). *N*s vary by subscale due to some participants responding to all items in some subscales but not in others; participant data were only included in the analysis if they provided responses for every item in the subscale, hence the differing ns. The *n* for the total score reflects the number of participants that provided responses for all 24 items.

Follow-up linear regression analyses revealed that, when controlling for general vocabulary size and corresponding emotion word frequencies, active EVs were not found to significantly predict BPD features. Regression coefficients for each of the covariates, in each of the models, are presented in Appendix B.1 (Table B.1).

4.2.2.4 Research Aim 2: Exploring the Context-Dependency of Associations Between BPD and Active EVs

Table 4.3 shows the correlation coefficients between BPD features and active EVs derived from the everyday behaviour essays. Correlation analyses uncovered that BPD features were positively associated with negative EV, but were not significantly associated with positive EV. Follow-up analyses showed that the association with negative EV was primarily driven by anger EV, as anger EV was significantly correlated with total BPD feature scores (*r* = .12, *p* = .023) as well as BPD affect (*r* = .11, *p* = .035) and identity (*r* = .18, *p* < .001) feature scores.

Table 4.3

Pearson Correlations between BPD Features and Emotion Vocabularies in Behaviour Essays

	Positive EV	Negative EV
BPD Total (*n* = 387)	-.05	.13*
Affect (*n* = 400)	-.08	.12*
Identity (*n* = 397)	-.06	.15**
Social (*n* = 398)	-.04	.09†
Self-harm (*n* = 400)	-.04	.04

***p* < .01, **p* < .05, †*p* < .10.

Note. All tests are two-tailed. The *n*s reflect the PAI-BOR data provided by participants (from a total of *n* = 530). *N*s vary by subscale due to some participants responding to all items in some subscales but not in others; participant data were only included in the analysis if they provided responses for every item in the subscale, hence the differing *n*s. The *n* for the total score reflects the number of participants that provided responses for all 24 items. The *n*s reported in this table (i.e., behaviour essays) are considerably smaller than in Table 4.2 (i.e., relationship essays) due to more participants completing and providing sufficient language data (i.e., ≥ 50 words) for the relationship essays than the behaviour essays.

As in the relationship essays, when controlling for general vocabulary size and corresponding emotion word frequencies in the follow-up linear regressions, active EVs did not significantly predict BPD features. The lack of significant associations between negative EV and BPD in these models can be explained as an effect of accounting for overall negative affect word frequencies, as the use of negative affect words in general significantly positively predicted BPD features in all regression models. Detailed regression results for the everyday behaviour essays are presented Appendix B.1 (Table B.2).

4.2.3 Discussion

Analyses of participant writing provided general support for our hypotheses that people with higher levels of BPD traits would exhibit larger negative EV and smaller positive EV – although the latter relationship was not evidenced in the everyday behaviour essays. The associations between BPD features and active EVs were not robust when controlling for general vocabulary size and the overall frequency of positive and negative affect words, which, for negative EV, appeared to be a direct result of accounting for negative affect word frequencies across both essay topics. The overlap of patterns found across the different topics provides an initial indication that the relationship between BPD features and active EVs may be insensitive to context, although more evidence is needed to confirm this. In Study 2, we sought to extend this investigation to a clinical population, while providing further clarity on the context-dependency of the associations between BPD and active EVs.

4.3 Study 2

4.3.1 Method

Study 2 was a secondary analysis of data previously collected for purposes unrelated to the present study's aims (i.e., the 'couple communication study'; see, e.g., Miano et al., 2017a, 2017b). For a brief overview, the couple communication study investigated various domains of social cognition, interpersonal functioning, affect, and behaviour in BPD, with the broad aim of providing greater understanding of interpersonal dysfunction in BPD. In this article, we only describe the study methods that are directly relevant to the present investigation (for a detailed description, including information relating to consent and ethics, see, e.g., Miano et al., 2017a, 2017b).

4.3.1.1 Participants

Participants (recruited in Germany) were mixed-gender romantic couples in which female partners were either diagnosed with BPD or did not have a clinical diagnosis (i.e., control couples). Participating couples were eligible for the study if they had been in their current relationship for at least three months and were not married or engaged. The final sample of participants comprised 64 couples in total – 30 couples in the BPD group and 34 control couples. Full inclusion/exclusion criteria and participant characteristics are described in detail in the above-referenced studies.

4.3.1.2 Measures

Borderline Pathology Symptoms. The short version of the Borderline Symptom List (BSL-23; Bohus et al., 2009) was used to assess borderline symptom severity. The BSL short-version is a 23-item self-report measure that assesses BPD symptomology severity using a 5-point Likert scale, with responses ranging from 0 to 4. Higher scores indicate greater borderline symptom severity within the last week. In the present sample, the average BSL score was 1.69 (*SD* = 0.67; α = .98).

Depressive Symptoms. The German version of the Beck Depression Inventory (BDI-II; Hautzinger et al., 2006) was used to measure the presence and severity of depressive symptoms. The BDI-II is a self-report measure composed of 21 items in total, with items rated on a 4-point Likert scale ranging from 0 to 3. Higher scores indicate greater severity of depressive symptoms within the last 2 weeks. The German version of the BDI-II has been well-validated (e.g., Kuhner et al., 2007). The average total BDI-II score in the present sample was 13.55 (*SD* = 15.17; α = .96).

4.3.1.3 Procedure

Following the verification of inclusion criteria, participants completed an online questionnaire in which a range of socio-psychological variables were assessed. Couples were then invited into the laboratory where they engaged in three different conversations with one another for six minutes each while being video recorded. In the first condition *(neutral condition)*, participants were asked to discuss their favourite film genre, which was designed as a non-emotive discussion topic. In the other two conditions, participants were asked to discuss topics of a more negative emotive nature, designed to induce feelings of threat and stress. Specifically, participants were asked to discuss a fear that was most relevant to them during the past year *(personally-threatening condition)* and plausible factors that could result in the couple ending their current relationship *(relationship-threatening condition)* – a situation that is likely to be particularly emotionally difficult for individuals with BPD. Following each conversation, couples separately completed a questionnaire – including a threat manipulation check – and prepared for the next conversation topic.

4.3.1.4 Pre-Processing and Language Analysis

Couples’ conversations were transcribed from the video recordings by trained research assistants in their original German language and translated to English for subsequent language analysis (for comparability with Study 1). Texts were translated to English using machine translation followed by manual inspection (see, e.g., Li et al., 2014; Windsor et al., 2019). Language data in the form of transcribed conversations were then separated by speaker (i.e., female and male partners) and pre-processed and analysed in exactly the same way as in Study 1. In total, language data were obtained from 128 individuals (64 couples) who each had three separate conversations with their partner, resulting in 384 individual texts to be analysed.

4.3.2 Results

4.3.2.1 Data Analysis

We primarily report the analyses conducted on the texts from the female partners in the main manuscript, as only the women had BPD diagnoses in the present sample, and we wanted to ensure a conceptually accurate analysis (and results). To address our first aim of examining the relationship between BPD and active EVs, we compared the active EVs of women with BPD and women without BPD (N = 64) via independent, two-tailed *t*-tests, with group (BPD vs. non-BPD) as the independent variable and the EVs as dependent variables. To test the robustness of any differences in EVs found between groups, we also conducted univariate ANCOVAs, controlling for (as in Study 1) general vocabulary size and the corresponding emotion word frequency (derived from LIWC). We also provide a statistical comparison of EVs between male partners (i.e., partners of women with BPD versus women without BPD) in Appendix B.4, to corroborate that the differences found between women with BPD and women without BPD are predominantly a result of personality pathology.

In addition to the group comparison (i.e., categorical) analytic approach, we adopted a dimensional approach, given that this is more in alignment with now widely supported contemporary models of dimensional psychopathology (Dalgleish et al., 2020), and borderline pathology in particular (e.g., Wright et al., 2016). Accordingly, we conducted the exact same analyses as in Study 1 (i.e., Pearson's correlations and linear regression analyses), using the total BPD symptom scores for all women in the sample. Note that we only conducted these dimensional analyses on the women in the sample to control for non-independence of data between partners.

To address our second aim of exploring the context-dependency of active EVs, we conducted 2 (group: BPD vs. non-BPD) x 3 (condition: neutral vs. personally-threatening vs. relationship-threatening) mixed ANCOVAs to examine differences in EVs between (across groups) and within (across conditions) participants. In all 2x3 ANCOVAs, "Group" represented the between-participants fixed factor and "Condition" the repeated measures variable, controlling for general vocabulary size and the corresponding emotion word frequency.

To test the robustness of findings, we conducted post-hoc analyses whereby we adopted a dyadic analytic perspective, following a well-established approach for analysing dyadic interactions that takes the dyadic interdependence of data into account (see Iida et al., 2018; Kenny et al., 2020). Specifically, we examined whether and how dyadic patterns in EVs differ between BPD and non-BPD groups. Refer to Appendix B.5 for all post-hoc dyadic analysis results.

4.3.2.2 Descriptive Analyses

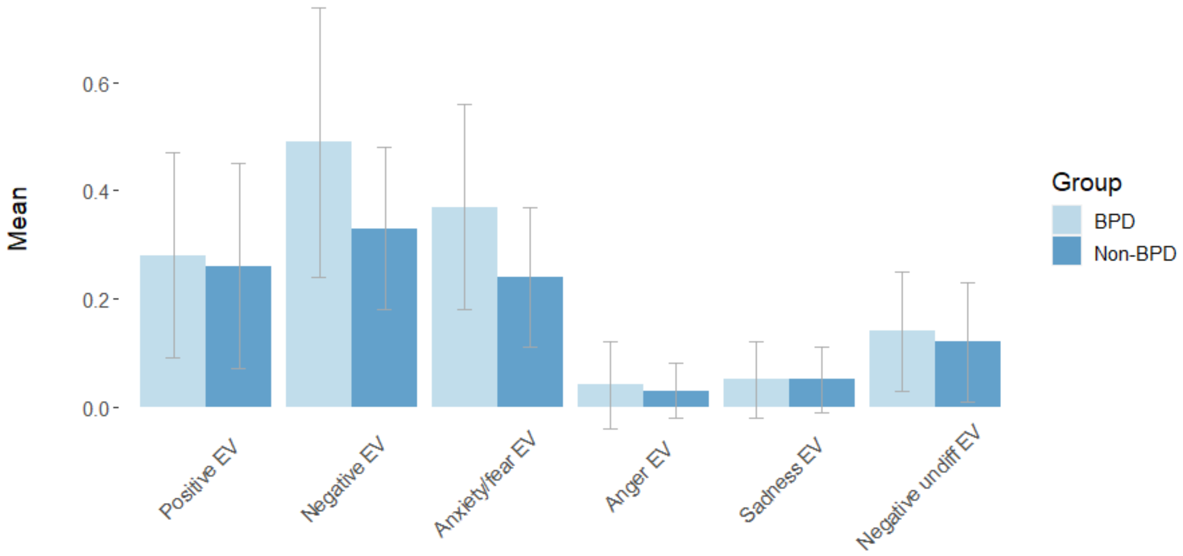
Calculated using LIWC, an average of 1.33% (*SD* = 0.48) of the women's language in the conversations were of emotive nature, including both negative (*M* = 0.80, *SD* = 0.40) and positive emotion (*M* = 0.53, *SD* = 0.28). In terms of the number of unique emotion words used, the mean positive EV was 0.27 (*SD* = 0.19) and mean negative EV was 0.41 (*SD* = 0.22). In the present sample, active EVs did not significantly correlate with corresponding emotion word frequencies. General vocabulary size correlated positively with both negative EV (*r* = .35, *p* = .005) and positive EV (*r* = .36, *p* = .004). Interestingly, paired *t*-tests revealed no significant differences in emotion word frequencies, active EVs, or general vocabulary size between female and male partners within each couple, generalised across the three conditions (see Appendix B.4 for more detailed results).

4.3.2.3 Research Aim 1: Examining the Relationship Between BPD and Active EVs

Figure 4.1 presents a comparison of average active EVs across all conditions between women with BPD and women without BPD. Independent *t*-tests revealed that women with BPD had considerably larger negative EVs (*t*(45) = -2.97, *p* = .005, *d* = -0.77) than women without BPD, but there was no difference in positive EV between these groups. Follow-up analyses revealed that group differences in negative EV were predominately a result of significantly larger anxiety/fear EVs among women with BPD (*t*(62) = -3.44, *p* = .001, *d* = -0.86).

Figure 4.1

Mean Emotion Vocabularies of Women with BPD Versus Women without BPD (N = 64)



Note. Error bars represent standard deviations.

Results from the univariate ANCOVAs controlling for general vocabulary size and corresponding emotion word frequencies confirmed, and thus established the robustness of, the group differences in EVs. The difference between women with BPD and women without BPD in negative EV remained significant when accounting for the control variables (results presented in Table 4.4), which was again driven by anxiety/fear EV (F(1,60 = 10.12, p = .002, np^2 = .14). The post-hoc analyses that take into account the non-independence of data also corroborate the results presented here (see Appendix B.4 and B.5).

Table 4.4

Differences in Emotion Vocabularies (EVs) Between Women with BPD and Women without BPD, Controlling for General Vocabulary and Emotion Word Frequencies (N = 64)

EV	Mean (*SD*)		*F*	*p*	*np*²	95% CI
	BPD (*N* = 30)	Non-BPD (*N* = 34)				
Positive EV	0.28 (0.19)	0.26 (0.19)	0.59	.446	.01	-.06 – .13
Negative EV	0.49 (0.25)	0.33 (0.15)	7.32	.009	.11	.04 – .23

Note. CI = confidence interval.

When adopting the dimensional analytic approach described above (in the Data Analysis section), the results show the exact same patterns as in the categorical analysis. Specifically, Pearson's correlations between women's total BPD symptoms and active EVs revealed that BPD symptoms were positively correlated with negative EV (*r* = .32, *p* = .009), with no significant association with positive EV. The association with negative EV was again driven by anxiety/fear EV (*r* = .32, *p* = .009). Further, these associations remained significant when controlling for general vocabulary size and overall negative affect word frequencies in the linear regression analyses. That is, negative EV – driven by anxiety/fear EV (β = 0.27, *t* = 2.03, *p* = .047) – significantly predicted BPD symptoms when accounting for the control variables (β = 0.28, *t* = 2.16, *p* = .034; see Appendix B.2, Table B.3 for full regression results).

4.3.2.4 Research Aim 2: Exploring the Context-Dependency of Associations Between BPD and Active EVs

Mixed ANCOVAs revealed significant group-by-condition interaction effects for negative EV but not for positive EV (see Table 4.5 for full group-by-condition interaction effects). Specifically, pairwise comparisons revealed that women with BPD had significantly larger negative EVs than women without BPD in the neutral film condition and the relationship-threatening condition, but not in the personally-threatening condition (see Table 4.5 for statistics). Moreover, women without BPD had significantly larger negative EVs in the personally-threatening condition compared to the neutral (*M* difference = 0.48, *SE* = 0.13, *p* < .001) and relationship-threatening conditions (*M* difference = 0.56, *SE* = 0.14, *p* < .001), whereas there were no significant differences in negative EV across the conditions in women with BPD. Non-significant interaction effects for positive EV are described in detail in Appendix B.3, along with a visual presentation of the results (see Figure B.1).

Table 4.5

Group by Condition Interaction Effects on Emotion Vocabularies (EVs), Controlling for General Vocabulary and Emotion Word Frequencies (N = 64)

EV	Condition	Mean (*SD*)		*F*	*p*	*np*^{2}	95% CI
		BPD (*N* = 30)	Non-BPD (*N* = 34)				
Positive EV	Film	0.55 (0.49)	0.39 (0.26)	5.09	.028	.08	.03 – .41
	Fear	0.10 (0.20)	0.15 (0.26)	0.43	.513	.01	-.17 – .08
	Separation	0.19 (0.32)	0.16 (0.28)	0.06	.803	.00	-.14 – .18
	Overall interaction			3.30	.074	.05	
Negative EV	Film	0.36 (0.28)	0.22 (0.25)	5.20	.026	.08	.02 – .29
	Fear	0.54 (0.42)	0.67 (0.84)	1.03	.314	.02	-.52 – .17
	Separation	0.39 (0.54)	0.13 (0.17)	6.38	.014	.10	.05 – .45
	Overall interaction			3.45	.048	.05	

Note. “Group” refers to the between-participants factor comparing EVs between women with BPD versus women without BPD. “Condition” refers to the within-participants factor comparing EVs across the three conditions (neutral, personally-threatening, relationship-threatening). Results presented show the overall group by condition interaction effects on the EVs (i.e., the “overall interaction” rows) as well as differences in EVs between women with BPD and women without BPD in each of the conditions. CI = confidence interval.

4.3.2.5 Post-Hoc Exploratory Analyses: Examining the Relationship Between Active EVs and Depression

Given that BPD is arguably reflective of general psychopathology (e.g., Gluschkoff et al., 2021; Wright et al., 2016), it could be presumed that the associations evidenced between BPD and active EVs could in fact be transdiagnostic (i.e., extend to various other mental health conditions), rather than necessarily being specific to BPD. Accordingly, to explore this, we ran additional exploratory analyses investigating the relationship between active EVs and depression – using the depression scores (derived from the BDI measure) of the women in the sample – to examine whether the associations found between active EVs and BPD also extend to depression, as a potential indicator that they may be transdiagnostic. To do this, we conducted the same analyses as in the dimensional analytic approach with BPD symptoms (i.e., Pearson's correlations and linear regressions), but instead using the BDI scores. As with the BPD analysis, we only conducted these analyses on the women in the sample to control for the non-independence of data between partners.

Bivariate Pearson's correlation analyses revealed that depression scores were significantly associated with larger negative EV (*r* = .35, *p* = .005), but there were no significant associations with positive EV. Follow-up analyses revealed that the association with negative EV was predominantly driven by anxiety/fear EV (*r* = .34, *p* = .006) and undifferentiated negative EV (*r* = .27, *p* = .032). Linear regression analyses – in which active EVs, general vocabulary, and corresponding emotion word frequencies were entered as predictors of depression scores – revealed that the association between depression and negative EV remained statistically significant when accounting for the control variables, but this was now primarily driven by only anxiety/fear EV. Specifically, larger negative EV (driven by larger anxiety/fear EV; β = 0.33, *t* = 2.44, *p* = .018) significantly predicted depression scores while accounting for the control variables (β = 0.33, *t* = 2.57, *p* = .013), thereby displaying the same pattern of results as with BPD. See Table B.7 in Appendix B.6 for full regression results.

4.3.3 Discussion

In Study 2, we extended the findings from the previous study to a clinical sample by analysing the emotion language of women with a BPD diagnosis in spoken conversations with their romantic partners, compared to non-clinical couples. The results from Study 2 largely replicated those found in Study 1; BPD was associated with larger negative EV (primarily driven by larger anxiety/fear EV), although this was true even after accounting for general vocabulary size and overall negative affect word frequencies in this sample, thereby providing further support to our hypotheses. However, BPD was not found to be associated with smaller positive EV, running counter to our hypothesis. Results from Study 2 confirm the initial suggestion from Study 1's findings that the relationship between BPD and active EVs is predominantly context-insensitive, as the associations found in Study 2 largely generalised across three types of conversation.

4.4 General Discussion

In the present research, we conducted two studies comprising different types of samples and language data modalities to examine the diversity of emotion word use (i.e., active emotion vocabularies [EVs]) associated with BPD. As expected, findings from both studies revealed BPD to be associated with larger negative EV (i.e., greater diversity in negative emotion word use), which was found to be true even when controlling for general vocabulary size and negative affect word frequencies in Study 2. However, contrary to our hypotheses, BPD was not found to be reliably associated with positive EV. The associations between BPD and active EVs were also context-insensitive, with findings generalising across a variety of topics in both studies. Our findings provide insight into how emotion is attended to, represented, and may be(come) dysregulated in BPD. Moreover, the present findings add to the broader language and emotion literature by extending the findings of Vine et al. (2020) to the context of psychopathology, while also generating insight into the context-dependency of active EVs.

Most importantly, the robust finding that BPD was associated with larger negative EV after controlling for the use of negative affect words in Study 2 means that this association cannot simply be explained by people with BPD being more likely to use greater negative language overall, despite this also being true (e.g., Lyons et al., 2018). Rather, the finding is specific to the *variety* of negative emotion words used by individuals manifesting BPD. This finding is consistent with linguistic theory (e.g., Zipf, 1949), according to which profoundly frequent, prolonged, and varied experience with (often intense) negative emotion, possibly combined with a preoccupation with negative emotion (e.g., Peters et al., 2017), is indexed by the negative emotion words that individuals with BPD spontaneously and habitually use (see Vine et al., 2020). Moreover, such broad negative EVs may contribute to the emotion dysregulation observed in BPD by driving emotional cascades and negative rumination cycles, subsequently exacerbating negative affect. For example, frequently using a wide range of negative emotion words in everyday life may result in greater attention to, and rumination around, these negative emotions.

The fact that we did not find BPD to be associated with larger negative EV when controlling for overall negative affect word frequencies in Study 1 suggests that, in non-clinical populations, BPD traits are more strongly associated with greater use of negative affect words in general. In contrast, in individuals with high severity of borderline pathology (i.e., clinically significant levels), negative EVs explain variance in BPD severity over and above general negative affect word frequencies (as demonstrated in Study 2). Following the logic outlined above, one potential explanation for such difference is that individuals with lower levels of BPD traits (i.e., those in the general population) may not have reached the level of experience with negative emotion that is frequent, intense, and overpowering enough to be reflected in their natural emotion vocabularies over and above the strong positive relationship with negative emotion word frequency (reflecting greater attention to negative emotion) in general. It is also worth emphasising that the same pattern of effects (i.e., BPD = larger negative EV) were still evident across both studies; these effects were simply stronger/more robust in Study 2, where participants suffer more severe borderline pathology (note that less than 4% of the Study 1 sample reached clinically significant levels of borderline pathology, according to the PAI-BOR manual; Morey, 1991).

Contrary to our hypotheses and findings from the Vine et al. (2020) study illustrating positive associations between positive EV and psychosocial health, BPD was not found to be associated with smaller positive EV after controlling for general vocabulary size and positive affect word frequencies. The fact that we did not find BPD to be reliably associated with smaller positive EV could be explained by the nature of emotional dysregulation in BPD. In particular, emotional problems in BPD will typically take the form of extreme and rapid fluctuations in mood (e.g., Carpenter & Trull, 2013). By definition, such extreme fluctuations in mood mean that individuals with BPD also frequently experience fluctuating periods of positive emotion (e.g., Russell et al., 2007). Thus, individuals manifesting BPD may have had sufficient encounters with positive emotion for them to not show differences in positive EVs compared to the general population, which is in alignment with research showing emotion dysregulation in BPD to be most prominently distinguished by heightened negative emotion (e.g., Chu et al., 2016). Future research is needed to confirm this hypothesis.

Interestingly, our findings indicate that the associations between BPD and active EVs were not sensitive to context. Findings from Study 2 provided the most support for this interpretation – the negative EVs of women with BPD did not differ when discussing a non-emotive film topic compared to topics of a more negative emotive nature (i.e., personal fears and relationship threats). In comparison, women without BPD typically had larger negative EVs when discussing personal fears compared to film and relationship threat topics, presumably because the topic of personal fears was the most emotive and psychologically threatening for this group. These results seem to indicate that in the general population, highly emotive contexts may draw out people's natural emotion vocabularies. In contrast, individuals manifesting BPD appear to frequently access and use a broad range of negative emotion words irrespective of context.

One possible explanation for the overall findings is that individuals with BPD, through frequent experience with and interest in negative emotion, have become 'experts' in this domain (see Vine et al., 2020). That is, they verbally represent their emotional experience using a greater diversity of (and, in this sense, more specific) labels (Hoemann et al., 2021). In adaptive forms of expertise, this type of verbal representation is often linked to the possession of broad and efficiently-structured domain knowledge (e.g., Bukach et al., 2006); here, diverse and specific concepts for emotion. Yet, accounts of expertise also stipulate context-specificity as a critical ingredient (Hoemann et al., 2021). Our finding that associations between BPD and negative EV did not differ based on the elicitation context suggests that individuals with BPD may have a maladaptive form of expertise in (negative) emotion, including over-attention to emotion-relevant information and inflexible implementation of emotion concepts.

In speaking to how emotion concepts are implemented, the present findings also have some bearing on emotion differentiation and emotion labelling research. Namely, we do not consider an association between BPD and large negative EV to be contrary to the established link between BPD and lower emotion differentiation (e.g., Fitzpatrick et al., 2019). It is certainly plausible that individuals with BPD are less able to differentiate between specific emotions when the options are made explicit, and at the same time use a wide range of words to spontaneously refer to negative emotion. Indeed, studies that have operationalised emotion differentiation using verbal behaviour have found labelling- and rating-based measures to be unrelated (Ottenstein & Lischetzke, 2019; Williams & Uliaszek, 2022). More broadly, there is a lack of consensus as to the role of emotion labelling in emotion regulation. While many theories and studies support the utility of emotion labels for reducing distress (e.g., Gross, 2015) – although sometimes only after a delay in follow-up, and not when observing the immediate effects (see Torre & Lieberman, 2018) – including in BPD (e.g., Linehan, 2014), other work has found emotion labelling to interfere with effective emotion regulation (e.g., Meier et al., in press; Nook et al., 2021; Vine et al., 2019). Indeed, the utility of emotion labels may vary based on the specific context of emotion regulation, such as the intensity of experienced distress (Levy-Gigi & Shamay-Tsoory, 2022). Further research is necessary to disentangle when, how, and for whom the use of more precise words may be beneficial for emotional functioning.

Taken together, the present findings show that BPD is associated with the spontaneous use of more varied negative emotive language – likely reflecting extensive experience with negative emotion and, potentially, a (maladaptive) type of expertise in which emotion concepts are not implemented in a context-sensitive way. These large, context-insensitive negative EVs emphasise the need for more regulation of the referenced negative emotions. Thus, it may be beneficial for therapeutic interventions to work with individuals manifesting BPD to encourage them to explicitly attend to the way in which they spontaneously refer to their emotions in everyday life, while simultaneously encouraging reference to negative emotion in a more context sensitive manner and attempting to incorporate a broader range of positive emotion words in their natural emotion vocabularies.

However, it should be acknowledged that the associations found between active EVs (i.e., larger negative EVs) and BPD also extended to depression when explored as a post-hoc question in Study 2, providing an initial indication that these patterns may in fact be shared across numerous mental health conditions, consistent with transdiagnostic approaches to mental health (e.g., Dalgleish et al., 2020). Moreover, this finding is also consistent with that of the Vine et al. (2020) study in which depression symptoms were found to be positively correlated with negative EV. Nevertheless, given that there are very high rates of comorbidity between BPD and depression (e.g., Beatson & Rao, 2013), and BPD and depression symptom levels were highly correlated in the present sample, future research is needed to probe associations with EVs for specificity across a broader range of psychopathologies, in distinct samples.

4.4.1 Limitations and Future Directions

Despite the strengths of the present research – including the consistency in findings across two studies comprising diverse samples and types of language data – it is not without limitations. First, Study 1 comprised a self-selected sample, with the assessment of BPD features done via self-report methods, of which are accompanied by various biases (e.g., sampling bias, demand characteristics). Moreover, since we adopted a dimensional approach to psychopathology, the sample largely represented BPD traits prevalent in the general population, rather than clinically significant levels of BPD.

Second, Study 2 is limited by the absence of a clinical control sample (i.e., a group of people diagnosed with a mental health condition other than BPD) to confirm the initial indication explored with depression that the associations found with active EVs may be transdiagnostic. Thus, it would be most informative for future research to investigate active EVs in other clinical groups, in distinct samples, to examine the potential transdiagnostic nature of associations with active EVs.

Third, another potential limitation of Study 2 surrounds the fact that language data were translated from German to the English language for analysis. It is a possibility that this translation process could have influenced the emotion vocabulary scores to some extent. Yet, even if this did occur, any translation effects should have influenced data from all individuals, and both groups (i.e., BPD versus non-BPD), to equal extents, and so should not have had any impact on the overall results.

Fourth, there are some constraints surrounding the method of calculating active EVs (see Vine et al., 2020). Given that the calculation of active EVs relies on a word-counting approach, this means that words counted as part of the EV program will not always be exclusive to the realm of emotions; such words will often have numerous meanings (e.g., "mad" can mean angry, irrational, or even enthusiastic). More generally, these types of automated emotion word-counting approaches do not account for the context in which emotion words are used. For example, the statements "this makes me so happy" and "this does not make me happy" would generate the same positive emotion word count (and positive EV score), despite conveying different meanings. Yet, in the context of this particular research, the semantic context of emotion labels is largely irrelevant. Regardless of the degree to which a person is experiencing a given emotion (e.g., "happy" versus "not happy"), the fact that individuals were *attending to* a particular affective state – through the lens of a particular affective concept (e.g., "happiness") rendered through natural language – was the central focus of the current work (see, e.g., Boyd & Schwartz, 2021; Pennebaker et al., 1997). Moreover, such constraints apply to all "bag-of-words" approaches, which have been widely well-established as meaningful indicators of a broad range of psychological constructs (e.g., Kennedy et al., 2022).

Finally, given the nature of the data, causal relationships cannot be inferred from the present findings. Further research comprising longitudinal data is needed to determine cause-and-effect relationships between emotion functioning and experience and active EVs.

4.4.2 Conclusion

In the present research, we conducted two studies to examine the relationship between BPD and active emotion vocabularies (EVs). Results from both studies revealed that BPD was associated with relatively large negative EV (i.e., using a broad variety of negative emotion words), even after controlling for general vocabulary size and negative affect word frequencies in Study 2, likely reflecting extensive experience and preoccupation with negative emotion. Moreover, the relationship between BPD and negative EV was largely insensitive to context. Taken together, these findings indicate that BPD is associated with extensive but inflexible attention to and knowledge of negative emotion, potentially contributing to emotion dysregulation. Our findings contribute to BPD theory as well as the broader language and emotion literature, and also have implications for clinical practice.

CHAPTER 5:

Suicidality and Deliberate Self-Harm in Borderline Personality Disorder: A Digital Linguistic Perspective

Charlotte Entwistle, Katie Hoemann, Sophie J. Nightingale, Ryan L. Boyd [7]

Abstract

Borderline personality disorder (BPD) is characterised by persistent behavioural regulation problems. In particular, BPD is strongly associated with engagement in suicidality and deliberate self-harm (DSH), which, concerningly, are the most predominant risk factors for completed suicide. Accordingly, in the present study, we leveraged modern natural language processing methods to better understand the nature of suicidality and DSH in BPD. To do this, we analysed data extracted from Reddit;

[7] The authors would like to thank all of the Lancaster University undergraduate research assistants of whom provided invaluable contributions to this project by thoroughly coding the Reddit data at various stages: Ellie Bater, Ondrej Bocek, Jodi Gordon, Fen Jackson-Argent, Zoe Kirk, Bogna Liziniewicz, Charlotte May, Johanna Merkle, Tanvi Tiwaskar, and Hannah Wrigley.

namely, BPD discussion forums. Specifically, we utilised natural language processing techniques to analyse data of 992 users who self-identified as having BPD (combined *N* posts = 66,786). Overall, the present findings generated much needed further insight into the psychosocial dynamics of suicidality and DSH in BPD, while also uncovering meaningful interactions between the online BPD community and suicidality and DSH behaviours. Our findings have important theoretical and practical implications, particularly with respect to helping to explain suicidality and DSH in BPD.

Keywords: borderline personality disorder, suicidality, self-harm, natural language processing, social media

5.1 Introduction

Borderline personality disorder (BPD) is a heterogenous construct, but is generally characterised by pervasive problems with affect regulation and functioning, developing and maintaining healthy relationships, identity stability, and adaptive behavioural regulation (American Psychiatric Association, 2013). Moreover, BPD is a particularly prevalent and high-risk disorder, in part, due to its problematic relationship with dysregulated behaviour. In particular, BPD is strongly associated with suicidality and deliberate self-harm (DSH), with these behaviours even conceptualised as key markers for the detection of BPD (see Reichl & Kaess, 2021). [8]

Alarmingly, prevalence rates of DSH have been reported to be around 90% among adults with BPD (Goodman et al., 2017), compared to 6% in adults in the general population (Klonsky, 2011). Similarly, suicidality is known to be strongly associated with BPD and is evidenced to be chronic in individuals with BPD (e.g., Paris, 2019); research has shown that around 75% of individuals with BPD attempt suicide at some point during their lives (Black et al., 2004). Such strong associations between BPD and self-harm are majorly concerning given that prior suicide attempts

[8] DSH is defined here as an intentional act of causing oneself physical injury without suicidal intent (Lauw et al., 2015), with common DSH behaviours including cutting, hitting oneself, substance abuse, and otherwise risky behaviours (e.g., dangerous driving; unsafe sex). By contrast, suicidality directly reflects risk of suicide, indicated by suicidal ideation (i.e., thoughts about suicide) and suicide attempts (i.e., attempting to end one's own life).

and DSH are the most important risk factors for completed suicide (e.g., Hawton et al., 2015). Coincidently, there are in fact disturbingly high rates of completed suicide in the BPD population, with reports reaching as high as 10% (Paris & Zweig-Frank, 2001). It is therefore crucial to study suicidality and DSH in BPD to better understand how such harmful behaviours can be lessened or prevented.

Notably, individuals who engage in suicidality and DSH, and those experiencing mental health problems more broadly, are ever-increasingly turning to online platforms to discuss their experiences and seek support (see, e.g., Tucker & Lavis, 2019). Subsequently, the disclosure and discussion of suicidality and DSH on online platforms provides researchers new and promising opportunities to study these high-risk behaviours at a large-scale and, importantly, in naturalistic contexts. In particular, the analysis of natural language provides a promising alternative approach to studying suicidality and DSH in BPD, given that one's language use can provide insight into their underlying psychology, values, motivations, and behaviours (e.g., Pennebaker, 2011; Schultheiss, 2013), making language analysis a clinically valuable tool. Accordingly, in the present study, we leverage modern natural language processing (NLP) methods to analyse large online BPD forums to investigate the psychosocial dynamics – that is, the internal, psychological and the external, relational processes that constitute our everyday lives – surrounding suicidality and DSH in BPD, while also examining online BPD community dynamics in relation to these behaviours.

Specifically, in this work, our goals are as follows:

1) Provide an overview of the psychosocial dynamics of suicidality and DSH in BPD, as evidenced by more traditional methods and NLP approaches.

2) Briefly discuss literature on effects of online mental health communities on engagement in suicidality and DSH.

3) Empirically investigate suicidality and DSH in BPD through the analysis of large-scale online BPD forums using NLP methods.

To our knowledge, this is the first study to integrate sophisticated computational linguistic methods with psychological theory to provide such a far-reaching, large-scale, naturalistic, psychologically insightful perspective on both suicidality and DSH *in situ*, and in a BPD population.

5.1.1 Psychosocial Dynamics of Suicidality and Deliberate Self-Harm in BPD

5.1.1.1 Trait-Level Risk Factors of Suicidality and DSH in BPD

Suicidality and DSH share overlapping provenance, including affect regulation (e.g., Hatkevich et al., 2019; Vansteelandt et al., 2017) and social functions (e.g., Muehlenkamp et al., 2013; Van Orden et al., 2010). Commensurate with this notion, research employing more traditional methods (e.g., self-report/laboratory studies) has identified consistent overlap in the trait-level psychosocial risk factors for each in the context of BPD. For instance, a recent literature review uncovered psychological distress, affective instability/dysregulation, impulsivity, and social dysfunction as predictors of both suicidality and DSH (Gee et al., 2020). Further, studies that have focused on BPD specifically have additionally found dysfunctional cognitive processes (e.g., negative rumination) to be major risk factors of both suicidality and DSH (e.g., Johnson et al., 2022).

Notably, although the general risk factors for suicidality and DSH in BPD are largely overlapping, there are also some key differences. For instance, an "all-or-nothing" (i.e., absolutist or dichotomous) thinking style has been evidenced to be a major risk factor for suicidality, but not for DSH (e.g., Halicka & Kiejna, 2018). Moreover, although affective dysregulation is a strong predictor of both suicidality and DSH in BPD, it differs in the form that it takes. Specifically, affective instability and heightened negative affect predict both suicidality and DSH (e.g., Gee et al., 2020), whereas dissociation (i.e., intense feelings of emptiness and "emotional numbness") is more reliably associated with DSH (see Al-Shamali et al., 2022, for a review).

Critically, although the research discussed here has provided valuable insights into the risk factors of suicidality and DSH in BPD, it is limited by a heavy reliance on self-report measures and otherwise non-naturalistic methods, which is problematic for numerous reasons, such as memory distortion issues and social desirability biases (see, e.g., Baldwin et al., 2019; Jeong et al., 2018). More generally, individuals participating in such studies may simply lack the necessary insight into their own internal processes to accurately report on psychosocial correlates of suicidality and DSH, which is even more likely in individuals with pathological personality traits who typically suffer major identity disturbance issues (see Entwistle et al., 2022, for a discussion).

Valuably, overcoming some of the constraints associated with overreliance on self-report measures, there has been some meaningful research that has leveraged naturalistic NLP techniques to better understand the risk factors and psychosocial correlates of suicidality and DSH. In general, two broad types of research have been carried out in this domain: research focused on predictive accuracy (i.e., NLP and machine learning tasks that aim to correctly identify the content of a message) and research directed towards psychological explanation (i.e., generating interpretable, psychologically meaningful results to develop better understanding of psychological phenomena). Most of the work to date that has applied NLP methods to the study of suicidality and DSH has focused on the predictive accuracy aspect, including identifying linguistic markers of these behaviours often measured using automated word counting programs, such as Linguistic Inquiry and Word Count (LIWC; Boyd et al., 2022). Importantly, the linguistic features revealed to be predictive of suicidal and DSH posts (i.e., at the document level) largely reflect the trait-level psychosocial risk factors, and often the same correlates evidenced from more traditional methods, such as internalisation, social disconnection, and relatively heightened negative emotion. That is, suicide and DSH relevant posts are marked by linguistic indicators of dysfunctional self-processes and mental distress – in particular, more self-focused language, or I-words (e.g., Sierra et al., 2022; Uban et al., 2021; a widely established indicator of mental distress and depression; see Tackman et al., 2019) – as well as linguistic indicators of affective dysregulation (i.e., more negative affect words, particularly anxiety, and fewer positive affect words; e.g., Aldhyani et al., 2022; Sierra et al., 2022; Uban et al., 2021), with these markers also found to predict suicidality risk at the person level (Ramírez-Cifuentes et al., 2020). Likewise, suicide-relevant (but not DSH) posts have additionally been revealed to be associated with linguistic markers of social dysfunction (i.e., fewer we-words, you-words, and social words; Aldhyani et al., 2022; Sierra et al., 2022) and cognitive style, including less analytic language (Aldhyani et al., 2022) and greater absolutist language (reflecting “all-or-nothing” thinking; Al-Mosaiwi & Johnstone, 2018).

Informatively, the findings discussed in this section illuminate the psychosocial correlates and risk factors of suicidality and DSH, with converging findings from both traditional and NLP methods, which has clinical utility in helping to identify individuals with BPD who may be at risk of or are currently engaging in suicidality and/or DSH. However, such findings do not provide insight into temporal dynamics surrounding suicidality and DSH, which would allow for understanding and explanation of *why* these behaviours might occur *when* they do.

5.1.1.2 Temporal Psychosocial Dynamics Surrounding Suicidality and DSH in BPD

Extending beyond trait-level risk factors, researchers have worked to determine *when* individuals with BPD are more likely to engage in suicidality and DSH. In particular, research on temporal components have focused considerable attention on the affective dynamics surrounding these behaviours — that is, fluctuations in mood and emotion and their relationship to self-harm. Indeed, there is strong empirical evidence highlighting the role of affective dysregulation as a temporal precursor to both suicidality and DSH in BPD, evidenced in both more traditional research (see a review by Reichl & Kaess, 2021) and NLP work (e.g., De Choudhury et al., 2016; Glenn et al., 2020; Sawhney et al., 2021). Generally speaking, research has frequently found heightened negative affect to precede both suicidality and DSH, which then typically decreases after the occurrence of these behaviours (see Kuehn et al., 2022, for a meta-analysis), which has largely been mirrored in research incorporating NLP methods (e.g., Coppersmith et al., 2016). Specifically, studies have typically shown negative emotive language (particularly sadness, anxiety, and anger) to increase in the three to two weeks before the suicide-relevant event and positive emotive language to decrease in the two to one weeks before the suicide-relevant event (e.g., Glenn et al., 2020; Sawhney et al., 2021), with one study also finding negative emotive language to decrease following the event (Coppersmith et al., 2016). Notably, this pattern of affective dynamics is most consistent for suicidality, as it has been repeatedly evidenced in BPD and non-BPD samples. In contrast, the affective dynamics surrounding DSH are less consistent. In particular, numerous studies have in fact found negative affect to increase both prior to *and* following DSH (e.g., Houben et al., 2017; Koenig et al., 2021). Moreover, research conducted specifically on BPD populations has revealed patterns of affective dynamics surrounding DSH that could be considered contrary to those found in non-BPD populations; that is, numerous studies have evidenced intense feelings of dissociation (or “emotional numbness”) to precede DSH, which typically eases following engagement in DSH (see Al-Shamali et al., 2022).

In related work, meaningful research has expanded the investigation of precursors to suicidality and DSH to include other, non-affective psychosocial processes. In particular, social dysfunction – typically in the form of isolation, rejection, or interpersonal conflict – has been established as a precursor to both suicidality (e.g., Gratz et al., 2022) and DSH (e.g., Snir et al., 2015) in individuals with BPD. Consistent with this, one study utilising NLP methods also revealed (markers of) social dysfunction (i.e., less socially connected and socially oriented language) to precede engagement in suicidality (De Choudhury et al., 2016). However, studies examining linguistic changes in a more precise and detailed way have, in contrast, failed to evidence indicators of social dysfunction as significant predictors of suicidality (Coppersmith et al., 2016; Glenn et al., 2020). In addition to social dysfunction, behavioural dysregulation – particularly impulsivity – has been uncovered as another key precursor to both suicidality (Selby et al., 2013) and DSH (Ammerman et al., 2017), which has again also been evidenced in NLP work on suicidality (De Choudhury et al., 2016). As for other linguistic markers revealed to precede engagement in suicidality, absolutist language (De Choudhury et al., 2016) and indicators of dysfunctional self-processes (e.g., heightened self-focused language; Coppersmith et al., 2016; De Choudhury et al., 2016) have been uncovered as additional temporal predictors of suicidality.

Valuably, the research discussed in this section has illustrated the promising potential of using NLP methods to better understand the psychosocial dynamics surrounding suicidality and DSH. Despite this, critical gaps in this literature remain. First, considerable research has been dedicated to using NLP to detect or predict suicidality or DSH from online data, however this work is often criticised due to its “black box” nature, with studies often lacking in construct validity and integration of psychological theory/perspective, thus restricting meaningful psychological insight, as proposed as critical limitations of work in this realm by recent literature reviews (Chancellor & De Choudhury, 2020; Yeskuatov et al., 2022). Second, none of the NLP studies discussed in this manuscript are in the context of BPD. Given the invaluable work that has been done using NLP techniques to study suicidality and DSH in general, and how NLP methods have opened up new understandings of these pathological behaviours, it would be exceedingly propitious to take the best aspects of this research –

naturalistic sampling, powerful NLP methods, and longitudinal perspectives – and bring them to bear on BPD. It can be seen from past research investigating the psychosocial dynamics of suicidality and DSH that these dynamics appear to differ in individuals with BPD; thus, examining these behaviours in a BPD population using sophisticated NLP methods could allow for further insight into the underlying psychology that may help to explain why suicidality and DSH are so prevalent in this population. Third, almost all of the studies discussed (bar one; Coppersmith et al., 2016) have not examined the "after-effects" of engagement in suicidality (or DSH); understanding the psychological consequences of engagement in suicidality and DSH is important as to create a fuller psychological picture of these harmful behaviours. Finally, the language-based research discussed in this section has exclusively focused on suicidality, with no psychologically insightful research having used NLP methods to investigate the psychosocial dynamics of DSH, which is problematic given that DSH is a very high-risk behaviour that is strongly predictive of completed suicide (e.g., Hawton et al., 2015).

5.1.2 Effects of Online Support Communities in Relation to Suicidality and Deliberate Self-Harm

In going a step beyond the vast majority of past work studying discussions of suicidality and DSH on online platforms, it is important to consider another critical dimension; that is, the dynamics of the support communities themselves. Research into the effects of online mental health support communities is currently in its infancy, despite this becoming a rapidly popular method of seeking mental health support (see, e.g., Tucker & Lavis, 2019). More specifically, the dynamics of online support communities are often empirically examined in a general sense, but not typically in the context of discrete mental health/clinical events, such as disclosures of suicidality or DSH. With regard to meaningful NLP research investigating online mental health support platforms in general, interestingly, one study examined features of mental health support seeking posts that generated more supportive community responses (De Choudhury & De, 2014). From this, it was uncovered that posts (i.e., mental health disclosures) that were more positive, more sociable, more self-focused, and shorter received greater community support (in the form of higher scores, or more "upvotes"), whereas posts that were more negative emotive and contained more swear words received less community support (in the form of lower scores, or less “upvotes”).

Informatively, other initial empirical work has focused attention towards understanding effects of online platforms on suicidality and DSH specifically, of which has presented a fairly mixed picture. In general, dedicated support platforms appear to show more positive effects, whereas social media platforms appear to show more negative effects (see Marchant et al., 2017, for a review). Building on this, in more detailed investigations, some studies have directly examined factors that influence effects of online platforms. In particular, one study by De Choudhury and colleagues (2016) examined features of interactions with the online community as predictors of future engagement on suicidality forum. Features of community interactions revealed to predict future engagement on a suicidality forum included lower language style matching with the mental health community, less engagement with the community (i.e., less posts made to the forum), and less support received from the community, in the form of less replies and upvotes in response to their posts.

Crucially, the research discussed in this section demonstrates how online support communities indeed have the potential to affect individuals’ psychological processes. However, research in this area is very limited. Most relevantly, there are no studies that have investigated the effects of online BPD platforms on their users, or that focus on suicidality and DSH specifically in the context of BPD, illustrating a major gap in the literature. It is critical to study possible effects of online BPD support platforms given that these communities have the potential to further drive harmful behaviours, such as DSH and suicidal behaviour, in a population of people who typically already frequently engage in such risky behaviours.

5.1.3 The Current Research

The research discussed in this article so far has uncovered major gaps in the literature base with respect to understanding suicidality and DSH in BPD via NLP. To summarise the most pivotal gaps, although a fair amount of empirical attention has leveraged NLP techniques to study suicidality, a large portion of this work is of a black-box nature (i.e., focused on predictive accuracy, rather than psychological explanation, per-se) and lacks incorporation of psychological theory, and no psychologically insightful research has applied these methods to investigations into DSH. Moreover, no studies to date have used NLP methods to investigate suicidality or DSH in BPD specifically. Finally, research into the effects of online mental health support platforms on mental health outcomes is scarce, and we are not aware of any research that has investigated the effects of online BPD platforms on their users. Accordingly, to address these critical gaps in the literature, in the present study we leverage modern, naturalistic NLP methods to analyse large online BPD forums in order to investigate the psychosocial dynamics of suicidality and DSH in BPD, while incorporating psychological theory and generating meaningful psychological insight. Additionally, we also examine online BPD community interactions in relation to suicidality and DSH. Thus, in this work, we integrate methodological approaches focused on predictive accuracy with psychological explanation approaches, developing a "middle-ground" relative to past work, allowing for improvements in psychological knowledge in a highly sophisticated and refined way. To achieve this, we analyse data extracted from the social media platform Reddit, specifically leveraging BPD discussion forums (i.e., "subreddits")**.** In particular, in this study, we aim to address three central research questions:

RQ1. In what ways are the psychosocial dynamics of suicidality and DSH in BPD evident in verbal behaviour?

RQ2. What features characterise the online BPD community's interaction with disclosures of suicidality and DSH?

RQ3. How might the online BPD community interact with the psychosocial dynamics preceding suicidality and DSH events to shape the outcome of these events?

We address the central research question – RQ1 – by examining associations between key linguistic features and frequencies of engagement in suicidality and DSH among individuals with BPD, providing new insight into the person-level psychosocial correlates of suicidality and DSH in this population via a naturalistic linguistic approach. Further, we address RQ1 more precisely by investigating changes in key linguistic features (i.e., linguistic trajectories) over the weeks preceding and following the occurrence of suicidality and DSH events, to generate better understanding of the *temporal* psychosocial dynamics (and linguistic predictors) in proximity to these behaviours in individuals with BPD.

Regarding RQ2 and RQ3, we address these research questions through examining interactions between the online BPD community – in the form of number of replies to submissions and post scores (i.e., the number of “upvotes” minus “downvotes”) – and suicidality and DSH disclosure posts themselves (RQ2), and with the key linguistic features found to temporally predict engagement in suicidality and DSH (RQ3). More specifically, determining whether posts that disclose engagement in suicidality or DSH are less or more likely to receive community support will provide novel insight into how online BPD communities interact with these behaviours (i.e., RQ2). Furthermore, analysing associations between community support and frequencies of linguistic features found to temporally precede suicidality and DSH in posts made to the platform will generate new understanding of how online BPD communities interact with the temporal psychosocial dynamics that surround these harmful behaviours (i.e., RQ3). Critically, the results from both RQ2 and RQ3 analyses should generate implications for how online BPD communities may potentially hamper or further drive engagement in suicidality and DSH in this population.

5.2 Methods

5.2.1 Data Overview

Data were collected from the widely used, publicly available social media platform Reddit. For a brief overview, Reddit is a very large, anonymous online discussion forum composed of numerous sub-forums (i.e., “subreddits”), each dedicated to a specific topic (e.g., sports teams, food, etc.). Within this forum structure, people can make initial posts (“submissions”) that are relevant to the given topic and respond to each other through chains of comments.

For the present study, data were collected from the *r/BPD* and *r/BorderlinePDisorder* subreddits, whereby people typically identify as having BPD and look to seek support from and build connections with others with BPD. These subreddits were selected as they are the dominant forums for people with BPD to discuss their disorder on Reddit, with the two subreddits combined comprising over 300,000 individual community members to date, making them among the largest (if not the largest) online BPD communities. All posts (including submissions and comments), along with the associated meta-data (e.g., number of replies to posts, post scores) made between October 2011 and August 2019 were extracted from the *r/BPD* and *r/BorderlinePDisorder* subreddits from a larger Reddit database (Baumgartner et al., 2020). In total, 607,559 posts from 52,369 individual users were extracted.

5.2.2 Data Refinement and Extraction

To address the aims of the present study and allow for linguistic trajectories to be examined, only users that had made multiple posts within the BPD subreddits were valuable to preserve. Thus, only users that had made a minimum of 10 posts within the BPD subreddits were retained. Given the large sample of data, it was expected that a minimum criterion of 10 posts would permit a large number of users to be included in the analysis while still allowing for meaningful assessment of individuals' psychosocial dynamics.

Given our goal of understanding suicidality and DSH in individuals with BPD, we sought to further refine our dataset by selecting posts by only individuals who self-identified as having BPD. Accordingly, posts made to the BPD subreddits were manually inspected and coded by expert raters to identify those who self-identified as having BPD, including statements such as "my BPD diagnosis" and "diagnosed with BPD", aided via an automated, custom-made "BPD diagnosis" dictionary to highlight posts containing such phrases – a commonly adopted approach to identifying users with mental health conditions on online platforms (e.g., Coppersmith et al., 2016). Each text was coded by two separate raters and any disagreements were clarified by a third researcher (the lead author). Only texts where it was very clear that the user self-identified as having BPD were classified as them having the disorder. We aimed to extract a total of 1,000 users with self-identified BPD, as to generate a large and representative sample. Although we initially achieved this sample size goal, further manual quality checks indicated that some of the 1,000 users initially identified as having BPD may in fact not have had an official BPD diagnosis; consequently, the coding process resulted in 992 users identified as having BPD, with a combined total of 97,787 posts made between them. This classification coding had good inter-rater reliability (α = .74).

As Reddit is largely an anonymous platform, we did not have access to users' demographic information to characterise the sample. Accordingly, posts made by the

992 users identified as having BPD were inspected and manually coded to extract various demographic characteristics (e.g., age, gender, country of residence). The demographic coding process was carried out in the same way as the BPD classification coding process, and also followed a demographics coding framework to ensure consistency (and again was aided via an automated, custom-made keyword dictionary).

In addition, as part of a larger investigation, our broad goals were to identify key (time-stamped) behaviours in users posts highly relevant to the construct of BPD, permitting detailed behavioural analytics to be carried out. Subsequently, a coding framework was developed to guide the coding of nine BPD-relevant behaviours/events in users posts, including self-harm behaviours, medication and therapy related behaviour, substance use, impulsive behaviour, social interactions, and emotion-relevant behaviours/events. The coding of BPD-relevant behaviours/events was consistent with the coding procedure described above, and was also guided by an automated dictionary stratification approach. That is, texts selected for manual coding were those which contained the most keywords and phrases (i.e., posts scoring highest on behaviour dictionaries, indicating greater likelihood that they would contain relevant information) related to a particular behaviour/event (for instance, examples of key words/phrases used for suicidality include “want to die” and “feel suicidal”). Resulting from this dictionary-based stratification process, of the 97,787 posts made by the 992 users with BPD, 9,106 were manually coded for BPD-relevant behaviours and events. Of the behaviours coded for, given the goals of the present work, it is those related to self-harming behaviour (i.e., suicidality and DSH) that are of interest for the present study. Suicidality was further broken down into suicide attempts and ideation, as well as past (i.e., longer than one week) and recent (i.e., within the same week for attempts and at the time of writing for ideation) occurrences of these behaviours. DSH was broken down into engagement in DSH and urge for DSH, and again into past and recent engagement. There was generally high inter-rater agreement for the coding of demographics and behaviours (see Appendix C.1, Table C.1, for a breakdown of the coding agreement percentages for the demographic and behavioural coding).

5.2.3 Language Pre-Processing and Analysis

Language data collected from Reddit were cleaned and prepared for analysis according to standard guidelines (Boyd, 2017) – formatting errors and common misspellings and elongations were corrected, URLs were simplified, quotes were removed, and all texts that contained fewer than 25 words were removed from the dataset to ensure validity of measurement, along with all duplicate texts. Following this cleaning process, 66,786 individual posts remained.

To address our research questions (particularly RQ1 and RQ3), following data pre-processing procedures, linguistic features were extracted using the latest version of the Linguistic Inquiry and Word Count (LIWC) software (Boyd et al., 2022). Briefly described, LIWC relies on an internal dictionary that maps words and phrases to psychologically meaningful categories, with the output produced by LIWC consisting of relative frequencies (i.e., percentages) of each category within each text. Despite its simplicity, LIWC has been extensively and well-validated as to its utility in modelling mental health relevant language (see, e.g., Soldaini et al., 2018; Spruit et al., 2022), as well as improving understanding of psychopathological constructs more generally (e.g., Lyons et al., 2018).

For the purposes of the present study, we were only interested in the linguistic features derived from the LIWC dictionary that are closely related to the construct of BPD (or psychopathology more broadly). Accordingly, we extracted linguistic features from LIWC that are most theoretically reflective of borderline pathology and well-established as indicators of psychopathology. More specifically, the selection of linguistic features for inclusion was made based on previous research findings that have evidenced the predictive validity of the linguistic features/categories in relation to suicidality, DSH, or psychopathology in general – most of which have been described in the introduction of this article – and that are relevant to the construct of BPD. In total, 16 linguistic features were selected for inclusion, of which we mapped onto four broad psychosocial dimensions based on consistent associations from previous research: self-processes, emotion processes, social processes, and cognitive processes (reflecting core areas of dysfunction in BPD; APA, 2013). A list of the selected LIWC variables and their mapping onto each of the four broad dimensions, along with references to previous research to justify their inclusion, can be seen in Table 5.1. Although some of the selected LIWC variables could have reasonably been mapped on to more than one psychosocial dimension, we have opted to map each variable on to their most conceptually relevant dimension for simplicity.

Table 5.1

LIWC Variables Included in the Current Study Mapped on to Broad Psychosocial Dimensions Related to Borderline Pathology, with Supporting Empirical Evidence

Psychosocial Dimension	LIWC Category	Example Words	Direction of Association	Example Reference(s)
Self-processes/ functioning	I-words	I, me, my	-	Sierra et al. (2022) Tackman et al. (2019) Uban et al. (2021)
	Negations	no, not, didn't	-	Coppersmith et al. (2015) Ramírez-Cifuentes et al. (2020)
Emotion processes/ functioning	Positive emotion	happy, excited, love	+	Glenn et al. (2020) Sierra et al. (2022)
	Negative emotion	sad, hate, hurt	-	De Choudhury et al. (2016) Sierra et al. (2022) Uban et al. (2021)
	Anxiety	anxious, fear, worry	-	De Choudhury et al. (2016) Lyons et al. (2018) Ramírez-Cifuentes et al. (2020)
	Sadness	depressed, cry, upset	-	Glenn et al. (2020) Lyons et al. (2018)
	Anger	Angry, mad, frustrated	-	Coppersmith et al. (2016) Glenn et al. (2020) Uban et al. (2021)
	Swear words	shit, fuck, damn	-	Coppersmith et al. (2015)
Social processes/ functioning	We-words	we, our, us	+	Coppersmith et al. (2015) Lyons et al. (2018) Sierra et al. (2022)

	You-words	you, your, yourself	+	De Choudhury et al. (2016) Sierra et al. (2022)
	Shehe-words	he, she, her	Mixed	De Choudhury et al. (2016) Lyons et al. (2018)
	They-words	they, their, them	Mixed	De Choudhury et al. (2016) Lyons et al. (2018)
	Affiliation	together, social, collectively	+	Ramírez-Cifuentes et al. (2020)
	Social references	you, we, her	+	Aldhyani et al. (2022) Sierra et al. (2022)
Cognitive processes/ functioning	Cognitive processes	think, puzzle, solve	Mixed	Aldhyani et al. (2022) Ramírez-Cifuentes et al. (2020)
	Absolutism/all-none	always, never, definitely	-	Al-Mosaiwi & Johnstone (2018) De Choudhury et al. (2016)

Note. This table shows the LIWC variables selected for inclusion in the present study and their mapping on to four broad psychosocial dimensions related to borderline pathology. Example words from the LIWC22 dictionary are presented for each of the linguistic categories. The direction of association column shows the direction (i.e., positive [+] versus negative [-]) of the associations between the LIWC variables and the broader psychosocial dimensions based on previous research (presented in the "Example References" column), generally in the context of deliberate self-harm and suicidality.

5.3 Results

5.3.1 Descriptive Analysis of Sample

With regard to the posting behaviour of the 992 users classified as having BPD, the total number of posts (including submissions and comments) made by each of these users to the BPD subreddits ranged from 1–1,958, with an average of 67.32 posts per user (*SD* = 106.58) and an average of 114.08 words per post (*SD* = 136.70). Regarding the length of time users spent posting to the BPD subreddits, the overall duration of time from users’ first to their last post ranged from 0 to 2,481 days (6 years, 9 months, 17 days), with an average duration of 392.29 days (1 year, 27 days; *SD* = 407.07). Refer to Appendix C.2 for descriptive illustrations of users’ posting behaviour.

The demographic coding procedure (described in the Methods section) applied to the BPD subreddits allowed us to extract a good amount of demographic information to characterise the sample (see Table 5.2 for detailed sociodemographic characteristics).

Table 5.2

Sociodemographic Characteristics of BPD Reddit Sample (N = 992)

Characteristic	Mean	*SD*
Age (*n* = 260; 26.21%)	27.87	8.06
	n	%
Gender (*n* = 394; 39.71%)		
Female	237	60.15
Male	146	37.06
Non-binary	11	2.79
Country of residence (*n* = 294; 29.64%)		
UK	87	29.59
US	75	25.51
Canada	55	18.71
Australia	29	9.86
Other European country	34	11.56
Other non-European country	14	4.76
Religion (*n* = 102; 10.28%)		
Non-religious/atheist	48	47.06
Religious	39	38.24
Spiritual	7	6.86
Agnostic	8	7.84
Relationship status (*n* = 700; 70.56%)		

Single	156	22.29
In a relationship	378	54.00
Married	140	20.00
Divorced/separated	26	3.71

Note. The *n*s and percentages provided alongside each of the demographic categories reflect the total number of users (and percentage of the sample) we were able to extract each demographic variable for.

As for the behavioural coding, Table C.2 in Appendix C.3 shows the frequencies of suicidality and DSH events manually coded for, broken down into past and recent occurrences. Overall (including all subcategories), 1,290 events (from 497 unique users) were captured for suicidality and 678 (from 319 unique users) for DSH.

5.3.2 RQ1: In What Ways are the Psychosocial Dynamics of Suicidality and DSH in BPD Evident in Verbal Behaviour?

5.3.2.1 Person-Level Linguistic Markers of Suicidality and Deliberate Self-Harm in BPD

Statistical Analysis. To examine the linguistic markers of suicidality and DSH in individuals with BPD – indicative of the trait-level psychosocial risk factors – in a descriptive fashion, we conducted two-tailed, bivariate Spearman's Rho correlation analyses between users' total frequencies of disclosures of suicidality and DSH (including disclosures of past and recent events) and mean scores (across all posts made to the BPD subreddits) for the 16 linguistic categories derived from LIWC. Spearman's Rho correlations were chosen over Pearson's due to the dataset comprising non-normally distributed frequency data. All posts coded for mention of suicidality or DSH were excluded from this analysis to ensure that the results are representative of users' general language and do not simply reflect language patterns specifically associated with suicidality/DSH disclosures (i.e., analyses were conducted at the person level rather than the document level).

Correlation Test Results. Results from the correlation analyses are presented in Table 5.3. In general, most of the significant correlations found were for the frequency of disclosures of recent (as opposed to past) occurrences of suicidality and DSH.

Linguistic indicators of dysfunctional self-processes (i.e., [excessive use of] I-words, or first-person singular pronouns) were evidenced to be positively associated with the frequency of both recent suicidality and DSH. However, negations (also an indicator of dysfunctional self-processes) were only found to be associated with recent suicidality.

Markers of affective dysfunction were also found to be associated with the frequency of disclosures of both suicidality and DSH. In particular, overall negative emotion, sadness, and anger words were all positively associated with the frequency of recent suicidality and DSH. In comparison, anxiety words were only associated with the frequency of past occurrences of these events. Swear words were found to be correlated with the frequency of recent suicidality only. No associations emerged for positive emotion words.

Regarding linguistic markers of social dysfunction, no statistically significant associations were found for these linguistic features, however you-words (i.e., second-person pronouns) and social references were negatively associated with frequencies of recent suicidality and DSH by trend.

As for cognitive functioning, negative associations emerged between cognitive processing language and frequencies of past suicidality and recent DSH. Absolutist language (indicative of cognitive dysfunction) was found to be positively associated with the frequency of recent suicidality only.

Table 5.3

Spearman's Rho Correlations Between Mean Language Variable Scores and Suicidality (Suicide Attempts and Ideation) and Deliberate Self-Harm (DSH) Frequencies (N = 992)

LIWC Variable	Past Suicidality	Recent Suicidality	Past DSH	Recent DSH
I	.02	.16***	.06†	.11***
Negations	.03	.07*	.04	.02
Positive emotion	.04	.05	.04	-.04
Negative emotion	.06†	.14***	.05	.12***
Anxiety	.07*	.05	.07*	.05
Sadness	.07*	.12***	.08*	.09**
Anger	.04	.13***	.04	.09**
Swear	.04	.11***	.03	.03
We	.02	-.03	.01	-.02
You	-.02	-.06†	-.02	-.06†
Shehe	-.02	-.03	.00	.03
They	.02	.03	.00	.02
Affiliation	-.01	-.04	-.04	.04
Social references	-.03	-.06†	-.06†	-.06†
Cognitive processes	-.09**	-.01	-.05	-.10**
Absolutism	.04	.11***	.00	.01

****p* < .001, ***p* < .01, **p* < .05, †*p* < .10.

Note. All tests are two-tailed. Language variable scores reflect users' mean LIWC22 category scores from the BPD subreddits, excluding posts coded for deliberate self-harm or suicidality of any nature (past or recent). Mean language scores were correlated with users' overall frequency of suicidality/DSH disclosures.

Variations of these correlation analyses – whereby outliers were removed, and users' total number of posts were controlled for – were also carried out; the results of which portray the same patterns as those reported here (see Appendix C.4 for these results tables).

5.3.2.2 Temporal Linguistic Trajectories Surrounding Suicidality and Deliberate Self-Harm in BPD

Subsetted Dataset. To better understand the temporal psychosocial dynamics surrounding suicidality and DSH in individuals with BPD, we investigated language changes in the weeks immediately preceding and following disclosures of recent suicidality and DSH. To do this, we created a subset of the dataset based on posts that were manually coded for disclosures of recent occurrences of suicidality and DSH. As shown in Table C.2 (Appendix C.3), this originally included 600 cases of recent suicidality (23 cases for suicide attempts and 577 for suicidal ideation) and 148 cases of recent DSH. From there, we extracted mean linguistic feature scores of all posts 3 weeks preceding suicidality/DSH events up until 3 weeks following (spanning 6 weeks in total), aggregated weekly, by user. The decision to set the time range as 3 weeks either side of the events was primarily based on previous research showing most of the psychological changes leading up to self-harm events, and suicidality events in particular, that portray themselves in language typically occur in the preceding 2 weeks (described as the “critical period” for suicidality; Millner et al., 2016), with the most critical changes occurring in the week immediately preceding (Glenn et al., 2020; Sawhney et al., 2021). We thus opted for a 3-week pre-event time frame based on this 2-week pre-suicidality critical period, with data 3 weeks prior to the event perceived as more reflective of users’ general (i.e., “baseline”) language use. The post-event time frame was simply matched with that of the pre-event period for consistency. Regarding the decision to aggregate data on a weekly level, this was largely based on the fairly sparse nature of the subsetted dataset (a common issue when utilising naturalistic social media data); there would not have been sufficient data to conduct meaningful statistical analyses if aggregating at a precision level greater than weekly (e.g., daily), as the majority of users did not post on a near-daily basis. Moreover, previous research has aggregated psychosocial precursors to suicidality data on a weekly level and generated psychologically insightful results (e.g., Selby et al., 2013).

Data Cleaning and Refinement. To clean the subsetted dataset for analysis, overlapping posts coded for recent suicidality or DSH (i.e., multiple posts disclosing recent engagement in suicidality/DSH in the same 6-week period) by the same user were removed, and multiple posts disclosing recent engagement in suicidality/DSH that were posted on the same day by the same user were merged to be the same event. After removing or merging overlapping posts, 453 cases of recent suicidality and 126 cases of recent DSH remained. We then refined the dataset further to ensure that all cases to be included had sufficient data for meaningful analysis, as many cases comprised large portions of missing data (i.e., no or minimal posts in the weeks surrounding the events). In developing the analysis inclusion criteria, we abided by the principle of identifying the minimum number of data points necessary to permit the observation of meaningful trends, resulting in the following criteria: a minimum of one post made in at least 2/3 of the weeks prior to the suicidality/DSH event and a minimum of one post made in at least 2/3 of the weeks following the event. Specifically, these criteria were agreed upon to ensure high-quality, meaningful data for subsequent analysis, with linguistic data from 2 separate weeks pre- and post-event being necessary to investigate temporal linguistic changes. This process meant that all cases included in subsequent analyses comprised linguistic data from a minimum of 4 of the 6 weeks surrounding the suicidality/DSH event. After applying the inclusion criteria, 159 cases of suicidality from 124 individual users (*N* complete observations = 827) and 43 cases of DSH from 40 individual users (*N* complete observations = 227) remained (some users had multiple occurrences of recent suicidality/DSH that did not overlap in time frames, and so were retained as individual cases), which were included in subsequent analyses.

Statistical Analysis. All weekly aggregated data (i.e., averaged linguistic feature scores) were assigned a time point in relation to the suicidality/DSH events, to allow for meaningful statistical analyses to be conducted; namely: -3 = 3 weeks before; -2 = 2 weeks before; -1 = 1 week before; 0 = day of event; 1 = 1 week after; 2 = 2 weeks after; 3 = 3 weeks after. Following this, generalised linear mixed models (GLMMs) were carried out to examine changes in language in proximity to occurrences of suicidality and DSH. GLMMs were selected for the analysis method given that they permit missing data and small sample sizes, and also allow us to nest posts within users and control for random user effects; GLMMs were used over LMMs due all DVs being non-normally distributed. In all GLMMs, time point was entered as the fixed effect variable (with 6 levels/time points, spanning 3 weeks before the event to 3 weeks after; to ensure clean results, the day of the event was not included) and the 16 LIWC variables were entered as dependant variables (DVs), with random effects of users controlled for and a log-link transformation applied (due to non-normal data distributions). Further, in cases where significant overall main effects of time were found, post-hoc pairwise comparisons were conducted to determine precisely when (i.e., between which time points) significant changes in language occurred. These analyses were carried out separately for suicidality and DSH. As a reminder, cases in which overlapping posts were coded for recent suicidality or DSH (i.e., multiple posts disclosing recent engagement in suicidality/DSH in the same 6-week period) by the same user were not included in the analyses, thus ensuring clean results with respect to the observation of distinct linguistic trajectories.

GLMM Descriptive Analyses. Prior to the main analyses, we conducted descriptive statistical analyses on the subsetted dataset in which we examined changes in the number of posts users made (i.e., posting frequency) to the BPD subreddits in proximity to suicidality and DSH events via GLMMs (as described above), with number of posts aggregated weekly entered as the DV. In addition to examining changes in posting frequency, we also investigated changes in the word count (i.e., length) of posts made to the BPD subreddits in proximity to suicidality and DSH events, in which average post word count (aggregated weekly) was entered as the DV. Refer to Appendix C.5 for all results from these descriptive analyses.

Main GLMM Results. Here, we present the main results from the GLMMs, segregated by the psychosocial domain (as in Table 5.1). Refer to Appendix C.6 for descriptive statistics and overall fixed effects of time point in proximity to suicidality (Table C.7) and DSH (Table C.8) for each of the 16 language variables.

Emotion Processes. Most of the significant linguistic changes that occurred in proximity to both suicidality and DSH events surrounded changes in emotion language. GLMMs revealed no significant fixed effects of time (in proximity to the events) for overall positive or negative emotion words in relation to both suicidality and DSH events (see Tables C.7 and C.8). However, meaningful changes emerged for several of the specific negative emotion categories.

For suicidality, significant overall fixed effects of time in proximity to suicidality were evidenced for anxiety (*F*(5, 821) = 7.85, *p* <.001), sadness (*F*(5, 821) = 6.04, *p* <.001), and swear words (*F*(5, 821) = 7.60, *p* <.001; see Table C.7 for descriptive statistics). Anxiety words sharply increased from 3 to 2 weeks before the suicidality event (*M* increase = 0.32, *SE* = 0.06, *t* = 5.59, *p* <.001), which remained at a heightened level in the week immediately preceding the event. Although still higher than baseline levels, there was a decrease in anxiety words in the week immediately following the suicidality event when compared to 2 weeks before (*M* decrease = -0.20, *SE* = 0.05, *t* = -3.74, *p* <.001), which stayed at the same frequency level 2 weeks after the event. Anxiety words increased again (back up to 2-weeks pre-event levels) 3 weeks after the event (*M* increase = 0.15, *SE* = 0.06, *t* = 2.51, *p* = .012). As for changes in sadness, there was a significant increase in sadness words in the week immediately preceding the suicidality event compared to 3 and 2 weeks before (*M* increase from 2 weeks before = 0.19, *SE* = 0.06, *t* = 3.44, *p* = .001). Sadness words dropped back down to baseline levels (i.e., 3-weeks pre-event) in the week immediately following the suicidality event (*M* decrease = -0.21, *SE* = 0.05, *t* = -3.84, *p* <.001), which stayed at the same frequency level 2- and 3-weeks post-event. No significant changes in anger words over the weeks surrounding suicidality events were found. Yet, swear words were evidenced to increase in the week immediately preceding the suicidality event compared to 3 and 2 weeks before (*M* increase from 3 weeks before = 0.18, *SE* = 0.05, *t* = 3.26, *p* = .001). Use of swear words significantly dropped (back to baseline levels) in the week immediately following the event (*M* decrease = -0.23, *SE* = 0.05, *t* = -4.36, *p* <.001), before sharply increasing again 2-weeks post-event (*M* increase = 0.26, *SE* = 0.05, *t* = 4.87, *p* <.001). Swear word use decreased again (back to baseline levels) 3 weeks following the suicidality event (*M* decrease = -0.19, *SE* = 0.05, *t* = -3.51, *p* <.001).

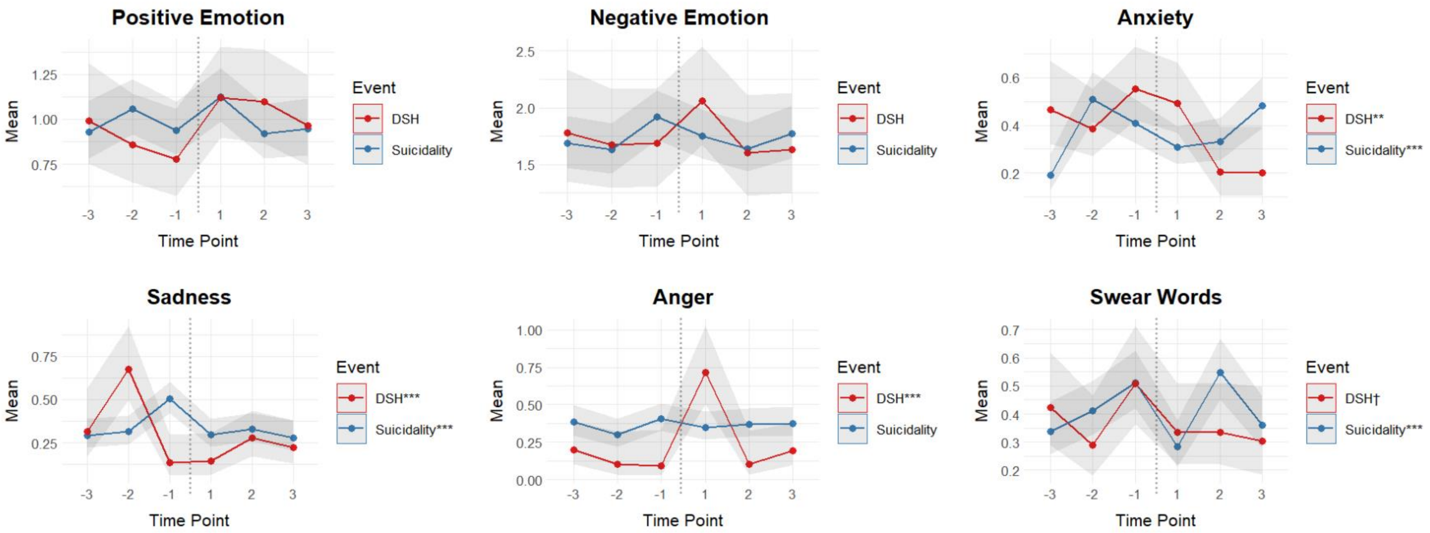
In terms of changes in emotion language surrounding DSH events, GLMMs revealed significant fixed effects of time in proximity to DSH for anxiety (*F*(5, 221) = 3.27, *p* = .007), sadness (*F*(5, 221) = 12.74, *p* <.001), and anger words (*F*(5, 221) = 12.33, *p* <.001; see Table C.8 for descriptive statistics). There were no significant changes in anxiety words across the weeks preceding the DSH event up until the week immediately following the event. However, anxiety word use significantly dropped 2 weeks following the event compared to all preceding weeks (e.g., *M* decrease from 3 weeks before = -0.26, *SE* = 0.11, *t* = -2.48, *p* = .014), which remained at a lower frequency 3-weeks post-event. Use of sadness words significantly increased from 3 to 2 weeks before the DSH event (*M* increase = 0.36, *SE* = 0.12, *t* = 3.02, *p* = 003), before sharply decreasing the week immediately preceding the event (*M* decrease = -0.54, *SE* =

0.11, *t* = -5.18, *p* <.001). Sadness words remained at a lower frequency over the 3 weeks following the DSH event. There were no changes in anger words across the weeks preceding the DSH event. However, anger words considerably increased in the week immediately following the event compared to the 3 preceding weeks (e.g., *M* increase from 1 week before = 0.62, *SE* = 0.13, *t* = 4.81, *p* <.001), which dropped back down to baseline levels by 2-weeks post-event (*M* decrease = -0.61, *SE* = 0.13, *t* = -4.77, *p* <.001), and stayed at this frequency level 3-weeks post-event. There was no significant overall effect of time in proximity to DSH on swear words.

Figure 5.1 illustrates the weekly changes in emotion language in proximity to both suicidality and DSH events.

Figure 5.1

GLMM Emotion Plots: Changes in Mean Emotion Language in Proximity to Recent Suicidality and Deliberate Self-Harm (DSH) Events



Note. The figure shows changes in mean emotion language category scores (derived from LIWC) per week (i.e., aggregated weekly) surrounding suicidality and DSH events. The dotted lines illustrate the point at which engagement in the event occurred (i.e., time point 0), thus dividing the figures by pre- and post-event. The shaded areas surrounding the means represent the error margins (95% confidence intervals). The means (and confidence intervals) have been estimated from the generalised linear mixed models (GLMMs), and thus are reflective of the repeated measures nature of the data (i.e., person-centered) while also controlling for random user effects. The indicators assigned to the suicidality and DSH keys show the statistical significance of the overall fixed effects of time in proximity to the events: ****p* < .001, ***p* < .01, **p* < .05, †*p* < .10. Time point labels: -3 = three weeks before event, -2 = two weeks before event, -1 = one week before event, 1 = one week after event, 2 = two weeks after event, 3 = three weeks after event.

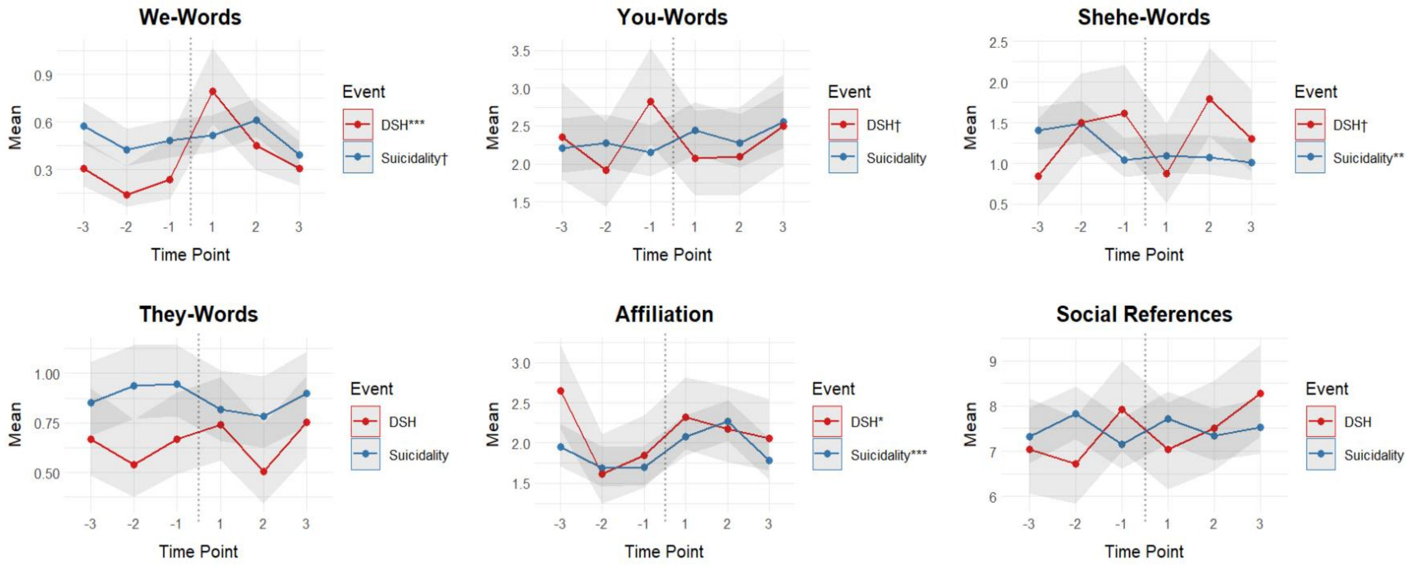
Social Processes. Of the linguistic categories related to social processes (see Table 5.1), for suicidality, GLMMs revealed significant overall fixed effects of time (in proximity to suicidality) for affiliation words (*F*(5, 821) = 4.20, *p* <.001) and shehe-words (i.e., third-person singular pronouns; *F*(5, 821) = 3.27, *p* = .006; see Table C.7 for descriptive statistics). There were no significant changes in affiliation words over the weeks preceding the suicidality event. However, there was an increase in affiliation words in the week immediately following the event compared to 2 weeks and 1 week before (*M* increase from 1 week before = 0.38, *SE* = 0.16, *t* = 2.37, *p* = .018), which remained at a slightly heightened frequency 2-weeks post-event. Affiliation word use dropped back down to pre-event frequency levels 3 weeks following the suicidality event (*M* decrease = -0.49, *SE* = 0.17, *t* = -2.89, *p* = .004). Shehe-word (i.e., third-person singular pronouns) use significantly decreased in the week immediately preceding the suicidality event compared to 3 and 2 weeks before (*M* decrease from 2 weeks before = -0.45, *SE* = 0.16, *t* = -2.75, *p* = .006), which remained at a lower frequency over the 3 weeks following the event. There were no other significant fixed effects of time in proximity to suicidality for any of the other language categories related to social processes.

With regard to DSH, GLMMs evidenced significant fixed effects of time in proximity to DSH for we-words (i.e., first-person plural pronouns; *F*(5, 221) = 12.58, *p* <.001) and affiliation words (*F*(5, 221) = 2.60, *p* = .026; see Table C.8 for descriptive statistics). Specifically, there was a marginally significant decrease in we-words from 3 to 2 weeks before the DSH event (*M* decrease = -0.16, *SE* = 0.08, *t* = -1.96, *p* = .052), which stayed at a similar frequency level in the week immediately preceding the event. We-words sharply increased in the week immediately following the DSH event compared to all pre-event weeks (e.g., *M* increase from 1 week before = 0.56, *SE* = 0.13, *t* = 4.22, *p* <.001), before dropping back down to baseline levels (i.e., 3-weeks pre-event) 2 weeks following the event (*M* decrease = -0.34, *SE* = 0.12, *t* = -2.84, *p* = .005), and remaining at a similar frequency level 3-weeks post-event. Affiliation words were also found to decrease from 3 to 2 weeks before the DSH event (*M* decrease = -1.03, *SE* = 0.32, *t* = -3.23, *p* = .001), remaining at a lower frequency in the week immediately preceding the event. Affiliation word use frequency did not significantly change from 1-week pre-event over the 3 weeks following the DSH event (but was similar to baseline levels by 1-week post-event). No other significant fixed effects of time in proximity to DSH were found for any of the other social language categories.

See Figure 5.2 for a visual display of weekly changes in social language in proximity to both suicidality and DSH events.

Figure 5.2

GLMM Social Plots: Changes in Mean Social Language in Proximity to Recent Suicidality and Deliberate Self-Harm (DSH) Events

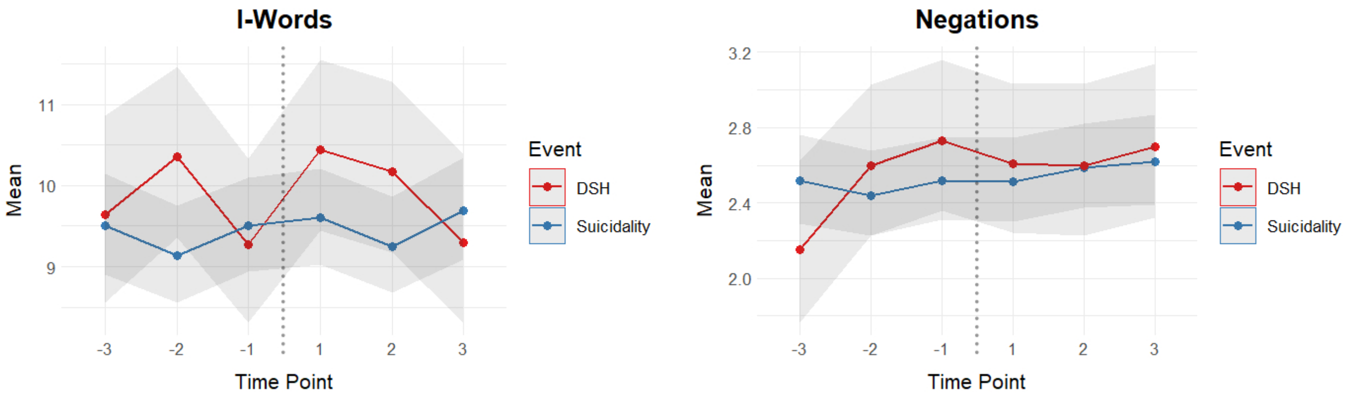


Note. The figure shows changes in mean social language category scores (derived from LIWC) per week (i.e., aggregated weekly) surrounding suicidality and DSH events. The dotted lines illustrate the point at which engagement in the event occurred (i.e., time point 0), thus dividing the figures by pre- and post-event. The shaded areas surrounding the means represent the error margins (95% confidence intervals). The means (and confidence intervals) have been estimated from the generalised linear mixed models (GLMMs), and thus are reflective of the repeated measures nature of the data (i.e., person-centered) while also controlling for random user effects. The indicators assigned to the suicidality and DSH keys show the statistical significance of the overall fixed effects of time in proximity to the events: ***p < .001, **p < .01, *p < .05, †p < .10. Time point labels: -3 = three weeks before event, -2 = two weeks before event, -1 = one week before event, 1 = one week after event, 2 = two weeks after event, 3 = three weeks after event.

Self-Processes. GLMMs evidenced no significant overall fixed effects of time (in proximity to the events) for I-words or negations (i.e., linguistic indicators of dysfunctional self-processes) in relation to both suicidality and DSH events (see Tables C.7 and C.8). Nevertheless, specific (mostly statistically non-significant) changes in I-words and negations over the weeks surrounding suicidality and DSH events can be seen in Figure 5.3.

Figure 5.3

GLMM Self-Processes Plots: Changes in Mean Self-Processes Language Indicators in Proximity to Recent Suicidality and Deliberate Self-Harm (DSH) Events

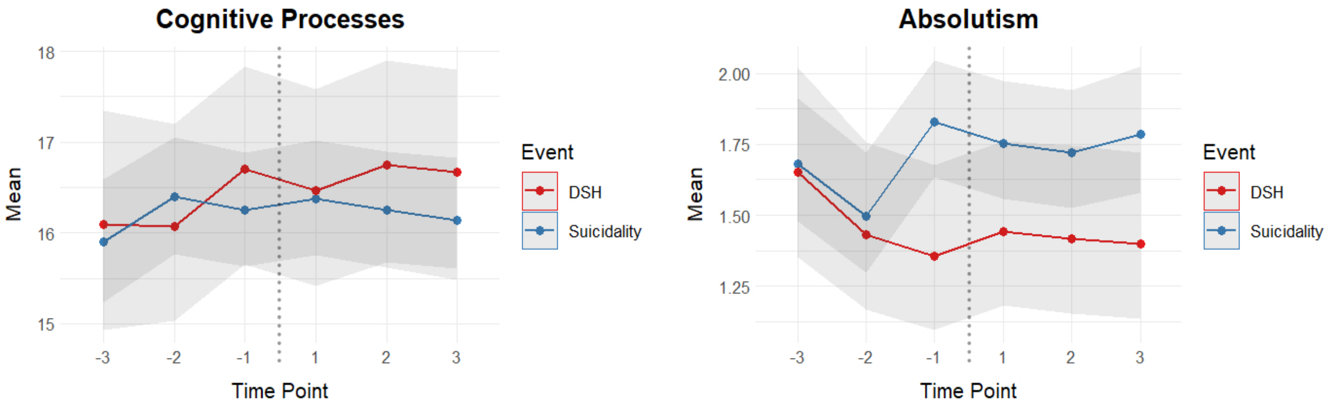


Note. The figure shows changes in mean linguistic indicators of self-processes scores (derived from LIWC) per week (i.e., aggregated weekly) surrounding suicidality and DSH events. The dotted lines illustrate the point at which engagement in the event occurred (i.e., time point 0), thus dividing the figures by pre- and post-event. The shaded areas surrounding the means represent the error margins (95% confidence intervals). The means (and confidence intervals) have been estimated from the generalised linear mixed models (GLMMs), and thus are reflective of the repeated measures nature of the data (i.e., person-centered) while also controlling for random user effects. There are no significance indicators assigned to suicidality and DSH keys in this figure as there were no statistically significant overall fixed effects of time in proximity to the events for these variables. Time point labels: -3 = three weeks before event, -2 = two weeks before event, -1 = one week before event, 1 = one week after event, 2 = two weeks after event, 3 = three weeks after event.

Cognitive Processes. As with the self-processes language categories, no significant overall fixed effects of time (in proximity to the events) were found for cognitive processing words or absolutist language in relation to both suicidality and DSH events (see Tables C.7 and C.8). Figure 5.4 displays the linguistic trajectories of these variables in proximity to suicidality and DSH events.

Figure 5.4

GLMM Cognition Plots: Changes in Mean Cognitive Language in Proximity to Recent Suicidality and Deliberate Self-Harm (DSH) Events



Note. The figure shows changes in mean cognitive language category scores (derived from LIWC) per week (i.e., aggregated weekly) surrounding suicidality and DSH events. The dotted lines illustrate the point at which engagement in the event occurred (i.e., time point 0), thus dividing the figures by pre- and post-event. The shaded areas surrounding the means represent the error margins (95% confidence intervals). The means (and confidence intervals) have been estimated from the generalised linear mixed models (GLMMs), and thus are reflective of the repeated measures nature of the data (i.e., person-centered) while also controlling for random user effects. There are no significance indicators assigned to suicidality and DSH keys in this figure as there were no statistically significant overall fixed effects of time in proximity to the events for these variables. Time point labels: -3 = three weeks before event, -2 = two weeks before event, -1 = one week before event, 1 = one week after event, 2 = two weeks after event, 3 = three weeks after event.

5.3.3 RQ2: What Features Characterise the Online BPD Community's Interaction with Disclosures of Suicidality and DSH?

In order to address RQ2, we leveraged Reddit community interaction variables from post meta-data; namely, post scores (i.e., the number of "upvotes" [or "likes"] minus "downvotes") and number of replies to posts, both of which have been utilised and conceptualised in previous related work as reflecting online community support (De Choudhury et al., 2016; De Choudhury & De, 2014).

5.3.3.1 Statistical Analysis

We first carried out a general person-level analysis whereby we ran Spearman's correlations (two-tailed) between users' mean post scores (averaged from all posts per user; *N* = 992) and mean number of replies to users' submissions (averaged per user, from submissions only; *N* = 880) and users' total frequency of posts disclosing engagement in suicidality and DSH (as separate variables; including both past and recent events). To build on these analyses further and more directly address RQ2 (i.e., at the document level), we ran a series of independent (two-tailed) *t*-tests comparing post scores and number of replies (for submissions only) between posts in which engagement in suicidality (*N* = 1,290 posts) or DSH (*N* = 678 posts) were disclosed versus posts in which engagement in suicidality (*N* = 7,816 posts) or DSH (*N* = 8,428 posts) were not disclosed, separately for disclosures of past and recent events. Specifically, in all *t*-tests, whether or not the post disclosed past/recent engagement in suicidality/DSH was entered as the IV and post score and number of replies were entered as the DVs. These *t*-tests were carried out on data that had been manually coded for suicidality/DSH events only (total *N* = 9,106 posts), as it was not known whether posts that were not manually inspected disclosed engagement in suicidality or DSH.

5.3.3.2 Person-Level Correlation Results

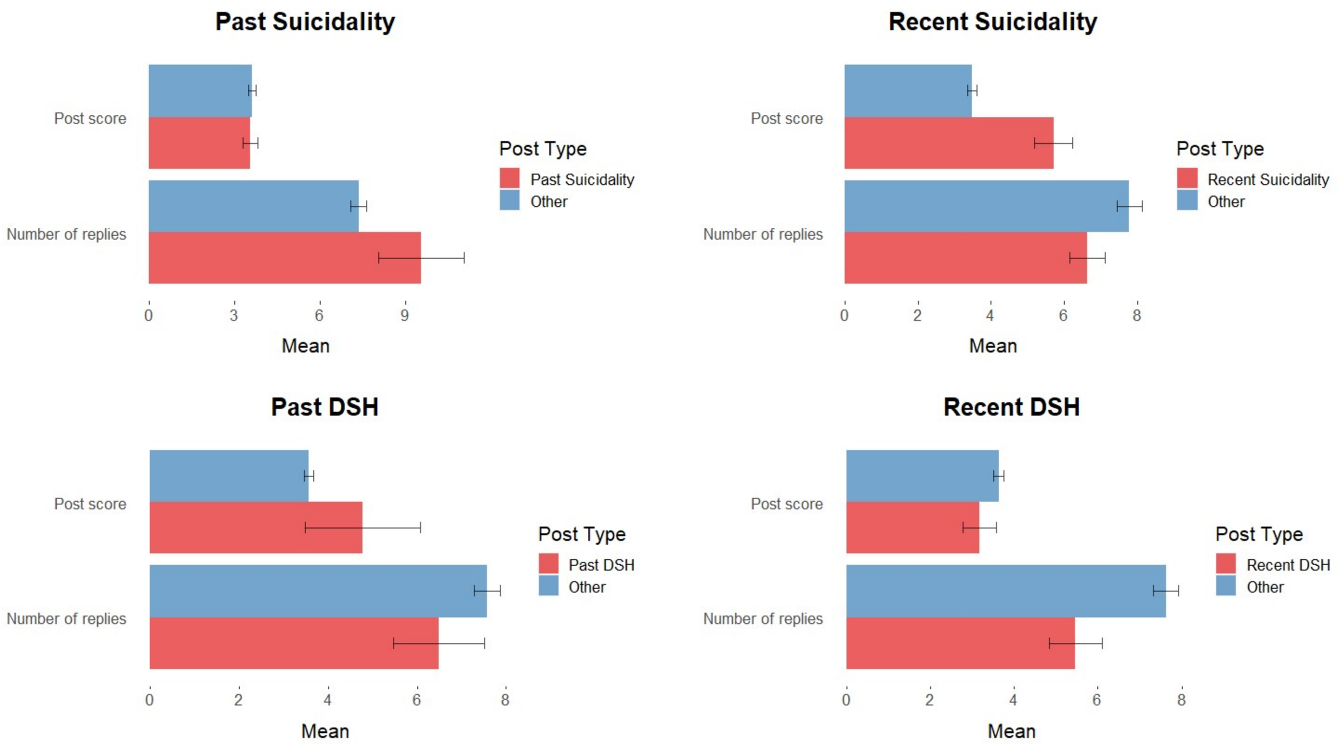
For post scores, Spearman's correlation analyses revealed a significant positive association between users' mean post scores and the frequency of posts disclosing recent engagement in suicidality (*r* = .10, *p* = .002). Associations between mean post scores and frequencies of posts disclosing past engagement in suicidality as well as both past and recent engagement in DSH were all non-significant (all *p*'s > .10). Regarding number of replies (to submissions only), results showed a significant positive association between users' mean number of replies received and the frequency of posts disclosing past engagement in suicidality (*r* = .11, *p* = .002). There were no other significant correlation results (all *p*'s > .10).

5.3.3.3 Document-Level T-Test Results

Results from the *t*-tests revealed no significant differences in post scores or number of replies to submissions between posts disclosing past engagement in either suicidality or DSH when compared to non-suicidality/DSH posts (i.e., posts that were manually coded as not disclosing engagement in suicidality/DSH). Yet, significant differences emerged when comparing non-suicidality/DSH posts to posts disclosing recent engagement in suicidality or DSH. Specifically, *t*-tests revealed that posts disclosing recent engagement in suicidality (*M* = 5.71, *SD* = 12.80) received significantly higher scores (i.e., up-votes) when compared to all other posts manually coded (*M* = 3.48, *SD* = 11.44; *t*(669) = 4.14, *p* <.001, *d* = .19). There was no significant difference found between these variables in the number of replies to submissions (*p* = .083). As for recent engagement in DSH, no differences were found when comparing post scores between posts disclosing recent engagement in DSH versus non-DSH posts (*p* = .631). However, submissions disclosing recent engagement in DSH (*M* = 5.47, *SD* = 4.76) were found to receive significantly fewer replies when compared to all other coded submissions (*M* = 7.62, *SD* = 8.71; *t*(80) = 3.04, *p* = .003, *d* = .25). See Figure 5.5 for an illustrative comparison of average post scores and number of replies received between posts disclosing past or recent engagement in suicidality/DSH versus non-suicidality/DSH posts.

Figure 5.5

Comparisons of Community Support Between Post Types (i.e., Whether Past/Recent Engagement in Deliberate Self-Harm [DSH]/Suicidality Was Disclosed; N = 9,106 Posts)



Note. This figure presents a visual display of the RQ2 *t*-test results, in which post scores and number of replies to submissions were compared between posts disclosing past/recent engagement in DSH/suicidality versus all other posts (i.e., posts not disclosing engagement in DSH/suicidality). Post scores reflect the number of upvotes a post receives subtracted by the number of downvotes, thereby reflecting the overall "rating" of the post. Number of replies are in relation to submissions only (i.e., not responses to comments). The figure is organised by the type of the event disclosed (i.e., suicidality or DSH) and whether the post related to past or recent engagement in the event. Error bars represent the standard errors.

5.3.4 RQ3: How Might the Online BPD Community Interact With the Psychosocial Dynamics Preceding Suicidality and DSH Events to Shape the Outcome of These Events?

As in RQ2, we utilised the post scores (i.e., the number of “upvotes” minus “downvotes”) and number of replies to submissions meta-data variables (operationalised as reflecting community support) to address this research question.

5.3.4.1 Statistical Analysis

We conducted Spearman’s correlation analyses (two-tailed) between the frequencies of key linguistic features within posts made to the BPD subreddits and community responses to the posts (i.e., post scores and number of replies to submissions), using the full BPD Reddit dataset (i.e., *N* = 66,786 posts). The language variables included in this analysis were those of which significant overall fixed effects of time were found in proximity to either suicidality or DSH events in the RQ1 analyses, which are: affiliation words, we-words, shehe words, anxiety, sadness, anger, and swear words. Examining the extent to which these linguistic features correlate with post scores and number of replies to submissions allows insight into how the online BPD community is interacting with the psychosocial dynamics (evident in language) surrounding suicidality and DSH in individuals with BPD.

5.3.4.2 Correlation Results

Overall, correlation results revealed no significant correlations between the number of replies to submissions and any of the key language variables. However, when looking at post scores, almost all of the included language variables showed positive correlations with post scores, with the exception of affiliation words (non-significant). Specifically, frequencies of we-words and shehe-words, as well as anxiety, sadness, anger, and swear words, positively correlated with higher scoring posts. Detailed correlation results are reported in Table 5.4.

Table 5.4

Spearman's Rho Correlations Between Key Linguistic Features and Community Support Variables

Language Variable	Post Score (*N* = 66,786)	Number of Replies (*N* = 7,307)
Affiliation	.01	-.01
We	.02***	-.01
Shehe	.02***	-.01
Anxiety	.03***	.02†
Sadness	.04***	.00
Anger	.04***	.02
Swear	.04***	-.01

****p* < .001, †*p* < .10.

Note. All tests are two-tailed. Post scores reflect the number of upvotes a post receives subtracted by the number of downvotes, thereby reflecting the overall "rating" of the post. Number of replies are in relation to submissions only (i.e., not responses to comments), hence the smaller *N*. These analyses were carried out on the full BPD Reddit dataset.

5.4 Discussion

In the present work, we leveraged modern NLP methods to analyse large online BPD discussion forums to generate better understanding of the psychosocial dynamics of suicidality and DSH in BPD, while also examining online BPD community interactions – and the potential impact of these interactions – in relation to these behaviours. To our knowledge, this is the first study to have integrated advanced computational linguistic methods with psychological theory to provide such a far-reaching, large-scale, naturalistic psychological perspective on both suicidality and DSH *in situ*, and in a BPD population. Overall, our findings have revealed key person-level linguistic markers of suicidality and DSH in BPD, which are largely consistent with previous literature. Further, the present findings have also shed light on the temporal psychosocial dynamics that surround suicidality and DSH in BPD, thereby generating much needed further understanding and explanation of why these behaviours occur when they do in individuals with BPD. Finally, we also evidenced meaningful interactions between the online BPD community and suicidality and DSH behaviours, in terms of both the disclosure of engagement in these behaviours on the platform as well as the temporal psychosocial dynamics that surround these harmful behaviours. Accordingly, the present work has major theoretical implications – including contributing to BPD theory as well as the broader NLP suicide/DSH literature base – and practical implications, such as the potential to leverage these findings to anticipate (and thus possibly prevent) engagement in suicidality and DSH in individuals with BPD.

5.4.1 Psychosocial Dynamics of Suicidality and Deliberate Self-Harm in BPD Evident in Verbal Behaviour

In addressing the central goal of the present work – that is, developing better understanding of the psychosocial dynamics of suicidality and DSH in BPD using NLP methods – we examined the person-level linguistic markers of suicidality and DSH in BPD by correlating frequencies of disclosures of suicidality and DSH made to the BPD discussion platform with frequencies of key linguistic features. More meaningfully, we also investigated linguistic trajectories across the weeks surrounding the occurrence of recent engagement in suicidality and DSH, to provide insight into the temporal psychosocial dynamics in proximity to engagement in these behaviours in individuals with BPD.

Promisingly, in examining the person-level linguistic markers of suicidality and DSH in BPD, the associations found were largely consistent with previous literature, including both more traditional and NLP research. For instance, as shown in previous research (e.g., Gee et al., 2020), overlapping linguistic markers were evidenced for both suicidality and DSH, including language indicative of dysfunctional self-processes (i.e., greater self-focused language; e.g., Sierra et al., 2022; Uban et al., 2021), emotion dysregulation (i.e., more negative emotive language in general, and sad and angry language in particular; e.g., Coppersmith et al., 2016; De Choudhury et al., 2016; Glenn et al., 2020), and social dysfunction (i.e., less references to other people; e.g., Sierra et al., 2022). Moreover, results from our study confirm the finding from previous research that absolutist/dichotomous thinking is associated with suicidality (e.g., Al-Mosaiwi & Johnstone, 2018), but not reliably associated with DSH (e.g., Halicka & Kiejna, 2018), as we found absolutist language (indicative of an all-or-nothing thinking style) to be associated with recent engagement in suicidality only. Importantly, the consistency between our findings and those of previous research allows us to confirm and further validate the psychosocial correlates and risk factors of suicidality and DSH in BPD using a naturalistic, behavioural approach that is less subject to bias (compared to self-report measures, for example), while also extending findings from past NLP suicide/DSH research to a BPD population.

That said, there were also some noteworthy discrepancies between the present findings and those of past work. For instance, in contrast to findings from previous NLP work showing suicidality/DSH to be associated with the use of relatively fewer positive emotion words (e.g., Sierra et al., 2022), we did not find any associations between positive emotion words and frequencies of engagement in suicidality and DSH in our study. One possible reason for this is that, in comparison to the present study, past NLP studies evidencing this result were not conducted on BPD populations. Accordingly, the nature of emotion dysregulation in BPD may explain such difference, given that individuals with BPD often experience fluctuating periods of (heightened or intense) positive emotion as part of their dysregulation (e.g., Russell et al., 2007), which may explain why we did not find fewer positive emotion words to be associated with more engagement in suicidality/DSH in our study. In addition to this, the present work also revealed another novel finding with respect to the linguistic correlates of suicidality; namely, the positive association between swear words and the frequency of recent engagement in suicidality, which could be perceived as a marker of a combination of anger, hostility, and/or impulsivity.

Taken together, the present findings highlight self-focused language, general negative emotive language – and sadness and anger words in particular – and fewer social references as linguistic markers of engagement in both suicidality and DSH in individuals with BPD, with greater use of negations, swear words, and absolutist language as additional markers of suicidality specifically. Interpretating this on a broader psychological level, such markers position dysfunctional emotion, social, and self processes as trait-level psychosocial correlates of both suicidality and DSH in BPD, with maladaptive cognitive processes as an additional correlate of suicidality. Importantly, such knowledge is of clinical value in that it generates better understanding of why engagement in suicidality and DSH is so common in individuals with BPD – with these psychosocial correlates being typical features of borderline pathology – which could also prove useful in guiding risk profiling in this population.

Nonetheless, knowledge that is even more invaluable surrounds insight into why individuals with BPD engage in suicidality and DSH *when* they do, as this knowledge could more directly aid the prevention of such behaviours. Vitally, results from our analyses examining the linguistic trajectories surrounding engagement in suicidality and DSH provide useful insight in this respect. Notably, consistent with previous NLP suicide/DSH research, our study revealed the emotion language categories to undergo the most prolific changes in proximity to both suicidality and DSH events, thus supporting dysfunctional emotion processes as the predominant causal factor in relation to these behaviours (e.g., Hatkevich et al., 2019; Vansteelandt et al., 2017). More precisely, regarding the specific changes in emotion language, the emotion language trajectories in proximity to suicidality were generally in alignment with those found in previous NLP research (e.g., Glenn et al., 2020); that is, negative emotive language increased in the weeks preceding engagement in suicidality and somewhat decreased again in the week immediately following. More specifically, anxiety word use increased 2 weeks before and sadness words increased 1 week before the suicidality event; sadness word use dropped back down to baseline levels by the week following the event, but anxiety words remained at a heightened level 3 weeks after the event, indicating long standing effects of suicidality on anxiety. However, in contrast to previous findings, we did not find increases in overall negative emotion or anger words or decreases in positive emotion words to precede suicidality, which could again potentially be due to the present study being conducted in a BPD population (as opposed to in past NLP work). Furthermore, our results revealed increases in swear words to immediately precede engagement in suicidality (before decreasing shortly after the event), which is a novel finding with respect to precursors to suicidality evident in language. Interestingly, as anger words did not increase at this time point, this increase in swear words prior to suicidality cannot be explained by heightened levels of anger. Rather, it could be perceived that such increases in swear words reflect increases in hostility and/or emotion-related impulsivity, of which may therefore precede suicidality in individuals with BPD.

Intriguingly, the emotion language trajectories surrounding DSH differed considerably from those in proximity to suicidality, and portray a less straightforward picture. In particular, sadness word use increased 2 weeks prior to engagement in DSH, but drastically decreased in the week immediately preceding the event. Combined with generally high levels of anxiety (indicated by relatively high frequencies of anxiety words), this decrease in sadness (words) could potentially reflect a period of "emotional-numbness" or dissociation immediately preceding engagement in DSH – a commonly experienced cause of distress among individuals with BPD, and a frequently evidenced precursor to DSH in this population (see Al-Shamali et al., 2022). Thus, our findings indicate support to those of more traditional research evidencing dissociation as a key precursor to DSH in individuals with BPD using naturalistic linguistic methods, which differs from findings in non-BPD populations showing clear patterns of heightened negative emotion preceding DSH (e.g., Koenig et al., 2021). As for positive emotion, as with suicidality, there were no changes in positive emotion words in proximity to DSH, implying that positive emotion words are not reliable markers of DSH or suicidality in individuals with BPD. Such differences between the present results and those of past NLP work conducted on more general non-BPD populations appear to illuminate distinguishing patterns of emotion-related precursors to DSH between BPD and non-BPD populations.

Moving on from the emotion-relevant precedents to engagement in DSH, other striking patterns of results surround the emotion language changes that occurred following engagement in DSH. That is, anxiety word use did not change (but remained relatively high) over the weeks leading up to the DSH event, but considerably decreased 2 weeks following engagement in DSH, indicating support for the notion that DSH serves affect regulation functions (i.e., reducing anxiety, in this case; Vansteelandt et al., 2017). However, despite decreases in anxiety, anger words were in fact revealed to sharply increase immediately following the DSH event. Interpretating the findings together, these patterns of emotion language trajectories appear to suggest that, while individuals with BPD may engage in DSH in an attempt to regulate their emotions (e.g., anxiety) and relieve dissociation/emotional-numbness, the act of engaging in DSH in fact appears to generate further emotion dysregulation – particularly in the form of heightened anger – indicating majorly maladaptive affect regulation–self-harm cycles in this population.

In addition to the observation of meaningful emotion language trajectories, other insightful findings were revealed when examining changes in social language in proximity to suicidality and DSH. Specifically, other-focused language (i.e., shehe-words, or third-person singular pronouns) decreased in the week immediately preceding engagement in suicidality, potentially indicating less orientation to others due to elevated psychological distress; a finding consistent with that of prior work examining linguistic precursors to suicidality (De Choudhury et al., 2016). Moreover, although other-focused language remained at a relatively lower frequency over the 3 weeks following engagement in suicidality, affiliation words (indicating social connectedness) were found to significantly increase immediately following the suicidality event, which remained at a higher level 2 weeks after the event. These patterns of linguistic trajectories provide some indication that suicidality may, at least to some extent, be serving a social function in BPD (i.e., improving feelings of social connectedness) – consistent with the interpersonal theory of suicide (e.g., Van Orden et al., 2010). Yet, stronger evidence of social functions emerged for DSH. In particular, socially connected language (i.e., we-words, or first-person plural pronouns, and affiliation words) significantly decreased 2 weeks prior to engagement in DSH, but then drastically increased (especially we-words) immediately following the DSH event. Although this increase in socially connected language was short-lived, these results still provide further support to the notion that DSH serves social functions in individuals with BPD (Muehlenkamp et al., 2013), through reviving feelings of social connectedness. Together, our findings confirm those from more traditional research evidencing social dysfunction as a key precursor to both DSH (e.g., Snir et al., 2015) and suicidality (e.g., Gratz et al., 2022) in individuals with BPD, albeit more strongly in DSH. Moreover, the increases in socially connected language following disclosures of suicidality and DSH appears to highlight the supportive nature of the online BPD community. Although, it is also possible that such social support could in fact be inadvertently reinforcing (disclosures of) engagement in these harmful behaviours, which will be discussed in more depth in the subsequent section.

As for the linguistic indicators of self- and cognitive processes, perhaps surprisingly, none of these language categories showed meaningful changes in proximity to engagement in either suicidality or DSH. As indicated by the results illustrating the person-level (i.e., non-temporal) linguistic markers of suicidality and DSH in BPD, it is highly likely that indicators of dysfunctional self-processes (particularly I-words, or self-focused language) are *consistently* prevalent in the natural language of individuals with BPD who engage in suicidality or DSH (reflecting consistently pervasive dysfunctional self-processes), and thus do not change (i.e., increase) in proximity to engagement in these events. Similarly, it could also be inferred that individuals with BPD who frequently engage in suicidality *consistently* use greater absolutist language (reflecting a persistent all-or-nothing thinking style), explaining why we did not find it to increase in the weeks preceding suicidality events. Such language variables are therefore informative in helping to explain why individuals with BPD (frequently) engage in suicidality/DSH on a general level (i.e., resulting from consistently prevalent dysfunctional self- and cognitive processes), but seemingly not for understanding why individuals with BPD engage in these behaviours *when* they do.

Taken in all, the present findings suggest that periods of elevated emotion dysregulation and social dysfunction (evident in language) are, to varying extents, key precursors to suicidality and DSH in BPD. To summarise and integrate, in simple terms, the specific linguistic markers found to precede suicidality and DSH in individuals with BPD, engagement in suicidality was preceded by co-occurring heightened anxiety words and decreased other-focused language (i.e., shehe-words or third-person singular pronouns), shortly followed by heightened sadness words and hostile language (i.e., swear words) as well as posting more frequently to the online mental health forum (refer to Appendix C.5). As for DSH, engagement in DSH was revealed to be preceded by heightened sadness words and decreased socially connected language (i.e., we-words and affiliation words), along with posting less frequently to the online mental health forum (see Appendix C.5), immediately followed by a sharp drop in sadness word use (potentially indicating dissociation). Provided such linguistic trajectories receive further validation, this knowledge could prove crucial with regard to the anticipation and prevention of engagement in suicidality and DSH in individuals with BPD. More broadly, such patterns highlight how dysfunctional emotion and social processes co-occur and most probably interact with one another in triggering engagement in both suicidality and DSH in individuals with BPD, illuminating a pivotal area for future research with respect to investigating interactions between dysfunctional emotion and social processes and their relations to self-harm in BPD.

5.4.2 Online BPD Community Interactions – and the Potential Impact of Such Interactions – in Relation to Suicidality and Deliberate Self-Harm

To investigate effects of interactions of the online BPD community – in the form of number of replies to submissions and post scores – on suicidality and DSH behaviours, we examined dynamics of the online community in relation to both the disclosure of these behaviours on the platform (RQ2) and the temporal psychosocial dynamics (evidenced via linguistic trajectories) that surround these harmful behaviours (RQ3). To re-emphasise, the overarching goal of the RQ2 and RQ3 analyses was to provide initial insight into whether and how online BPD communities may potentially hamper or further drive engagement in suicidality and DSH in individuals with BPD.

Informatively, results from the RQ2 analyses revealed disclosures of recent engagement in suicidality – but not DSH – to be associated with more community support, primarily in the form of upvotes (or post scores). Specifically, such conclusion is evidenced by results showing that individuals who more frequently disclosed recent engagement in suicidality on the online platform received higher post scores, on average, as well as the finding that posts disclosing recent engagement in suicidality received higher scores when compared to non-suicidality posts. In contrast, disclosures of engagement in DSH were not revealed to generate greater community support. In fact, posts disclosing recent engagement in DSH were found to receive fewer replies when compared to non-DSH posts. In integrating these results, such findings suggest that the online BPD community is displaying relatively more supportive behaviour in response to disclosures of (recent) engagement in suicidality (i.e., feeling suicidal), but in fact appears to be displaying less support to disclosures of (recent) engagement in DSH. Usefully, such findings help to explain the increase in socially connected language (i.e., affiliation words) following the disclosure of recent engagement in suicidality on the platform. However, these findings do not help to explain why this increase in socially connected language also occurred after the disclosure of recent engagement in DSH, of which was not associated with greater community support.

Presumably, it is likely that the increase in socially connected language following engagement in DSH is instead due to factors external to the online platform, such as receiving more social support offline (e.g., from friends and family) in response to the self-injurious behaviour.

Most importantly, findings from the RQ2 analyses generate multiple potential hypotheses. First, one possible hypothesis is that the elevated support provided by the online community in response to disclosures of suicidality may help to prevent (or lessen the likelihood of) future engagement in suicidality through providing much needed social support, reassurance, and a sense of social connectedness and belonging; as implied in previous work (De Choudhury et al., 2016). Alternatively, a second, more concerning potential hypothesis that warrants consideration surrounds the possibility that the greater social support provided by the online community in response to disclosures of suicidality could in fact be inadvertently reinforcing (disclosures of) engagement in suicidality, and subsequently further driving engagement in suicidality. Indeed, some empirical support for this hypothesis can be seen from research – particularly fMRI research – showing how “likes” (or upvotes) on social media posts tap into the reward pathway in the brain, with posts with large numbers of likes evidenced to generate activation in brain areas implicated in reward processing, social cognition and social memory, imitation, and attention (Sherman et al., 2016). Based on this evidence, it is certainly possible that the relatively greater number of upvotes provided to posts disclosing recent engagement in suicidality, even though unintentional, could positively reinforce engagement in suicidality. Specifically, such reinforcement could impact those who made the post(s) disclosing suicidality (i.e., those posting about feeling suicidal), as well as those passively viewing suicidality posts, as passive viewing also has the potential to tap into social cognition and reward pathways in the brain, potentially (subconsciously) motivating passive viewers to also (disclose) engage(ing) in suicidality to receive the same positive social response. Although this latter hypothesis is possible, it remains speculative at present, as there is currently no known research that has tested this directly.

Valuably, the results from the RQ3 analyses shed further light on whether and how dynamics of the online BPD community may be (unintentionally) driving, or hampering, engagement in suicidality and DSH. In particular, although there were no significant associations between the number of replies to posts and any of the relevant language variables (i.e., those in which significant fixed effects of time were found in the RQ1 analyses), posts scores were revealed to be positively correlated with frequencies of all linguistic features (i.e., we-words [first-person plural pronouns] and shehe-words [third-person singular pronouns], and anxiety, sadness, anger, and swear words) except for affiliation words. Put simply, the results here illustrate that the online BPD communities are typically scoring posts higher when they display more socially connected and socially oriented language (i.e., we-words and shehe-words, respectively), but also when they display more negative emotive and hostile language (i.e., anxiety, sadness, anger, and swear words). The finding that posts displaying more socially connected/oriented language generate greater support from online mental health communities is supported by previous research (De Choudhury & De, 2014). In contrast, our finding that posts displaying more negative emotive and hostile language also generate greater community support directly contradicts findings from this previous study, as in the previous study (De Choudhury & De, 2014), posts displaying more negative emotive and hostile language (i.e., swear words) generated less supportive responses (i.e., lower post scores) from the online mental health community. Such discrepancies could be explained by differences in the community population studied, as the present study was conducted on online BPD communities, whereas the De Choudhury and De (2014) study investigated a general mental health community. Accordingly, these differences suggest that in (online) general mental health communities, posts expressing negativity and hostility are directly unsupported, whereas displays of such negativity in fact appear to be supported in online BPD communities.

Interpretating the RQ2 and RQ3 (and RQ1) findings together, these patterns of results appear to indicate mixed effects regarding the influence of online BPD communities on suicidality and DSH. In particular, the BPD communities appear to be displaying somewhat of a hampering effect on engagement in suicidality and DSH with respect to implicitly discouraging expressions of social-disconnectedness in posts (through providing more support and attention to posts displaying more social-connectedness), of which was evidenced as a precursor to both suicidality and DSH (but more so for DSH). However, the communities also appear to be portraying a driving effect on engagement in suicidality and DSH through implicitly rewarding (and, in turn, potentially driving) displays of emotion dysregulation (i.e., heightened negative emotion) and hostility (i.e., swearing) – of which were evidenced to precede both suicidality and DSH in RQ1 – by rewarding these more negative posts with more upvotes; an effect that is seemingly specific to BPD communities. Thus, put simply, the online BPD community may be unintentionally and implicitly driving engagement in suicidality and DSH by rewarding displays of emotion dysregulation and hostility that precede disclosures of engagement in these behaviours on the platform. When combined with the results showing that disclosures of suicidality typically generate more community support (and thus are potentially being rewarded/reinforced), it seems that the online BPD communities might be having (albeit unintentional) harmful effects on suicidality in individuals with BPD. Yet, the effects of the communities on engagement in DSH appear more mixed, especially as social disconnectedness/dysfunction was indicated (via language) to have a stronger role in anticipating this behaviour (relative to suicidality), which is discouraged by the communities in the form of less upvotes to posts displaying social disconnection, and disclosures of recent engagement in DSH appear to be unreinforced by the communities (in the form of less replies to posts). Although it might be tempting conclude from these findings that the online BPD communities may potentially be (inadvertently) having harmful effects on engagement in suicidality but no or possibly beneficial effects on engagement in DSH in individuals with BPD, such interpretations remain fully speculative at present, as effects of community interactions on future engagement in suicidality and DSH have not been directly tested; this is a crucial area for future research.

5.4.3 Implications

Findings from the present study have generated numerous theoretical and practical implications. In terms of theoretical implications, the present work contributes to BPD theory in that our findings provide a better and more complete understanding of the psychosocial dynamics of suicidality and DSH in BPD – behaviours that are highly dangerous and strongly associated with borderline pathology (see, e.g., Reichl & Kaess, 2021) – using a large-scale, naturalistic language-based approach. Moreover, the present work also contributes to suicide and DSH theory more broadly, as our findings provide further insight into the trait-level psychosocial correlates of suicidality and DSH, and, vitally, the temporal psychosocial dynamics that surround engagement in these behaviours. In the process, we addressed critical gaps in the literature base by leveraging NLP methods to investigate suicidality and DSH specifically in a BPD population, and while integrating a psychological theoretical perspective to generate meaningful psychological insight; something that has not been done before. Further, we additionally addressed key gaps in the literature by examining the “after-effects” (evident in language) of engaging in suicidality and DSH, as there was previously very little research in this domain. Finally, the present study generated some initial insight into the effects of online BPD communities – of which are widely-used communities that have not, to our knowledge, been empirically investigated before – on suicidality and DSH, of which has sparked some critical hypotheses to be tested in future research.

Regarding the practical implications of the present work, our findings provided much needed insight into the psychosocial markers of suicidality and DSH in a BPD population. Importantly, such insight is of major clinical value given that these psychosocial markers could potentially be monitored by clinicians to identify individuals with BPD who are likely to (frequently) engage in these harmful behaviours, upon which appropriate psychological interventions could be provided. Even more valuably, the present work also generated critical insight into the temporal psychosocial dynamics that precede engagement in suicidality and DSH, which is vital knowledge for clinicians working with individuals with BPD with regard to anticipating when these individuals are likely to engage in such harmful behaviours, and subsequently helping to prevent or lessen the likelihood of self-harming behaviours occurring. Further, additional clinically meaningful findings generated from this study surround the results highlighting periods of heightened emotion dysregulation and social dysfunction as the predominant precedents to suicidality and DSH in individuals with BPD, thus illuminating the most critical areas of dysfunction in BPD to be targeted through therapeutic intervention. Vitally, given that DSH and prior suicide attempts are seemingly the most prominent risk factors for completed suicide (e.g., Hawton et al., 2015), the prevention of suicidality and DSH in individuals with BPD should help to lessen the high risk of completed suicide in this population.

5.4.4 Limitations

Despite the many strengths of the current dataset – including the dataset comprising a large, naturalistic sample of individuals with (expertly annotated) self-identified BPD – it is also inherently accompanied by some biases. Namely, given that all data analysed were collected from Reddit – BPD subreddits in particular – this means that it is possible that our findings may be specific to disclosures of suicidality and DSH on Reddit, and therefore may not generalise to other contexts. Thus, our findings relating to effects of the online BPD communities on suicidality and DSH may be exclusive to the particular BPD Reddit communities studied, rather than being generalisable to other online BPD communities. Yet, given that the combined total of community members of the BPD subreddits investigated in the present study reaches over 300,000 to date, it is likely that the communities investigated are some of the largest – if not the largest – online BPD communities, indicating that it is a representative community to study. Nonetheless, in terms of the relationships evidenced between language and suicidality and DSH in the present study more generally, given that these relationships have been evidenced in the context of the BPD subreddits specifically, they may not be generalisable to language more broadly (e.g., if examining language data from text messages). However, our findings regarding the linguistic markers of suicidality and DSH were largely consistent with previous studies comprising various different language data sources, and are also in alignment with broader BPD suicidality/DSH theory.

Further, another related limitation of the present sample surrounds the fact that the sample comprised individuals with BPD who, at least at one point, were frequent users of Reddit (specifically the BPD subreddits), suggesting that our sample may not be representative of the broader population of individuals with BPD of whom do not (frequently) use (or have not used) Reddit. Moreover, our classification of individuals as having BPD (i.e., the 992 users that comprise our sample) relied upon users' self-identified BPD statements, and thus have not been clinically verified as having a BPD diagnosis. Nevertheless, this is a commonly utilised approach to identifying individuals with a mental health condition online (e.g., Coppersmith et al., 2016), and we also ensured that we were conservative and thorough in our classification process.

In terms of analytic limitations, our analyses were somewhat constrained by the fairly sparse and inconsistent data (i.e., posts) surrounding occurrences of recent engagement in suicidality and DSH, which meant that we could not aggregate our data at a precision level greater than weekly. Accordingly, we were not able to examine changes in language in proximity to suicidality and DSH events at a level of high precision (e.g., daily changes); this highlights an important area for future research.

Moreover, one other analysis-related limitation that occurred as a result of the nature of our dataset is that the exact timing of the suicidality/DSH events was not always clear from users' posts. Thus, when posts were recorded as an occurrence of "recent engagement in suicidality/DSH", this sometimes will have been several days out from the exact day of the occurrence. This was generally not an issue for the coding of recent suicidality events, as posts disclosing recent engagement in suicidality were largely reflected by users disclosing feeling suicidal at the moment of writing the post (e.g., "I feel so suicidal"), or more generally on the same day of writing (e.g., "I haven't been able to stop feeling suicidal today"). However, disclosures of recent engagement in DSH were often less precise with regard to the exact occurrence of the event. Because of this, we took the approach of recording the disclosure of engagement in the behaviour as the day of the event, treating the disclosure itself as the behaviour. Although, posts were only recorded as recent engagement if it was clear from the text that the user had engaged in this behaviour very recently – also note that we undertook a rigorous and conservative manual coding process when coding for suicidality and DSH (refer to the Methods section). Accordingly, the general linguistic trajectories evidenced should therefore be accurate, especially given that they were aggregated on a weekly level (rather than daily, for instance).

Finally, another central limitation of the present work surrounds the fact that we relied on users of the BPD discussion platform being honest and accurate in their disclosures of engagement in suicidality and DSH. Posts in which users disclosed engaging in suicidality/DSH were taken as the ground truth. Yet, it is of course possible that users might occasionally not have been truthful in disclosing engagement in these behaviours. However, given that Reddit is an anonymous platform, and that the BPD subreddits in particular encourage open, honest, and supportive discussion, it would only be in users' best interests to be fully open and truthful in their Reddit posts, especially given that the platform is primarily used for support-seeking, building connections, and the honest discussion of sensitive topics.

5.4.5 Conclusion

In the present work, we leveraged modern natural language processing methods to investigate the psychosocial dynamics of suicidality and DSH in BPD, while also examining online BPD community interactions – and the potential impact of these interactions – in relation to these behaviours. To our knowledge, this is the first study to have provided such a far-reaching, large-scale, naturalistic perspective on both suicidality and DSH *in situ* in a BPD population, while also incorporating meaningful psychological theoretical perspective. Our findings have important theoretical and practical implications, such as the potential to leverage these findings to anticipate (and thus possibly prevent) engagement in suicidality and DSH in individuals with BPD. Moreover, the present study generated some novel initial insight into the effects of online BPD communities on suicidality and DSH in BPD, of which has sparked some critical hypotheses to be tested in future research. We are confident that future empirical work will further improve and refine upon our analyses, generating deeper and more precise insights into the relationship between natural language and suicidality and DSH in BPD, as well as the influence of online BPD communities.

CHAPTER 6:

General Discussion

Personality pathology is a particularly high-risk construct and has high economic costs to society, and thus constitutes a serious societal problem worldwide (e.g., Tyrer et al., 2010). Since we currently still know little with regard to some of the fundamental aspects of personality pathology, the present work sought to improve understanding through the application of computational language analysis. Accordingly, the central, overarching research question guiding this entire thesis was: *How can personality pathology be better understood through the computational analysis of natural language?* To address this question, four empirical studies were conducted (portrayed in three articles) – each with their own aims and research questions – that leveraged computational language analysis methods to better understand personality pathology: Study 1 (Paper 1; Chapter 3) focused on interpersonal dysfunction; Studies 2 and 3 (Paper 2; Chapter 4) focused on emotion dysregulation; and Study 4 (Paper 3; Chapter 5) focused on behavioural dysregulation.

With regard to the specific thesis research questions that guided each of the study chapters, Chapter 3 (Paper 1; Study 1) was guided by the research question: *How can language provide insight into the characterising dimensions of interpersonal dysfunction in PD?* (i.e., RQ1 of the thesis). Study 1 addressed this research question by analysing people's natural language use in written essays about their interpersonal relationships to uncover core social-cognitive dimensions, of which were subsequently used to characterise interpersonal dysfunction in BPD. Chapter 4 (Paper 2; Studies 2 and 3) aimed to address the thesis research question (RQ2): *To what extent, and how, is emotion dysregulation in PD reflected in natural emotion vocabularies?* To achieve this, Studies 2 and 3 analysed natural language use in written essays (Study 2) and spoken conversations between women diagnosed with BPD and their romantic partners (Study 3) to describe the natural emotion vocabularies (i.e., the variety of emotion words actively used) associated with BPD, providing insight into the maladaptive emotion processes that may contribute to emotion dysregulation in PD. Finally, Chapter 5 (Paper 3; Study 4) aimed to address the thesis research question (RQ3): *How can natural language be used to better understand, and subsequently prevent, self-harm in PD?* In a larger-scale naturalistic study, Study 4 addressed this research question by examining the natural language of individuals with self-identified BPD on online BPD discussion forums. These data were subsequently analysed to better understand the psychosocial dynamics (evident in language) of self-harm (i.e., suicidality and deliberate self-harm [DSH]) in BPD, uncovering key relationships between language (reflective of psychosocial processes) and engagement in these maladaptive behaviours. Additionally, this work revealed meaningful interactions between the online BPD community and engagement in suicidality and DSH in this population.

6.1 Summary and Integration of Findings

Here, I provide an integrative overview of the key findings from the present work and discuss how they have enabled better understanding of the nature of personality pathology (the central thesis research question). In terms of providing better understanding of the nature of social dysfunction in PD (RQ1), in first study (Study 1; Chapter 3), using language analysis methods, we uncovered four social-cognitive dimensions that characterised social dysfunction in BPD: (less) *Connectedness/Intimacy*; *Immediacy; Social Rumination;* and *Negative Affect*. Critically, although our findings that emotion dysregulation, problems with intimacy, and immediacy or impulsivity characterise social dysfunction in BPD are consistent with previous research findings (e.g., Euler et al., 2019; Jeung & Herpertz, 2014; Koenigsberg et al., 2001), our finding that social dysfunction in BPD is characterised by social rumination appears to be a novel finding in this realm. Intriguingly, it was revealed that problems with intimacy and affect appear to characterise social dysfunction across a range of problematic interpersonal constructs (i.e., the Dark Triad traits), whereas immediacy and social rumination are more specific to BPD. Furthermore, our findings indicated that social rumination may distinguish interpersonal dysfunction in BPD from other problematic interpersonal constructs, as social rumination was evidenced to be exclusive to BPD (relative to the Dark Triad traits), suggesting that this may be a critical component in understanding the prolonged and pervasive social impairments typically experienced by individuals manifesting BPD. Importantly, the present findings addressed the lack of consensus surrounding the core psychological features that characterise social dysfunction in BPD, providing much needed clarity to this theoretical knowledge base, while also illuminating a fundamental component that may differentiate social impairments in BPD from social dysfunction more broadly. On a more general level, this work illuminates how computational analysis of natural language can be used to describe fundamental components of personality pathology.

Moving on to improving knowledge of the emotion dysregulation facet of PD (RQ2), our findings from Studies 2 and 3 (Chapter 4) describing the natural emotion vocabularies (EVs) associated with BPD provide important insight in this respect. In particular, in two studies, BPD was revealed to be associated with relatively large negative EV (i.e., greater diversity in negative emotion word use), and this relationship was generally insensitive to context. Subsequently, the spontaneous, yet inflexible, use of a relatively broad range of negative emotion words among individuals manifesting BPD likely reflects extensive experience and preoccupation with negative emotion (in alignment with linguistic theory; e.g., Zipf, 1949) and a (maladaptive) form of expertise in negative emotion (see Vine et al., 2020) in which emotion concepts are not implemented in a context-sensitive way. Such interpretations therefore indicate possible mechanisms underpinning maladaptive emotion processes in BPD, of which may directly contribute to emotion dysregulation. Our findings thus generated critical insight into one of the defining features of BPD – emotion dysregulation – therefore providing an invaluable contribution to the personality pathology knowledge base. Moreover, this work demonstrates how a broad range of language-based methods can be leveraged to generate a deeper and more precise understanding of core features of personality pathology, beyond solely focusing on general word frequencies.

As for the behavioural dysregulation that typically accompanies PD, our findings from the final study (Study 4; Chapter 5) generated much needed insight into the psychosocial dynamics surrounding engagement in dangerous maladaptive (self-harming) behaviours (i.e., suicidality and DSH) in BPD (RQ3). Specifically, our findings revealed key person-level psychosocial correlates of suicidality and DSH in BPD, including an indication that expressions of hostility (evidenced via hostile language) may be associated with engagement in suicidality in this population (a novel finding). Furthermore, the present findings generated insight into the temporal psychosocial dynamics (evident in language) surrounding engagement in suicidality and DSH in BPD, illuminating individualised, psychological changes that occur in the weeks immediately leading up to engagement in these harmful behaviours in this population. In particular, affective and social processes were revealed – via linguistic indicators – to undergo the most prolific changes (i.e., indicators of greater emotion dysregulation and social disconnectedness) preceding engagement in both suicidality and DSH. Such findings thus shed light on the combination of maladaptive psychological processes that may result in individuals with BPD, or PD more broadly, engaging in self-harm. As for other stand-out contributions, our study provided initial insight into how psychological processes are affected following suicidality and DSH in individuals with BPD – an area that has received little empirical attention. Importantly, this work additionally evidenced meaningful interactions between the online BPD community and suicidality and DSH behaviours, in terms of both the disclosure of these behaviours on the platform as well as the temporal psychosocial dynamics that surround these behaviours. Accordingly, these findings generated initial insight into the effects of online BPD communities on suicidality and DSH, which sparked some critical hypotheses to be tested in future research. Essentially, the present study generated a pivotal, fuller understanding of the nature of self-harm – a form of dysregulated behaviour strongly associated with PD – in BPD, thus adding a significant contribution to personality pathology theory in respect to the behavioural dysregulation aspect.

Integrating the present findings together, through leveraging computational naturalistic linguistic methods, all of the studies conducted generated better understanding of fundamental features of PD, including three central areas of dysfunction: social dysfunction, emotion dysregulation, and dysregulated behaviour, thereby demonstrating the potential of language analysis to provide novel insight into the nature of personality pathology. More specifically, this work has provided further support to the conceptualisation of (B)PD as predominately defined by social, emotional, and behavioural dysfunction via a naturalistic behavioural approach, as our integrated findings highlight the prominence of these dysfunctions to BPD. Subsequently, the present findings have provided a significant contribution to the PD literature base, using methodology not typically applied to this field, thus developing understanding from an alternative perspective. Given that personality pathology is a global problem that results in major costs to society, and widespread suffering to those directly affected by PD, the personality pathology knowledge that has resulted from the present work is of crucial importance. Undoubtably, the present findings have sparked numerous meaningful theoretical and practical implications, which will be discussed in the following section.

6.2 Implications

In this section, I integrate the findings of the studies conducted to bring together the main contributions and implications of the present work for theory (Section 6.2.1) and practice (Section 6.2.2).

6.2.1 Theoretical Implications

The present work generated numerous theoretical implications, including implications for personality pathology theory – as well as more general personality and psychopathology related theory – and implications for linguistic theory and the broader language literature.

6.2.1.1 Informing Theoretical Models of Personality Pathology

Findings from the present work that have generated better understanding of the nature of personality pathology subsequently have implications for theoretical models of PD. In particular, findings from all of this research support the conceptualisation of BPD as primarily defined by emotion dysregulation (e.g., Crowell et al., 2009; Linehan, 1993; Sauer-Zavala & Barlow, 2014). Specifically, findings from Study 1 imply that one of the fundamental characterising components of social dysfunction in BPD is emotion dysregulation (i.e., the *Negative Affect* component), which is consistent with the notion that affective dysregulation is a fundamental driver of interpersonal dysfunction – and potentially all dysfunctions – in BPD (e.g., Euler et al., 2019; Lazarus et al., 2014). Building on this, Study 4 uncovered indicators of emotion dysregulation to undergo the most prolific changes (i.e., increases) in proximity to engagement in both suicidality and DSH, positioning dysfunctional emotion processes as the predominant causal factor in relation to these behaviours (e.g., Hatkevich et al., 2019; Vansteelandt et al., 2017). Moreover, also emphasising the centrality of emotion dysregulation to

BPD, Studies 2 and 3 illuminate how maladaptive emotion processes are prevalent, longstanding, and continuous in individuals manifesting BPD, given that BPD was found to be associated with an active use of a broad range of negative emotion words (reflecting extensive experience and preoccupation with negative emotion; see Vine et al., 2020), irrespective of context. Such findings therefore highlight the pervasive, context-insensitive nature of emotion dysregulation in BPD. Taken together, the present work provides further support to the theoretical consensus that emotion dysregulation is the most central, defining feature of BPD – and possibly PD more broadly – and may in fact be the underlying driver of all other areas of dysfunction in BPD, as proposed in numerous theoretical BPD models (e.g., Linehan's, 1993, biosocial theory of BPD).

Focusing on emotion-relevant theories of PD more specifically, Studies 2 and 3 explicitly connect with clinical theoretical models of BPD in helping to explain the relationship between emotion processes and functioning and the construct of BPD. In particular, Linehan's (1993) biosocial theory of BPD places emotion dysregulation right at the centre of dysfunction in BPD, in which emotion (and emotion dysregulation) is operationalised broadly to encompass aspects such as emotion-related cognitive processes, biochemistry and physiology. In such theory, cognitive processes linked to emotion are proposed as pivotal to emotion functioning, with the way in which an individual understands and thinks about their emotions playing a major role in the subsequent impact of these emotions. For instance, according to Linehan's theory, negative rumination about experiences of negative emotion would consequently result in exacerbated negative affect. In a similar vein, the Emotional Cascade Model of BPD (Selby & Joiner, 2009) also emphasises the importance of maladaptive cognitive processes related to emotion – particularly negative rumination – in moderating the impact of such emotions. Specifically, this theory proposes that, following emotion activation, (negative) ruminative processes result in a positive feedback cycle that increases emotional intensity. Thus, both prominent clinical theories suggest that the way in which individuals with PD think about their (negative) emotions will greatly influence their experiences of said emotions. In interacting with these theoretical models, findings from Studies 2 and 3 can be interpretated to highlight how such broad negative emotion vocabularies associated with BPD may contribute to emotion dysregulation by driving emotional cascades and negative rumination cycles, subsequently exacerbating negative affect. For instance, frequently using a wide range of negative emotion words in everyday life may result in greater attention to, and subsequently rumination around, these negative emotions. The present findings have therefore provided a meaningful addition to emotion-relevant clinical models of BPD by indicating possible mechanisms underpinning prominent theories of emotion dysregulation in BPD.

As with the present work supporting the conceptualisation of BPD (or PD more broadly) as predominantly defined by emotion dysregulation, findings from this research also support the established link between personality pathology and social dysfunction (e.g., Hill et al., 2008; Miano et al., 2020), highlighting social impairments to be an additional fundamental feature (even if driven by affective dysregulation) of PD. That is, Study 1 revealed BPD to be significantly associated with *all* four social-cognitive components in ways indicative of social dysfunction, therefore demonstrating the centrality of social impairments to BPD. Comparatively, any given Dark Triad trait (as well as the Dark Triad as a whole construct) was only found to correlate with a maximum of two of the four social-cognitive components indicating social dysfunction. Such findings imply that social dysfunction is less of a pervasive issue for subclinical problematic personalities than it is for pathological forms of personality. In addition, findings from Study 4 provide further support to the notion that social dysfunction is a central feature of PD, as they revealed increases in indicators of social disconnectedness to precede engagement in both suicidality and DSH in individuals with BPD. Consistent with broader PD theory (e.g., Hill et al., 2008), the present findings support social dysfunction as an additional central feature of PD, and, as with emotion dysregulation, a possible key driver of other forms of dysfunction in PD – such as dysregulated behaviour – thus informing the nature and structure of personality pathology.

An additional meaningful contribution of the present work surrounds adding further support to the consensus that PDs – and BPD in particular – largely stem from invalidating, traumatic, or otherwise adverse early experiences (e.g., Bozzatello et al., 2021; Linehan, 1993). Namely, Study 1 uncovered characterising components of social dysfunction in BPD that suggest early adverse experiences play a fundamental role in the interpersonal dysfunction experienced by individuals with BPD. That is, the component highlighting problems with intimacy (i.e., [lack of] *Social Connectedness/Intimacy*) as a characterising feature of social dysfunction in BPD likely originates from insecure attachment styles – of which are strongly associated with PD (e.g., Lorenzini & Fonagy, 2013) – given that intimacy problems are a typical presentation of insecure attachment. Moreover, insecure attachment (and, in turn, problems with intimacy) is closely related to adverse childhood experiences (e.g., Erozkan, 2016). Accordingly, the finding that social dysfunction in BPD is characterised by problems with intimacy suggests that early adverse experiences and childhood environments may be at the root of such interpersonal dysfunction, mediated by insecure attachment styles. Relatedly, the *Social Rumination* component also revealed to characterise social dysfunction in BPD in Study 1 additionally points to early adverse or traumatic experiences explaining such dysfunction, with this component reflecting the negatively valenced past-orientated nature of individuals with BPD (e.g., Miano et al., 2020), indicating that social impairments may result from an over-fixation on processing (early) past (traumatic or negative) relationships and events.

Furthermore, supporting the notion that PDs originate from early adverse experiences in a more indirect way, findings from Studies 2 and 3 also shed some light in this respect. In particular, given that the broad active negative EVs found to be associated with BPD are perceived as reflective of extensive experience with negative emotion (e.g., Zipf, 1949), it is most plausible that such extensive experience with negative emotion will have originated from childhood experiences. In alignment with this hypothesis, the fact that the associations between BPD and negative EVs were stronger and more robust in the sample comprising individuals with diagnosed BPD (Study 3) compared to those with lower levels of BPD traits in the general population (Study 2) suggests that individuals with pathological levels of BPD have considerably more intense and exhaustive experience with negative emotion, of which would have presumably developed at an earlier point in life (i.e., childhood) to have become so extensive. Indeed, in Linehan's (1993) biosocial theory of BPD, BPD was proposed to develop from an invalidating early environment, resulting in the child not learning how to efficiently understand, label, or regulate their emotions, thus triggering affective dysfunction (the defining feature of BPD). The present findings are therefore in alignment with Linehan's (1993) biosocial theory of BPD – and numerous other conceptualisations of BPD – in suggesting that BPD typically develops in response to early adverse experiences, which subsequently results in core areas of dysfunction (i.e., affective and social dysfunction).

Undoubtably, findings generated from the present work have stimulated numerous implications for personality pathology theory, including developing a better understanding of the nature of PD and explicitly informing theoretical models of PD. Given that personality pathology lies at the intersection between general personality and psychopathology (as emphasised in Chapter 1.1), the present work also has implications for general personality and psychopathology theory.

6.2.1.2 Implications for General Personality Theory

In adopting the now more widely supported dimensional, trait-based approach to personality pathology (i.e., along a continuum of normal-abnormal personality; for empirical evidence see, e.g., Wright et al., 2016), PD classification is based on a global evaluation of personality functioning, in comparison to arbitrary symptom thresholds. Subsequently, this spectrum-based approach to personality dysfunction implies that better understanding of pathological forms of personality will also have implications for understanding of non-pathological, or less extreme, forms of personality (dysfunction), and vice-versa, given that normative and pathological forms of personality are not perceived as being qualitatively, or categorically, different (see, e.g., Hopwood et al., 2018). Based on this operationalisation, all of the findings generated from the present work also inherently have implications for general, or normative, personality theory.

Providing some support to the dimensional approach to personality pathology, our findings from Study 1 revealed that several of the social-cognitive dimensions found in the context of personality pathology majorly overlapped with social-cognitive dimensions evidenced in a study conducted in the realm of normative personality (Pennebaker & King, 1999). Accordingly, similarities between our findings and findings from general personality research suggest that such social-cognitive dimensions may explain social (dys)function in PD and in normative personality more broadly, thereby supporting the current consensus that PDs are dimensional in nature (e.g., Wilmot et al., 2019). In addition, findings from Studies 2 and 3 also provide some support to the dimensional approach to personality pathology (and psychopathology more generally). In particular, although the relationship between BPD and larger negative EVs was found to be stronger and more robust in the clinical sample of individuals diagnosed with BPD (Study 3), this relationship was also evidenced in the general population sample (Study 2), suggesting that reflections of experience with negative emotion in emotion vocabularies may be “dose dependent”, rather than reflecting categorically different emotion processes between these groups. Furthermore, when employing a dimensional analytic approach (i.e., correlating BPD symptom levels with active EV scores) in Study 3 (comprising the clinical BPD sample), the results revealed the exact same relationships between BPD and active EVs as when employing the categorical analytic approach (i.e., comparing active EV scores between BPD and non-BPD groups). Taken together, findings from the present work provide further support to the move towards dimensional approaches to personality pathology, thereby generating additional implications for theoretical models of PD. Moreover, support for dimensional conceptualisations of personality pathology inherently makes our findings relevant to and informative for general personality literature and theory.

6.2.1.3 Implications for Broader Psychopathology Theory

Similar to the generalisability of the present work to the realm of normative personality, our findings also naturally have a bearing on the broader field of psychopathology. In particular, given its heterogenous nature combined with high rates of comorbidity, BPD has been conceptualised by some scholars as reflective of general psychopathology (see, e.g., Gluschkoff et al., 2021; Wright et al., 2016). Combining this with the dimensional conceptualisation of personality pathology, and psychopathology more broadly (e.g., Hengartner & Lehmann, 2017), all of the present findings are thus informative for psychopathology literature and theory, in the broadest sense.

Most relevantly, findings from Studies 2 and 3, as well as Study 4, have generated explicit implications for the psychopathology knowledge base. Specifically, in providing an initial glance into the potential transdiagnostic nature of the relationship between active EVs and BPD via examining these associations in depression in Study 3, the associations found between active EVs (i.e., larger negative EVs) and BPD were indeed revealed to extend to depression. Accordingly, the generalisability of the findings to depression provides an initial indication that these EV patterns may in fact be shared across numerous mental health conditions, consistent with transdiagnostic approaches to mental health (e.g., Dalgleish et al., 2020). Further, this finding is also consistent with that of related research in which depression symptoms were found to be positively correlated with negative EV (Vine et al., 2020). Overall, such findings indicate that the broad range of negative emotion words found to be spontaneously used in individuals manifesting BPD may be typical of individuals with various forms of psychological distress associated with extensive experience with negative emotion, rather than necessarily specific to BPD. Subsequently, the theoretical implications generated from Studies 2 and 3 may be generalisable to the broader psychopathology knowledge base. Yet, further research in other samples and clinical groups is undoubtably required to confirm the potential transdiagnostic nature of the patterns evidenced from the present work.

In a similar vein, findings from Study 4 also have implications for psychopathology theory, and suicidality and DSH theories specifically. Although Study 4 focused on investigating psychosocial dynamics of suicidality and DSH in a BPD population, it is likely that many of the findings are generalisable to suicidality and DSH more generally, which is of importance given that suicidality and DSH are associated with all forms of psychopathology (see, e.g., Kaess, 2022). In support of the generalisability of the present findings, a large portion of the findings regarding the person-level psychosocial correlates of and temporal psychosocial dynamics surrounding engagement in suicidality and DSH were consistent with those found in non-BPD populations (e.g., Glenn et al., 2020). Such consistency in results suggests that the present findings are insightful for suicidality and DSH theory in general, providing a greater and more complete understanding of these harmful behaviours. For instance, our findings provide support to the theoretical and empirical notion that both suicidality and DSH serve affect regulation (e.g., Hatkevich et al., 2019) and social functions (e.g., Muehlenkamp et al., 2013), thus providing a significant contribution to the self-harm knowledge base. Nonetheless, it is important to point out that some meaningful discrepancies emerged between findings from the present study conducted in a BPD population relative to findings from previous research on suicidality/DSH in non-BPD populations (also evidenced in past work), indicating that there may in fact be some distinguishing psychosocial dynamics of suicidality and DSH between BPD and non-BPD populations.

6.2.1.4 Implications for Linguistic Theory and the Broader Language Literature

Valuably, findings from all of the studies carried out have also generated meaningful implications for linguistic theory and the broader language literature. That is, given that computational language analysis has been employed in all studies to generate insightful findings, all of the present work demonstrates the potential of language analysis methods to develop greater psychological understanding. Accordingly, all of the present findings contribute to the broader language literature. With regard to specific implications, the similarities between the findings from Study 1 regarding the core social-cognitive dimensions uncovered and the social dimensions found in the traditional Pennebaker and King (1999) study highlight how language-based dimensions of thought can, at least to some extent, be reliably replicated across samples. Such overlap therefore establishes computational language analysis as a reliable technique in identifying fundamental psychosocial dimensions. Likewise, overlap in the linguistic markers of suicidality/DSH evidenced in Study 4 with linguistic markers of these behaviours found in previous NLP research demonstrates the reliability of the predictive utility of computational language analysis in the domain of psychopathology. Additionally, the results from Studies 2 and 3 were also consistent with the patterns of associations found between active EVs and psychosocial functioning and experiences in the previous Vine et al. (2020) study. The present research has therefore provided a valuable contribution to the language literature in further evidencing the reliability of language analysis techniques.

As for implications that directly relate to linguistic theory, findings from Studies 2 and 3 generate support for linguistic theory in proposing that language is reflective of experience, and one’s underlying psychology more broadly (e.g., Pennebaker, 2011; Zipf, 1949). More specifically, linguistic theory proposes that the emotion words one spontaneously uses in everyday life should correspond with one’s typical or frequent experiences. Accordingly, this theory was confirmed by findings from both Studies 2 and 3, given that BPD – a construct strongly characterised by heightened negative affect (e.g., Chu et al., 2016) – was found to be associated with relatively frequent use of negative emotion words in general, as well as the spontaneous use of a broad variety of negative emotion words. Such findings thus provide clear support to the notion that habitual experience is reflected in natural language, thereby providing further validation to linguistic theory. Moreover, the results from Study 4 also provide some support to linguistic theory in confirming the face validity of computational linguistic measures. In particular, the psychosocial correlates of and temporal psychosocial dynamics surrounding suicidality and DSH found in the present study were consistent with findings in this realm from more traditional research (e.g., Gee et al., 2020; Halicka & Kiejna, 2018), as well as suicidality/DSH theory (e.g., Vansteelandt et al., 2017; Van Orden et al., 2010). Accordingly, findings from the present work have thus supported linguistic theory in demonstrating the validity of computational language analysis methods, as well as empirically supporting the basic theoretical principles behind the methodology.

Importantly, Study 4 additionally demonstrated how language analysis can be used in conjunction with other (non-verbal) behavioural measures to generate insight into the effects of online communities on individuals' psychological processes and behaviour. Such work therefore adds a significant contribution to the broader language literature in further demonstrating the versatility of computational language analysis in understanding psychological phenomena. Taken in all, findings from the present work have generated numerous important implications for theory, including personality pathology, general personality, and psychopathology theory, as well as linguistic theory and the broader language literature. Vitally, these theoretical implications subsequently inform and progress into practical implications.

6.2.2 Practical Implications

Regarding practical implications of the present work, given that this research has all been conducted in the context of BPD, the findings have important implications for clinical practice. Furthermore, our findings also have implications regarding guidance on the use of online PD support platforms.

6.2.2.1 Informing Clinical Practice

Informing Psychological Therapy. Findings from the present work are of clinical relevance given that they have the potential to inform therapeutic interventions. For instance, findings from Study 1 revealed four fundamental components of social dysfunction in BPD, which could, in turn, be precisely targeted in therapeutic interventions to improve social functioning in BPD, or potentially PD more broadly. In particular, the social-cognitive component found to be associated with BPD exclusively – *Social Rumination* – suggests that the treatment of longstanding patterns of social dysfunction (as typically seen in PD) may benefit from directing therapeutic attention towards individuals' past relationships and experiences, with the aim of shifting their focus from past negative or traumatic relationships to developing new healthy relationships. Consistent with this idea, Schema Therapy (ST) – a commonly employed therapeutic treatment for BPD (see Sempértegui et al., 2013) – somewhat incorporates this notion of identifying and shifting the attentional focus from past traumatic experiences to being more present-day orientated, to improve functioning. Yet, ST primarily focuses on improving general cognitive functioning, with less emphasis placed on past relationships and interpersonal functioning specifically. Adapting the focus of ST to place more emphasis on the role of past relationships and improving social functioning may prove effective in treating longstanding patterns of social dysfunction in personality pathology. Additionally, the other core characterising dimensions of social dysfunction in BPD uncovered in this study could also be specifically targeted through therapeutic intervention in a similar fashion.

Relatedly, findings from Studies 2 and 3 also generated meaningful implications for therapeutic interventions. Specifically, given that BPD was found to be associated with the spontaneous use of more varied negative emotive language – likely reflecting extensive experience with negative emotion and, potentially, a (maladaptive) type of expertise in which emotion concepts are not implemented in a context-sensitive way – these large, context-insensitive negative EVs emphasise the need for more regulation of the referenced negative emotions. Thus, it may be helpful for clinicians to work with individuals manifesting (B)PD to encourage them to explicitly attend to the way in which they spontaneously refer to their emotions in everyday life. Indeed, developing self-understanding and better control over emotions are important personal recovery goals for individuals with BPD (Katsakou et al., 2012). Moreover, although the present findings revealed no associations between BPD and positive EV, given the positive association found between positive EVs and psychosocial health in Vine et al. (2020), it may be beneficial to attempt to expand the range of positive emotion words actively used by individuals with PD through therapeutic treatment, but this requires further investigation.

Other important implications of the present work with respect to informing psychological therapy can be drawn from the Study 4 findings. Valuably, this study provided insight into the psychosocial markers of engagement in suicidality and DSH in BPD, of which could be monitored by trained clinicians to identity individuals with (B)PD who are likely to be (frequently) engaging in these harmful behaviours, upon which appropriate psychological interventions could be provided to treat such self-harming behaviour. Even more usefully, Study 4 also generated much needed insight into the temporal psychosocial dynamics that precede engagement in suicidality and DSH in individuals with BPD, which is critical knowledge for clinicians working with people with BPD as these psychosocial dynamics could also be monitored (e.g., through explicit assessment or passive observation) to anticipate when these individuals are likely to engage in such harmful behaviours. Consequently, knowledge surrounding when individuals with PD are likely to engage in suicidality or DSH could help to prevent or lessen the likelihood that these individuals will go on to engage in these behaviours. In addition, other clinically meaningful findings generated from this work surround those indicating that it is the emotion dysregulation and social dysfunction features of BPD that predominately precede engagement in suicidality and DSH in this population, thus illuminating *the* most critical areas of dysfunction in PD to be targeted through therapeutic intervention; focusing therapeutic attention towards affective and social dysfunction in individuals with PD who self-harm could subsequently help to reduce engagement in self-harm in this population. Vitally, given that engagement in DSH and prior suicide attempts are arguably the most prominent risk factors for completed suicide (e.g., Hawton et al., 2015), the prevention of engagement in suicidality and DSH in individuals with BPD should help to lessen the high risk of completed suicide in this population.

Detecting Personality Pathology (Features) from Language. In addition to informing therapeutic interventions, the present work has also generated clinical implications with respect to the possibility of identifying personality pathology (features) from natural language. For instance, Study 1 uncovered core language-based social dimensions of thought that reflect social dysfunction in BPD, of which could subsequently be used in detecting the presence of social dysfunction from natural language data in a social context. Usefully, given that the *Social Rumination* language-based component was associated with BPD exclusively, the prevalence of this component in socially-relevant language could be utilised as an additional identifying linguistic marker of BPD. Likewise, Studies 2 and 3 revealed emotion language patterns associated with BPD – namely, large, context-insensitive negative EVs – which could also be used in the identification of BPD from natural language. Yet, as indicated in Study 3, these patterns of emotion language may in fact be transdiagnostic, and so further research on this is required before active EV patterns could be incorporated in the identification of PD from language.

Rather than uncovering language patterns that could be incorporated in the identification of PD in a general sense, informatively, Study 4 revealed linguistic markers of maladaptive (self-harming) behaviours associated with PD. In particular, findings from this study have further validated, and in some cases uncovered, person-level linguistic markers of suicidality and DSH specifically in a BPD population. Such markers could in turn be utilised by clinical practitioners to identify (i.e., risk profile) individuals with BPD who are more likely to be (more frequently) engaging in suicidality and DSH from examining their natural language. Even more invaluably, the present study also provided novel insight into temporal linguistic precursors to engagement in suicidality and DSH in BPD, which could again be used by clinical practitioners to anticipate when individuals with BPD are likely to (or are at risk of) engage in these harmful behaviours from their natural language and potentially prevent them from occurring.

Taken together, the present findings illuminate patterns of verbal behaviour associated with B(PD), which could, pending further investigation, be incorporated alongside other digital behavioural traces to identify PD from natural language, or even potentially predict the onset of PDs. Importantly, the ability to reliably, accurately, and unobtrusively detect the presence of PD from language would help to reduce the problem of misdiagnoses and allow for appropriate treatment to be provided. The implications would be even more impactful if it were possible to predict, with good accuracy, whether an individual is likely to develop a PD before its occurrence as this would allow for early intervention, which is strongly associated with better outcomes for people with mental health conditions (e.g., Birchwood et al., 2000). In terms of sources of language data that could be examined to identify personality pathology (features) from natural language, these could include (provided consent has been granted): social media posts, text messages, emails, blog posts, diary posts, and practitioner prompted free-response writing tasks.

6.2.2.2 Informing Expert Clinical Guidelines on the Use of Online PD Support Platforms

Further to the numerous implications for clinical practice generated from the present work, findings from Study 4 specifically also have implications regarding informing clinical guidelines on the use of online PD support platforms. That is, Study 4 generated some initial insight into the effects of online BPD communities – of which are widely-used communities that have not been empirically investigated before – on suicidality and DSH in BPD. In particular, our findings provide an initial indication of the possibility that online BPD communities (or at least, those studied on Reddit in the present work) may be inadvertently having harmful effects on engagement in suicidality through reinforcing behaviour (i.e., upvotes), but no or possibly even beneficial effects on engagement in DSH, in individuals with BPD. However, it is important to note that such interpretations remain fully speculative at present, as effects of community dynamics on future engagement in suicidality and DSH have not been directly tested; this is a therefore a crucial area for future research. Nevertheless, this research provides an essential initial building block for future research to develop and more explicitly and precisely examine the effects of online BPD communities in relation to engagement in harmful behaviours. If future research indeed finds online BPD communities to be having (likely inadvertently) harmful effects on suicidality and/or DSH, such findings would have direct implications for informing clinical guidelines on the usage of such online support platforms.

6.3 Strengths and Limitations

Although many of the strengths and limitations of the present research have been indicated throughout this thesis, in this section, I explicitly highlight the key points. Furthermore, I also discuss some of the central ethical considerations relevant to the present work, particularly with respect to the collection and analysis of Reddit data in Study 4.

6.3.1 Key Strengths and Contributions of the Present Work

First and foremost, one of the central strengths of the present work surrounds the relatively novel methodological application of computational language analysis to the study of personality pathology. As can be seen from the book chapter presented in the introduction of this thesis (Chapter 1.1) and the literature review chapter (Chapter 2), there is currently limited work that has applied computational linguistic techniques to the study of personality pathology. Valuably, in employing this novel methodological approach, the present research has allowed for new insights into personality pathology via naturalistic, individualised, behavioural methods that are less subject to biases when compared to more traditional methods, such as self-report. For instance, when employing natural language analysis methods, individuals will typically not be aware that they are being evaluated based on their language patterns, or sometimes even that they are being empirically studied at all (as in Study 4), leaving little room for demand characteristics or social desirability biases. This methodological advantage is even more valuable when investigating personality pathology specifically, given that PD is associated with major identity disturbance and self-awareness issues; employing a naturalistic methodological approach thus takes the pressure away from participants in comprising the self-awareness and self-knowledge required to accurately report on their characteristics, thoughts, feelings, and experiences. Taken in all, the present work brings together the success of computational language analysis in the fields of psychopathology and personality psychology (emphasised in Chapter 1.1) to illuminate the promising potential of these methods in generating a better understanding of the complex nature of personality pathology.

Building on this further, the current research demonstrated the far-reaching potential of computational language analysis methods by going beyond the limits of previous language-based work applied to the study of PD, of which predominantly involved simply counting general word frequencies and correlating them with personality pathology features (see literature review; Chapter 2). That is, the present work utilised a range of varied, more sophisticated, and psychologically insightful applications of computational language analysis to better understand personality pathology. Specifically, in Study 1 we progressed a step further from counting general word frequencies to incorporating such linguistic features in a factor analysis. Importantly, although common in the field of normative personality psychology (e.g., Pennebaker & King, 1999), language-based factor analytic techniques had not been applied to the domain of PD before, and this approach subsequently allowed for fundamental social dimensions of thought to be uncovered that help to better understand social dysfunction in PD. Additionally, in Studies 2 and 3 we also went beyond simply counting general word frequencies to more precisely examining variability in (emotion) word use (i.e., active EVs) – a recently developed language-based measure of emotion processes (introduced in Vine et al., 2020) that had not previously been applied to the field of psychopathology. Finally, in Study 4 we advanced beyond examining associations between language and personality pathology features to using linguistic features to track changes in psychosocial processes in proximity to suicidality and DSH, which generated important clinical implications. Moreover, Study 4 also demonstrated how language-based measures can be combined with non-verbal behavioural measures to provide insight into (maladaptive) behavioural patterns in personality pathology. Thus, the present work exemplified some of the limitless possible applications of computational language analysis in improving our understanding of personality pathology, by going beyond simply examining associations between PDs and language (based on counting general word frequencies).

In addition to methodological strengths relating to measurement tools, other methodological strengths of the present work surround the samples used, particularly the clinical diversity of the samples investigated (i.e., using a combination of non-clinical and clinical samples). That is, the present research comprised mixtures of non-clinical (Study 1 and Study 2), clinically verified (Study 3), and self-identified clinical samples (Study 4). Critically, the inclusion of clinically diverse samples permits a more complete picture of personality pathology, ranging normative–problematic–pathological personality traits present in the general population, up to clinically diagnosed PD. Moreover, the broad range of levels of personality pathology in the present samples is consistent with the dimensional, continuum-based approach to PD, and is thus in alignment with the theoretical and empirical consensus that PDs are dimensional in nature (e.g., Hopwood et al., 2018). The inclusion of clinical PD samples in the current research is also of major value, in that it ensures our findings are relevant and generalisable to individuals with clinical levels of personality pathology. Accordingly, the range of levels of (borderline) personality pathology across the samples in the present research subsequently enhances the generalisability of our findings to the broad construct of BPD.

Relatedly, a further strength of the samples investigated in the current research is that they are generally large in size, particularly when compared to traditional sample sizes of research in the PD (or mental health more broadly) field. Specifically, Study 1 and Study 2 comprised a large non-clinical sample of 530 individuals (this was the same sample), and Study 4 comprised an even larger clinical sample of 992 individuals with (expertly annotated) self-identified BPD. Notably, a sample size of almost 1,000 individuals with clinical PD is rare to achieve, especially using traditional methods. Thus, the sample in Study 4 presents one of the largest (if not *the* largest) expertly annotated clinical PD datasets, which will be made available to other researchers upon request, thereby making a significant contribution to research in the field.

Moving away from the methodological strengths of the present research, other important strengths and contributions relate to the inclusion of psychological theory and perspective throughout the work carried out. More specifically, a psychological perspective was adopted throughout all aspects of this research, and we ensured to relate all of our work and findings to psychological theories where meaningful, as can be seen in the three papers (Chapters 3, 4, and 5), and in this general discussion chapter. The incorporation of psychological theory and perspective is important given that the aim of the present work was to generate psychologically insightful findings with respect to personality pathology, making it essential to ensure that the present work was connected with psychological theory as much as possible. Moreover, adopting a psychological perspective is especially important when utilising non-traditional, large-scale computational approaches – such as the NLP methods employed in the present work – to avoid such work becoming of a “black-box” nature, and to ensure transparency and accessible understanding. In particular, Study 4 addressed limitations of previous work leveraging NLP methods to study self-harm that was of a more black-box nature by investigating the linguistic trajectories surrounding suicidality and DSH while incorporating a psychological perspective and ensuring to generate psychologically insightful results. Accordingly, the present research leveraged the strengths of the relatively novel method (with respect to typical methods used in the domain of PD) of computational language analysis, while also incorporating psychological perspective and ensuring that the work and findings relate to psychological theory, subsequently permitting psychologically and clinically meaningful findings from a different methodological perspective.

6.3.2 Key Limitations and Challenges in the Present Work

Despite the many invaluable strengths and contributions of the present work highlighted in the previous section, this research also has some limitations that warrant consideration. Although each of the studies conducted come with their own specific limitations – outlined in each of the individual papers (i.e., Chapters 3, 4, and 5) – in this section, I focus on the most central, overarching limitations of the research as a whole (i.e., limitations that apply to most or all of the research carried out).

One of the central limitations of the present work is that all of the studies conducted have solely focused on BPD, and have not involved the study of other forms of (clinical) personality pathology. This focus means that the present findings are constrained in their ability to generalise beyond BPD to the broader construct of personality pathology. Yet, as emphasised in other sections of this thesis, the heterogeneous nature of BPD (e.g., Cavelti et al., 2021) – combined with high comorbidity rates – has prompted scholars to conceptualise BPD as reflective of general personality pathology (see, e.g., Wright et al., 2016), with BPD evidenced to strongly map on to a "general factor" of personality pathology (e.g., Sharp et al., 2015). Thus, this conceptualisation implies that findings in the context of BPD should be generalisable to the broader construct of PD. Moreover, the core features examined in the present research (i.e., emotional, social, and behavioural dysfunction) are common to all forms of personality pathology (APA, 2013). Subsequently, investigating such disturbances in individuals manifesting BPD – of whom often suffer comorbidity with other PD types – should naturally allow for insight into personality pathology in a broad sense. The notion that research in the context of BPD should be generalisable to PD more broadly is further supported by dimensional approaches to personality pathology (and psychopathology), in claiming that types of PDs (e.g., BPD vs. ASPD) are not categorically different from one another, and personality pathology thus should instead be evaluated based on core personality dysfunction and global level of severity. This dimensional conceptualisation of personality pathology therefore supports our focus on arguably the most generalisable and representative form of PD – BPD – in the present work, and indicates that our findings are likely *broadly* generalisable to other forms of personality pathology. It is also true that studying BPD in its own right is meaningful and important, in order to provide better understanding of this complex, prevalent, and high-risk construct.

A related limitation of the present research with regard to the study population and the samples used surrounds the fact that none of the studies conducted included clinical control groups (i.e., a group of people diagnosed with a mental health condition other than BPD) to confirm the specificity of the findings to BPD (versus other clinical groups). Notably, the lack of a clinical control group in the present work means that our results may not be specific to BPD, and in fact could potentially be transdiagnostic (i.e., shared across numerous mental health conditions). Indeed, the possibility that our findings may be transdiagnostic is made more likely by the conceptualisation of BPD as reflective of "general psychopathology" (see, e.g., Gluschkoff et al., 2021). Consistent with this, findings from Study 3 revealed that patterns of emotion vocabularies associated with BPD also extended to depression in this sample, thus indicating that these findings may be transdiagnostic. Yet, findings from Study 1 evidenced core social-cognitive dimensions found to be associated with BPD exclusively when compared to Dark Triad traits. Moreover, although Study 4 did not comprise a clinical control group, previous literature in the domain of suicidality and DSH in non-BPD populations illuminated some meaningful differences when compared with the present findings in a BPD population. Thus, it appears that although some of the present findings could potentially be transdiagnostic, other results appear to be more distinct to BPD. Whether the findings from the present work are transdiagnostic or distinct to BPD does not necessarily pose a limitation of the current research, but is a meaningful question in and of itself, and warrants consideration in future research.

In terms of methodological limitations, all of the studies, to varying extents, relied on an automated word-counting approach (i.e., LIWC) when analysing language data, despite such linguistic frequency variables being subsequently incorporated in more sophisticated analyses (thus going beyond the realms of past related work). Accordingly, the present research is constrained by the same limitations that typically apply to such computational word-counting approaches. In particular, one of the main limitations of automated word-counting approaches surrounds the fact that words counted as part of the automated program will often have numerous meanings (e.g., the word "alone" can reflect feeling lonely or being isolated, but it could also be used to indicate a positive expression of independence). Despite this, these types of automated word-counting approaches do not account for the context in which words are used. For example, the statements "this makes me so happy" and "this does not make me happy" would generate the same positive emotion word count, despite conveying different, and potentially contradictory, meanings. Yet, in employing this methodological approach, the semantic context of words is often irrelevant, as regardless of the meaning of a particular word, the reference to the concept in itself highlights that attention has been directed towards such concept rendered in natural language (see, e.g., Boyd & Schwartz, 2021; Pennebaker et al., 1997). Moreover, such constraints apply to all "bag-of-words" approaches, which have been well-established as meaningful indicators of a broad range of psychological constructs (e.g., Kennedy et al., 2022).

Moving on to analytic limitations, the majority of the studies conducted (i.e., Studies 1, 2, and 3) were constrained by the cross-sectional nature of the datasets, meaning that causal relationships could not be inferred. For instance, although Studies 2 and 3 uncovered meaningful relationships between natural emotion vocabularies and BPD, the datasets did not allow us to determine cause-and-effect relationships between emotion functioning and experience and active EVs. Yet, although some of the Study 4 analyses were correlational in nature – thus not allowing causality to be inferred – the longitudinal analyses examining linguistic trajectories in proximity to self-harm permitted insight into temporal psychosocial dynamics surrounding self-harm in BPD, thereby inferring potentially causal relationships in this regard.

A more generic limitation of the present work surrounds the fact that almost all of the studies (with the exception of Study 3) were conducted exclusively in the English language, which is a typical limitation of the broader computational language analysis field. It is therefore unclear whether and how the present findings would translate to other languages. For example, we do not know whether the language-based social-cognitive dimensions found to be associated with BPD in Study 1 would be replicated in non-English languages. Thus, the present findings, and the subsequent knowledge generated, are largely limited in their generalisability to other languages and cultures. However, promisingly, findings from Study 3 conducted in the German language (in Germany) were consistent with those in Study 2 that comprised an English-language sample, providing an initial indication that the present work may be generalisable to, at least some, non-English languages.

Finally, possibly the most important challenge with respect to the present work, and in relation to using computational language analysis methods to improve understanding and treatment of personality pathology (and psychopathology) more broadly, surrounds overcoming the barriers between empirical (theoretical) research and clinical practice. Notably, this need to bridge the gap between theoretical science and practice is a particular challenge given that there has traditionally been resistance among clinicians to adopt such computational methods (see, e.g., Goldberg et al., 2020). Thus, greater collaboration – through sharing knowledge and developing mutual understanding – between academic scientists and clinical practitioners is essential for moving forward. The development of such collaborations would undoubtably improve the likelihood of the implementation of new technologies, such as computational linguistic methods, into clinical practice, and would also help to ensure findings from clinical work utilising these computational methods – such as those generated from the present research – are incorporated into and have a meaningful impact on clinical practice.

6.3.3 Ethical Considerations

As with all scientific research, the ethical implications of the present research is an important topic and has been given serious consideration throughout all aspects of this work. All research that has the potential to influence people's lives comes with inherent ethical challenges, and the present work is no exception. Accordingly, throughout my PhD, I frequently consulted with experts in advanced computational methods and online technologies to help to inform my decisions, to ensure that the present research was carried out as ethically as possible.

Given that Studies 1, 2, and 3 involved the collection of new data via recruiting people to participate in the research, these studies underwent the typical rigorous university ethical approval procedures, and were all subsequently approved by the Faculty of Science and Technology Research Ethics Committee (FSTREC) at Lancaster University. As Study 3 involved a secondary data analysis of data already collected in previous research (i.e., the 'couple communication study'; see, e.g., Miano et al., 2017a, 2017b), this study was also approved by the ethics committee of the Freie Universität Berlin. With regard to particular ethical procedures undertaken, in Studies 1, 2, and 3 all participants received sufficient information about the studies prior to participating, and subsequently provided informed consent to participate in the research. Moreover, participants were all informed that they could withdraw themselves from the studies at any time, and were also provided with a sufficient debrief at the end of each of the studies. In Studies 1 and 2 (this was the same sample of participants) specifically, participants remained fully anonymous throughout all aspects of the research, and we ensured (as agreed with participants) that we did not share any of the participants' raw language data, to maintain confidentiality and anonymity. Maintaining participants' anonymity was not possible for Study 3, given that this study involved participants having face-to-face conversations with their romantic partners in the laboratory while being video recorded. However, anonymity was ensured in Study 3 with respect to the secondary data analysis carried out as part of the present work, as the text from the video recordings were transcribed and only the transcribed language data were subsequently analysed; the researchers directly involved in the analysis of these data did not have access to the video recordings, just the transcribed conversations. As with Studies 1 and 2, we have ensured not to share any of the raw language data (or any other identifiable data) from Study 3, to protect participant anonymity.

In comparison to Studies 1, 2, and 3, Study 4 did not involve the collection of new data from primary sources (i.e., recruiting new participants), as this study solely utilised online data in the public domain; specifically, Reddit data. Given that Study 4 did not involve the collection of new data, this naturally triggered a less exhaustive ethical approval procedure via the Faculty of Science and Technology Research Ethics Committee (FSTREC) at Lancaster University (note that this study still required and was granted formal approval from the ethics committee). Nevertheless, it is of course of vital importance to ensure that research conducted using naturalistic, online data sources adheres to best practices with respect to ethics. Although there remains to be insufficient standardised guidance regarding ethical procedures in relation to the use of online naturalistic data, we made sure to take all relevant and important ethical issues into consideration when carrying out this research. In particular, one of the most salient ethical concerns when working with online data, particularly social media data (even when in the public domain), surrounds anonymity and confidentiality (i.e., ensuring that data is non-identifiable). Such ethical concerns are somewhat naturally lessened when utilising Reddit data specifically, as this is a largely anonymous platform, with users not required to provide their name or any other personal details or identifiable information to the site. Even if users do include a full name in their username, there is no way of knowing whether it is their real name. The majority of data collected from Reddit should therefore be inherently anonymous and non-identifiable. Yet, steps were still taken to further mitigate concerns around anonymity and confidentiality in the use of Reddit data in Study 4, which will be discussed later in this section.

The other major ethical concern that requires consideration when working with social media data – including semi-anonymous platforms such as Reddit – surrounds privacy and involuntary measurement (i.e., lack of informed consent), as the surveillance of digital data without informed consent inherently threatens individual privacy. Although many of the computational social science and clinical communities argue the acceptability of studying users of online platforms without explicitly gaining informed consent (see, e.g., The British Psychological Society, 2021) – given that such data is already in the public domain, of which users are made aware of prior to using the platforms – this remains an ever-growing, prominent debate (e.g., Benton et al., 2017; Morant et al., 2021; Tušl et al., 2022) that warrants careful consideration of the risks and ethical challenges.

Notably, the fine ethical line between preserving individual privacy and leveraging digital data with the potential for large-scale positive societal impact is defined less by what data is collected and more by how the data is used and protected. To this end, throughout the present research, I made sure to take steps to mitigate the described ethical concerns by ensuring that the management of the Reddit data utilised was strongly guided by relevant ethical principles (e.g., as outlined in Benton et al., 2017), particularly making sure that all data was fully de-identified and well-protected. More specifically, regarding the publication and dissemination of findings, usernames have not been and will never be shared. Rather, if necessary, pseudonyms would be used to ensure that all data is fully anonymous and non-identifiable. Further, raw language data has not been and will never be shared or made publicly available on the basis of protecting confidentiality and privacy. Instead, unidentifiable, quantified versions of the language data (i.e., LIWC scores) have been made available, for accessibility to data. Throughout the present work, I have also ensured that all data is well-protected, through storing data electronically on secure encrypted computers under password protected folders. Finally, I have (and continue to) frequently discussed and collaborated with experts on ethics in computational science and digital data to ensure that my research is in adherence with the most up-to-date ethical guidance.

In addition to abiding by typical ethical guidelines relating to the use of online social media data for research, I have also attempted to involve the online communities studied (i.e., the BPD Reddit communities in Study 4) directly in the research. That is, I consolidated with the moderators of the BPD subreddits (i.e., *r/BPD* and *r/BorderlinePDisorder*) throughout various stages of this project, to ensure that they were accepting of the way in which we were utilising the data from the communities. Such a collaborative approach with those closely connected with the online BPD communities adds a layer of public involvement to the present work, and also ensures that the work is as ethical as possible. Taken in all, I have attempted to guarantee that all of the present research abided by relevant ethical guidelines, with all projects granted ethical approval. Although there are significantly less ethical guidelines available regarding the use of online social media data in the public domain, I made sure to consider relevant ethical principles throughout all stages of the current research that utilised social media data (i.e., Study 4), including involving the online communities themselves to some extent.

6.4 Conclusions

Personality disorders are currently some of the most prevalent and high-risk mental health conditions, and yet remain poorly understood. Accordingly, the central goal of this thesis was to demonstrate how personality pathology can be better understood through the computational analysis of natural language, given that word use allows insight into individuals' broad constellation of thinking, feelings, and behaviours. In addressing the central goal, this thesis presents three research articles – comprising four empirical studies – that each leveraged computational language analysis to better understand personality pathology. Each of the research papers focused on a different core feature of PD. In particular, the core personality pathology features investigated were interpersonal dysfunction (Paper 1 – Study 1; Chapter 3), emotion dysregulation (Paper 2 – Studies 2 and 3; Chapter 4), and behavioural dysregulation (Paper 3 – Study 4; Chapter 5).

Subsequently, findings from the present research have generated better understanding of fundamental features of personality pathology, including new insight into characterising dimensions of social dysfunction (Study 1; Chapter 3), maladaptive emotion processes that may drive emotion dysregulation (Studies 2 and 3; Chapter 4), and psychosocial dynamics (evident in language) relating to suicidality and deliberate self-harm (Study 4; Chapter 5) in PD. Integrating the present findings together, all of the studies conducted have generated better understanding of and novel insight into fundamental features of personality pathology via a naturalistic linguistic approach. Undoubtably, the present findings have, in turn, stimulated numerous implications for personality pathology theory – including developing a better understanding of the nature of PD and explicitly informing theoretical models of PD – as well as implications for the broader fields of general personality psychology, psychopathology, and linguistic theory. Vitally, these theoretical implications subsequently inform and progress into practical implications, particularly with respect to informing clinical practice and clinical guidelines on the use of online PD support platforms. More broadly, this research highlights how language can provide implicit and unobtrusive insight into personality and psychological processes underlying personality pathology at a large-scale, using an individualised, behavioural approach.

Notably, the present work also has implications for research with respect to the identification of new avenues of empirical investigation in need of exploration. That is, all of the studies carried out each sparked their own recommendations for future research. For instance, Study 1 highlighted the value of analysing natural language data in a social context to better understand personality pathology, and subsequently directed attention towards the possibility of utilising language analysis techniques that take into account the context of words to assess the conceptual level of the language of individuals with PD, to generate insight into self and other representations in PD. Moreover, Studies 2 and 3 sparked new research questions with respect to the potential transdiagnostic nature of the associations found between BPD and active EVs in the present research (as initially indicated with depression). Thus, it would be most informative for future research to investigate active EVs in various other clinical groups, in distinct samples, to examine this potential transdiagnostic nature of associations with active EVs. In addition, Study 4 also uncovered numerous critical areas in need of further research. Most importantly, findings from Study 4 generated new hypotheses to be tested in future research with respect to effects of online PD communities on engagement in self-harm. Given that effects of community interactions on future engagement in self-harm were not tested directly in the present work, this enlightens a crucial area for future research, of which has life-saving potential.

In terms of addressing some of the key limitations of the present work, it would be most informative for future research to incorporate language analysis techniques in the study of other forms of personality pathology (and other forms of psychopathology more broadly), rather than solely focusing on BPD, to determine how generalisable findings are across different types of PD (or mental health conditions). Building on this, examining the extent to which the present findings extend to other clinical groups would help to determine how transdiagnostic versus distinct to (B)PD our findings are. Importantly, this knowledge may also shed further light on the dimensional versus categorical nature of PDs, and psychopathology in general. In addition, it would be useful for future research to attempt to replicate the present findings in languages other than English, to determine how generalisable our findings are to other languages and cultures. On a more general level, any empirical research applying computational language analysis to the study of personality pathology would be most valuable, given the limited work in this realm to date, and given the potential of such methods in improving understanding of PD. Moreover, leveraging a broader range of linguistic analytic techniques, beyond simply counting general word frequencies and associating them with PD features, will help to further maximise the reach and impact of the application of language analysis methods to the field of personality pathology.

Over recent decades, the growth in advanced language-based methods and powerful statistical techniques has allowed rich sources of data to be analysed in ways that permit better understanding of psychopathologies, including personality pathology. Critically, this research direction provides an opportunity to ensure that both empirical research and clinical practice can reciprocally inform and enhance each other. In moving forward, communication and interdisciplinary collaborations between empirical research and clinical practice are essential. For such promising methods to reach their maximum potential and have a real-world impact, it is important that they lead to insights that can directly inform clinical practice. Taken in all, I hope that this thesis will inspire researchers and clinicians to come together and take advantage of the many benefits that the application of computational language analysis can bring to personality pathology research and practice.

Consolidated Bibliography

Aafjes-van Doorn, K., Kamsteeg, C., Bate, J., & Aafjes, M. (2020). A scoping review of machine learning in psychotherapy research. *Psychotherapy Research.* 1–25. https://doi.org/10.1080/10503307.2020.1808729.

Aktas, N., Bodt, E. de, Bollaert, H., & Roll, R. (2016). CEO narcissism and the takeover process: From private initiation to deal completion. *Journal of Financial and Quantitative Analysis, 51*(1), 113–137. https://doi.org/10.1017/S0022109016000065

Allport, G. W. (1937). *Personality: A Psychological Interpretation.* New York: Holt.

Al-Mosaiwi, M., & Johnstone, T. (2018). In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science*, 2167702617747074. https://doi.org/10.1177/2167702617747074

Al-Shamali, H. F., Winkler, O., Talarico, F., Greenshaw, A. J., Forner, C., Zhang, Y., Vermetten, E., & Burback, L. (2022). A systematic scoping review of dissociation in borderline personality disorder and implications for research and clinical practice: Exploring the fog. *The Australian and New Zealand Journal of Psychiatry*, *56*(10), 1252–1264. https://doi.org/10.1177/00048674221077029

Aldhyani, T. H. H., Alsubari, S. N., Alshebami, A. S., Alkahtani, H., & Ahmed, Z. A. T. (2022). Detecting and analyzing suicidal ideation on social media using deep learning and machine learning models. *International Journal of Environmental Research and Public Health*, *19*(19), 12635. https://doi.org/10.3390/ijerph191912635

Altamura, A. C., Buoli, M., Caldiroli, A., Caron, L., Cumerlato Melter, C., Dobrea, C., Cigliobianco, M., & Zanelli Quarantini, F. (2015). Misdiagnosis, duration of untreated illness (DUI) and outcome in bipolar patients with psychotic symptoms: A naturalistic study. *Journal of Affective Disorders*, *182*, 70–75. https://doi.org/10.1016/j.jad.2015.04.024

Amad, A., Ramoz, N., Thomas, P., Jardri, R., & Gorwood, P. (2014). Genetics of borderline personality disorder: Systematic review and proposal of an integrative model. *Neuroscience & Biobehavioral Reviews, 40*, 6–19. https://doi.org/10.1016/j.neubiorev.2014.01.003

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®).* American Psychiatric Pub.

Anderson, B., Goldin, P. R., Kurita, K., & Gross, J. J. (2008). Self-representation in social anxiety disorder: Linguistic analysis of autobiographical narratives. *Behaviour Research and Therapy*, *46*(10), 1119–1125. https://doi.org/10.1016/j.brat.2008.07.001

Arevian, A. C., Bone, D., Malandrakis, N., Martinez, V. R., Wells, K. B., Miklowitz, D. J., & Narayanan, S. (2020). Clinical state tracking in serious mental illness through computational analysis of speech. *PLoS ONE*, *15*(1), e0225695. https://doi.org/10.1371/journal.pone.0225695

Arntz, A., Hawke, L. D., Bamelis, L., Spinhoven, P., & Molendijk, M. L. (2012). Changes in natural language use as an indicator of psychotherapeutic change in personality disorders. *Behavior, Research and Therapy*, *50*(3), 191–202. https://doi.org/10.1016/j.brat.2011.12.007

Bagroy, S., Kumaraguru, P., & De Choudhury, M. (2017). A social media based index of mental well-being in college campuses. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI Conference*, 1634–1646. https://doi.org/10.1145/3025453.3025909

Baldwin, J. R., Reuben, A., Newbury, J. B., & Danese, A. (2019). Agreement between prospective and retrospective measures of childhood maltreatment: A systematic review and meta-analysis. *JAMA Psychiatry*, *76*(6), 584–593. https://doi.org/10.1001/jamapsychiatry.2019.0097

Barr, K. R., Townsend, M. L., & Grenyer, B. F. S (2020). Using peer workers with lived experience to support the treatment of borderline personality disorder: A qualitative study of consumer, carer and clinician perspectives. *Borderline Personality Disorder and Emotion Dysregulation*, *7*, 20. https://doi.org/10.1186/s40479-020-00135-5

Barrett, L. F., Gross, J., Christensen, T. C., & Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition and Emotion*, *15*(6), 713–724. https://doi.org/10.1080/02699930143000239

Bateman, A. W., Gunderson, J., & Mulder, R. (2015). Treatment of personality disorder. *The Lancet, 385*(9969), 735–743. https://doi.org/10.1016/S0140-6736(14)61394-5

Beatson, J. A., & Rao, S. (2013). Depression and borderline personality disorder. *The Medical Journal of Australia, 199*(6 Suppl), S24–S27. https://doi.org/10.5694/mja12.10474

Beck, A.T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive Therapy of Depression.* Guilford Press.

Benton, A., Coppersmith, G., Dredze, M. (2017). Ethical research protocols for social media health research. *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing,* Valencia, Spain. 94–102.

Ben-Zeev, D., Scherer, E. A., Wang, R., Xie, H., & Campbell, A. T. (2015). Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. *Psychiatric Rehabilitation Journal, 38*(3), 218–226. https://doi.org/10.1037/prj0000130

Berenson, K. R., Dochat, C., Martin, C. G., Yang, X., Rafaeli, E., & Downey, G. (2018). Identification of mental states and interpersonal functioning in borderline personality disorder. *Personality Disorders*, *9*(2), 172–181. https://doi.org/10.1037/per0000228

Berenson, K. R., Gregory, W. E., Glaser, E., Romirowsky, A., Rafaeli, E., Yang, X., & Downey, G. (2016). Impulsivity, rejection sensitivity, and reactions to stressors in borderline personality disorder. *Cognitive Therapy and Research*, *40*(4), 510–521. https://doi.org/10.1007/s10608-015-9752-y

Birchwood, M. J., Fowler, D., & Jackson, C. (2000). *Early Intervention in Psychosis: A Guide to Concepts, Evidence and Interventions*. Chichester: John Wiley & Sons.

Birnbaum, M. L., Ernala, S. K., Rizvi, A. F., Arenare, E., R. Van Meter, A., De Choudhury, M., & Kane, J. M. (2019). Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from Facebook. *Npj Schizophrenia, 5*(1), 17. https://doi.org/10.1038/s41537-019-0085-9

Black, D. W. (2019). Antisocial personality disorder: Epidemiology, clinical manifestations, course and diagnosis. *UpToDate*. Retrieved March, 2020, from https://www.uptodate.com/contents/antisocial-personality-disorder-epidemiology-clinical-manifestations-course-and-diagnosis

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

Black, D. W., Gunter, T., Loveless, P., Allen, J., & Sieleni, B. (2010). Antisocial personality disorder in incarcerated offenders: Psychiatric comorbidity and quality of life. *Annals of Clinical Psychiatry: Official Journal of the American Academy of Clinical Psychiatrists, 22*(2), 113–120.

Boghrati, R., Hoover, J., Johnson, K. M., Garten, J., & Dehghani, M. (2018). Conversation level syntax similarity metric. *Behavior, Research Methods, 50,* 1055–1073. https://doi.org/10.3758/s13428-017-0926-2

Bogolyubova, O., Panicheva, P., Tikhonov, R., Ivanov, V., & Ledovaya, Y. (2018). Dark personalities on Facebook: Harmful online behaviors and language. *Computers in Human Behavior, 78,* 151–159. https://doi.org/10.1016/j.chb.2017.09.032

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

Borelli, J. L., Sohn, L., Wang, B. A., Hong, K., DeCoste, C., & Suchman, N. E. (2019). Therapist–client language matching: Initial promise as a measure of therapist–client relationship quality. *Psychoanalytic Psychology, 36*(1), 9–18. https://doi.org/10.1037/pap0000177

Boyd, R. L. (2017). Psychological text analysis in the digital humanities. In S. Hai-Jew (Ed.), *Data Analytics in Digital Humanities* (pp. 161–189). Springer International Publishing. https://doi.org/10.1007/978-3-319-54499-1_7

Boyd, R. L. (2020). *BUTTER: Basic Unit-Transposable Text Experimentation Resource*. https://www.butter.tools

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *The Development and Psychometric Properties of LIWC-22*. Austin, TX: University of Texas at Austin. https://www.liwc.app

Boyd, R. L., Pasca, P., & Conroy-Beam, D. (2019). You're only Jung once: Building generalized motivational systems theories using contemporary research on language. *Psychological Inquiry, 30*(2), 93–98. https://doi.org/10.1080/1047840X.2019.1633122

Boyd, R. L., Pasca, P., & Lanning, K. (2020). The personality panorama: Conceptualizing personality through big behavioural data. *European Journal of Personality*. https://doi.org/10.1002/per.2254

Boyd, R. L., & Schwartz, H. A. (2021). Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field. *Journal of Language and Social Psychology*. https://doi.org/10.1177/0261927X20967028

Boyd, R. L., Wilson, S. R., Pennebaker, J. W., Kosinski, M., Stillwell, D. J., & Mihalcea, R. (2015). Values in words: Using language to evaluate and understand personal values. *Proceedings of the Ninth International AAAI Conference on Web and Social Media*, 31–40. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10482

Bozzatello, P., Rocca, P., Baldassarri, L., Bosia, M., & Bellino, S. (2021). The role of trauma in early onset borderline personality disorder: A biopsychosocial perspective. *Frontiers in Psychiatry, 12,* 721361. https://doi.org/10.3389/fpsyt.2021.721361

Brinkley, C. A., Newman, J. P., Harpur, T. J., & Johnson, M. M. (1999). Cohesion in texts produced by psychopathic and nonpsychopathic criminal inmates. *Personality and Individual Differences*, *26*(5), 873–885. https://doi.org/10.1016/S0191-8869(98)00189-5

Buckholdt, K. E., Parra, G. R., Anestis, M. D., Lavender, J. M., Jobe-Shields, L. E., Tull, M. T., & Gratz, K. L. (2015). Emotion regulation difficulties and maladaptive behaviors: Examination of deliberate self-harm, disordered eating, and substance misuse in two samples. *Cognitive, Therapy, and Research, 39*(2), 140–152. https://psycnet.apa.org/doi/10.1007/s10608-014-9655-3

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: The power of an expertise framework. *Trends in Cognitive Sciences*, *10*(4), 159–166. https://doi.org/10.1016/j.tics.2006.02.004

Bulbena-Cabre, A., Bassir Nia, A., & Perez-Rodriguez, M. M. (2018). Current knowledge on gene-environment interactions in personality disorders: An update. *Current Psychiatry Reports, 20*, 74.

https://doi.org/10.1007/s11920-018-0934-7

Cavelti, M., Lerch, S., Ghinea, D., Fischer-Waldschmidt, G., Resch, F., Koenig, J., & Kaess, M. (2021). Heterogeneity of borderline personality disorder symptoms in help-seeking adolescents. *Borderline Personality Disorder and Emotion Dysregulation, 8*(1), 9. https://doi.org/10.1186/s40479-021-00147-9

Cahalan, S. (2012). *Brain on Fire: My Month of Madness.* Free Press: New York.

Carey, A. L., Brucks, M. S., Küfner, A. C. P., Holtzman, N. S., große Deters, F., Back, M. D., Donnellan, M. B., Pennebaker, J. W., & Mehl, M. R. (2015). Narcissism and the use of personal pronouns revisited. *Journal of Personality and Social Psychology*, *109*(3), e1–e15. https://doi.org/10.1037/pspp0000029

Carpenter, R. W., & Trull, T. J. (2013). Components of emotion dysregulation in borderline personality disorder: A review. *Current Psychiatry Reports, 15*(1), 335. https://doi.org/10.1007/s11920-012-0335-2

Carr, D. S., & Francis, A. (2009). Childhood familial environment, maltreatment and borderline personality disorder symptoms in a non-clinical sample: A cognitive behavioural perspective. *Clinical Psychologist*, *13*(1), 28–37. https://doi.org/10.1080/13284200802680476

Carter, P. E., & Grenyer, B. F. S. (2012). Expressive language disturbance in borderline personality disorder in response to emotional autobiographical stimuli. *Journal of Personality Disorders*, *26*(3), 305–321. https://doi.org/10.1521/pedi.2012.26.3.305

Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. *Journal of Abnormal and Social Psychology*, *38*(4), 476–506. https://doi.org/10.1037/h0054116

Cavelti, M., Lerch, S., Ghinea, D., Fischer-Waldschmidt, G., Resch, F., Koenig, J., & Kaess, M. (2021). Heterogeneity of borderline personality disorder symptoms in help-seeking adolescents. *Borderline Personality Disorder and Emotion Dysregulation, 8*(1), 9. https://doi.org/10.1186/s40479-021-00147-9

Chad, R. H., Moore, A. C., & Meehan, K. B. (2018). Borderline personality disorder. In Zeigler-Hill, V., Shackelford, T. (Eds) *Encyclopedia of Personality and Individual Differences*. Springer, Cham. https://doi.org/10.1007/978-3-319-28099-8_573-1

Chancellor, S., & De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: A critical review. *NPJ Digital Medicine, 3,* 43. https://doi.org/10.1038/s41746-020-0233-7

Chanen, A., Sharp, C., Hoffman, P., & Global Alliance for Prevention and Early Intervention for Borderline Personality Disorder (2017). Prevention and early intervention for borderline personality disorder: A novel public health priority. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA), 16*(2), 215–216. https://doi.org/10.1002/wps.20429

Charlson, F. J., Ferrari, A. J., Santomauro, D. F., Diminic, S., Stockings, E. Scott, J. G., McGrath, J. J., & Whiteford, H. A. (2018). Global epidemiology and burden of schizophrenia: Findings from the global burden of disease study 2016. *Schizophrenia Bulletin, 44*(6), 1195–1203. https://doi.org/10.1093/schbul/sby058

Choi, M., Aiello, L., Varga, K. Z., & Quercia, D. (2020). Ten social dimensions of conversations and relationships. In *Proceedings of The Web Conference 2020 (WWW'20)*, 1514–1525. https://doi.org/10.1145/3366423.3380224.

Chu, C., Victor, S. E., & Klonsky, E. D. (2016). Characterizing positive and negative emotional experiences in young adults with borderline personality disorder symptoms. *Journal of Clinical Psychology, 72*(9), 956–965. https://doi.org/10.1002/jclp.22299

Chung, C., & Pennebaker, J. (2008). Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality, 42*, 96–132. https://doi.org/10.1016/j.jrp.2007.04.006.

Clark, L. A., Cuthbert, B., Lewis-Fernández, R., Narrow, W. E., Reed, G. M. (2017). Three approaches to understanding and classifying mental disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). *Psychological Science in the Public Interest, 18*(2), 72-145. doi:10.1177/1529100617727266

Cleckley, H. M. (1976). *The Mask of Sanity: An Attempt to Clarify Some Issues about the So-called Psychopathic Personality*. MO: Mosby.

Coid, J., & Ullrich, S. (2010). Antisocial personality disorder is on a continuum with psychopathy. *Comprehensive Psychiatry, 51*(4), 426–433. https://doi.org/10.1016/j.comppsych.2009.09.006

Compton, W. M., Conway, K. P., Stinson, F. S., Colliver, J. D., & Grant, B. F. (2005). Prevalence, correlates, and comorbidity of DSM-IV antisocial personality syndromes and alcohol and specific drug use disorders in the United States: Results from the national epidemiologic survey on alcohol and related conditions. *The Journal of Clinical Psychiatry*, *66*(6), 677–685. https://doi.org/10.4088/jcp.v66n0602

Coppersmith, G. A., Dredze, M., Harman, C., & Hollingshead, K. (2015). From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 1–10. https://doi.org/10.3115/v1/W15-1201

Coppersmith, G. A., Ngo, K., Leary, R., & Wood, A. (2016). Exploratory analysis of social media prior to a suicide attempt. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, 106–117. Association for Computational Linguistics. http://dx.doi.org/10.18653/v1/W16-0311

Crego, C., & Widiger, T. A. (2022). Core traits of psychopathy. *Journal of Personality Disorders*, *13*(6), 674–684. https://doi.org/10.1037/per0000550

Cristea, I. A., Gentili, C., Cotet, C. D., Palomba, D., Barbui, C., & Cuijpers, P. (2017). Efficacy of psychotherapies for borderline personality disorder: A systematic review and meta-analysis. *JAMA Psychiatry*, *74*(4), 319–328. https://doi.org/10.1001/jamapsychiatry.2016.4287

Crowell, S. E., Beauchaine, T. P., & Linehan, M. M. (2009). A biosocial developmental model of borderline personality: Elaborating and extending Linehan's theory. *Psychology Bulletin*, *135*(3), 495–510. https://doi.org/10.1037/a0015616

Cutler, A. D., Carden, S. W., Dorough, H. L., & Holtzman, N. S. (2021). Inferring grandiose narcissism from text: LIWC versus machine learning. *Journal of Language and Social Psychology*, *40*(2), 260–276. https://doi.org/10.1177/0261927X20936309

Dalgleish, T., Black, M., Johnston, D., & Bevan, A. (2020). Transdiagnostic approaches to mental health problems: Current status and future directions. *Journal of Consulting and Clinical Psychology, 88*(3), 179–195. https://doi.org/10.1037/ccp0000482

De Choudhury, M., & De, S. (2014). Mental health discourse on Reddit: Self-disclosure, social support, and anonymity. *Proceedings of the International AAAI Conference on Web and Social Media, 8*(1), 71–80. https://doi.org/10.1609/icwsm.v8i1.14526

De Choudhury, M., Kiciman, E., Dredze, M., Coppersmith, G., & Kumar, M. (2016). Discovering shifts to suicidal ideation from mental health content in social media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI Conference,* 2098–2110. https://doi.org/10.1145/2858036.2858207

Dean, H. J., & Boyd, R. L. (2020). Deep into that darkness peering: A computational analysis of the role of depression in Edgar Allan Poe's life and death. *Journal of Affective Disorders, 266*, 482–491. https://doi.org/10.1016/j.jad.2020.01.098

Derks, Y. P. M. J., Westerhof, G. J., & Bohlmeijer, E. T. (2017). A meta-analysis on the association between emotional awareness and borderline personality pathology. *Journal of Personality Disorders*, *31*(3), 362–384. https://doi.org/10.1521/pedi_2016_30_257

DeWall, C. N., Pond Jr., R. S., Campbell, W. K., & Twenge, J. M. (2011). Tuning in to psychological change: Linguistic markers of psychological traits and emotions over time in popular U.S. song lyrics. *Psychology of Aesthetics, Creativity, and the Arts*, *5*(3), 200–207. https://doi.org/10.1037/a0023195

Digman, J. M. (1990). Personality structure: Emergence of the Five-Factor model. *Annual Review of Psychology, 41*(1), 417–440. https://doi.org/10.1146/annurev.ps.41.020190.002221

Dixon-Gordon, K. L., Gratz, K. L., Breetz, A., & Tull, M. (2013). A laboratory-based examination of responses to social rejection in borderline personality disorder: The mediating role of emotion dysregulation. *Journal of Personality Disorders*, *27*(2), 157–171. https://doi.org/10.1521/pedi.2013.27.2.157

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

Dorough, H. (2018). *Vulnerable narcissism and first-person singular pronoun use* (Dissertation). Retrieved from https://digitalcommons.georgiasouthern.edu/honors-theses/377

Dunlop, W. L., Karan, A., Wilkinson, D., & Harake, N. (2020). Love in the first degree:

Individual differences in first-person pronoun use and adult romantic attachment styles. *Social Psychological and Personality Science, 11*(2), 254–265. https://doi.org/10.1177/1948550619847455

Edwards, T., & Holtzman, N. S. (2017). A meta-analysis of correlations between depression and first person singular pronoun use. *Journal of Research in Personality*, *68*, 63–68. https://doi.org/10.1016/j.jrp.2017.02.005

Eichler, M. (1965). The application of verbal behavior analysis to the study of psychological defense mechanisms: Speech patterns associated with sociopathic behaviour. *The Journal of Nervous and Mental Disease, 141*(6), 658–663.

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preoţiuc Pietro, D., Asch, D. A., & Schwartz, H. A. (2018). Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences*, 201802331. https://doi.org/10.1073/pnas.1802331115

Ellison, W. D., Rosenstein, L. K., Morgan, T. A., & Zimmerman, M. (2018). Community and clinical epidemiology of borderline personality disorder. *The Psychiatric Clinics of North America*, *41*(4), 561–573. https://doi.org/10.1016/j.psc.2018.07.008

Ellouze, M., Mechti, S., & Belguith, L. H. (2021). Approach based on ontology and machine learning for identifying causes affecting personality disorder disease on Twitter. Proceedings in *Knowledge Science, Engineering and Management: 14th International Conference (KSEM 2021)*, 659–669. https://doi.org/10.1007/978-3-030-82153-1_54

Entwistle, C., Marceau, E., & Boyd, R. L. (2022). Personality disorder and verbal behavior. In M. Dehghani & R. L. Boyd (Eds.), *The Handbook of Language Analysis in Psychology* (pp. 335–356). The Guilford Press.

Erkoreka, L., & Navarro, B. (2017). Vulnerable narcissism is associated with severity of depressive symptoms in dysthymic patients. *Psychiatry Research*, *257*, 265–269. https://doi.org/10.1016/j.psychres.2017.07.061

Erozkan, A. (2016). The link between types of attachment and childhood trauma. *Universal Journal of Educational Research 4*(5), 1071–1079. https://doi.org/10.13189/ujer.2016.040517

Euler, S., Nolte, T., Constantinou, M., & Griem, J., Montague, P., & Fonagy, P. (2019). Interpersonal problems in borderline personality disorder: Associations with mentalizing, emotion regulation, and impulsiveness. *Journal of Personality Disorders*. 1–17. Advance online publication. https://doi.org/10.1521/pedi_2019_33_427

Eysenck, H. J. (1991). Dimensions of personality: 16, 5 or 3? Criteria for a taxonomic paradigm. *Personality and Individual Differences*, *12*(8), 773–790. https://doi.org/10.1016/0191-8869(91)90144-Z

Fast, L. A., & Funder, D. C. (2008). Personality as manifest in word use: Correlations with self-report, acquaintance report, and behavior. *Journal of Personality and Social Psychology, 94*(2), 334.

Fineberg, S. K., Deutsch-Link, S., Ichinose, M., McGuinness, T., Bessette, A. J., Chung, C. K., & Corlett, P. R. (2015). Word use in first-person accounts of schizophrenia. *The British Journal of Psychiatry*, *206*(1), 32–38. https://doi.org/10.1192/bjp.bp.113.140046

Fitzpatrick, S., Ip, J., Krantz, L., Zeifman, R., & Kuo, J. R. (2019). Use your words: The role of emotion labeling in regulating emotion in borderline personality disorder. *Behaviour, Research and Therapy*, *120*, 103447. https://doi.org/10.1016/j.brat.2019.103447

Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy, 55*, 316–340. http://dx.doi.org/10.1037/pst0000172

Frankenburg, F. R., & Zanarini, M. C. (2004). The association between borderline personality disorder and chronic medical illnesses, poor health-related lifestyle choices, and costly forms of health care utilization. *The Journal of Clinical Psychiatry, 65*(12), 1660–1665. https://doi.org/10.4088/JCP.v65n1211

Freud, S. (1891). *On Aphasia: A Critical Study [Zur Auffassung der Aphasien: eine kritische Studie]*. Stengel, E., translator. New York: International UP.

Freud, S. (1915). *The Unconscious*. SE, 14: 159–215.

Friborg, O., Martinsen, E. W., Martinussen, M., Kaiser, S., Overgård, K. T., & Rosenvinge, J. H. (2014). Comorbidity of personality disorders in mood disorders: A meta-analytic review of 122 studies from 1988 to 2010. *Journal of Affective Disorders*, *152–154*, 1–11. https://doi.org/10.1016/j.jad.2013.08.023

Gabbard, G. (2012). Treatment Resistance in Personality Disorders. In *Management of*

Treatment-Resistant Major Psychiatric Disorders. Oxford, UK: Oxford University Press. Retrieved from https://oxfordmedicine.com/view/10.1093/med/9780199739981.001.1/med-9780199739981-chapter-0013.

Garcia, D., & Sikström, S. (2014). The dark side of Facebook: Semantic representations of status updates predict the Dark Triad of personality. *Personality and Individual Differences*, *67*, 92–96. https://doi.org/10.1016/j.paid.2013.10.001

Garofalo, C., Velotti, P., Callea, A., Popolo, R., Salvatore, G., Cavallo, F., & Dimaggio, G. (2018). Emotion dysregulation, impulsivity and personality disorder traits: A community sample study. *Psychiatry Research*, *266*, 186–192. https://doi.org/10.1016/j.psychres.2018.05.067

Gawda, B. (2010a). Language of love and hate of persons diagnosed with antisocial personality. *Bulletin De La Societe Des Sciences Medicales Du Grand-Duche De Luxembourg*, *1*(1), 157–165.

Gawda, B. (2010b). Syntax of emotional narratives of persons diagnosed with antisocial personality. *Journal of Psycholinguistic Research*, *39*(4), 273–283. https://doi.org/10.1007/s10936-009-9140-4

Gawda, B. (2013). The emotional lexicon of individuals diagnosed with antisocial personality disorder. *Journal of Psycholinguistic Research*, *42*(6), 571–580. https://doi.org/10.1007/s10936-012-9237-z

Gawda B. (2022). The differentiation of narrative styles in individuals with high psychopathic deviate. *Journal of Psycholinguistic Research*, *51*(1), 75–92. https://doi.org/10.1007/s10936-021-09824-w

Gee, B. L., Han, J., Benassi, H., & Batterham, P. J. (2020). Suicidal thoughts, suicidal behaviours and self-harm in daily life: A systematic review of ecological momentary assessment studies. *Digital Health*, *6*, 2055207620963958. https://doi.org/10.1177/2055207620963958

Gibbon, S., Khalifa, N. R, Cheung, N.H.-Y., Völlm, B. A., & McCarthy, L. (2020). Psychological interventions for antisocial personality disorder. *Cochrane Database of Systematic Reviews, 9,* CD007668. https://doi.org/10.1002/14651858.CD007668.pub3.

Gilbert, F., Daffern, M., Talevski, D., & Ogloff, J. (2013). Understanding the personality disorder and aggression relationship: an investigation using contemporary aggression theory. *Journal of Personality Disorders, 29*(1), 100–114. https://doi.org/10.1521/pedi_2013_27_077

Glenn, J. J., Nobles, A. L., Barnes, L. E., & Teachman, B. A. (2020). Can text messages identify suicide risk in real time? A within-subjects pilot examination of temporally sensitive markers of suicide risk. *Clinical Psychological Science, 8*(4), 704–722. https://doi.org/10.1177/2167702620906146

Gluschkoff, K., Jokela, M., & Rosenström, T. (2021). General psychopathology factor and borderline personality disorder: Evidence for substantial overlap from two nationally representative surveys of U.S. adults. *Journal of Personality Disorders, 12(*1), 86–92. https://doi.org/10.1037/per0000405

Golbeck, J. A. (2016). Predicting personality with social media. *AIS Transactions on Replication Research, 2*(2), 1–10.

Goldberg, S. B., Flemotomos, N., Martinez, V. R., Tanana, M. J., Kuo, P. B., Pace, B. T., Villatte, J. L., Georgiou, P. G., Van Epps, J., Imel, Z. E., Narayanan, S. S., & Atkins, D. C. (2020). Machine learning and natural language processing in psychotherapy research: Alliance as example use case. *Journal of Counseling Psychology, 67*(4), 438–448. https://doi.org/10.1037/cou0000382

Gonzales, A., Hancock, J., & Pennebaker, J. (2010). Language style matching as a predictor of social dynamics in small groups. *Communication Research, 37*, 3–19. https://doi.org/10.1177/0093650209351468.

Goodman, M., Tomas, I. A., Temes, C. M., Fitzmaurice, G. M., Aguirre, B. A., & Zanarini, M. C. (2017). Suicide attempts and self-injurious behaviours in adolescent and adult patients with borderline personality disorder. *Personality and Mental Health*, *11,* 157–163. https://doi.org/10.1002/pmh.1375

Gottschalk, L. A. (2000). The application of computerized content analysis of natural language in psychotherapy research now and in the future. *American Journal of Psychotherapy, 54*(3), 305–311. https://doi.org/10.1176/appi.psychotherapy.2000.54.3.305

Gratz, K. L., Kiel, E. J., Mann, A. J. D., & Tull, M. T. (2022). The prospective relation between borderline personality disorder symptoms and suicide risk: The mediating roles of emotion regulation difficulties and perceived burdensomeness. *Journal of Affective Disorders, 313*, 186–195. https://doi.org/10.1016/j.jad.2022.06.066

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry, 26*(1), 1–26. https://psycnet.apa.org/doi/10.1080/1047840X.2014.940781

Guntuku, S. C., Sherman, G., Stokes, D. C., Agarwal, A. K., Seltzer, E., Merchant, R. M., & Ungar, L. H. (2020). Tracking mental health and symptom mentions on Twitter during COVID-19. *Journal of General Internal Medicine, 35*(9), 2798–2800. https://doi.org/10.1007/s11606-020-05988-8

Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: An integrative review. *Current Opinion in Behavioral Sciences*, *18*, 43–49. https://doi.org/10.1016/j.cobeha.2017.07.005

Halicka, J., & Kiejna, A. (2018). Non-suicidal self-injury (NSSI) and suicidal: Criteria differentiation. *Advances in Clinical and Experimental Medicine: Official Organ Wroclaw Medical University*, *27*(2), 257–261. https://doi.org/10.17219/acem/66353

Hall, M., & Caton, S. (2017). Am I who I say I am? Unobtrusive self-representation and personality recognition on Facebook. *PLOS ONE, 12*(9), e0184417. https://doi.org/10.1371/journal.pone.0184417

Hancock, J. T., Woodworth, M., & Boochever, R. (2018). Psychopaths online: The linguistic traces of psychopathy in email, text messaging and Facebook. *Media and Communication, 6*(3), 83–92. https://doi.org/10.17645/mac.v6i3.1499

Hancock, J. T., Woodworth, M. T., & Porter, S. (2013). Hungry like the wolf: A word-pattern analysis of the language of psychopaths. *Legal and Criminological Psychology*, *18*(1), 102–114. https://doi.org/10.1111/j.2044-8333.2011.02025.x

Hatkevich, C., Penner, F., & Sharp, C. (2019). Difficulties in emotion regulation and suicide ideation and attempt in adolescent inpatients. *Psychiatry Research*, *271*, 230–238. https://doi.org/10.1016/j.psychres.2018.11.038

Hautzinger, M., Keller, F., & Kühner, C. (2006). *BDI-II. Beck Depressions Inventar Revision—Manual*. Frankfurt, Germany: Harcourt Test Services.

Hawton, K., Bergen, H., Cooper, J., Turnbull, P., Waters, K., Ness, J., & Kapur, N.

(2015). Suicide following self-harm: Findings from the Multicentre Study of self-harm in England, 2000-2012. *Journal of Affective Disorders, 175*, 147–151. https://doi.org/10.1016/j.jad.2014.12.062

Hegarty, B. D., Marceau, E. M., Gusset, M., & Grenyer, B. F. S. (2019). Early treatment response in psychotherapy for depression and personality disorder: Links with core conflictual relationship themes. *Psychotherapy Research*, 30(1), 112–123. https://doi.org/10.1080/10503307.2019.1609114

Hengartner, M. P., & Lehmann, S. N. (2017). Why psychiatric research must abandon traditional diagnostic classification and adopt a fully dimensional scope: Two solutions to a persistent problem. *Frontiers in Psychiatry*, *8*, 101. https://doi.org/10.3389/fpsyt.2017.00101

Hicks, B. M., Clark, D. A., & Durbin, C. E. (2017). Person-centered approaches in the study of personality disorders. *Personality Disorders*, *8*(4), 288–297. https://doi.org/10.1037/per0000212

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. *Psychological Medicine*, *38*(1), 135–146. https://doi.org/10.1017/S0033291707001626

Hoemann, K., Lee, Y., Kuppens, P., Gendron, M., & Boyd, R. L. (2023). Emotional granularity is associated with daily experiential diversity. *Affective Science*. https://doi.org/10.1007/s42761-023-00185-2

Hoemann, K., Nielson, C., Yuen, A., Gurera, J. W., Quigley, K. S., & Barrett, L. F. (2021). Expertise in emotion: A scoping review and unifying framework for individual differences in the mental representation of emotional experience. *Psychological Bulletin*, *147*(11), 1159–1183. https://doi.org/10.1037/bul0000327

Holtzman, N., Tackman, A., Carey, A., Brucks, M., Küfner, A., Deters, F., Back, M., Pennebaker, J., Sherman, R., & Mehl, M. (2019). Linguistic markers of grandiose narcissism: A LIWC analysis of 15 samples. *Journal of Language and Social Psychology*, *38*(5-6), 773–786. https://doi.org/10.1177%2F0261927X19871084

Hopwood, C. J., & Bleidorn, W. (2018). Stability and change in personality and personality disorders. *Current Opinion in Psychology, 21,* 6–10. https://doi.org/10.1016/j.copsyc.2017.08.034

Hopwood, C. J., Kotov, R., Krueger, R. F., Watson, D., Widiger, T. A., Althoff, R. R., Ansell, E. B., Bach, B., Michael Bagby, R., Blais, M. A., Bornovalova, M. A., Chmielewski, M., Cicero, D. C., Conway, C., De Clercq, B., De Fruyt, F., Docherty, A. R., Eaton, N. R., Edens, J. F., Forbes, M. K., … Zimmermann, J. (2018). The time has come for dimensional personality disorder diagnosis. *Personality and Mental Health, 12*(1), 82–86. https://doi.org/10.1002/pmh.1408

Horn, A. B., & Meier, T. (2022). Language in close relationships. In M. Dehghani & R. L. Boyd (Eds.), *The Handbook of Language Analysis in Psychology* (pp. 335–356). The Guilford Press.

Horvath, A. O., & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology, 38*(2), 139–149. https://doi.org/10.1037/0022-0167.38.2.139

Iida, M., Seidman, G., & Shrout, P. E. (2018). Models of interdependent individuals versus dyadic processes in relationship research. *Journal of Social and Personal Relationships*, *35*(1), 59–88. https://doi.org/10.1177/0265407517725407

Ireland, M. E., & Pennebaker, J. W. (2010). Language style match-ing in writing: Synchrony in essays, correspondence, and poetry. *Journal of Personality and Social Psychology, 99,* 549–571. https://doi.org/10.1037/a0020386

Jeong, H., Yim, H. W., Lee, S., Lee, H. K., Potenza, M. N., Kwon, J., Koo, H. J., Kweon, Y., Bhang, S., & Choi, J. (2018). Discordance between self-report and clinical diagnosis of Internet gaming disorder in adolescents. *Scientific Reports, 8,* 10084. https://doi.org/10.1038/s41598-018-28478-8

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

John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality, 2*(3), 171–203. https://doi.org/10.1002/per.2410020302

Johnson, S. L., Robison, M., Anvar, S., Swerdlow, B. A., & Timpano, K. R. (2022).

Emotion-related impulsivity and rumination: Unique and conjoint effects on suicidal ideation, suicide attempts, and nonsuicidal self-injury across two samples. *Suicide and Life-Threatening Behavior*, *52*(4), 642–654. https://doi.org/10.1111/sltb.12849

Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment*, *22*(2), 420–432. https://doi.org/10.1037/a0019265

Kaess M. (2022). Self-harm: A transdiagnostic marker of psychopathology and suicide risk during the COVID-19 pandemic? *European Child & Adolescent Psychiatry, 31*(7), 1–3. https://doi.org/10.1007/s00787-022-02044-0

Kahn, J., Tobin, R., Massey, A., & Anderson, J. (2007). Measuring emotional expression with the Linguistic Inquiry and Word Count. *The American Journal of Psychology, 120,* 263–286. https://doi.org/10.2307/20445398.

Kalokerinos, E. K., Erbas, Y., Ceulemans, E., & Kuppens, P. (2019). Differentiate to regulate: Low negative emotion differentiation is associated with ineffective use but not selection of emotion-regulation strategies. *Psychological Science, 30*(6), 863–879. https://doi.org/10.1177/0956797619838763

Katsakou, C., Marougka, S., Barnicot, K., Savill, M., White, H., Lockwood, K., & Priebe, S. (2012). Recovery in borderline personality disorder (BPD): A qualitative study of service users' perspectives. *PloS One, 7*(5), e36517. https://doi.org/10.1371/journal.pone.0036517

Kennedy, B., Ashokkumar, A., Boyd, R. L., & Dehghani, M. (2022). Text analysis for Psychology: Methods, principles, and practices. In M. Dehghani & R. L. Boyd (Eds.), *The Handbook of Language Analysis in Psychology* (pp. 3–62). The Guilford Press. https://doi.org/10.31234/osf.io/h2b8t

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2020). *Dyadic Data Analysis*. The Guilford Press.

Khawaja, M. A., Chen, F., & Marcus, N. (2014). Measuring cognitive load using linguistic features: Implications for usability evaluation and adaptive interaction design. *International Journal of Human-Computer Interaction, 30*(5), 343–368. https://doi.org/10.1080/10447318.2013.860579

Khazbak, M., Wael, Z., Ehab, Z., Gerorge, M., & Eliwa, E. (2021). MindTime: Deep learning approach for borderline personality disorder detection, *2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, 337–344, https://doi.org/10.1109/MIUCC52538.2021.9447620.

Kim, Y. R., & Tyrer, P. (2010). Controversies surrounding classification of personality disorder. *Psychiatry Investigation, 7*(1), 1–8. https://doi.org/10.4306/pi.2010.7.1.1

Klonsky, E. (2011). Non-suicidal self-injury in United States adults: Prevalence, sociodemographics, topography and functions. *Psychological Medicine, 41*(9), 1981–1986. https://doi.org/10.1017/S0033291710002497

Koenig, J., Klier, J., Parzer, P., Santangelo, P., Resch, F., Ebner-Priemer, U., & Kaess, M. (2021). High-frequency ecological momentary assessment of emotional and interpersonal states preceding and following self-injury in female adolescents. *European Child & Adolescent Psychiatry*, *30*(8), 1299–1308. https://doi.org/10.1007/s00787-020-01626-0

Koenigsberg, H. W., Harvey, P. D., Mitropoulou, V., New, A. S., Goodman, M., Silverman, J., Serby, M., Schopick, F., & Siever, L. J. (2001). Are the interpersonal and identity disturbances in the borderline personality disorder criteria linked to the traits of affective instability and impulsivity?. *Journal of Personality Disorders*, *15*(4), 358–370. https://doi.org/10.1521/pedi.15.4.358.19181

Kramer, A. D., & Chung, C. K. (2011). Dimensions of self-expression in Facebook status updates. *ICWSM*. https://doi.org/10.1609/icwsm.v5i1.14140

Kuehn, K. S., Dora, J., Harned, M. S., Foster, K. T., Song, F., Smith, M. R., & King, K. M. (2022). A meta-analysis on the affect regulation function of real-time self-injurious thoughts and behaviours. *Nature Human Behaviour*, *6*(7), 964–974. https://doi.org/10.1038/s41562-022-01340-8

Kühner, C., Bürger, C., Keller, F., & Hautzinger, M. (2007). Reliabilität und Validität des revidierten Beck-Depressionsinventars (BDI-II). Befunde aus deutschsprachigen Stichproben [Reliability and validity of the Revised Beck Depression Inventory (BDI-II). Results from German samples]. *Der Nervenarzt*, *78*(6), 651–656. https://doi.org/10.1007/s00115-006-2098-7

Kulkarni, V., Kern, M. L., Stillwell, D., Kosinski, M., Matz, S., Ungar, L., Skiena, S., &

Schwartz, H. A. (2018). Latent human traits in the language of social media: An open-vocabulary approach. *PLOS ONE, 13*(11), e0201703. https://doi.org/10.1371/journal.pone.0201703

Lauw, M., How, C. H., & Loh, C. (2015). PILL Series. Deliberate self-harm in adolescents. *Singapore Medical Journal*, *56*(6), 306–309. https://doi.org/10.11622/smedj.2015087

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

Le, M., Woodworth, M., Gillman, L., Hutton, E., & Hare, R. (2017). The linguistic output of psychopathic offenders during a PCL-R interview. *Criminal Justice and Behavior*, *44*(4), 551–565. https://doi.org/10.1177/0093854816683423

Leavitt, J. (2019). *A lexical analysis of the child attachment interview in an adolescent sample: Markers and correlates of borderline personality disorder* (Unpublished doctoral thesis). Retrieved from https://uh-ir.tdl.org/handle/10657/4681

Leichsenring, F., Leibing, E., Kruse, J., New, A. S., & Leweke, F. (2011). Borderline personality disorder. *The Lancet*, *377*(9759), 74–84. https://doi.org/10.1016/S0140-6736(10)61422-5

Levy, K. N. (2005). The implications of attachment theory and research for understanding borderline personality disorder. *Development and Psychopathology*, *17*(4), 959–986. https://doi.org/10.1017/s0954579405050455

Levy-Gigi, E., & Shamay-Tsoory, S. (2022). Affect labeling: The role of timing and intensity. *Plos One, 17*(12), e0279303. https://doi.org/10.1371/journal.pone.0279303

Li, H., Graesser, A. C., & Cai, Z. (2014). Comparison of Google translation with human translation. *Proceedings of the Twenty-Seventh International Florida Artificial Intelligence Research Society Conference*, 190–195. https://www.aaai.org/ocs/index.php/FLAIRS/FLAIRS14/paper/view/7864

Li, Y., Masitah, A., & Hills, T. T. (2020). The Emotional Recall Task: Juxtaposing recall and recognition-based affect scales. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(9), 1782. https://psycnet.apa.org/doi/10.1037/xlm0000841

Linehan, M. M. (1993). *Cognitive-Behavioral Treatment of Borderline Personality Disorder.* Guilford Press.

Linehan, M. M. (2014). *DBT Skills Training Manual (2nd Ed.).* Guilford Press.

Lorenzini, N., & Fonagy, P. (2013). Attachment and personality disorders: A short review. *FOCUS: The Journal of Lifelong Learning in Psychiatry, 11*(2) 155–166. https://doi.org/10.1176/appi.focus.11.2.155

Luborsky, L., & Crits-Christoph, P. (1998). *Understanding Transference: The Core Conflictual Relationship Theme Method* (2nd Ed.). American Psychological Association.

Lyons, M. (2019). *The Dark Triad of Personality: Narcissism, Machiavellianism, and Psychopathy in Everyday Life.* Elsevier Academic Press.

Lyons, M., Aksayli, N. D., & Brewer, G. (2018). Mental distress and language use: Linguistic analysis of discussion forum posts. *Computers in Human Behavior, 87,* 207–211. https://doi.org/10.1016/j.chb.2018.05.035

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research, 30,* 457–500. https://doi.org/10.1613/jair.2349

Ma-Kellams, C., Or, F., Baek, J. H., & Kawachi, I. (2016). Rethinking suicide surveillance: Google Search data and self-reported suicidality differentially estimate completed suicide risk. *Clinical Psychological Science, 4*(3), 480–484. https://doi.org/10.1177/2167702615593475

Marceau, E. M., Meuldijk, D., Townsend, M. L., Solowij, N., & Grenyer, B. F. S. (2018). Biomarker correlates of psychotherapy outcomes in borderline personality disorder: A systematic review. *Neuroscience & Biobehavioral Reviews, 94,* 166–178. https://doi.org/10.1016/j.neubiorev.2018.09.001

Marchant, A., Hawton, K., Stewart, A., Montgomery, P., Singaravelu, V., Lloyd, K., Purdy, N., Daine, K., & John, A. (2017). A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown. *PloS One, 12*(8), e0181722. https://doi.org/10.1371/journal.pone.0181722

Marcus, D. K., & Zeigler-Hill, V. (2016). Understanding the dark side of personality: Reflections and future directions. In *The dark side of personality: Science and practice in social, personality, and clinical psychology* (pp. 363–374). American Psychological Association. https://doi.org/10.1037/14854-019

Marko, K., & Leibetseder, I. (2023). Linguistic indicators of psychopathy and malignant narcissism in the personal letters of the Austrian Killer Jack Unterweger. *Forensic Sciences, 3*(1), 45–68. http://dx.doi.org/10.3390/forensicsci3010006

Martindale, C. (1975a). The grammar of altered states of consciousness: A semiotic reinterpretation of aspects of psychoanalytic theory. *Psychoanalysis and Contemporary Science, 4,* 331–354.

Matz, S. C., Gladstone, J. J., & Stillwell, D. (2017). In a world of big data, small effects can still matter: A reply to Boyce, Daly, Hounkpatin, and Wood (2017). *Psychological Science*, *28*(4), 547–550. https://doi.org/10.1177/0956797617697445

McMurran, M., Huband, N., & Overton, E. (2010). Non-completion of personality disorder treatments: A systematic review of correlates, consequences, and interventions. *Clinical Psychology Review*, *30*(3), 277–287. https://doi.org.ezproxy.uow.edu.au/10.1016/j.cpr.2009.12.002

Meier, T., Stephens, J. E., & Haase, C. M. (In press). Feelings in words: Emotion word use and cardiovascular reactivity in marital interactions. *Emotion*.

Mergenthaler E., & Kächele H. (1988). The Ulm Textbank Management System: A tool for psychotherapy research. In Dahl H., Kächele H., & Thomä H. (eds.), *Psychoanalytic Process Research Strategies*. Springer: Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-74265

Miano, A., Dziobek, I., & Roepke, S. (2017a). Understanding interpersonal dysfunction in borderline personality disorder: A naturalistic dyadic study reveals absence of relationship-protective empathic inaccuracy. *Clinical Psychological Science, 5*(2), 355–366. https://doi.org/10.1177/2167702616683505

Miano, A., Dziobek, I., & Roepke, S. (2020). Characterizing couple dysfunction in borderline personality disorder. *Journal of Personality Disorders*, *34*(2), 181–198. https://doi.org/10.1521/pedi_2018_32_388

Miano, A., Fertuck, E. A., Roepke, S., & Dziobek, I. (2017b). 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

Miller, J. D., Vize, C., Crowe, M. L., & Lynam, D. R. (2019). A critical appraisal of the dark-triad literature and suggestions for moving forward. *Current Directions in Psychological Science, 28*(4), 353–360. https://doi.org/10.1177/0963721419838233

Millner A. J., Lee M. D., & Nock M. K. (2016). Describing and measuring the pathway to suicide attempts: A preliminary study. *Suicide and Life-Threatening Behavior*, 47, 353–369. https://doi.org/10.1111/sltb.12284

Molendijk, M. L., Bamelis, L., van Emmerik, A. A. P., Arntz, A., Haringsma, R., & Spinhoven, P. (2010). Word use of outpatients with a personality disorder and concurrent or previous major depressive disorder. *Behaviour Research and Therapy*, *48*(1), 44–51. https://doi.org/10.1016/j.brat.2009.09.007

Morant, N., Chilman, N., Lloyd-Evans, B., Wackett, J. and Johnson, S. (2021) ‘Acceptability of using social media content in mental health research: A reflection. Comment on “Twitter users’ ’ views on mental health crisis resolution team care compared with stakeholder interviews and focus groups: Qualitative analysis”””, *JMIR Mental Health*, 8(8), 1–2. https://doi.org/10.2196/32475.

Morey, L. C. (1991). *The Personality Assessment Inventory Professional Manual.* Psychological Assessment Resources

Muehlenkamp, J. J., Ertelt, T. W., Miller, A. L., & Claes, L. (2011). Borderline personality symptoms differentiate non-suicidal and suicidal self-injury in ethnically diverse adolescent outpatients. *Journal of Child Psychology and Psychiatry, 52*, 148–155. https://doi.org/10.1111/j.1469-7610.2010.02305.x

Muris, P., Merckelbach, H., Otgaar, H., & Meijer, E. (2017). The malevolent side of human nature: A meta-analysis and critical review of the literature on the dark triad (narcissism, Machiavellianism, and psychopathy). *Perspectives on Psychological Science, 12*(2), 183–204. https://doi.org/10.1177/1745691616666070

Neuman, Y., & Cohen, Y. A. (2014). Vectorial semantics approach to personality assessment. *Scientific Reports, 4,* 4761. https://doi.org/10.1038/srep04761

Neumann, C. S., Johansson, P. T., & Hare, R. D. (2013). The Psychopathy Checklist-Revised (PCL-R), low anxiety, and fearlessness: A structural equation modeling analysis. *Personality Disorders*, *4*(2), 129–137. https://doi.org/10.1037/a0027886

Newson, J. J., Hunter, D., & Thiagarajan, T. C. (2020). The heterogeneity of mental health assessment. *Frontiers in Psychiatry, 11*, 76. https://doi.org/10.3389/fpsyt.2020.00076

Nguyen, T., Phung, D., & Venkatesh, S. (2013). Analysis of psycholinguistic processes and topics in online autism communities. *Proceedings - IEEE International Conference on Multimedia and Expo*. https://doi.org/10.1109/ICME.2013.6607615

Nook, E. C., Satpute, A. B., & Ochsner, K. N. (2021). Emotion naming impedes both cognitive reappraisal and mindful acceptance strategies of emotion regulation. *Affective Science*, *2*(2), 187–198. https://doi.org/10.1007/s42761-021-00036-y

Oltmanns, T. F., & Balsis, S. (2011). Personality disorders in later life: Questions about the measurement, course, and impact of disorders. *Annual Review of Clinical Psychology, 7*, 321–349. https://doi.org/10.1146/annurev-clinpsy-090310-120435

Onraet, E., Hiel, A. V., & Dhont, K. (2013). The relationship between right-wing ideological attitudes and psychological well-being. *Personality and Social Psychology Bulletin. 39*(4), 509–522. https://doi.org/10.1177/0146167213478199

Ottenstein, C., & Lischetzke, T. (2019). Development of a novel method of emotion differentiation that uses open-ended descriptions of momentary affective states. *Assessment*, 107319111983913. https://doi.org/10.1177/1073191119839138

Paris, J., & Zweig-Frank, H. (2001). A 27-year follow-up of patients with borderline personality disorder. *Comprehensive Psychiatry*, *42*(6), 482–487. https://doi.org/10.1053/comp.2001.26271

Park, A., & Conway, M. (2017). Longitudinal changes in psychological states in online health community members: Understanding the long-term effects of participating in an online depression community. *Journal of Medical Internet Research, 19*(3), e71. https://doi.org/10.2196/jmir.6826

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism and psychopathy. *Journal of Research in Personality, 36*(6), 556–563. https://doi.org/10.1016/S0092-6566(02)00505-6

Pennebaker, J. W. (2011). *The Secret Life of Pronouns: What Our Words Say About Us*. Bloomsbury Press/Bloomsbury Publishing.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The Development and Psychometric Properties of LIWC2015*. Retrieved from https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf

Pennebaker, J. W., & Ireland, M. E. (2011). Using literature to understand authors: The case for computerized text analysis. *Scientific Study of Literature, 1*(1), 34–48. https://doi.org/10.1075/ssol.1.1.04pen

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77*(6), 1296–1312. https://doi.org/10.1037/0022-3514.77.6.1296

Pennebaker, J. W., Mayne, T. J., & Francis, M. E. (1997). Linguistic predictors of adaptive bereavement. *Journal of Personality and Social Psychology, 72*(4), 863–871. https://doi.org/10.1037/0022-3514.72.4.863

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*, 547–577. https://doi.org/10.1146/annurev.psych.54.101601.145041

Peters, J. R., Eisenlohr-Moul, T. A., Upton, B. T., Talavera, N. A., Folsom, J. J., & Baer, R. (2017). Characteristics of repetitive thought associated with borderline personality features: A multimodal investigation of ruminative content and style. *Journal of Psychopathology and Behavioral Assessment*, *39*(3), 456–466. https://doi.org/10.1007/s10862-017-9594-x

Preotiuc-Pietro, D., Carpenter, J., Giorgi, S., & Ungar, L. (2016). Studying the dark triad of personality through Twitter behavior. *The 25th ACM International,* 761–770. https://doi.org/10.1145/2983323.2983822

Ramírez-Cifuentes, D., Freire, A., Baeza-Yates, R., Puntí, J., Medina-Bravo, P., Velazquez, D. A., Gonfaus, J. M., & Gonzàlez, J. (2020). Detection of suicidal ideation on social media: Multimodal, relational, and behavioral analysis. *Journal of Medical Internet Research, 22*(7), e17758. https://doi.org/10.2196/17758

Raskin, R., & Shaw, R. (1988). Narcissism and the use of personal pronouns. *Journal of Personality*, *56*(2), 393–404. https://doi.org/10.1111/j.1467-6494.1988.tb00892.x

Rathner, E.-M., Djamali, J., Terhorst, Y., Schuller, B. W., Cummins, N., Salamon, G., Hunger-Schoppe, C., & Baumeister, H. (2018). How did you like 2017? Detection of language markers of depression and narcissism in personal narratives. *INTERSPEECH*. https://doi.org/10.21437/Interspeech.2018-2040

Reichl, C., & Kaess, M. (2021). Self-harm in the context of borderline personality disorder. *Current Opinion in Psychology, 37*, 139–144. https://doi.org/10.1016/j.copsyc.2020.12.007

Roepke, S., Vater, A., Preißler, S., Heekeren, H. R., & Dziobek, I. (2013). Social cognition in borderline personality disorder. *Frontiers in Neuroscience*, *6*, 195. https://doi.org/10.3389/fnins.2012.00195

Ronningstam, E., & Baskin-Sommers, A. R. (2013). Fear and decision-making in narcissistic personality disorder—A link between psychoanalysis and neuroscience. *Dialogues in Clinical Neuroscience*, *15*(2), 191–201.

Rosenbach, C., & Renneberg, B. (2015). Remembering rejection: Specificity and linguistic styles of autobiographical memories in borderline personality disorder and depression. *Journal of Behavior, Therapy and Experimental Psychiatry*, *46*, 85–92. https://doi.org/10.1016/j.jbtep.2014.09.002

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

Sauer-Zavala, S., & Barlow, D. H. (2014). The case for borderline personality disorder as an emotional disorder: Implications for treatment. *Clinical Psychology: Science and Practice, 21*(2), 118–138. https://doi.org/10.1111/cpsp.12063

Sawhney, R., Joshi, H., Nobles, A., & Shah, R. R. (2021). Tweet classification to assist human moderation for suicide prevention. *Proceedings of the International AAAI Conference on Weblogs and Social Media, 15*, 609–620.

Schimpff, A. (2019). “We” but not “me”: A sociolinguistic study of the speaker-exclusive first-person plural pronoun “we”. *Lifespans & Styles, 5*(1), 1–15. http://journals.ed.ac.uk/lifespansstylesusive

Schneider, B., Schnabel, A., Wetterling, T., Bartusch, B., Weber, B., & Georgi, K. (2008). How do personality disorders modify suicide risk? *Journal of Personality Disorders*, *22*(3), 233–245. https://doi.org/10.1521/pedi.2008.22.3.233

Schultheiss, O. (2013). Are implicit motives revealed in mere words? Testing the marker-word hypothesis with computer-based text analysis. *Frontiers in Psychology, 4,* 748. https://doi.org/10.3389/fpsyg.2013.00748

Schwartz, H. A., Eichstaedt, J., Kern, M. L., Park, G., Sap, M., Stillwell, D., Kosinski, M., & Ungar, L. (2014). Towards assessing changes in degree of depression through Facebook. *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality,* 118–125. https://www.aclweb.org/anthology/W14-3214

Seah, T. H. S., & Coifman, K. (2021). Emotion differentiation and behavioral dysregulation in clinical and non-clinical samples: A meta-analysis. *Emotion, 22*(7), 1686–1697. https://doi.org/10.1037/emo0000968

Selby, E. A., & Joiner, T. E., Jr (2009). Cascades of emotion: The emergence of borderline personality disorder from emotional and behavioral dysregulation. *Review of General Psychology: Journal of Division 1, of the American Psychological Association*, *13*(3), 219. https://doi.org/10.1037/a0015687

Selby, E. A., Yen, S., & Spirito, A. (2013). Time varying prediction of thoughts of death and suicidal ideation in adolescents: Weekly ratings over 6-month follow-up. *Journal of Clinical Child and Adolescent Psychology*, *42*(4), 481–495. https://doi.org/10.1080/15374416.2012.736356

Sempértegui, G. A., Karreman, A., Arntz, A., & Bekker, M. H. J. (2013). Schema therapy for borderline personality disorder: A comprehensive review of its empirical foundations, effectiveness and implementation possibilities. *Clinical Psychology Review, 33*(3), 426–447, https://doi.org/10.1016/j.cpr.2012.11.006.

Seppala, J., Vita, I., Jamsa, T., Miettunen, J., Isohanni, M., Rubinstein, K., Feldman, Y., Grasa, E., Corripio, I., Berdun, J., D'Amico, E., & Bulgheroni, M. (2019). Smartphone and wearable sensors-based m-Health approach for psychiatric disorders and symptoms – a systematic review and link to m-RESIST project (Preprint). *JMIR Mental Health, 6*(2), e9819. https://doi.org/10.2196/mental.9819

Shah, R., & Zanarini, M. C. (2018). Comorbidity of borderline personality disorder: Current status and future directions. *The Psychiatric Clinics of North America, 41*(4), 583–593. https://doi.org/10.1016/j.psc.2018.07.009

Sharp, C., Pane, H., Ha, C., Venta, A., Patel, A. B., Sturek, J., & Fonagy, P. (2011). Theory of mind and emotion regulation difficulties in adolescents with borderline traits. *Journal of the American Academy of Child and Adolescent Psychiatry, 50*(6), 563–573.e1. https://doi.org/10.1016/j.jaac.2011.01.017

Sharp, C., Wright, A. G., Fowler, J. C., Frueh, B. C., Allen, J. G., Oldham, J., & Clark, L. A. (2015). The structure of personality pathology: Both general ('g') and specific ('s') factors?. *Journal of Abnormal Psychology, 124*(2), 387–398. https://doi.org/10.1037/abn0000033

Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine, 49*(9), 1–23. https://doi.org/10.1017/S0033291719000151

Sherman, L. E., Payton, A. A., Hernandez, L. M., Greenfield, P. M., & Dapretto, M. (2016). The power of the like in adolescence: Effects of peer influence on neural and behavioral responses to social media. *Psychological Science, 27*(7), 1027–1035. https://doi.org/10.1177/0956797616645673

Sierra, G., Andrade-Palos, P., Bel-Enguix, G., Osornio-Arteaga, A., Cabrera-Mora, A., García-Nieto, L., & Sierra-Aparicio, T. (2022). Suicide risk factors: A language analysis approach in social media. *Journal of Language and Social Psychology*, *41*(3), 312–330. https://doi.org/10.1177/0261927X211036171

Snir, A., Rafaeli, E., Gadassi, R., Berenson, K., & Downey, G. (2015). Explicit and inferred motives for nonsuicidal self-injurious acts and urges in borderline and avoidant personality disorders. *Personality Disorders*, *6*(3), 267–277. https://doi.org/10.1037/per0000104

Soldaini, L., Walsh, T., Cohan, A., Han, J., & Goharian, N. (2018). Helping or hurting? Predicting changes in users’ risk of self-harm through online community interactions. In *Proceedings of the Fifth Workshop on Computational Linguistics* *and Clinical Psychology: From Keyboard to Clinic*, 194–203. Association for Computational Linguistics. http://dx.doi.org/10.18653/v1/W18-0621

Sonnenschein, A. R., Hofmann, S. G., Ziegelmayer, T., & Lutz, W. (2018). Linguistic analysis of patients with mood and anxiety disorders during cognitive behavioral therapy. *Cognitive Behaviour Therapy, 47*(4), 315–327. https://doi.org/10.1080/16506073.2017.1419505

Southward, M. W., Heiy, J. E., & Cheavens, J. S. (2019). Emotions as context: Do the naturalistic effects of emotion regulation strategies depend on the regulated emotion?. *Journal of Social and Clinical Psychology*, *38*(6), 451–474. https://doi.org/10.1521/jscp.2019.38.6.451

Spruit, M., Verkleij, S., de Schepper, K., & Scheepers, F. (2022). Exploring language markers of mental health in psychiatric stories. *Applied Sciences, 12*(4), 2179. http://dx.doi.org/10.3390/app12042179

Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J. W., Patel, V., & Silove, D. (2014). The global prevalence of common mental disorders: A systematic review and meta-analysis 1980-2013. *International Journal of Epidemiology, 43*(2), 476–493. https://doi.org/10.1093/ije/dyu038

Steffen, A., Nübel, J., Jacobi, F., Bätzing, J., & Holstiege, J. (2020). Mental and somatic comorbidity of depression: A comprehensive cross-sectional analysis of 202 diagnosis groups using German nationwide ambulatory claims data. *BMC Psychiatry, 20*(1), 142. https://doi.org/10.1186/s12888-020-02546-8

Stelmack, R. M., & Stalikas, A. (1991). Galen and the humour theory of temperament. *Personality and Individual Differences*, *12*(3), 255–263. https://doi.org/10.1016/01918869(91)90111-N

Sumner, C., Byers, A., Boochever, R., & Park, G. J. (2012). Predicting dark triad personality traits from Twitter usage and a linguistic analysis of Tweets. *2012 11th International Conference on Machine Learning and Applications*, 386–393. https://doi.org/10.1109/ICMLA.2012.218

Suvak, M. K., Litz, B. T., Sloan, D. M., Zanarini, M. C., Barrett, L. F., & Hofmann, S. G. (2011). Emotional granularity and borderline personality disorder. *Journal of Abnormal Psychology, 120*(2), 414–426. https://doi.org/10.1037/a0021808

Tackman, A. M., Baranski, E. N., Danvers, A. F., Sbarra, D. A., Raison, C. L., Moseley,

S. A., Polsinelli, A. J., & Mehl, M. R. (2020). ‘Personality in its Natural Habitat’ revisited: A pooled, multi-sample examination of the relationships between the Big Five personality traits and daily behaviour and language use. *European Journal of Personality, 34,* 753–776. https://doi.org/10.1002/per.2283

Tackman, A., Sbarra, D., Carey, A., Donnellan, M., Horn, A., Holtzman, N., Edwards, T., Pennebaker, J., & Mehl, M. (2019). Depression, negative emotionality, and self-referential language: A multi-lab, multi-measure, and multi-language-task research synthesis. *Journal of Personality and Social Psychology, 116*(5), 817–834. https://doi.org/10.1037/pspp0000187

Tausczik, Y., & Pennebaker, J. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29,* 24–54. https://doi.org/10.1177/0261927X09351676

The British Psychological Society (2021). *Ethics Guidelines for Internet-Mediated Research*. https://doi.org/10.53841/bpsrep.2021.rep155

Thompson, R. J., Springstein, T., & Boden, M. (2021). Gaining clarity about emotion differentiation. *Social and Personality Psychology Compass, 15*(3), e12584. https://psycnet.apa.org/doi/10.1111/spc3.12584

Tolboll, K. (2019). Linguistic features in depression: A meta-analysis. *Language Works, 4*(2), 39–59. Retrieved from https://tidsskrift.dk/lwo/article/view/117798

Torre, J. B., & Lieberman, M. D. (2018). Putting feelings into words: Affect labeling as implicit emotion regulation. *Emotion Review, 10*(2), 116–124. https://doi.org/10.1177/1754073917742706

Trull, T. J., Vergés, A., Wood, P. K., Jahng, S., & Sher, K. J. (2012). The structure of Diagnostic and Statistical Manual of Mental Disorders (4th edition, text revision) personality disorder symptoms in a large national sample. *Journal of Personality Disorders, 3*(4), 355–369. https://doi.org/10.1037/a0027766

Tucker, I. M., & Lavis, A. (2019). Temporalities of mental distress: Digital immediacy and the meaning of ‘crisis’ in online support. *Sociology of Health and Illness, 41*, 132–146. https://doi.org/10.1111/1467-9566.12943

Tušl, M., Thelen, A., Marcus, K., Peters, A., Shalaeva, E., Scheckel, B., Sykora, M., Elayan, S., Naslund, J. A., Shankardass, K., Mooney, S. J., Fadda, M., & Gruebner, O. (2022). Opportunities and challenges of using social media big data to assess mental health consequences of the COVID-19 crisis and future major events. *Discover Mental Health, 2*(1), 14. https://doi.org/10.1007/s44192-022-00017-y

Tyrer, P., Mulder, R., Crawford, M., Newton-Howes, G., Simonsen, E., Ndetei, D., Koldobsky, N., Fossati, A., Mbatia, J., & Barrett, B. (2010). Personality disorder: A new global perspective. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, *9*(1), 56–60. https://doi.org/10.1002/j.2051-5545.2010.tb00270.x

Tyrer, P., Reed, G. M., & Crawford, M. J. (2015). Classification, assessment, prevalence, and effect of personality disorder. *The Lancet*, *385*(9969), 717–726. https://doi.org/10.1016/S0140-6736(14)61995-4

Uban, A., Chulvi-Ferriols, M. A., & Rosso, P. (2021). An emotion and cognitive based analysis of mental health disorders from social media data. *Future Generation Computer Systems, 124,* 480–494. https://doi.org/10.1016/j.future.2021.05.032

Underberg, J. E., Gollwitzer, A., Oettingen, G., & Gollwitzer, P. M. (2019). The best words: Linguistic indicators of grandiose narcissism in politics. *Journal of Language and Social Psychology, 39*(2), 271–281. https://doi.org/10.1177/0261927X19883898

Van Orden, K. A., Witte, T. K., Cukrowicz, K. C., Braithwaite, S. R., Selby, E. A., & Joiner, T. E., Jr (2010). The interpersonal theory of suicide. *Psychological Review*, *117*(2), 575–600. https://doi.org/10.1037/a0018697

Vansteelandt, K., Houben, M., Claes, L., Berens, A., Sleuwaegen, E., Sienaert, P., & Kuppens, P. (2017). The affect stabilization function of nonsuicidal self injury in borderline personality disorder: An ecological momentary assessment study. *Behaviour, Research and Therapy*, *92*, 41–50. https://doi.org/10.1016/j.brat.2017.02.003

Videler, A. C., Hutsebaut, J., Schulkens, J. E. M., Sobczak, S., & van Alphen, S. P. J. (2019). A life span perspective on borderline personality disorder. *Current Psychiatry Reports*, *21*(7), 51. https://doi.org/10.1007/s11920-019-1040-1

Vine, V., Bernstein, E. E., & Nolen-Hoeksema, S. (2019). Less is more? Effects of exhaustive vs. minimal emotion labelling on emotion regulation strategy planning. *Cognition and Emotion, 33*(4), 855–862. https://doi.org/10.1080/02699931.2018.1486286

Vine, V., Boyd, R. L., & Pennebaker, J. W. (2020). Natural emotion vocabularies as windows on distress and well-being. *Nature Communications, 11,* 4525. httsps://doi.org/10.1038/s41467-020-18349-0

Wampold, B. E. (2015). How important are the common factors in psychotherapy? An update. *World Psychiatry, 14*(3), 270–277. https://doi.org/10.1002/wps.20238

Wang, B., Wu, Y., Taylor, N., Lyons, T., Liakata, M., Nevado-Holgado, A., & Saunders, K. (2020). Learning to detect bipolar disorder and borderline personality disorder with language and speech in non-clinical interviews. Proceedings of *International Speech Communication Association*, 437–441. http://dx.doi.org/10.21437/Interspeech.2020-3040

Werner, K. B., Few, L. R., & Bucholz, K. K. (2015). Epidemiology, comorbidity, and behavioral genetics of antisocial personality disorder and psychopathy. *Psychiatric Annals*, *45*(4), 195–199. https://doi.org/10.3928/00485713-20150401-08

Williams, G. E., & Uliaszek, A. A. (2022). Measuring negative emotion differentiation via coded descriptions of emotional experience. *Assessment*, *29*(6), 1144–1157. https://doi.org/10.1177/10731911211003949

Wilmot, M. P., Haslam, N., Tian, J., & Ones, D. S. (2019). Direct and conceptual replications of the taxometric analysis of Type A behavior. *Journal of Personality and Social Psychology, 116*(3), 12–26. https://doi.org/10.1037/pspp0000195

Windsor, L. C., Cupit, J. G., & Windsor, A. J. (2019). Automated content analysis across six languages. *PLOS ONE, 14*(11), e0224425. https://doi.org/10.1371/journal.pone.0224425

Winsper, C. (2018). The aetiology of borderline personality disorder (BPD): Contemporary theories and putative mechanisms. *Current Opinion in Psychology, 21,* 105–110. https://doi.org/10.1016/j.copsyc.2017.10.005

Winsper, C., Bilgin, A., Thompson, A., Marwaha, S., Chanen, A. M., Singh, S. P., Wang, A., & Furtado, V. (2020). The prevalence of personality disorders in the community: a global systematic review and meta-analysis. *British Journal of Psychiatry, 216*(2), 69–78. https://doi.org/10.1192/bjp.2019.166

Witt, S. H., Streit, F., Jungkunz, M., Frank, J., Awasthi, S., Reinbold, C. S., Treutlein, J., Degenhardt, F., Forstner, A. J., Heilmann-Heimbach, S., Dietl, L., Schwarze, C. E., Schendel, D., Strohmaier, J., Abdellaoui, A., Adolfsson, R., Air, T. M., Akil, H., Alda, M., Alliey-Rodriguez, N., … Rietschel, M. (2017). Genome-wide association study of borderline personality disorder reveals genetic overlap with bipolar disorder, major depression and schizophrenia. *Translational Psychiatry, 7*(6), e1155. https://doi.org/10.1038/tp.2017.115

Wojcik, S. P., Hovasapian, A., Graham, J., Motyl, M., & Ditto, P. H. (2015). Conservatives report, but liberals display, greater happiness. *Science, 347*(6227), 1243–1246. https://doi.org/10.1126/science.1260817

Woodworth, M., & Porter, S. (2002). In cold blood: Characteristics of criminal homicides as a function of psychopathy. *Journal of Abnormal Psychology, 111*(3), 436–445. https://doi.org/10.1037/0021-843X.111.3.436

Wright, A. G. C. (2017). The current state and future of factor analysis in personality disorder research. *Personality Disorders: Theory, Research, and Treatment, 8*(1), 14–25. https://doi.org/10.1037/per0000216

Wright, A. G. C., Hopwood, C. J., Skodol, A. E., & Morey, L. C. (2016). Longitudinal validation of general and specific structural features of personality pathology. *Journal of Abnormal Psychology, 125*(8), 1120–1134. https://doi.org/10.1037/abn0000165

Wright, A. G. C., Scott, L. N., Stepp, S. D., Hallquist, M. N., & Pilkonis, P. A. (2015). Personality pathology and interpersonal problem stability. *Journal of Personality Disorders, 29*(5), 684–706.

Wright, A. G. C., & Simms, L. J. (2016). Stability and fluctuation of personality disorder features in daily life. *Journal of Abnormal Psychology, 125*(5), 641–656. https://doi.org/10.1037/abn0000169

Wright, A. G. C, & Zimmerman, J. (2015). At the nexus of science and practice: Answering basic clinical questions in personality disorder assessment and diagnosis with quantitative modeling techniques. In Huprich, S., (Eds.), *Personality Disorders: Toward Theoretical and Empirical Integration in Diagnosis and Assessment*, pp. 109–144. American Psychological Association.

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality, 44*(3), 363–373. https://dx.doi.org/10.1016%2Fj.jrp.2010.04.001

Yeskuatov, E., Chua, S. L., & Foo, L. K. (2022). Leveraging Reddit for suicidal ideation detection: A review of machine learning and natural language processing techniques. *International Journal of Environmental Research and Public Health*, *19*(16), 10347. https://doi.org/10.3390/ijerph191610347

Yuan, C., Hong, Y., & Wu, J. (2020). Does Facebook activity reveal your dark side? Using online language features to understand an individual's dark triad and needs. *Behaviour & Information Technology, 41*, 292–306. https://doi.org/10.1080/0144929X.2020.1805513

Zajenkowski, M., & Szymaniak, K. (2021). Narcissism between facets and domains. The relationships between two types of narcissism and aspects of the Big Five. *Current Psychology, 40,* 2112–2121. https://doi.org/10.1007/s12144-019-0147-1

Zaki, L. F., Coifman, K. G., Rafaeli, E., Berenson, K. R., & Downey, G. (2013). Emotion differentiation as a protective factor against nonsuicidal self-injury in borderline personality disorder. *Behavior Therapy*, *44*(3), 529–540. https://doi.org/10.1016/j.beth.2013.04.008

Zeigler-Hill, V., & Marcus, D. K. (Eds.). (2016). *The Dark Side of Personality: Science and Practice in Social, Personality, and Clinical Psychology*. American Psychological Association. https://doi.org/10.1037/14854-000

Zhang, S., Fingerman, K. L., & Birditt, K. S. (2023). Detecting narcissism from older adults' daily language use: A machine learning approach. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *78*(9), 1493–1500. https://doi.org/10.1093/geronb/gbad061

Zimmermann, J., Brockmeyer, T., Hunn, M., Schauenburg, H., & Wolf, M. (2017). First-person pronoun use in spoken language as a predictor of future depressive symptoms: Preliminary evidence from a clinical sample of depressed patients. *Clinical Psychology & Psychotherapy*, *24*(2), 384–391. https://doi.org/10.1002/cpp.2006

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology.* Addison-Wesley Press, Inc.

Zomick, J., Levitan, S. I., & Serper, M. (2019). Linguistic analysis of schizophrenia in Reddit posts. *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology,* 74–83. http://dx.doi.org/10.18653/v1/W19-3009

Appendices

Appendix A:

Supplemental Material for "Uncovering the Social-Cognitive Contributors to Social Dysfunction in Borderline Personality Disorder Through Language Analysis" (Paper 1; Chapter 3)

A.1 Analysis of Missing Data

To check for possible sampling biases, we conducted an analysis of missing (i.e., excluded) data. In particular, we examined whether there were any differences between those who were excluded due to not meeting the minimum word count threshold (i.e., minimum of 50 words) for the relationship writing task and those who were included on key outcome variables: BPD features and Dark Triad traits. To test this, we ran independent, two-tailed *t*-tests comparing differences between those excluded due to writing < 50 words for the writing task (*n* = 67) and those included (*n* = 530) in BPD features and Dark Triad traits.

The results from the *t*-tests revealed no significant differences between those included versus those excluded in BPD features (*p* = .071). However, those who were excluded (and thus wrote less words in their relationship essays) were found to have significantly higher levels of Dark Triad traits, for all traits, compared to those included. Specifically, compared to those included, individuals who were excluded were found to score higher on overall Dark Triad traits (*t*(71) = 3.61, *p* <.001, *d* = 0.59), Machiavellianism (*t*(75) = 2.73, *p* = .008, *d* = 0.43), narcissism (*t*(581) = 3.38, *p* <.001, *d* = 0.45), and psychopathy (*t*(77) = 3.86, *p* <.001, *d* = 0.56).

It is not clear, however, that these differences would alter any interpretation of our results – the only conclusion that can be drawn from this analysis is that individuals scoring higher on *exclusively* Dark Triad traits (i.e., equally high BPD traits but especially high Dark Triad traits) may have lacked the patience or interest to conscientiously complete the writing task as instructed compared to individuals within our final sample.

A.2 Sociodemographic and Descriptive Statistics

Table A.1

Sociodemographic Characteristics of Participants and Descriptive Statistics for Problematic Personality Self-Report Measures (N = 530)

Characteristic	Mean	*SD*
Age (*n* = 528)	26.22	8.41
	n	%
Gender (*n* = 520)		
Female	379	72.88
Male	126	24.23
Non-binary	15	2.88
Ethnicity (*n* = 521)		
Asian	40	7.68
Black	10	1.92
Hispanic or Latino	37	7.10
Mixed	34	6.53
White	394	75.62
Other	6	1.15
Marital status (*n* = 521)		
Single	268	51.44
Married/partnered	237	45.49
Divorced/separated	16	3.07
Education level (*n* = 522)		
Less than high school	17	3.26

High school/some college	260	49.81
College	74	14.18
University/postgraduate degree	171	32.76
Employment status (*n* = 524)		
Unemployed	117	22.33
Student	172	32.82
Employed	213	40.65
Self-employed	16	3.05
Retired	6	1.15
Measure	Mean	*SD*
BPD features (*n* = 498)	37.24	13.03
Affect	10.42	4.19
Identity	10.46	4.15
Social	9.62	3.74
Self-harm	6.69	4.33
Dark Triad (*n* = 512)	31.80	10.15
Psychopathy	10.36	4.08
Narcissism	10.93	4.08
Machiavellianism	10.49	4.56

A.3 LIWC Descriptive Statistics

Table A.2

Descriptive Statistics for LIWC Categories Included in the Principal Component Analysis (N = 530)

LIWC Category	Mean	*SD*	Range	Minimum	Maximum	Skewness
WC	211.60	186.22	3747.00	51.00	3798.00	13.76
ppron	15.39	3.49	24.24	1.47	25.71	-0.37
i	12.58	3.10	22.22	0.00	22.22	-0.40
we	0.55	0.85	5.77	0.00	5.77	2.31
you	0.13	0.48	5.63	0.00	5.63	6.97
shehe	0.65	1.12	6.12	0.00	6.12	2.32

they	1.48	1.30	7.93	0.00	7.93	1.15
ipron	5.00	2.21	15.15	0.00	15.15	0.56
article	3.80	1.56	8.77	0.00	8.77	0.21
prep	14.06	2.68	18.02	5.71	23.73	0.08
auxverb	10.29	2.43	16.18	1.47	17.65	-0.08
adverb	6.42	2.27	14.08	0.00	14.08	0.18
conj	8.49	2.15	13.18	2.82	16.00	0.31
negate	2.52	1.53	9.52	0.00	9.52	0.78
verb	18.23	3.06	20.67	7.69	28.36	-0.04
adj	5.56	2.07	12.38	0.00	12.38	0.29
compare	3.01	1.62	10.77	0.00	10.77	0.85
interrog	1.48	1.04	6.06	0.00	6.06	0.69
number	0.80	0.88	5.71	0.00	5.71	1.54
quant	2.76	1.47	9.41	0.00	9.41	0.85
affect	7.84	2.84	21.93	1.54	23.47	1.03
posemo	4.55	2.30	19.12	0.00	19.12	1.36
negemo	3.11	1.90	11.27	0.00	11.27	0.85
anx	0.78	0.92	6.12	0.00	6.12	1.92
anger	0.77	0.89	6.78	0.00	6.78	1.78
sad	0.57	0.71	4.40	0.00	4.40	1.70
social	12.61	3.31	19.60	4.07	23.67	0.28
family	1.30	1.34	8.99	0.00	8.99	1.96
friend	1.39	1.14	7.46	0.00	7.46	1.56
female	0.67	1.03	5.85	0.00	5.85	2.02
male	0.81	1.29	7.88	0.00	7.88	2.28
cogproc	15.67	3.84	28.70	1.69	30.39	-0.03
insight	3.02	1.66	10.42	0.00	10.42	0.73
cause	1.78	1.17	6.19	0.00	6.19	0.70
discrep	1.53	1.23	7.27	0.00	7.27	1.14
tentat	4.39	2.07	13.64	0.00	13.64	0.65
certain	1.57	1.15	6.25	0.00	6.25	0.99
differ	5.20	2.00	12.12	0.00	12.12	0.42
percept	2.11	1.35	8.16	0.00	8.16	0.95
see	0.39	0.57	5.26	0.00	5.26	2.67
hear	0.50	0.68	3.90	0.00	3.90	1.82
feel	1.12	1.00	5.71	0.00	5.71	1.10
bio	1.48	1.29	8.00	0.00	8.00	1.23
body	0.19	0.39	3.15	0.00	3.15	2.99
health	0.80	0.84	5.00	0.00	5.00	1.27
sexual	0.12	0.36	2.73	0.00	2.73	4.12
ingest	0.12	0.32	2.42	0.00	2.42	3.55
drives	10.50	3.31	23.73	2.94	26.67	0.62
affiliation	5.31	2.60	18.33	0.00	18.33	0.78
achieve	1.56	1.17	6.15	0.00	6.15	0.96
power	1.69	1.18	7.04	0.00	7.04	0.90
reward	1.90	1.29	8.96	0.00	8.96	1.23
risk	0.81	0.84	5.71	0.00	5.71	1.52

focuspast	2.33	2.02	11.38	0.00	11.38	1.22
focuspresent	14.96	3.43	20.83	3.79	24.62	-0.15
focusfuture	0.69	0.75	4.17	0.00	4.17	1.39
relativ	11.29	3.29	18.88	2.70	21.58	0.30
motion	1.15	0.94	5.83	0.00	5.83	1.10
space	5.55	1.98	12.22	0.00	12.22	0.17
time	4.69	2.18	12.99	0.00	12.99	0.52
work	1.30	1.19	6.94	0.00	6.94	1.17
leisure	0.79	0.86	8.82	0.00	8.82	2.41
home	0.55	0.70	7.61	0.00	7.61	2.95
money	0.22	0.44	3.08	0.00	3.08	2.86
relig	0.05	0.18	1.26	0.00	1.26	4.30
death	0.08	0.24	1.94	0.00	1.94	3.89
informal	0.60	0.78	5.23	0.00	5.23	2.13
swear	0.11	0.33	3.27	0.00	3.27	4.17
netspeak	0.11	0.40	4.96	0.00	4.96	7.40
assent	0.09	0.26	2.09	0.00	2.09	3.96

Note. Mean values represent the mean percentage of total words used.

A.4 Bivariate Correlation Results

Table A.3

Correlations between Social-Cognitive Components and BPD Features and Dark Triad Traits, Without Controlling for Age and Gender

	Connectedness/ Intimacy	Immediacy	Social Rumination	Negative Affect
BPD Features (*n* = 483)	-.12**	.12**	.14**	.20***
Overall Dark Triad (*n* = 497)	-.13**	.09*	.00	.09*
Machiavellianism (*n* = 497)	-.13**	.10*	-.01	.12**
Narcissism (*n* = 497)	-.04	.02	-.03	.11*
Psychopathy (*n* = 497)	-.15***	.10*	.05	-.02

****p* < .001, ***p* < .01, **p* < .05.
Note. All tests are two-tailed.

Appendix B:

Supplemental Material for “Natural Emotion Vocabularies and Borderline Personality Disorder” (Paper 2; Chapter 4)

B.1 Regression Results Tables for Study 1 (of Paper 2)

Table B.1

Regression Coefficients for Emotion Vocabularies (EVs) and Covariates Predicting BPD Features in Relationship Essays (N = 498)

EV Model	Predictors	β	*t*	*p*
Positive EV	General vocabulary	-0.08	-1.72	.086
	LIWC posemo	-0.07	-1.43	.154
	Positive EV	-0.06	-1.27	.205
Negative EV	General vocabulary	-0.11	-2.41	.016
	LIWC negemo	0.18	3.27	.001
	Negative EV	0.02	0.36	.722

Note. The “EV Model” column shows the EV (along with the covariates) included in the particular regression model; two regression models were conducted in total (i.e., one model for each EV). The predictor “LIWC posemo” refers to the total number of positive affect words and “LIWC negemo” refers to the total number of negative affect words, calculated from LIWC2015.

Table B.2

Regression Coefficients for Emotion Vocabularies (EVs) and Covariates Predicting BPD Features in Behaviour Essays (N = 387)

EV Model	Predictors	β	*t*	*p*
Positive EV	General vocabulary	-0.08	-1.63	.103
	LIWC posemo	0.03	0.62	.539
	Positive EV	-0.06	-1.01	.313
Negative EV	General vocabulary	-0.10	-1.93	.055
	LIWC negemo	0.15	2.11	.035
	Negative EV	0.03	0.42	.675

Note. The "EV Model" column shows the EV (along with the covariates) included in the particular regression model; two regression models were conducted in total (i.e., one model for each EV). The predictor "LIWC posemo" refers to the total number of positive affect words and "LIWC negemo" refers to the total number of negative affect words, calculated from LIWC2015.

B.2 Regression Results Table for Dimensional Analyses in Study 2 (of Paper 2)

Table B.3

Regression Coefficients for Emotion Vocabularies (EVs) and Covariates Predicting BPD Symptoms in Female Partners in Study 2 (N = 64)

EV Model	Predictors	β	*t*	*p*
Positive EV	General vocabulary	0.12	0.90	.371
	LIWC posemo	-0.25	-1.96	.055
	Positive EV	0.16	1.21	.233
Negative EV	General vocabulary	0.14	1.03	.306
	LIWC negemo	0.07	0.55	.582
	Negative EV	0.28	2.16	.034
Anxiety/fear EV	General vocabulary	0.11	0.82	.414
	LIWC negemo	0.04	0.29	.776
	Anxiety/fear EV	0.27	2.03	.047
Anger EV	General vocabulary	0.23	1.79	.079
	LIWC negemo	0.08	0.64	.523
	Anger EV	0.03	0.21	.833
Sadness EV	General vocabulary	0.24	1.84	.071
	LIWC negemo	0.08	0.63	.532
	Sadness EV	0.01	0.07	.945
Undifferentiated negative EV	General vocabulary	0.20	1.46	.148
	LIWC negemo	0.08	0.62	.541
	Undifferentiated negative EV	0.14	1.08	.286

Note. The "EV Model" column shows the EV (along with the covariates) included in the particular regression model; six regression models were conducted in total (i.e., one model per EV). Regression analyses for the negative EV subcategories were conducted as follow-up specificity tests following the predictive effects of overall negative EV. The predictor "LIWC posemo" refers to the total number of positive affect words and "LIWC negemo" refers to the total number of negative affect words, calculated from LIWC2015.

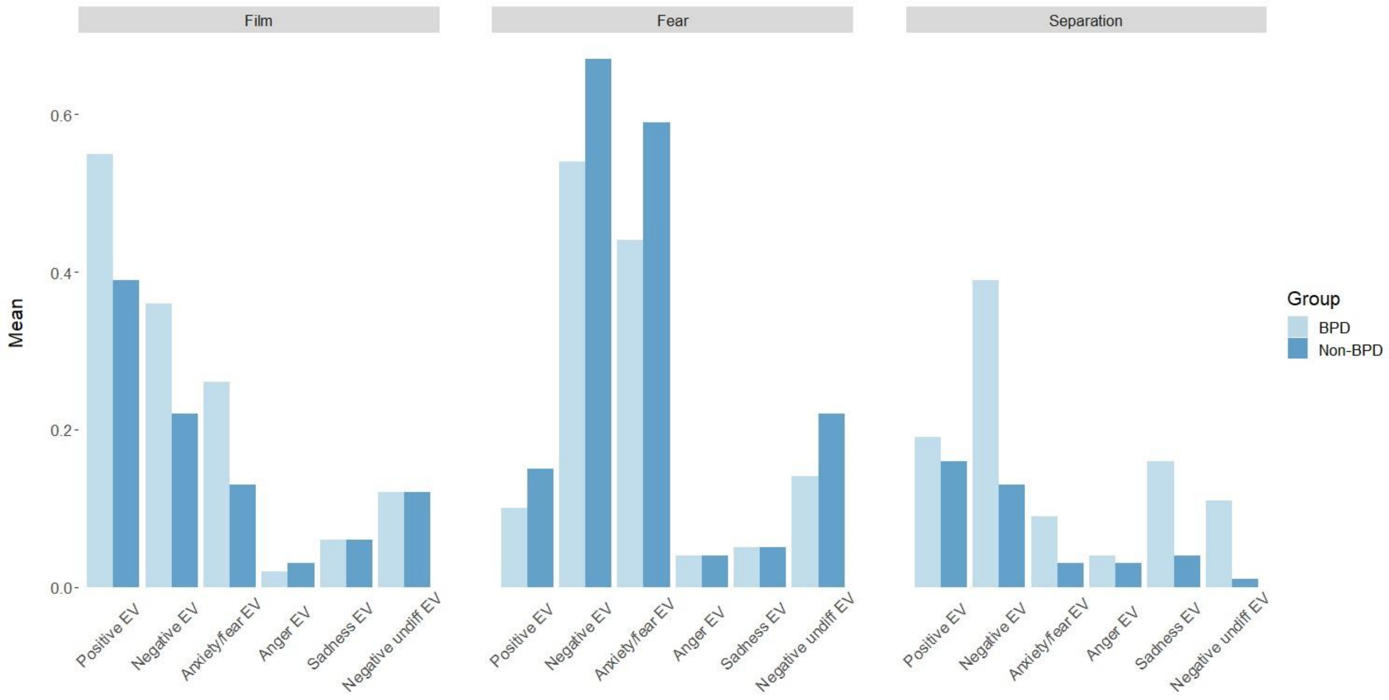
B.3 Group by Condition Interaction Effects on Emotion Vocabularies in Study 2 (of Paper 2)

Here, we discuss the 2 (group: women with BPD vs. women without BPD) x 3 (condition: neutral vs. personally-threatening vs. relationship-threatening) interaction effects on women's positive EVs, as only the significant interaction effects on negative EVs were reported in the main manuscript (see Figure B.1 for a comparison of EVs between women with BPD and women without BPD, split by condition).

As reported in the main body of the manuscript, significant group-by-condition interactions emerged for negative EV (refer to the main manuscript for the post-hoc analysis results for these interaction effects). Nonetheless, the interaction effect on positive EV was only marginally non-significant (*p* = .074). The interaction effect on positive EV was driven by women with BPD having significantly larger positive EVs than women without BPD in the neutral film condition, but not in the two threatening conditions. Further, women without BPD had larger positive EVs in the neutral condition compared to the personally-threatening (*M* difference = 0.22, *SE* = 0.07, *p* = .005) and relationship-threatening conditions (*M* difference = 0.20, *SE* = 0.07, *p* = .009), but there was no significant difference in positive EV between the personally-threatening and relationship-threatening conditions (*p* = .762). Likewise, women with BPD also had larger positive EVs in the neutral condition compared to the personally-threatening (*M* difference = 0.48, *SE* = 0.08, *p* < .001) and relationship-threatening conditions (*M* difference = 0.40, *SE* = 0.08, *p* < .001), and there was again no significant difference in positive EV between the personally-threatening and relationship-threatening conditions (*p* = .235).

Fig. B.1

Mean Emotion Vocabularies (EVs) of Women with BPD Versus Women without BPD Split by Condition (N = 64)



B.4 Emotion Vocabularies of Male Partners in Study 2 (of Paper 2)

Statistical comparisons of EVs between men with BPD partners and men without BPD partners (i.e., “control men”) in Study 2 are reported here. Focusing on the male partners allows for exploration into the effects of being in an intimate relationship with an individual with BPD on one’s own emotional experiences and functioning. Analyses comparing EVs in male partners have been conducted in the same way as with the women in the sample. Specifically, analyses include independent *t*-tests (two-tailed) and ANCOVAs comparing EVs between men with BPD partners and control men, as well as 2x3 mixed ANCOVAs looking at the group-by-condition interaction effects on EVs, while controlling for general vocabulary and corresponding emotion word frequencies (derived from LIWC). In addition, we also ran basic statistical tests comparing female and male partners (within each couple) in their emotive language, via two-tailed, paired *t*-tests.

B.4.1 Descriptive Analyses for Male Partners

With regard to the frequency of emotion words expressed by men in the conversations (calculated using LIWC), an average of 1.30% (*SD* = 0.60) of the men's language were of emotive nature, which included both negative (*M* = 0.82, *SD* = 0.44) and positive emotion (*M* = 0.48, *SD* = 0.34). The frequency of emotive language used by men was similar to that of their female partners, which was confirmed by *t*-tests showing no significant differences between male and female partners in the frequency of overall positive (*p* = .331) or negative affect words used (*p* = .723). As for the number of unique emotion words used by the men in the sample, the average positive EV was 0.26 (*SD* = 0.20) and negative EV was 0.36 (*SD* = 0.18). As with the frequency of emotion words, *t*-tests revealed no significant differences between male and female partners (within each couple) in any of the EVs (all *p*'s > .10).

Like with the female partners, none of the active EVs significantly correlated with the corresponding emotion word frequencies in men (Pearson's correlations all *p*'s > .05). Positive EV significantly correlated with general vocabulary size (*r* = .33, *p* = .007), but not negative EV. There was also no significant difference in general vocabulary size between male and female partners in the sample (*p* = .954).

B.4.2 BPD vs. Non-BPD Comparison of EVs in Male Partners

Figure B.2 presents a comparison of average active EVs (across all conditions) between men with BPD partners and men without BPD partners. Independent *t*-tests revealed no significant differences between men with BPD partners and control men in positive or negative EV. Confirming the results of the *t*-tests, there remained no significant differences between men with BPD partners and control men in positive or negative EV when controlling for general vocabulary and corresponding emotion word frequencies in univariate ANCOVAs (see Table B.4 for full ANCOVA results).

Table B.4

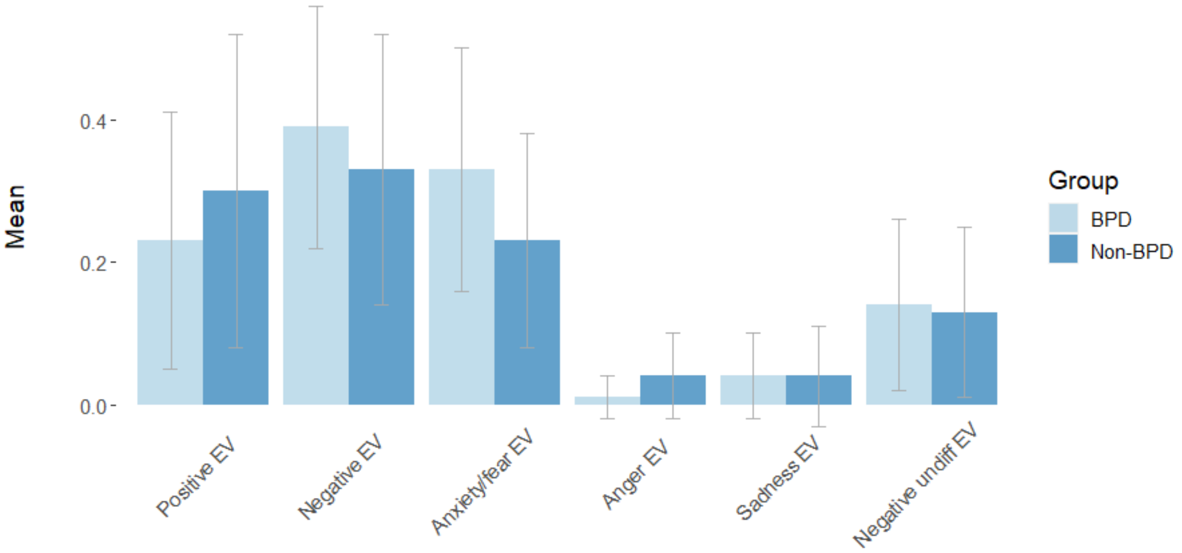
Differences in Emotion Vocabularies (EVs) Between Men with BPD Partners and Men without BPD Partners, Controlling for General Vocabulary and Emotion Word Frequencies (N = 64)

EV	Mean (*SD*) BPD (*N* = 30)	 Non-BPD (*N* = 34)	*F*	*p*	*np*²	95% *CI*
Positive EV	0.23 (0.18)	0.30 (0.22)	2.27	.137	.04	-.17 – .02
Negative EV	0.39 (0.17)	0.33 (0.19)	1.80	.185	.03	-.03 – .15

Note. CI = confidence interval.

Fig. B.2

Mean Emotion Vocabularies (EVs), Across All Conditions, of Men with BPD Partners Versus Men without BPD Partners (*N* = 64)



Note. Error bars represent standard deviations.

B.4.3 Group-by-Condition Interaction Effects on EVs in Male Partners

Figure B.3 presents a comparison of mean EVs between men with BPD partners and control men in each of the three conditions. 2 (group: BPD vs. non-BPD) x 3 (condition: neutral vs. personally-threatening vs. relationship-threatening) mixed ANCOVAs revealed a significant interaction effect for positive EV (*p* = .012) and a by-trend interaction effect for negative EV (*p* = .077) in male partners (see Table B.5 for full interaction results for both EVs across the three conditions).

Table B.5

Group by Condition Interaction Effects on Emotion Vocabularies (EVs) in Male Partners, Controlling for General Vocabulary and Emotion Word Frequencies (N = 64)

		Mean (*SD*)					
EV	Condition	BPD (*N* = 30)	Non-BPD (*N* = 34)	*F*	*p*	*np*²	95% CI
Positive EV	Film	0.32 (0.24)	0.48 (0.38)	3.44	.069	.05	-.32 – .01
	Fear	0.14 (0.17)	0.13 (0.22)	0.00	.949	.00	-.10 – .11
	Separation	0.24 (0.35)	0.10 (0.18)	3.02	.087	.05	-.02 – .26
	Overall interaction			4.62	.012	.07	
Negative EV	Film	0.26 (0.22)	0.23 (0.23)	0.42	.518	.01	-.07 – .14
	Fear	0.43 (0.30)	0.59 (0.52)	2.48	.120	.04	-.39 – .05
	Separation	0.21 (0.30)	0.14 (0.24)	1.21	.276	.02	-.06 – .21
	Overall interaction			2.74	.077	.04	

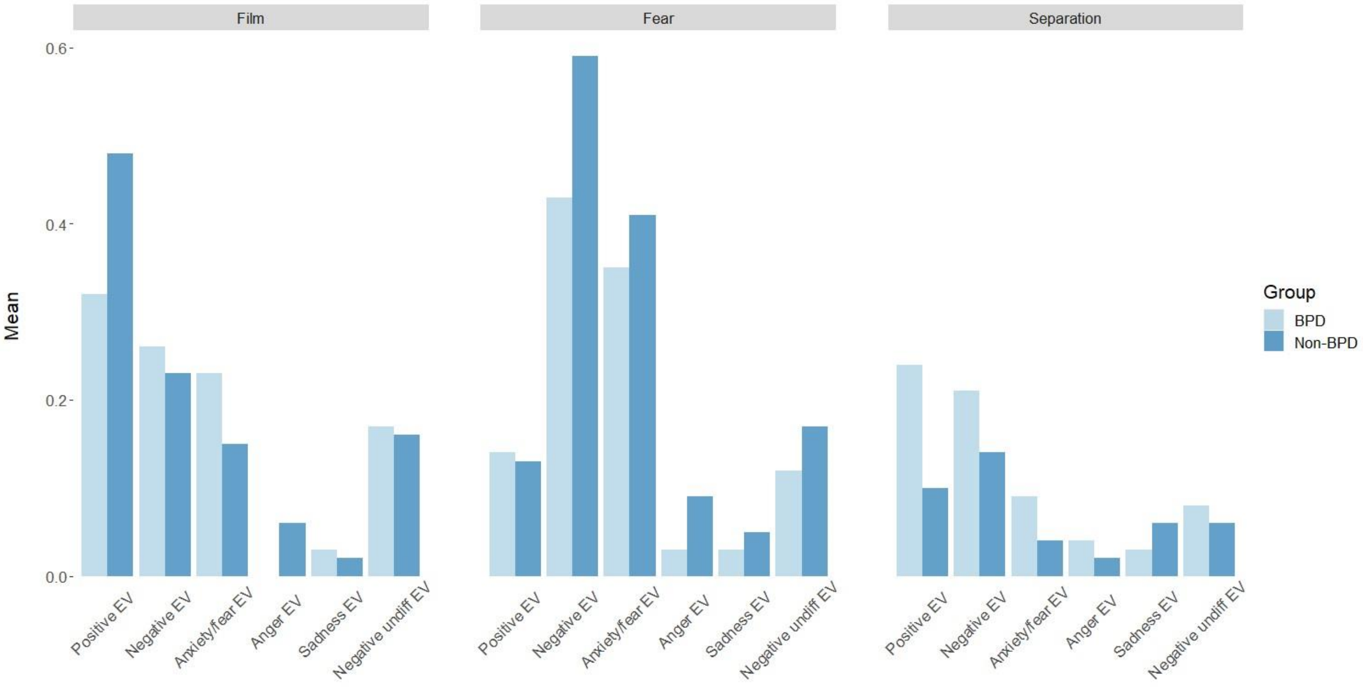
Note. "Group" refers to the between-participants factor comparing EVs between men with BPD partners versus control men. "Condition" refers to the within-participants factor comparing EVs across the three conditions (neutral, personally-threatening, relationship-threatening). Results presented show the overall group by condition interaction effects on the EVs (i.e., the "overall interaction" rows) as well as differences in EVs between men with BPD partners and control men in each of the conditions. CI = confidence interval.

With regard to the interaction effect on positive EV, post-hoc analyses revealed differences in the conditions between men with BPD partners and control men by trend only (all *p*'s > .05). Relative to control men, men with BPD partners had smaller positive EVs in the neutral film condition (*p* = .069), but larger positive EVs in the relationship-threatening condition (*p* = .087), with no difference between groups in the personally-threatening condition (*p* = .949). Further, control men had significantly larger positive EVs in the neutral condition than both the personally-threatening (*M* difference = 0.34, *SE* = 0.07, *p* < .001) and relationship-threatening conditions (*M* difference = 0.36, *SE* = 0.06, *p* < .001), with no significant difference in positive EV between the two threatening conditions (*p* = .661). Similarly, men with BPD partners also had significantly larger positive EVs in the neutral condition compared to the personally-threatening condition (*M* difference = 0.19, *SE* = 0.07, *p* = .010), but not compared to the relationship-threatening condition (*p* = .189). There was also no significant difference in positive EV between the two threatening conditions in men with BPD partners (*p* = .100).

As for the interaction effect on negative EV, there were no significant differences in negative EV between men with BPD partners and control men in any of the conditions (all *p*'s > .10). Yet, control men had significantly larger negative EVs in the personally-threatening condition than the neutral (*M* difference = 0.37, *SE* = 0.08, *p* < .001) and relationship-threatening conditions (*M* difference = 0.46, *SE* = 0.09, *p* < .001), with no difference in positive EV between the neutral and relationship-threatening conditions (*p* = .133). Likewise, men with BPD partners also had larger negative EVs in the personally-threatening condition compared to the relationship-threatening condition (*M* difference = 0.21, *SE* = 0.09, *p* = .021) and compared to the neutral condition by trend (*M* difference = 0.16, *SE* = 0.09, *p* = .074); there was again no difference in negative EV between the neutral and relationship-threatening conditions (*p* = .407).

Fig. B.3

Mean Emotion Vocabularies (EVs) of Men with BPD Partners Versus Men without BPD Partners Split by Condition (*N* = 64)



B.5 Post-Hoc Dyadic Analyses (Study 2, of Paper 2)

B.5.1 Dyadic Data Analysis

To explore the dyadic nature of active EVs, we first examined how dyadic patterns in active EVs (i.e., dyadic discrepancies within the couple) may differ between BPD and non-BPD groups. Specifically, we compared dyadic discrepancies in EVs (female partner’s EVs – male partner’s EVs), as well as dyadic mean EV scores, between BPD and non-BPD groups through independent *t*-tests (two-tailed); a well-established approach for analysing dyadic interactions that takes the dyadic interdependence of the data into account (see Iida et al., 2018; Kenny et al., 2020).

Further, we explored the dyadic patterns in active EVs in more detail by comparing these patterns between BPD and non-BPD couples across the three conditions (i.e., neutral vs. personally-threatening vs. relationship-threatening), to reveal exactly where the main differences lie. Group-by-condition interaction effects on dyadic patterns of EVs were examined via 2 (group: BPD vs. non-BPD) x 3 (condition: neutral vs. personally-threatening vs. relationship-threatening) mixed ANCOVAs, with the EVs added as the dependent variables. General vocabulary size, corresponding emotion word frequencies, and age were also controlled for. Such ANCOVAs were conducted separately for dyadic discrepancies in EVs (i.e., female partner's EVs – male partner's EVs) and average EVs of couples.

Finally, to examine potential dyad-by-group interaction effects on EVs, 2 (partner: female vs. male) x 2 (group: BPD vs. non-BPD) mixed ANCOVAs were conducted, comparing differences in EVs within couples (between female and male partners) and between groups, while controlling for general vocabulary, corresponding emotion word frequencies, and age. Specifically, "Partner" represented the repeated measures variable and "Group" the between-participants fixed factor.

B.5.2 Dyadic Patterns of Emotion Vocabularies Between BPD and Non-BPD Groups

When examining dyadic patterns in EVs, *t*-tests revealed no significant differences in dyadic discrepancies in EVs (i.e., female partner's EVs – male partner's EVs) between BPD and non-BPD groups for positive or negative EV.

With regard to couple-means of EVs and potential differences between BPD and non-BPD groups, *t*-tests replicated the results found when looking at the women only. Namely, couples in the BPD group had significantly larger negative EV (*t*(62) = -2.92, *p* = .005, *d* = -0.73) than couples in the non-BPD group, which was driven by larger anxiety/fear EV (*t*(62) = -4.05, *p* < .001, *d* = -1.02).

B.5.3 Group-by-Condition Interaction Effects on Dyadic Patterns of Emotion Vocabularies

B.5.3.1 Dyadic Discrepancies in EVs

With regard to dyadic discrepancies, a significant group-by-condition interaction effect emerged for positive EV only (*F*(2, 101) = 5.92, *p* = .005, *np*² = .10). Post hoc analyses revealed a greater difference between female and male partners in positive EV in BPD couples (*M* difference = 0.24, *SD* = 0.49) compared to non-BPD couples (*M* difference = -0.09, *SD* = 0.45) in the neutral film condition only (*M* difference = 0.33, *SE* = 0.12, *p* = .007). To illustrate such differences in more detail, in the neutral film condition, women with BPD (*M* = 0.55, *SD* = 0.49) had significantly larger positive EVs than their male partners (*M* = 0.32, *SD* = 0.24; *t*(29) = 2.63, *p* = .013, *d* = 0.48), whereas there was no significant difference in positive EV in this condition between partners in non-BPD couples (*p* = .266). There were no significant differences in dyadic discrepancies in positive EV between BPD and non-BPD couples in the personally-threatening (*p* = .830) or relationship-threatening conditions (*p* = .246).

B.5.3.2 Couple-Level EV Averages

Group-by-condition interaction effects on EVs at the couple level largely mirrored those found when looking at women only. Specifically, significant interaction effects emerged for negative EV, but not positive EV (as when looking at the women only).

In terms of the interaction effect for negative EV (*F*(2, 95) = 6.42, *p* = .004, *np*² = .10), couples with BPD (*M* = 0.30, *SD* = 0.36) had significantly larger negative EVs than non-BPD couples (*M* = 0.13, *SD* = 0.17) in the relationship-threatening condition (*M* difference = 0.19, *SE* = 0.07, *p* = .009). Couples with BPD (*M* = 0.31, *SD* = 0.21) also had larger negative EVs than non-BPD couples (*M* = 0.23, *SD* = 0.22) by trend in the neutral film condition (*M* difference = 0.09, *SE* = 0.05, *p* = .074). However, couples with BPD (*M* = 0.48, *SD* = 0.29) had smaller negative EVs than non-BPD couples (*M* = 0.63, *SD* = 0.53) by trend in the personally-threatening condition (*M* difference = -0.20, *SE* = 0.11, *p* = .076). Moreover, in non-BPD couples, they had significantly larger negative EVs in the personally-threatening condition compared to both the neutral (*M* difference = 0.43, *SE* = 0.08, *p* < .001) and relationship-threatening conditions (*M* difference = 0.53, *SE* = 0.09, *p* < .001), as well as larger negative EVs by trend in the neutral condition than the relationship-threatening condition (*M* difference = 0.10, *SE* = 0.06, *p* = .088). In contrast, couples with BPD only had larger negative EVs in the personally-threatening condition than the neutral condition by trend (*M* difference =

0.14, *SE* = 0.08, *p* = .095); there were no other differences between conditions in couples with BPD (all *p*'s > .10).

B.5.4 Dyad-by-Group Interaction Effects on Emotion Vocabularies

Table B.6 shows the full dyad-by-group interaction effects on the EVs (also see Figure B.4 for a comparison of mean EVs between partners split by group). Regarding the main effects, overall, there were no significant differences between female and male partners in any of the EVs. The main effects of group (BPD vs. non-BPD) on negative EV (driven by anxiety/fear EV) remained significant at the dyad level (as when comparing the women only), in which dyadic interaction effects were accounted for. Specifically, BPD couples (*M* = 0.44, *SD* = 0.15) had significantly larger negative EVs than non-BPD couples (*M* = 0.33, *SD* = 0.14; *F*(1, 56) = 6.08, *p* = .017, np^2 = .10). This was again primarily driven by anxiety/fear EV, as BPD couples (*M* = 0.35, *SD* = 0.13) had significantly larger anxiety/fear EVs than non-BPD couples (*M* = 0.23, *SD* = 0.10; *F*(1, 56) = 15.00, *p* < .001, np^2 = .21).

Table B.6

Dyad-by-Group Interaction Effects on Emotion Vocabularies (EVs), Controlling for General Vocabulary, Emotion Word Frequencies, and Age (N = 128)

EV	Partner	Mean (*SD*) BPD (*N* = 30)	 Non-BPD (*N* = 34)	*F*	*p*	np^2	95% CI
Positive EV	Female	0.28 (0.19)	0.26 (0.19)	0.54	.464	.01	-.06 – .13
	Male	0.23 (0.18)	0.30 (0.22)	3.15	.081	.05	-.20 – .01
	Overall interaction			4.96	.030	.08	
Negative EV	Female	0.49 (0.25)	0.33 (0.15)	6.26	.015	.10	.02 – .22
	Male	0.39 (0.17)	0.33 (0.19)	1.58	.215	.03	-.03 – .15
	Overall interaction			1.27	.264	.02	

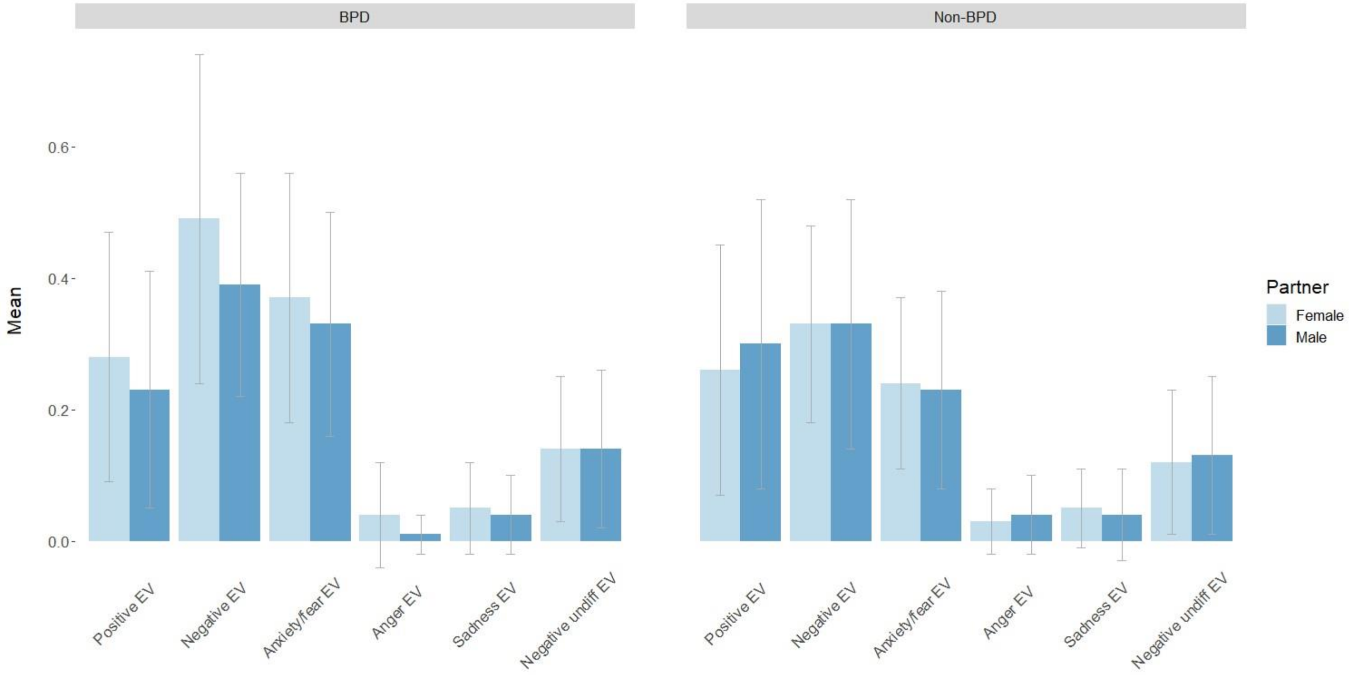
Note. “Group” refers to the between-participants factor comparing EVs between BPD and non-BPD couples. “Partner” refers to the within-participants factor comparing EVs between female and male partners in each couple. Results presented show the overall partner by group interaction effects on the EVs (i.e., the “overall interaction” rows) as well as differences in EVs between BPD and non-BPD couples for each partner in the couple. CI = confidence interval.

Significant dyad-by-group interaction effects emerged for positive EV only, which was investigated further through post-hoc analyses. In particular, men with BPD partners were found to have smaller positive EVs than control men by trend, whereas there were no differences in positive EVs between women with BPD and women without BPD. Further, in BPD couples, women with BPD had larger positive EVs than their male partners by trend (*M* difference = 0.08, *SE* = 0.04, *p* = .064), whereas there were no differences in positive EVs between female and male partners in non-BPD couples (*p* = .175).

Although there was no significant overall dyad-by-group interaction effect on negative EV, pairwise comparisons revealed some noteworthy distinctions. That is, women with BPD had significantly larger negative EVs than non-BPD women, but there was no significant difference in negative EV between men with BPD partners and control men. Further, in the BPD group, women with BPD had larger negative EVs than their partners (*M* difference = 0.09, *SE* = 0.04, *p* = .050), but there were no differences in negative EV between non-BPD women and their partners (*p* = .648).

Fig. B.4

Mean Emotion Vocabularies (EVs) of Female Versus Male Partners Split by Group (*N* = 128)



Note. Error bars represent standard deviations.

B.6 Regression Results Table for Post-Hoc Exploratory Analyses with Depression (Study 2, of Paper 2)

Table B.7

Regression Coefficients for Emotion Vocabularies (EVs) and Covariates Predicting Depression Scores in Female Partners in Study 2 (N = 64)

EV Model	Predictors	β	*t*	*p*
Positive EV	General vocabulary	0.07	0.54	.212
	LIWC posemo	-0.31	-2.43	.018
	Positive EV	0.12	0.93	.358
Negative EV	General vocabulary	0.06	0.43	.667
	LIWC negemo	0.03	0.23	.822
	Negative EV	0.33	2.57	.013
Anxiety/fear EV	General vocabulary	0.03	0.20	.839
	LIWC negemo	-0.01	-0.09	.929
	Anxiety/fear EV	0.33	2.44	.018
Anger EV	General vocabulary	0.16	1.21	.233
	LIWC negemo	0.06	0.42	.675
	Anger EV	0.13	1.00	.322
Sadness EV	General vocabulary	0.18	1.35	.182
	LIWC negemo	0.04	0.34	.738
	Sadness EV	0.02	0.15	.883
Undifferentiated negative EV	General vocabulary	0.11	0.79	.432
	LIWC negemo	0.04	0.30	.763
	Undifferentiated negative EV	0.24	1.85	.069

Note. The "EV Model" column shows the EV (along with the covariates) included in the particular regression model; six regression models were conducted in total (i.e., one model per EV). Regression analyses for the negative EV subcategories were conducted as follow-up specificity tests following the predictive effects of overall negative EV. The predictor "LIWC posemo" refers to the total number of positive affect words and "LIWC negemo" refers to the total number of negative affect words, calculated from LIWC2015.

Appendix C:

Supplemental Material for "Suicidality and Deliberate Self-Harm in Borderline Personality Disorder: A Digital Linguistic Perspective" (Paper 3; Chapter 5)

C.1 Borderline Personality Disorder Subreddit Inter-Rater Coding Agreement

Table C.1

Borderline Personality Disorder Subreddit Coding Agreement Percentages

Coded Variable	Agreement Percentage
BPD classification	94.77
Demographics	95.67
Age	97.96
Gender	92.54
Ethnicity	98.82
Country of residence	96.24
Relationship status	92.60
Religion	98.15
Behavioural categories	93.91
Suicidality	91.52
Deliberate self-harm	94.14

C.2 BPD Reddit Sample Posting Behaviour – Descriptive Illustrations

Figure C.1

Frequency of Users' (N = 992) Posts to BPD Subreddits Over Time

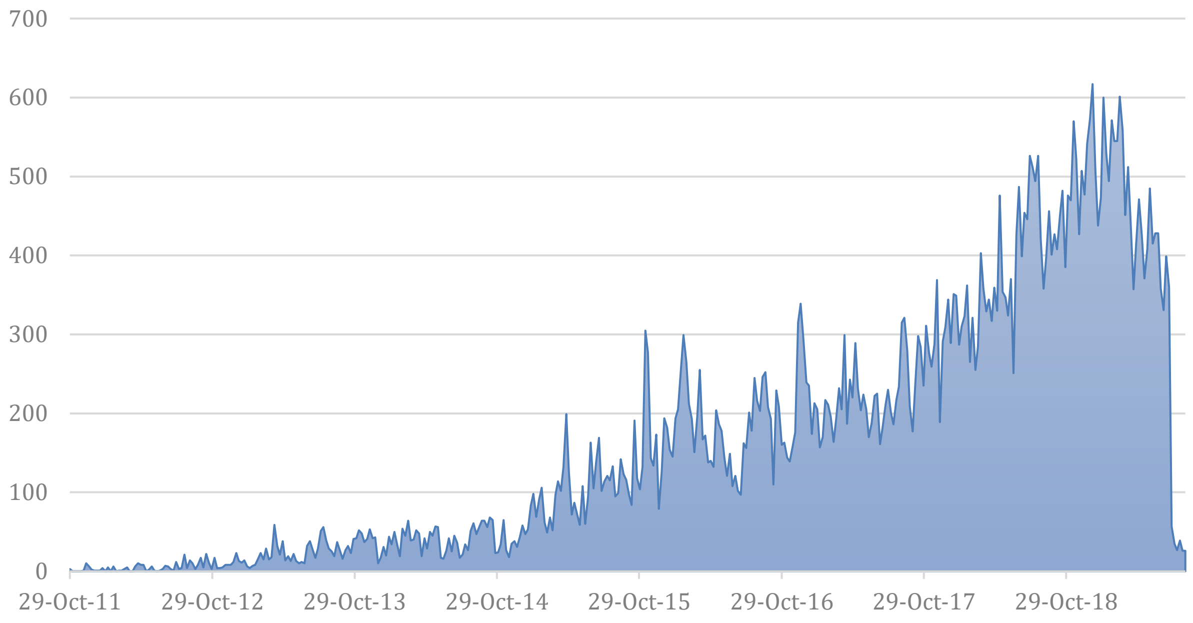
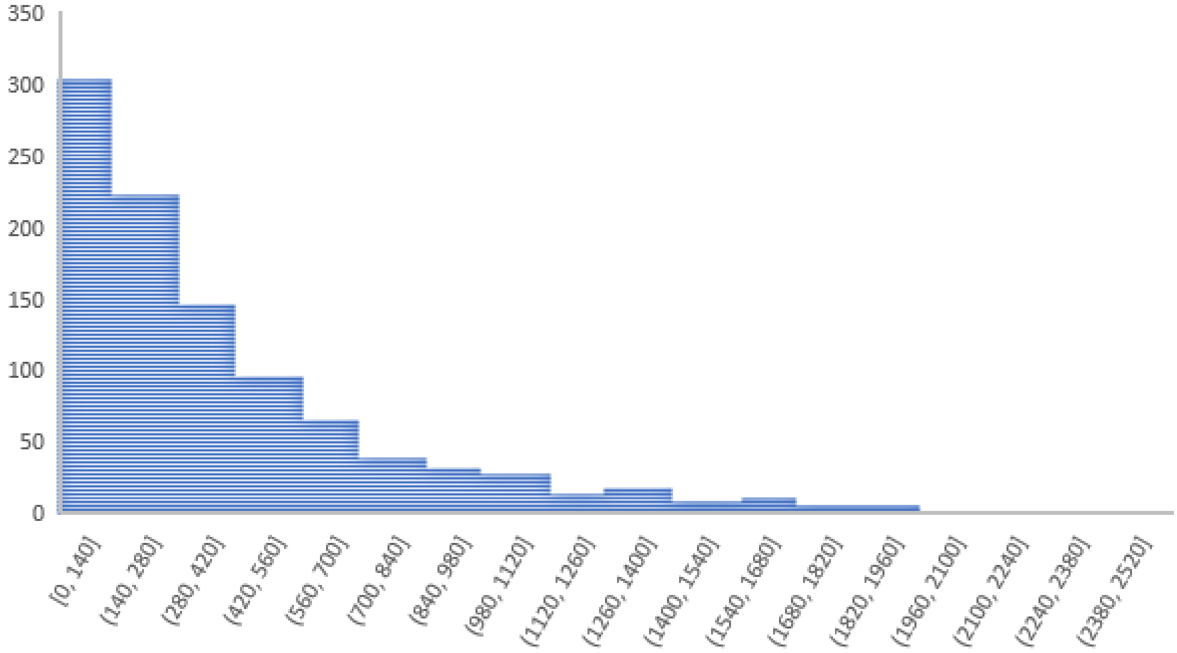


Figure C.2

Histogram of Frequency of Days Between Users' (N = 992) First and Last Posts to the BPD Subreddits



C.3 BPD Reddit Sample Behavioural Coding Frequencies

Table C.2

Frequencies of Suicidality and Deliberate Self-Harm (DSH) Events Manually Coded (N = 992 Users)

Behaviour/Event	*N*	% Of Behavioural Category	% Of Coded Submissions (*n* = 9,106)	% Of Total Submissions (*n* = 66,786)
Suicidality	1,290		14.17	1.93
Past suicide attempt	225	17.44	2.47	0.34
Past suicidal ideation	465	36.05	5.11	0.70
Recent suicide attempt	23	1.78	0.25	0.03
Recent suicidal ideation	577	44.73	6.34	0.86
Deliberate self-harm	678		7.45	1.02
Past DSH	504	74.34	5.53	0.75
Recent DSH	148	21.83	1.63	0.22
Urge for DSH	26	3.83	0.29	0.04

Note. The "*N*" column reflects the total number of suicidality and DSH events coded.

C.4 RQ1 Variations of Correlation Analyses – Results Tables

Table C.3

Spearman's Rho Correlations Between Mean Language Variable Scores and Suicidality (Suicide Attempts and Ideation) and Deliberate Self-Harm (DSH) Frequencies, with Outliers Removed (N = 992)

LIWC Variable	Past Suicidality	Recent Suicidality	Past DSH	Recent DSH
I	.02	.16***	.06†	.11***
Negations	.03	.07*	.04	.02
Positive emotion	.04	.05	.04	-.04
Negative emotion	.06†	.14***	.05	.12***
Anxiety	.07*	.05	.07*	.05
Sadness	.08*	.12***	.08*	.09**
Anger	.03	.13***	.04	.09**
Swear	.04	.10**	.03	.03
We	.02	-.03	.01	-.02
You	-.02	-.06†	-.02	-.06†
Shehe	-.02	-.03	.00	.03
They	.02	.03	.00	.02
Affiliation	-.01	-.04	-.04	.04
Social references	-.03	-.06†	-.06†	-.06†
Cognitive processes	-.08**	-.01	-.05	-.10**
Absolutism	.04	.11***	.00	.01

****p* < .001, ***p* < .01, **p* < .05, †*p* < .10.

Note. All tests are two-tailed. Language variable scores reflect users' mean LIWC22 category scores from the BPD subreddits, excluding posts coded for DSH or suicidality of any nature. Mean language scores were correlated with users' overall frequency of suicidality/DSH, after removing outliers. One outlier was removed for each of the following measures: past suicidality, recent suicidality, and past DSH.

Table C.4

Spearman's Rho Correlations Between Mean Language Variable Scores and Suicidality (Suicide Attempts and Ideation) and Deliberate Self-Harm (DSH) Frequencies, Controlling for Users' Total Number of Posts (N = 992)

LIWC Variable	Past Suicidality	Recent Suicidality	Past DSH	Recent DSH
I	.12***	.18***	.12***	.16***
Negations	.01	.11***	.02	.02
Positive emotion	-.00	.00	-.00	-.03
Negative emotion	.04	.10**	.03	.09**
Anxiety	.03	.00	.02	.03
Sadness	.05†	.07*	.02	.08*
Anger	.02	.10**	.04	.06†
Swear	.03	.13***	.04	.06†
We	-.05	-.05	-.04	-.04
You	-.12***	-.10**	-.08**	-.09**
Shehe	-.03	-.05	-.02	.02
They	-.02	.01	-.02	-.02
Affiliation	-.02	-.03	-.02	.04
Social references	-.11***	-.09**	-.08**	-.07*
Cognitive processes	-.06†	-.03	-.04	-.11***
Absolutism	.02	.12***	.02	-.01

***p < .001, **p < .01, *p < .05, †p < .10.

Note. All tests are two-tailed. Language variable scores reflect users' mean LIWC22 category scores from the BPD subreddits, excluding posts coded for DSH or suicidality of any nature (past or recent). Mean language scores were correlated with users' overall frequency of suicidality/DSH disclosures, while controlling for users' overall number of posts.

C.5 RQ1 Subsetted Dataset – Descriptive Analyses

As part of RQ1, we conducted descriptive statistical analyses on the subsetted dataset in which we examined changes in the number of posts users made (i.e., posting frequency) to the BPD subreddits in proximity to suicidality and DSH events via GLMMs (as described in the main manuscript), with number of posts aggregated weekly entered as the DV. These analyses were carried out using the full subsetted dataset prior to further data refinement (i.e., 453 cases of recent suicidality and 126 cases of recent DSH), given that cases comprising weeks/time points with 0 posts are not classified as missing data in this analysis (i.e., there is no missing data). See Table C.5 for the estimated means and standard errors for number of posts made per week surrounding suicidality and DSH events.

Table C.5

Estimated Means and Standard Errors (SE) for Number of Posts Per Week in Proximity to Suicidality and Deliberate Self-Harm (DSH) Events

	Suicidality (*N* cases = 453; *N* observations = 2,718)		DSH (*N* cases = 126; *N* observations = 756)	
Time point	Mean	*SE*	Mean	*SE*
3 weeks before	2.10	0.18	3.35	0.42
2 weeks before	2.39	0.19	2.36	0.35
1 week before	3.68	0.23	3.36	0.42
1 week after	3.90	0.24	4.93	0.54
2 weeks after	2.71	0.20	2.49	0.36
3 weeks after	2.18	0.18	2.35	0.35

Note. The means and standard errors presented here have been estimated from the GLMMs, and thus are in accordance with the repeated measures nature of the data (i.e., person-centered) while also controlling for random user effects. *N* cases reflects the total number of distinct suicidality/DSH events in the analysis; *N* observations reflects the total number of observations included in the analysis (i.e., the total number of weeks surrounding suicidality/DSH that contain data).

The GLMM conducted for suicidality revealed a significant, large fixed effect of time point (in proximity to suicidality) on posting frequency (*F*(5, 2712) = 31.03, *p* <.001). This effect resulted from a significant increase in the number of posts made in the week immediately preceding the suicidality event compared to 3 and 2 weeks before (*M* increase from 2-weeks pre-event = 1.29, *SE* = 0.21, *t* = 6.24, *p* <.001). Posting frequency remained heightened 1 week after the suicidality event, but significantly decreased by 2-weeks post-event (*M* decrease = -1.19, *SE* = 0.21, *t* = -5.79, *p* <.001); decreasing further (and returning to baseline levels) by 3-weeks post-event (*M* decrease = -0.53, *SE* = 0.20, *t* = -2.70, *p* = .007).

There was also a significant overall fixed effect of time point in proximity to DSH on posting frequency (*F*(5, 750) = 14.34, *p* <.001). Regarding specific changes, there was a drop in posting frequency 2 weeks preceding the DSH event compared to 3 weeks before (*M* decrease = -1.00, *SE* = 0.40, *t* = -2.51, *p* = .012), which returned to baseline levels (i.e., 3 weeks pre-event) by the week immediately preceding the event. Further, there was a significant increase in posting frequency in the week immediately following the DSH event compared to each of the 3 weeks preceding the event (*M* increase from 1 week before = 1.57, *SE* = 0.41, *t* = 3.80, *p* <.001), which sharply decreased again by 2-weeks post-event (*M* decrease = -2.44, *SE* = 0.45, *t* = -5.43, *p* < .001) and remained at a similar level 3-weeks post-event.

In addition to examining changes in posting frequency, we also investigated changes in the word count (i.e., length) of posts made to the BPD subreddits in proximity to suicidality and DSH events. For this analysis, average post word count (aggregated weekly) was entered as the DV. We used the refined version of the subsetted dataset for this analysis (i.e., 159 cases for suicidality and 43 cases for DSH), as used with all other linguistic variables, to ensure that all cases have sufficient, high-quality data. See Table C.6 for the estimated means and standard errors for post word count for each week surrounding suicidality and DSH events.

Table C.6

Estimated Means and Standard Errors (SE) for Post Word Count (Aggregated Weekly) in Proximity to Suicidality and Deliberate Self-Harm (DSH) Events

	Suicidality (*N* cases = 159; *N* observations = 827)		DSH (*N* = cases 43; *N* observations = 227)	
Time point	Mean	*SE*	Mean	*SE*
3 weeks before	108.56	5.94	162.27	15.03
2 weeks before	100.60	5.57	60.74	7.84
1 week before	110.34	5.66	92.22	9.61
1 week after	91.34	5.36	81.74	9.17
2 weeks after	97.42	5.52	65.00	8.13
3 weeks after	100.73	5.81	58.33	7.77

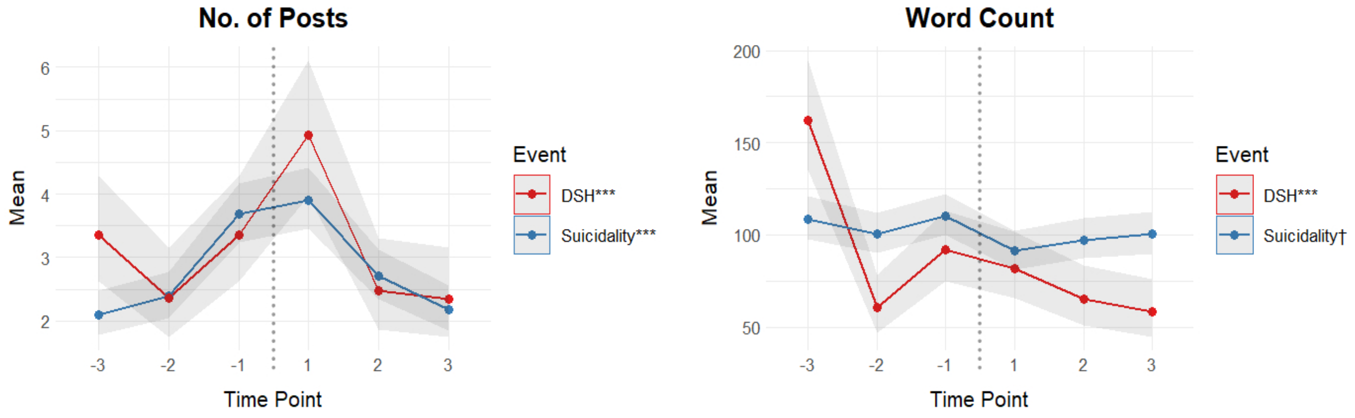
Note. The means and standard errors presented here have been estimated from the GLMMs, and thus are in accordance with the repeated measures nature of the data (i.e., person-centered) while also controlling for random user effects. *N* cases reflects the total number of distinct suicidality/DSH events in the analysis; *N* observations reflects the total number of observations included in the analysis (i.e., the total number of weeks surrounding suicidality/DSH that contain data).

GLMMs revealed no statistically significant overall fixed effect of time point in proximity to suicidality on post length ($F(5, 821) = 2.19$, $p = .054$). However, a significant fixed effect of time point in proximity to DSH on post length was evidenced ($F(5, 221) = 39.75$, $p <.001$). This effect largely stemmed from considerably shorter post lengths in all weeks surrounding the DSH event when compared to 3-weeks pre-event (e.g., *M* decrease from 3- to 2-weeks pre-event = -101.53, *SE* = 12.59, $t = -8.07$, $p <.001$). Yet, post length was found to significantly increase from 2 weeks to 1 week before the DSH event (*M* increase = 31.48, *SE* = 9.07, $t = 3.47$, $p = .001$), which stayed around the same level in the week immediately following the event. Post length significantly decreased again 3 weeks after the DSH event (compared to 1-week post-event; *M* decrease = -23.42, *SE* = 9.18, $t = -2.55$, $p = .011$).

See Figure C.3 for a visual display of weekly changes in posting frequency and post length (i.e., word count) in proximity to suicidality and DSH events.

Figure C.3

GLMM Descriptive Plots: Changes in Posting Frequency and Post Length in Proximity to Recent Suicidality and Deliberate Self-Harm (DSH) Events



Note. The figure shows changes in the mean number of posts made to BPD subreddits (i.e., posting frequency) and the mean word count of posts (i.e., post length) per week (i.e., aggregated weekly) surrounding suicidality and DSH events. The dotted lines illustrate the point at which engagement in the event occurred (i.e., time point 0), thus dividing the figures by pre- and post-event. The shaded areas surrounding the means represent the error margins (95% confidence intervals). The means (and confidence intervals) have been estimated from the generalised linear mixed models (GLMMs), and thus are reflective of the repeated measures nature of the data (i.e., person-centered) while also controlling for random user effects. The indicators assigned to the suicidality and DSH keys show the statistical significance of the overall fixed effects of time in proximity to the events: ****p* < .001, ***p* < .01, **p* < .05, $^{\dagger}p$ < .10. Time point labels: -3 = three weeks before event, -2 = two weeks before event, -1 = one week before event, 1 = one week after event, 2 = two weeks after event, 3 = three weeks after event.

C.6 Main GLMM Results Tables for RQ1

Table C.7

GLMM Descriptive Statistics and Fixed Effects of Time in Proximity to Suicidality (Aggregated Weekly) on Language Variables (N Cases = 159; N Observations = 827)

	Time Point (Estimated Means (*SE*))							
LIWC Variable	-3	-2	-1	1	2	3	*F*	*p*
I	9.50	9.13	9.50	9.60	9.25	9.70	0.65	.662
	(0.32)	(0.30)	(0.30)	(0.30)	(0.30)	(0.30)		
Negations	2.52	2.44	2.52	2.51	2.59	2.62	0.34	.891
	(0.12)	(0.12)	(0.11)	(0.11)	(0.11)	(0.12)		
Positive emotion	0.93	1.06	0.94	1.13	0.92	0.95	1.39	.224
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)		
Negative emotion	1.69	1.63	1.92	1.75	1.64	1.77	1.19	.315
	(0.12)	(0.11)	(0.11)	(0.11)	(0.11)	(0.12)		
Anxiety	0.19	0.51	0.41	0.31	0.33	0.48	7.85	<.001
	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)		
Sadness	0.29	0.31	0.50	0.30	0.33	0.28	6.04	<.001
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)		
Anger	0.38	0.30	0.41	0.35	0.37	0.37	0.57	.720
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)		
Swear	0.34	0.41	0.51	0.28	0.55	0.36	7.60	<.001
	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)		
We	0.58	0.42	0.48	0.51	0.61	0.39	2.09	.065
	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)		
You	2.21	2.28	2.15	2.44	2.28	2.56	0.88	.492
	(0.18)	(0.18)	(0.17)	(0.18)	(0.18)	(0.19)		
Shehe	1.41	1.49	1.05	1.10	1.08	1.01	3.27	006
	(0.14)	(0.13)	(0.12)	(0.12)	(0.12)	(0.13)		
They	0.85	0.94	0.95	0.82	0.79	0.90	0.57	.725
	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)	(0.10)		
Affiliation	1.95	1.69	1.70	2.08	2.27	1.78	4.20	<.001
	(0.13)	(0.13)	(0.12)	(0.13)	(0.13)	(0.13)		
Social references	7.32	7.82	7.16	7.72	7.34	7.53	0.85	.516
	(0.31)	(0.30)	(0.29)	(0.29)	(0.29)	(0.31)		
Cognitive processes	15.90	16.40	16.25	16.37	16.25	16.14	0.35	.885
	(0.34)	(0.33)	(0.32)	(0.32)	(0.33)	(0.34)		
Absolutism	1.68	1.50	1.83	1.75	1.72	1.79	1.29	.265
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)		

Note. The means and standard errors reported here have been estimated from the generalised linear mixed models (GLMMs), and thus are reflective of the repeated measures nature of the data (i.e., person-centered) while also controlling for random user effects. *N* cases reflect the total number of distinct suicidality events in the analysis; *N* observations reflect the total number of observations included in the analysis (i.e., the total number of weeks surrounding suicidality that contain data). Time point labels: -3 = three weeks before event, -2 = two weeks before event, -1 = one week before event, 1 = one week after event, 2 = two weeks after event, 3 = three weeks after event. *SE* = standard error.

Table C.8

GLMM Descriptive Statistics and Fixed Effects of Time in Proximity to Deliberate Self-Harm (Aggregated Weekly) on Language Variables (N Cases = 43; N Observations = 227)

	Time Point (Estimated Means (*SE*))							
LIWC Variable	-3	-2	-1	1	2	3	*F*	*p*
I	9.65 (0.58)	10.35 (0.53)	9.27 (0.51)	10.44 (9.25)	10.17 (0.53)	9.29 (0.52)	1.51	.189
Negations	2.15 (0.22)	2.60 (0.20)	2.73 (0.20)	2.61 (0.20)	2.60 (0.20)	2.70 (0.21)	0.97	.436
Positive emotion	0.99 (0.14)	0.86 (0.12)	0.78 (0.12)	1.12 (0.13)	1.10 (0.13)	0.97 (0.13)	1.28	.273
Negative emotion	1.78 (0.25)	1.67 (0.22)	1.69 (0.22)	2.06 (0.22)	1.61 (0.22)	1.64 (0.22)	0.65	.661
Anxiety	0.47 (0.09)	0.39 (0.07)	0.55 (0.08)	0.49 (0.07)	0.21 (0.07)	0.20 (0.07)	3.27	.007
Sadness	0.31 (0.09)	0.68 (0.11)	0.13 (0.05)	0.14 (0.06)	0.28 (0.06)	0.22 (0.06)	12.74	<.001
Anger	0.20 (0.07)	0.10 (0.06)	0.09 (0.06)	0.72 (0.13)	0.10 (0.06)	0.19 (0.07)	12.33	<.001
Swear	0.42 (0.08)	0.29 (0.07)	0.51 (0.09)	0.33 (0.07)	0.34 (0.07)	0.30 (0.08)	2.23	.052
We	0.30 (0.07)	0.14 (0.06)	0.24 (0.09)	0.79 (0.12)	0.45 (0.10)	0.31 (0.07)	12.58	<.001
You	2.35 (0.32)	1.91 (0.28)	2.83 (0.32)	2.07 (0.28)	2.09 (0.29)	2.50 (0.30)	1.95	.088
Shehe	0.85 (0.26)	1.51 (0.26)	1.62 (0.26)	0.87 (0.24)	1.80 (0.27)	1.30 (0.25)	2.15	.061
They	0.67 (0.11)	0.54 (0.10)	0.67 (0.10)	0.74 (0.11)	0.50 (0.10)	0.75 (0.11)	1.16	.329
Affiliation	2.65 (0.26)	1.62 (0.22)	1.84 (0.23)	2.33 (0.23)	2.18 (0.24)	2.05 (0.23)	2.60	.026
Social references	7.03 (0.53)	6.72 (0.49)	7.93 (0.51)	7.04 (0.48)	7.50 (0.51)	8.28 (0.51)	1.85	.104
Cognitive processes	16.09 (0.61)	16.08 (0.55)	16.71 (0.55)	16.47 (0.55)	16.75 (0.56)	16.67 (0.55)	0.35	.882
Absolutism	1.65 (0.17)	1.43 (0.15)	1.36 (0.15)	1.44 (0.15)	1.42 (0.15)	1.40 (0.15)	0.47	.802

Note. The means and standard errors reported here have been estimated from the generalised linear mixed models (GLMMs), and thus are reflective of the repeated measures nature of the data (i.e., person-centered) while also controlling for random user effects. *N* cases reflect the total number of distinct DSH events in the analysis; *N* observations reflect the total number of observations included in the analysis (i.e., the total number of weeks surrounding DSH that contain data). Time point labels: -3 = three weeks before event, -2 = two weeks before event, -1 = one week before event, 1 = one week after event, 2 = two weeks after event, 3 = three weeks after event. *SE* = standard error.

Appendices Bibliography

Hautzinger M., Keller F., Kühner C. (2006). *Beck Depressions-Inventar (BDI-II)* [Beck Depression Inventory]. Frankfurt, Germany: Harcourt.

Iida, M., Seidman, G., & Shrout, P. E. (2018). Models of interdependent individuals versus dyadic processes in relationship research. *Journal of Social and Personal Relationships*, *35*(1), 59–88. https://doi.org/10.1177/0265407517725407

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2020). *Dyadic Data Analysis*. The Guilford Press.

Kühner, C., Bürger, C., Keller, F., & Hautzinger, M. (2007). Reliabilität und Validität des revidierten Beck-Depressionsinventars (BDI-II). Befunde aus deutschsprachigen Stichproben [Reliability and validity of the Revised Beck Depression Inventory (BDI-II). Results from German samples]. Der Nervenarzt, 78(6), 651–656. https://doi.org/10.1007/s00115-006-2098-7