The University Of Sheffield.

Paranoia and Mistrust:

Epidemiological, Experimental, and Clinical Investigations

By:

Anton P. Martinez

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Sheffield

Faculty of Science

Department of Psychology

Submission Date: 5th June 2023

ACKNOWLEDGMENTS

I started this PhD on April 2019, and I am submitting this thesis on June 2023. Throughout this time, I had the chance to witness several social and historical events such as four different UK prime ministers, Brexit, a global pandemic, a war in Europe, a cost of living crisis, the death of the British queen, and the coronation of the new British king, just to name a few. Thus, doing a PhD during these tumultuous times was indeed challenging and I could have not done it without the support of amazing people that were there for me.

First, I want to acknowledge my family, my parents Marcela and Antonio, as well as my siblings Agustín, Lara, Elisa, and Dante. Your unconditional love and support have been the driving force that has made everything possible.

In this category, I would also like to acknowledge my dear Naima. Thanks for providing me with a second home in Europe, I am incredibly grateful for your love and friendship.

Second, I want to acknowledge my cohort friends made during this period. I want to particularly thank Amber and Emily, for always being there for me from the very start of the PhD. Also, Haifa, Fidan, and Ben. Thanks for the amazing times we spent together, and for the fun, love, and support that made PhD work way easier. Also, I want to thank those who were there at the very beginning, Eema, Danielle, Kat, Stan, Meghan, Mel, Harriet, Siham, Zhuo, Wenyi, Shuangke , and Bixuan as well as those who were there for me at the very end, Sonia, Anna, Catalina, Dom, and James. I also want to thank Gina, for her support and friendship, particularly during pandemic times.

Third I want to acknowledge my collaborators who provided guidance as well as supervision. I would like to thank Prof. Mark Shevlin, Prof. Philip Hyland, Prof. Carmen Valiente, Dr. Stephen Kellet, Dr. Azzam Alsuhibani, Kubra Ulusoy, and my dear friend Max Agostini. I would also like to thank the members of the COVID-19 Psychological Research

Consortium. Your help and advice have significantly contributed to the research presented in this thesis which in turn has been an excellent learning opportunity for me. I would also like to thank my dear colleague Dr. Almudena Trucharte for being my partner in crime when studying this research topic. I hope our friendship and passion for psychosis research will lead to many projects and great times together.

I would also like to thank Jayne, for always being helpful and supportive with regards to PhD as well as personal matters.

Fourth, I want to thank my second and third supervisors Prof. Elizabeth Milne and Dr. Georgina Rowse for their help, guidance, and supervision throughout this period. Particularly to Prof. Milne for training me in EEG techniques even at the late stages of my PhD. It has been an amazing experience and I have had a lot of fun learning new ways of doing psychological research.

Finally, I want to thank Prof. Richard Bentall. I have been incredibly privileged to have you as a primary supervisor throughout my PhD. Since the first time we met, you have constantly supported me, from always backing up my (at times crazy) ideas to being there at challenging times where obstacles needed to be overcome. You have provided me with career opportunities that are seldom given to PhD students. I cannot quantify the amount of knowledge I have acquired throughout our weekly meetings, feedback, and advice. Your passion for psychological research is contagious and you always managed to keep me motivated every time we discussed our projects. I will be forever thankful for your unconditional generosity and support.

DECLARATION AND NOTE ON INCLUSION OF PUBLISHED WORK

I, Anton P. Martinez, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not previously been presented for an award at this, or any other, university.

This thesis is in a publication format, and contains the following published work:

Chapter I: Parts of the narrative review presented in this chapter were written as contributions in two different chapter within two edited books:

Martinez, A. P., Huddy, V., & Bentall, R.P. (2023). The psychology of paranoid beliefs. In D. Turkington & K. Hardy (Eds.), *Decoding Delusions* (pp.79 -112). American Psychiatric Pub.

Martinez, A.P., & Bentall, R.P. (2023). Paranoia. In Friedman, H. S. & J. Hooley (Eds.). *Encyclopaedia of Mental Health*. Academic Press.

Chapter II: Research presented in this chapter has been published and is available online at *Frontiers in Psychology*:

Martinez, A. P., Shevlin, M., Valiente, C., Hyland, P., & Bentall, R. P. (2022). Paranoid beliefs and conspiracy mentality are associated with different forms of mistrust: A three-nation study. Frontiers in Psychology, 6186.

Chapter III: Research presented in this chapter has been published and is available online at *Psychology and Psychotherapy: Theory, Research, and Practice*:

Martinez, A. P., Agostini, M., Al-Suhibani, A., & Bentall, R. P. (2021). Mistrust and negative self-esteem: Two paths from attachment styles to paranoia. Psychology and Psychotherapy: Theory, Research and Practice, 94(3), 391-406.

Conference Dissemination

Chapter II: Findings from this chapter have been presented as part of a symposium at the European Conference on Schizophrenia Research (ECSR):

Martinez, A. P., and Bentall, R.P. (2021). Paranoid beliefs and conspiracy mentality are associated with different forms of mistrust: A three-nation study. ECSR, Berlin, Germany (Online).

Chapter III: Findings from this chapter have been presented as an open paper at the European Association of Behavioural and Cognitive Therapies (EABCT):

Martinez, A. P. (2021). Mistrust and negative self-esteem: Two paths from attachment styles to paranoia. EABCT, Belfast, Northern Ireland.

Chapter IV: Findings from this chapter have been presented as an open paper at the European Association of Behavioural and Cognitive Therapies (EABCT):

Martinez, A. P., and Trucharte, A. (2022). Effect of a classical conditioning intervention on social processes associated with paranoid beliefs: evidence from two experimental studies. EABCT, Barcelona, Spain.

Note. At the beginning of each chapter, an overview of contributors to the research such as supervisors, co-authors, and collaborators, is provided. As this thesis follows a publication format, there is a certain degree of repetition in methodological details and background literature across the chapters.

ABSTRACT

Paranoia entails the core rigid and unfounded belief that other people are intentionally trying to harm the integrity of the believer. This belief system is better conceptualized as a continuum, with ordinary social evaluative concerns at the end of the spectrum and hypervigilance as well as persecutory feelings at the other end. Several psychological models have been developed throughout the last decades to explore the underlying mechanisms of paranoia. Insecure attachment styles as well as negative self-esteem have been proposed as potential candidates for better explaining paranoid interpretations. However, the role of key interpersonal aspects such as mistrust has been less studied. Given the social nature of paranoid beliefs, this thesis aims to explore the role of different features of mistrust and its relationship with paranoid beliefs in non-clinical and clinical samples by employing epidemiological, experimental, and psychophysiological designs. To begin, the first chapter offers a thorough narrative review of conceptualizations of paranoia and states the objectives of this thesis. Following this, the second chapter (study 1) provides results from a large representative and international sample revealing the specific association of a tendency toward mistrusting unfamiliar faces and paranoia in contrast to related constructs, such as conspiracy mentality. The third chapter (study 2) reveals the role of a bias toward mistrust between insecure attachment styles and paranoia in a large representative UK sample. Subsequently, the fourth chapter (study 3) shows using a convenience sample, that manipulating relational schemas through an evaluative conditioning task performed online did not influence trustworthiness judgments, Lastly, the fifth chapter (study 4) explores the temporal dynamics of untrustworthiness judgements in clinically paranoid and non-clinical samples using event-related-potential techniques revealing differences in face processing between the groups. The final chapter synthesises these findings and reflects on the theoretical, methodological, and clinical implications of this thesis by also discussing general limitations, strengths, and future directions.

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I - General Introduction

Contributions

Anton P. Martinez (conceptualization, writing – original draft)

Vyv Huddy (writing, & reviewing)

Richard P. Bentall (supervision, writing –reviewing & editing– original draft)

Parts of the current chapter were written as contributions in two publications:

Martinez, A. P., Huddy, V., & Bentall, R.P. (2023). The psychology of paranoid beliefs. In D. Turkington & K. Hardy (Eds.), *Decoding Delusions* (pp.79 -112). American Psychiatric Pub.

Martinez, A.P., & Bentall, R.P. (2023). Paranoia. In Friedman, H. S. & J. Hooley (Eds.). *Encyclopaedia of Mental Health*. Academic Press.

Summary

Paranoid (sometimes called persecutory) delusions are one of the most common symptoms of severe mental illness (Bentall et al., 2001; Bebbington & Freeman, 2017), present in more than ninety percent of first-episode schizophrenia patients (Moutoussis et al., 2007). However, because of the importance of suspiciousness and mistrust in ordinary social relationships, the study of these kinds of beliefs is not limited to clinical research but also expands to social and behavioural disciplines. This chapter will explore conceptualizations of paranoia from the initial years of psychiatry through to the theoretical advances resulting from psychological research during the last decades. An up-to-date review of theories about the origins and mechanisms of paranoid beliefs will be provided by summarising the available evidence from clinical, experimental and epidemiological studies. Finally, the aims of each of the remaining chapters that encompass this thesis will be outlined.

1.1. Defining Paranoia

1.1.1. *A journey throughout the history of psychiatry*

Paranoia is a word often used in ordinary life to refer to persons who feel excessively suspicious, distrustful or persecuted for no apparent reason. The term is derived from the Greek words *para* ('beyond', 'beside') and *nous* ('mind', 'intellect'), creating the word *paranoia*, which was used in Greek literature to describe people who were 'out of their minds'. However, it was not until the eighteenth century that it was transformed into a formal clinical concept (Dowbiggin, 2000). French and German psychiatrists, for example, J. E. D. Esquirol (1772-1840), J. C. A. Heinroth (1773-1843) and Karl Kahlbaum (1828 -1899), used the word to describe a limited form of insanity in which abnormal beliefs affected only circumscribed areas of functioning, leaving intact other domains of reasoning and judgment (Dowbiggin, 2000; Lewis, 1970).

At the beginning of the 20th century, Emil Kraepelin (1856-1927; considered one of the most influential physicians of modern psychiatry) defined the term *dementia praecox* as a chronic mental disease that was marked by a deterioration of the psychic functions (e.g., volition, affect, intellect), which led to the degradation of personality (Dowbiggin, 2000; Kendler, 988; Lewis, 1970). According to Kraepelin, paranoia was a different entity from *dementia praecox* conceptualising the former as a deviant form of personality development characterised by a stable uniformly connected system of coherent non-bizarre delusions without marked mental deterioration (Dowbiggin, 2000; Kendler, 1988; Lewis, 1970). Kraepelin also made distinctions regarding aetiology and course between *dementia praecox* and paranoia. Whereas the former was considered a 'natural disease' that led to profound mental disturbances, the latter was conceptualised as a disorder with a chronic non-remitting course caused by the interaction of personality development with life experiences (Dowbiggin, 2000; Kendler, 2016; Lewis, 1970).

This view of paranoia as a 'reactive' or psychogenic psychosis bore some similarities to the psychoanalytic approach developed by Sigmund Freud (Freud, 1911/1950). Freud's theory was developed in his analysis of the case of the German high court judge *Daniel Schreber,* who he never met (the analysis was based on Schreber's autobiography; Schreber, 1903/1955), and attributed the judge's highly disorganized and religiously-themed delusions to the unconscious conflicts caused by his repressed homosexual desires. On the other hand, Euguen Bleuler - who had a close relationship with psychoanalytic circles and was the first to coin the term schizophrenia- viewed paranoia as a 'situational psychosis' conceptualising it as a psychic formation that gives rise to the exaggeration of normal processes that cannot be corrected (Dowbiggin, 2000; Kendler, 2016).

1.1.2. *The DSM era: towards a medical classification of mental health*

These psychoanalytic views had an impact on North American psychiatrists who, by the middle of the 20th century, had become highly influential in international psychiatry (Strand, 2011). Nonetheless, a perceived need to reform psychiatric nosology was rising in the United States given the high rate of occupational psychiatrists working in hospitals after the Second World War (Dowbiggin, 2000; Strand, 2011). This led US psychiatrists to create standardized diagnostic manuals by categorising mental disorders for administrative, governmental and educational purposes, which culminated in the first edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-I).* This manual was heavily influenced by a psychoanalytical approach that embraced a very broad formulation of schizophrenia which was seen as having a paranoid subtype (Dowbiggin, 2000).

The poor reliability of psychiatric diagnoses during this period (Spitzer & Fleiss, 1974), and the wish of some psychiatrists to enhance their credibility as medical practitioners led, ultimately, to the neoKraepelinian movement, which aimed to return psychiatry to the principles developed by Kraepelin in the late 19th century (Blashfield, 1984). The

neoKraepelinians conceived psychiatric disorders as biological conditions which, they believed, would ultimately be explained in terms of genetic and neurochemical processes (Guze, 1989). This conception led clinicians to focus on the need to operationalize psychiatric diagnoses as a means to accelerate scientific advances in these areas. Having a manual describing distinct diagnostic categories was fundamental for psychiatrists who wished to conduct randomised control trials (RCTs) to test the effects of pharmacological treatments (Dowbiggin, 2000; Strand, 2011). Perhaps the most important achievement of the neoKraepelinians was the publication of the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 1980) which, for the first time, provided precise operational criteria for each diagnostic condition (Bentall, 2004).

The widespread adoption of DSM-III in 1980 encouraged the view of mental disorders as discrete and explicit diagnostic categories making credible the claim that mental illnesses were medical diseases (Strand, 2011). In this edition of the manual, paranoia was included as a diagnosis with the stipulation that a diagnosis of schizophrenia must be excluded, a criterion that was maintained in all subsequent editions until the present day. The DSM underwent further revisions and the paranoia diagnosis broadened into *delusional disorder (DD)* which included additional delusional themes such as erotomania, grandiose, jealousy, and somatic. In the present DSM-5 (American Psychiatric Association, 2013) the term refers to the presence of non-bizarre delusions, absence of disorganised speech, and minimal hallucinations which, if present, are congruent with the delusional content. Nonetheless, persecutory delusions are also present in schizophrenia spectrum disorder (SSD) and (in contrast to *DD*) they tend to co-occur with other symptoms such as hallucinations and disorganized speech and behaviour. Another differentiation between SSD and DD is that the former typically leads to marked social and occupational dysfunction whereas in the latter, the patient can be high functioning and its demeanour is not considered bizarre.

1.1.3. *Challenging the categorical model: The idea of paranoia as a continuum*

The medical approach to mental disorders assumes that psychopathological syndromes are caused by an underlying common causal process – a medical condition - so that the symptoms of a disorder are explained by the latent disease (Borsboom & Cramer, 2013). However, this model struggles to account for the overlap of symptoms from different disorders and the consequent high rate of comorbidity (patients meeting the criteria for two or more categorical diagnoses) observed in everyday clinical and research practice (Brown & Barlow, 2005; Renard et al., 2017). Paranoid beliefs, for example, can be present in DD, Schizophrenia, Bipolar Disorder, Major Depression, Paranoid Personality Disorder, Autism and Obsessive-compulsive disorders (American Psychiatric Association, 2013; Bentall et al., 2009). Thus, the notion that mental health symptoms are caused by discrete underlying neurobiological abnormalities has gradually been replaced by a more complex picture in which symptoms are regarded as the consequences of interactions between various psychological, social, and biological mechanisms. At the same time, the idea that there is a clear dividing line between psychiatric disorders and healthy functioning has been increasingly questioned.

This new way of thinking has been strengthened by epidemiological studies which have estimated that paranoid beliefs are experienced by at least 10-15% of the general population (Freeman, 2007), ranging in severity from less pathological to more dysfunctional forms (Elahi et al., 2017). The continuum approach is also supported by numerous studies of social and cognitive factors, which have generated similar findings from patients with paranoid delusions and ordinary people who score high on paranoia scales. Freeman (2007), for example, has proposed a hierarchal 'pyramid' of paranoia severity where social evaluative concerns and ideas of reference experienced by many lie at the base and irrational beliefs about severe threats, experienced by very few people, are positioned at the top. Consistent with this picture, large survey studies carried out on the general population have found that the most endorsed

paranoid beliefs concern interpersonal sensitivity and mistrust whereas ideas of reference and persecution are less frequent but very common in clinical populations (Bebbington et al., 2013; Bell & O'Driscoll, 2018).

Another perspective within the continuum approach is regarding paranoid beliefs as traits characteristic of certain types of personalities. Personality traits are believed to reflect a stable set of core beliefs that guide interpersonal behaviour and which, when maladaptive, can lead to disproportionate emotional and behavioural responses in specific situations (Beck et al., 2018). From this perspective, paranoid traits would reflect an exaggeration of normally adaptive personality traits which would lead to impairment in self and interpersonal functioning (Bach & First, 2018). This conceptualization of paranoia is in line with clinical criteria of different types of personality disorders from psychiatric classification manuals. For example, the DSM-V describes paranoid personality disorder (PPD) as characterized by excessive mistrust, unfounded suspiciousness, bearing grudges, and persistent interpretations of others' intentions as malicious, often leading to interpersonal conflicts such as unjustified jealousy, hostility, and litigious behaviours (American Psychiatric Association, 2013). Paranoid beliefs are also present in schizotypal personality disorder (SPD) whereas stress-related paranoid ideation is considered a symptom of borderline personality disorder (BPD) which is usually transient and triggered by interpersonal conflicts (American Psychiatric Association, 2013). Although described in manuals of psychiatric classification, there is robust evidence showing that personality disorders are best represented dimensionally rather than categorically (Ahmed et al., 2012) and some diagnostic manuals such as the ICD-11 have adopted this approach (World Health Organization, 2018).

An important point to consider when thinking about paranoia from a dimensional approach is the differentiation of a delusion from a maladaptive trait. For example, persecutory delusions are considered an extreme manifestation of paranoia where these beliefs are firmly

held and coupled with high degrees of preoccupation and distress, usually experienced during active psychotic episodes. On the other hand, a paranoid trait can be portrayed as a general tendency to perceive others' intentions as negative while adopting a mistrustful disposition towards others. Following this line of thought, it is also important to distinguish a paranoid trait from a paranoid state. Being hypervigilant in a particular threatening environment can be considered a sensible response to a specific situation and some researchers even regard these types of responses as an adaptive evolutionary strategy (Nettle & Haselton, 2006).

1.1.4. *Current conceptualisations of paranoia*

In recent years, psychological researchers have developed a parsimonious conceptualization of paranoia that considers the heterogeneity of its manifestation as well as its characterization as a continuum rather than a category. From this viewpoint, the central aspect of paranoia is the unfounded core belief that an imminent threat against the believer is intentionally orchestrated by others (Bebbington et al., 2013; Bell & O'Driscoll, 2018; Brown, Waite, & Freeman, 2019; Freeman, 2016; Murphy, Bentall, Freeman, O'Rourke, & Hutton, 2018; Trotta, Kang, Stahl, & Yiend, 2021). This type of belief is characterised by high degrees of conviction (i.e., how strongly the belief is held), preoccupation (i.e., fixation on the belief) and distress (i.e., negativity associated with the belief; Combs et al., 2006), but each of these dimensions varies according to the severity of the experience (Elahi et al.,2017; Freeman, 2007). Depending on the severity, paranoia can be manifested in different ways, from social evaluative concerns and suspiciousness, through self-referential processing and persecutory ideation, culminating in delusions (Freeman, 2007, 2016). Thus, the expectation of intended harm reflects different psychological and behavioural processes. For example by being wary of other people's intentions (i.e., mistrust), perceived vulnerability in the presence of others (i.e., interpersonal sensitivity), interpreting innocuous experiences as personally targeted messages (i.e., self-referential beliefs) or performing safety behaviours to prevent significant

harm from others (i.e., anticipation of threat; Bebbington et al., 2013; Bell & O'Driscoll, 2018; Bentall et al., 2009; Freeman, 2007). In short, paranoia is currently conceived as a wide spectrum which varies not only in how persecutory beliefs are experienced (e.g., rigid, distressing) but also in the way these beliefs can be exhibited behaviourally.

An important caveat to this description concerns the extent to which people believe that they deserve to be persecuted. Trower and Chadwick (1995) proposed that paranoia occurs in two forms: 'poor-me' in which the individual believes that they are the innocent victims of the malign intentions of others, and 'bad-me' in which the person believes that they deserve to be persecuted. Bad me paranoia is much less common in psychiatric patients than poor me paranoia (Fornells-Ambrojo & Garety, 2005) and longitudinal studies have shown that patients tend to be either stably poor-me or fluctuate between bad-me and poor-me, with fluctuations occurring over periods of a few hours or days (Melo et al., 2006; Udachina et al., 2012). These results can be interpreted as paranoia having different profiles which may indicate either defensive (poor-me) or powerlessness (bad-me) processes (Trower & Chadwick, 1995) highlighting the need for comprehensive psychological models of paranoid beliefs to understand its aetiology and maintenance.

1.2. Psychological models of paranoia

The shift in focus from construing paranoia as the symptom of an underlying illness to considering it as a phenomenon in its own right, and the development of the continuum approach has led to the development of several psychological models. These models attempt to identify cognitive and emotional factors involved in the development and maintenance of paranoid beliefs to get a better psychological conceptualization when generating clinical interventions.

1.2.1. *Anomalous Perception Model*

Perhaps the simplest psychological model was proposed in the early 1970s by the American psychologist Brendan A. Maher (1974,1999). Maher claimed that all delusions were generated as an explanation of abnormal perceptual experiences, which provided the evidence on which patients make delusional inferences. From this point of view, Maher argued that the reasoning processes involved in delusions were not impaired but rather were the same as those present in non-clinical populations. The idea that aberrant perceptions may play a critical role was partially inspired by psychological models of schizophrenia from the period, which saw deficits in filtering perceptual information as a central feature of the disorder (Hemsley, 1993). Maher held that aberrant perceptual experiences would inevitably lead to a state of uncertainty and anxiety, prompting a search for an explanation as a way to ameliorate these negative feelings (Bell et al., 2006). The content of the resulting delusional explanations was believed to reflect the patient's cultural background and personal experiences (Maher, 1999). For example, Maher argued, based on Freud's analysis of the Schreber case, that the persecutory content of the judge's beliefs was a consequence of the guilt associated with his repressed homosexuality, which served as an explanation of his unusual somatic experiences (Maher, 1974).

Empirical studies of both nonclinical samples (Chapman & Chapman, 1988) and also clinical samples (Bell et al., 2008) have not consistently supported the idea of a strong relationship between anomalous experiences (e.g., hallucinations) and delusions in general of the kind Maher's theory predicted. Though some types of rare delusional systems can be accounted for in this way (notably the Capgras syndrome, in which the individual believes a loved one has been replaced by an impostor, and which seems to be caused by an impairment in the ability to recognise the familiarity of faces; Ellis, Young, Quayle, & De Pauw, 1997). In the case of paranoia, there is strong evidence that late-onset deafness sometimes plays a role.

Early studies showed that progressive hearing loss in elderly people is associated with paranoid beliefs (Cooper & Curry, 1976) and an association between paranoia and hearing difficulties across the lifespan has subsequently been confirmed in several epidemiological studies (Stefanis et al., 2006; Thewissen et al., 2005). One possible explanation for this association is that people with hearing loss develop suspicions about why people are no longer communicating with them.

Several other factors that seem to contribute to paranoia can also be interpreted within an anomalous perception framework. One is the impact of cannabis consumption on paranoid thinking, which has been demonstrated experimentally (Freeman et al., 2015). Another is the observed association between sleep disturbances (e.g., insomnia) and paranoid beliefs, which has been widely reported in clinical and non-clinical populations (Blanchard et al., 2020; Freeman et al., 2020) and supported by experimental studies, in which participants have been deliberately deprived of sleep (Kahn-Greene et al., 2007; Reeve et al., 2018). However, randomized control trials that have been conducted to reduce paranoia by targeting insomnia through cognitive behavioural treatment have led to inconsistent results, with some studies showing an effect (Freeman et al., 2017; Myers et al., 2011) but others showing no impact on paranoid symptoms, despite an improvement in insomnia (Freeman et al., 2015). Finally, the possible role of dissociative experiences should be considered within an anomalous perception framework. These experiences, which are common consequences of severe traumatic events either in adulthood and childhood (Dalenberg et al., 2012), involve a disturbance of the integration of identity (memory, personality, emotion) which can lead to feelings of detachment from the self (depersonalization) or surroundings (derealisation; American Psychiatric Association, 2013). While a robust link between dissociation and auditory hallucinations has been established (Pilton et al., 2015), some researchers have argued that this kind of strange experience could lead to delusional interpretations fuelled by worry and anxiety (Černis et al.,

2021; Freeman et al., 2013), and a recent meta-analysis has provided some support for this hypothesis (Longden et al., 2020).

1.2.2. *Paranoia as a defence model*

By the late 1980s, the idea that anomalous experiences could be the sole cause of delusional thinking was no longer widely accepted. Given the nature of persecutory delusions (malevolent intentions from others), some researchers began to consider the psychological processes by which patients judged their position in the social universe (Kaney & Bentall, 1989, 1992). From this perspective, social reasoning biases seemed to be a likely candidate, particularly the attributions (explanations) that people made about interpersonal events (Bentall et al., 2001; Kaney & Bentall, 1992). Bentall and colleagues published a series of studies reporting an association between self-serving biases (the tendency to make external, especially other-blaming attributions for negative events and internal, self-blaming attributions for positive events) and persecutory delusions (Kaney & Bentall, 1989; Kinderman & Bentall, 1996). Although some later studies have not found this effect a recent meta-analysis of 56 studies reported that a self-serving bias was especially marked in schizophrenia patients with persecutory delusions (Müller et al., 2021).

In light of this observation, Bentall et al. (1994) hypothesised that this specific attributional style was a cognitive bias that served to protect a vulnerable self against potentially threatening information, an account that seemed to capture the clinical presentation of many paranoid patients (i.e., poor-me paranoia). In this model, the bias is conceived to be a psychological defence that serves to ward off negative thoughts about the self and thereby reduce an otherwise unpleasant discrepancy between current beliefs about the actual self and beliefs about the ideal self. Thus, in order to operationalise this model and be empirically tested, one would predict that people with paranoia would show explicit self-esteem instability as well

as a discrepancy between implicit and explicit self-esteem, being the former lower than the latter (Murphy et al., 2018).

Many studies were carried on to test these predictions but the results have not consistently supported the defence model. For example, many researchers have noted that low self-esteem is common in paranoid patients, which seems inconsistent with the model (Garety & Freeman, 1999). Indeed, a recent meta-analysis of studies with both clinical and nonclinical participants reported a consistent association between negative self-beliefs and paranoia, although a complication is that some of this effect may have been due to depression, which is a common comorbidity of persecutory delusions (Humphrey et al., 2021). While some studies have reported an association between implicit negative self-esteem and paranoia other studies have reported that implicit levels of self-esteem do not differ between patients with persecutory delusions and control samples (Kesting & Lincoln, 2013; Tiernan et al., 2014). A recent meta-analysis of 64 studies testing the defence theory found an association between self-serving bias and paranoia (as reported in Müller et al.2021) as well as high levels of self-esteem fluctuation in paranoid patients, although little evidence was found supporting a discrepancy between implicit and explicit self-esteem measures (Murphy et al., 2018).

1.2.3. *The cognitive (direct) model of paranoia*

A model of paranoia that has been developed over many years by Daniel Freeman, Philippa Garety and colleagues, was inspired in part by Maher's earlier anomalous perception account but also incorporated additional cognitive and emotional components (Freeman, 2016; Freeman et al., 2002). The model stresses how the patient's appraisals of their own experiences can be affected by emotional and cognitive biases, leading directly to paranoid explanations. Persecutory delusions are therefore regarded as threat beliefs that arise from a search for the meaning of anomalous internal or external experiences. Once established, threat beliefs are, in turn, maintained by various cognitive and behavioural factors (Garety and Freeman, 1999;

Freeman et al., 2002), including "safety behaviours" such as the avoidance of situations which might provide disconfirmatory evidence about paranoid hypotheses (Freeman et al., 2007). Hence, the model attempts to account for both the formation of paranoid beliefs as a direct reflection of the psychology of the individual rather than as a defence, and also how they are maintained once formed (Freeman et al., 2002; Freeman, 2007).

An early attempt to characterize a cognitive bias that has subsequently been incorporated into this model was made by Hemsley and Garety (1986), who argued that the reasoning style of deluded patients deviated from Bayesian probabilistic inference. This meant that patients would reach a conclusion without seeking the necessary information or rendering them unable to change their minds when presented with new information that is incongruent with an established belief system (Garety et al., 1991). This type of bias has been termed a *data-gathering* or *jumping-to-conclusion bias* and many studies have measured it, typically using the 'beads task'. In this task, participants are shown two jars, one with predominantly red beads but some blue, and the other with the ratio (typically 85:15) reversed. Participants are presented with a bead and given the choice of deciding which jar it came from or requesting another bead (seeking more evidence); the process continues until the participant makes a decision (several variations of this procedure have been tried in different studies). Studies have consistently found strong associations between delusions and reaching a hasty decision compared to controls, such that some deluded patients make a decision after seeing only one or two beads (Garety & Freeman, 2013; Soet al., 2016; McLean et al., 2017). Note that this bias is not believed to be specific to paranoia but to delusions more generally. However, some studies have found that the data-gathering bias is especially strong in paranoid patients when they are assessed with probability reasoning tasks that employ social and emotional salient stimuli rather than beads (Dudley et al., 1997; Lincoln et al., 2011; Young & Bentall, 1997). More than thirty years after this task was first used, some questions about the precise

psychological mechanisms involved remain. For example, some studies have shown that the bias is closely correlated with and may be indistinguishable from broader measures of executive function and general intelligence (Bentall et al., 2009; Tripoli et al., 2021).

A similar concept which has garnered attention more recently is the bias against disconfirmatory information (BADE; Woodward et al.,2007; Woodward et al., 2008). This is measured using a task in which participants are shown a picture and then asked to choose which of four hypotheses most likely describes the events depicted. Further pictures are presented and the participants are asked to re-evaluate the hypotheses. A BADE is demonstrated when information which is inconsistent with an initial hypothesis does not lead to a change in the participant's appraisal of the likelihood of the hypotheses. Again the BADE has been consistently associated with delusions (McLean et al., 2017) although not specifically with paranoia. One significant aspect of this observation (as with most versions of the beads task) is that the content of the task is unrelated to patients' delusional beliefs, suggesting that the cognitive biases of deluded patients are general and not restricted to when they reason about information related to their delusions.

1.2.4. *The role of attachment and early victimization experiences*

As noted earlier, there is consistent evidence that paranoid beliefs are associated with negative beliefs about the self (Humphrey et al., 2021) and attributional biases (Murphy et al., 2018; Müller et al., 2021). The defence model and the direct cognitive model both incorporate these biases, although the way that they then lead to paranoia is different in the two models. Whereas in the defence model, the individual uses attributions to reduce the distress associated with awareness of negative beliefs about the self; in the direct model these biases cause feelings of vulnerability and help shape paranoid interpretations of anomalous experiences. Most researchers believe that these biases are likely to be mediating mechanisms that help to explain

the relationship between paranoia and adverse life experiences such as experiences of victimisation and disrupted early relationships with caregivers.

Recent research has expanded the understanding of emotional processes in paranoia by focusing on attachment processes. Attachment styles were conceptualised by the British psychoanalyst John Bowly (1969) as internalised representations of early interpersonal experiences with primary caregivers (or attachment figures). These representations take the form of relational models of the self and others and influence the way close interpersonal relations are established (Shaver & Mikulincer, 2005). Depending on the availability of the attachment figures throughout childhood, these styles can be developed into adulthood as secure or insecure. While the former would reflect a sense of confidence that emotional needs would be met by significant others in times of need the latter would reflect the contrary (Mikulincer, 1995). Insecure attachment styles, in turn, can be categorised as being avoidant or anxious (Shaver & Mikulincer, 2005). When threatened, people with an anxious attachment style tend to express an excessive need for closeness, revealing fear of abandonment and worries about separation. In contrast, when facing a distressing situation, people with an avoidant attachment style tend to suppress a need for closeness through emotional and physical distancing reflected in an excessive sense of autonomy (Berry et al., 2007; Dewitte et al., 2008).

Bentall and Fernyhough (2008) proposed that an insecure attachment process might explain the relationship between early adverse experiences, particularly disrupted bonds with caregivers, and paranoid beliefs. Pickering et al. (2008) subsequently demonstrated that subclinical paranoid beliefs were associated with both anxious and avoidant styles and that this relationship was mediated by negative beliefs about the self, the anticipation of threats and the perception of others as powerful and controlling. Later, Wickham et al. (2015) showed that paranoid symptoms (but not hallucinations) in a clinical sample were also associated with both insecure styles and that this effect was mediated by low self-esteem (see also Ringer et al.,

2014). These associations between insecure attachment and paranoia have since been replicated many times in both clinical and nonclinical samples (for a meta-analysis, see Murphy et al., 2020). Moreover, in a longitudinal (experience sampling) study with clinical participants, fluctuations in attachment-related cognitions were shown to predict fluctuations in paranoid beliefs, and this effect was much stronger than the association between fluctuations in self-esteem and paranoia (Sitko et al., 2016). Given the interpersonal nature of paranoid beliefs, it is not surprising that core beliefs about the self and others (i.e., attachment styles) can be considered a precursor to feelings of vulnerability as well as a disposition of perceiving others as unreliable (Humphrey et al., 2021).

1.3. Towards a biopsychosocial integration of paranoia

1.3.1. *Socio-environmental and genetic factors*

Challenges to the notion that mental health symptoms are explained by latent medical disorders and the development of psychological models for its conceptualization have led to a renewed focus on the role of social and environmental factors in their aetiology. For example, Read et al. (2008) have proposed a biopsychosocial model of psychotic experiences, arguing that early traumatic events might impact brain development, leading to dysregulation of the brain's stress regulation mechanisms (e.g., hypothalamic-pituitary-adrenal axis). This view is in line with the diathesis-stress model which states that psychosocial stressors might act upon pre-existing genetic vulnerabilities and thus elicit psychotic symptoms (Walker & Diforio, 1997). It seems likely that research on the interactions between social factors and biological mechanisms will lead to a better understanding of the aetiology of paranoid experiences.

In the case of paranoia, studies have found that this type of belief is associated with living in harsh urban environments, poverty, and belonging to ethnic minority groups (Bosqui et al., 2014; Vassos et al., 2012; Wicks et al., 2005). Furthermore, there is considerable evidence establishing a link between childhood interpersonal trauma and the probability of

experiencing psychotic symptoms (Varese et al., 2012), particularly between paranoia and emotional neglect, bullying, and early attachment disruptions (Bentall et al., 2012; Wickham et al., 2015). Paranoia has also been shown to be associated with victimization, for example, discrimination, in adult life (Janssen et al, 2003). It is hypothesized that these experiences contribute to developing paranoid beliefs because repeatedly being rejected, criticized, or discriminated against, particularly during childhood, can lead to a heightened sensitivity to threatening social cues (DeWall & Bushman, 2011; Tso et al., 2015). Moreover, these kinds of experiences are often associated with feelings of inferiority, powerlessness, vulnerability, and a negative view of the self, which can foment distrustful and threatening expectations regarding other people's intentions (Freeman & Garety, 2014).

Humans have likely evolved to form strong ties with others because, in the early stages of our evolutionary history, living in groups no doubt facilitated cooperation and protective strategies that enhanced our ability to meet our reproductive and survival needs. In contrast, being isolated from a group would have reduced the likelihood of surviving and reproducing (DeWall & Bushman, 2011). Hence, the pursuit of acceptance by others is considered one of the most basic human needs and is associated with feelings of success, enhanced social status and positive affect (Baumeister & Leary, 1995). Consistent with this account, several authors argue that identifying with a social group can act as a protective factor against mental illness (Jetten, et al., 2017). Hence, studies have found that having a strong sense of belonging to various social identity groups such as one's ethnicity, nationality, neighbourhood, family, and friends is associated with low levels of paranoid beliefs (Greenaway et al., 2019; Elahi et al., 2018; McIntyre et al., 2018; Thomas et al., 2017). It has been argued that this effect occurs because having a strong identity boosts positive psychological states such as self-esteem and perceived control, ameliorating mechanisms associated with paranoia such as negative self-schemas and an external locus of control (McIntyre et al., 2018).

However, social factors cannot fully explain the development of psychotic symptoms. While many studies have explored the interplay between environmental and genetic factors and the risk of being diagnosed with a psychotic disorder (e.g., schizophrenia, bipolar disorder, schizoaffective disorder; Davies et al., 2020), very little research has explored this interplay in the aetiology of specific symptoms such as paranoia. One study employing a large sample of monozygotic and dizygotic adolescents twins found that the risk of experiencing paranoia was 50% heritable in the general population, suggesting that both environmental and genetic factors are equally important, but without revealing how these factors interact (Zavos et al., 2014).

1.3.2. *Social cognition*

Social cognition involves the study of the processing of social stimuli or in other words, how people make sense of themselves and other people from social perception to making attributions about other people's intentions (Fiske & Taylor, 2017). Given the complexity of social information (thoughts, feelings, behaviours, and attitudes toward the self and others) social psychology researchers argue that there is more than one core psychological process involved when processing social stimuli. For example, the processing of unfamiliar faces is done within hundreds of milliseconds after stimuli presentation (Todorov et al., 2008) with neuroimaging studies showing that the neural processing of faces is more sensitive to emotional face stimuli (e.g., anger) than non-expressive face stimuli (Hinojosa et al., 2015). On the other hand, processes such as reflecting upon our and others' mental states (i.e., mentalizing) require the allocation of cognitive resources indicating that it is a higher-order type of processing (Firth & Firth, 2011; Cane et al., 2016). These examples would fall into two main types of psychological processes: implicit (fast, automatic, intuitive, associative thinking) and explicit (slow, deliberative, rational, and propositional thinking; Kahneman, 2013; De Houwer et al., 2009). This framework is also known as *dual-process theory* with *system 1* (implicit) having

early evolutionary roots than *system 2* (explicit) which is linked to language, reflective consciousness, and higher-order cognitive functions (Evans, 2008).

Several studies have been carried out on people with schizophrenia looking at socio-cognitive factors such as facial perception and mentalizing processes. Results have shown that this clinical population shows difficulties in judging facial affective features as well as properly inferring other people's intentions reflected in aberrant neural activation when measured with neuroimaging techniques (Green et al., 2015). Although most of these studies have adopted a categorical approach to psychiatric conditions, some researchers have found particular associations between a tendency to over-attribute intentions to external agents (hyper mentalizing) and paranoid delusions (Firth, 2004; Ciaramidaro et al., 2014). Moreover, some studies have found that people with paranoid schizophrenia were inclined to misperceive ambiguous facial expressions as threatening as reflected in increased amygdala activation (Green & Phillips, 2004). Engagement of the amygdala is involved in the preferential processing of social stimuli so that potential threats can be rapidly detected which in turn is mediated by the dopaminergic system (Lipka et al., 2011). Some authors argue that exposure to early victimization experiences influences brain development and hence information processing during future stressful events by sensitizing neurobiological pathways related to the original traumatic response (Read et al., 2003; Kindt et al., 2005). Thus, given the well-established association between early victimization experiences and paranoid symptoms (Coughlan & Cannon, 2017), it would be reasonable to hypothesize that the processing of environmental information (e.g., social stimuli) in people with paranoia will be overtaken by the implicit system (e.g., fast, associative) leading to automatic interpretations that others are dangerous and threatening.

1.3.3. *Human cooperation and paranoia*

Although present in other organisms in the animal kingdom, cooperation is probably one of the most complex as well as distinct characteristics of the human species. Simply defined as prosocial behaviours that usually involve a personal cost to pay for the benefit of another party (Rand & Nowak, 2013), cooperation is thought to have strong evolutionary roots in humans. Anthropologists and palaeontologists argue that cooperative strategies would have been favoured by natural selection given our unique reproductive and developmental patterns (i.e., short fertility intervals between births; extended childhood dependence) as well as our lack of in-built tools to survive as foragers (Apicella & Silk, 2019; Henrich & Muthukrishna, 2021). This strong interdependence benefited pro-social behaviours such as cooperative hunting, food sharing, division of labour, and alloparenting (i.e., care of offspring by non-kin members) resulting in high advantages to our hunter-gatherer ancestors which in turn set the foundation for socio-psychological processes (Jaeggi et al., 2010 ; Apicella & Silk, 2019). For example, it is argued that primates (particularly humans) are more sensitive to processing signals of need which seems to depend on mentalizing capacities as well as emotional abilities such as empathy and sympathy (Jaeggi et al.,2010; Seyfarth & Cheney, 2013). Moreover, it is thought that the development of human language as well as the growth in social group sizes led to mechanisms of identifying non-cooperative members who in turn would be faced with social punishment (e.g., ostracism, violence; Jaeggi et al.,2010; Henrich & Muthukrishna, 2021). These social mechanisms have been considered the basis of psychosocial processes such as spreading rumours (i.e., gossiping), the importance of social reputation, as well as being sensitive to the presence and size of an audience (i.e., heightened sense of being observed; Gintis et al., 2001; Jaeggi et al.,2010).

Behavioural economists have paid considerable attention to the construct of cooperation and have developed scientific methods to study how individuals make economic

decisions when presented with different options that have clear costs and benefits for themselves and other people (Dijk & De Dreu, 2021). These types of methods have been conceptualized within the game theory framework, which provides a mathematical framework for understanding social interactions between sets of players employing cooperative or competitive strategies (i.e., making choices that could reward both players, punish both, or favour one over the other; Rand & Nowak, 2013). One of the most well-known game theory tasks is the "prisoner's dilemma game" in which two players are presented with a set of options where, if both cooperate, they get a higher payoff than if both defect. However, defecting while the other party cooperates gets the highest payoff for the defector and the lowest payoff for the co-operator, highlighting a dilemma between acting out of self-interest or cooperatively (Bergmuller et al., 2010; Rand & Nowak, 2013). Another task within the game theory paradigm is the "ultimatum game" where two players are provided with an endowment and player A needs to decide how to distribute it between themselves and the other player. Player B can accept the terms of player A or can reject them, resulting in the loss of the endowment for both players. A variation of this same game is the "dictator game" in which player A is provided with the endowment and can decide how to split it however, player B does not have any say on players' A decision (Singer, 2008; Thielman et al., 2020). Behavioural scientists argue that a rejection outcome from player B or a prejudicial offer from player A can reflect punitive attitudes either to penalise the opponent's past behaviours (i.e., tit-for-tat) or to sanction unfair actions from the counterpart (Bland et al.,2017).

Psychologists have been interested in how individual differences can predict cooperative behaviours in game theory paradigms with studies showing that personality traits such as agreeableness have been associated with prosocial behaviours in game theory tasks (Thielman et al., 2020). Given that paranoid beliefs are negatively associated with agreeableness traits (Freeman et al., 2012) one would hypothesise that people with this belief

would lack the perceptions of others as holding cooperative or helpful intentions (Huddy et al., 2014). A study employing the Prisoner's Dilemma revealed that high paranoia in non-clinical samples was associated with competing rather than cooperating and that this decision was predicted by state paranoia and distrust of their counterpart rather than maximizing their own payoffs (i.e., self-interest; Ellett et al., 2013). These results have been replicated in clinical populations showing the role of distrust-based competition between state paranoia and competitive behaviour in participants with persecutory delusions (Ellett et al., 2023). Conversely, large non-clinical studies using alternative game theory paradigms (i.e., Ultimatum and Dictator Game) have found that reduced cooperation was better explained by self-interest rather than distrust (Raihani & Bell, 2017). Ellet et al. (2023) argue that the different designs of game paradigms influence how the social interaction is set up either by creating uncertainty about the other party's decisions (prisoners' dilemma) or waiting for the turns of each player (ultimatum game; Ellet et al.,2023). Thus, different game theory paradigms probably measure specific aspects of cooperation, and hence can inform different psychosocial processes that may be involved in the non-cooperative behaviours observed in people with paranoid beliefs.

1.3.4. *The Role of Trust*

Almost all our everyday interactions with other people, from small monetary transactions to voting for a political party, involve a certain degree of trust. Thus, trust is a key facet of social behaviour, and is probably rooted in evolutionary mechanisms, given that our ancestors would have had to rely on each other in order to survive in harsh and stressful environments (Simpson, 2007). Trust is usually conceived as the willingness to accept vulnerability to the actions of others, which involves positive expectations about the intentions of the party being trusted, the trustee (Lewicki, Tomlinson, & Gillespie, 2006). These positive expectations are held in the absence of knowledge that the other party will act accordingly

(Lewicki & Brinsfield, 2012) and, thus, the act of trust involves placing personal welfare in the hands of someone else in conditions of uncertainty (Hatzakis, 2009; Simpson, 2007).

Trust has generally been conceptualized as a multidimensional construct, which includes initial perceptual judgements based on first impressions of often unfamiliar faces (Oosterhof & Todorov, 2008) and more general beliefs about the degree to which people can be trusted (i.e., interpersonal trust; Lewicki & Brinsfield, 2012; Simpson, 2007). Whereas the former involves fast and automatic evaluations (Todorov, Pakrashi, & Oosterhof, 2009) the latter encompasses a more stable set of cognitive, affective and behavioural dispositions about the reliability of other people's intentions (Lewicki et al., 2006). Some authors argue that trust processes become activated by uncertain situations and that, in harsh environments, adopting a tendency to perceive and/or believe that others are untrustworthy allows people to minimize the risk of making the most costly error, and hence avoid getting hurt by someone who was initially trusted (Haselton & Nettle, 2006; Hatzakis, 2009).

Several studies have explored the trustworthiness judgements about faces made by patients with schizophrenia (Baas et al., 2008; Couture et al., 2008; Strauss et al., 2012) but only a few of them have looked at this construct specifically in relation to paranoia (Buck et al., 2016; Hooker et al., 2011; Haut & MacDonald, 2010; Pinkham et al., 2008; Tremeau et al., 2016). Although some of these studies have found positive associations between paranoia and mistrust judgements (Buck et al., 2016; Pinkham et al., 2008) others have found no association at all (Haut & MacDonald, 2010; Trémeau et al., 2016). The inconsistency of these findings must be interpreted in the context of the heterogeneity of symptom presentation in the participating patients. Moreover, one study found that paranoid patients reported low levels of trust only after they had been primed beforehand with threatening images, supporting the notion that trustworthiness judgements are context-dependent, particularly in clinical populations (Hooker et al., 2011). Similar research in subclinical paranoia has been less

extensive. One study reported differences in trustworthiness ratings between high and low-paranoid participants (Kirk et al., 2013) whereas another did not (Hillmann et al., 2017).

The relationship between more general interpersonal trust and paranoia has been more extensively studied in nonclinical populations, with most studies reporting medium to large associations (Axelrod et al., Widiger et al., 1997; Furnham & Crump, 2015; Greenaway et al., 2019; Kong, 2017; Kramer, 1994; Murphy et al., 2012; Wickham et al., 2014; Wickham et al., 2014b). Some of these studies have been conducted in the context of psychometric validation of instruments assessing subclinical paranoia (Axelrod et al., 1997; Barreto Carvalho et al., 2017; Furnham & Crump, 2015) or to explore the mediational role of trust between social adversities (e.g., trauma, social deprivation) and paranoid beliefs (Murphy et al., 2012; Wickham et al., 2014; Wickham et al., 2014b). Other studies have considered trust as a potential mechanism for promoting prosocial behaviours (e.g., cooperation, coordination) and thus reducing paranoid cognitions (Greenaway et al., 2019; Kong, 2017; Kramer, 1994). Hence, interpersonal trust, operationalized as a personality trait or a set of cognitions that guide social behaviour, seems to show a more stable association with paranoia when measured in non-clinical populations.

Mistrust is considered a core feature of paranoid thinking and is regarded as a subcomponent of the paranoia spectrum (Bebbington et al., 2013; Bell & O'Driscoll, 2018). However, it seems that the role of trustworthiness in the nonclinical population and interpersonal trust in the clinical population has not been extensively studied. Moreover, findings regarding the association between trustworthiness and paranoia in clinical populations seem to be more inconsistent than the association between interpersonal mistrust and paranoia in non-clinical populations. Thus, it is not clear if problems of trust act as a precursor of paranoid beliefs or vice versa as there is no psychological model of paranoia that includes mistrust as a potential underlying mechanism or as an outcome. Given the social and relational

character of paranoia and the role of trust in everyday interaction, more studies should focus on exploring the relationship between different forms of trust and paranoid beliefs by exploring psychological underlying mechanisms.

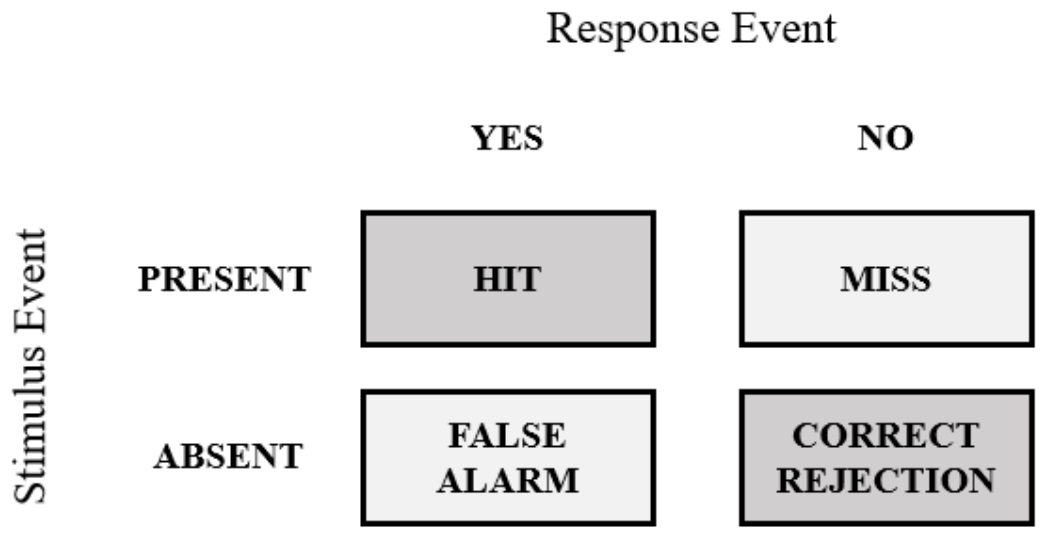
1.3.5. *Signal Detection Theory*

Every decision made, particularly in uncertain situations, involves the assessment of the costs and benefits associated with the outcome of that specific decision (Kantowitz et al., 2014). For example, when buying a plane ticket, deciding whether or not to pay an extra fee for fast-track security can involve the benefit of avoiding standard security queues and making it to the flight on time or the risk of losing the flight due to a very long security queue. This type of situation implicates a balancing of costs and benefits, which can be influenced by different factors such as travelling season, the size of the airport, or previous experiences. One can consider those factors and make up one's mind hoping that the benefits of that decision will outweigh the costs. Other examples encompass perceptual spheres (such as deciding if a specific stimulus is present or not), cognitive spheres (recognising new versus old presented information), and social spheres (a jury discriminating between innocent and guilty defendants). To understand the underlying process by which people arrive at different decisions experimental psychologists have adopted a mathematical approach to study this phenomenon known as *signal detection theory*.

Signal detection theory (SDT) informs analytical techniques to measure to what extent an individual can discriminate between two different types of stimuli, particularly under uncertain conditions (Stanislaw & Todorov, 1999). Thus, to implement SDT, experimental tasks are usually designed in a way so that participants are instructed to detect a signal (stimuli) from noise (no stimuli). For example, a classic signal detection task involves presenting a series of trials where some include the presentation of pink noise sound (noise trials) and others where a voice sound is presented over the pink noise sound (signal trials). After the presentation of

each trial participants are asked whether they have detected the voice over the pink background noise by answering “yes” or “no”. From this design, four possible behavioural outcomes can be drawn: detecting a voice sound on signal trials (hit), detecting a voice sound on noise trials (false alarm), not detecting a voice sound on signal trials (miss), and not detecting a voice sound on noise trials (correct rejection; see Figure 1.1.). Based on the distribution of responses to the trials two types of parameters can be calculated, sensitivity (indicating the level of accuracy in discriminating signal from noise trials) and bias (tendency to respond *yes* or *no*; Kantowitz et al., 2014).

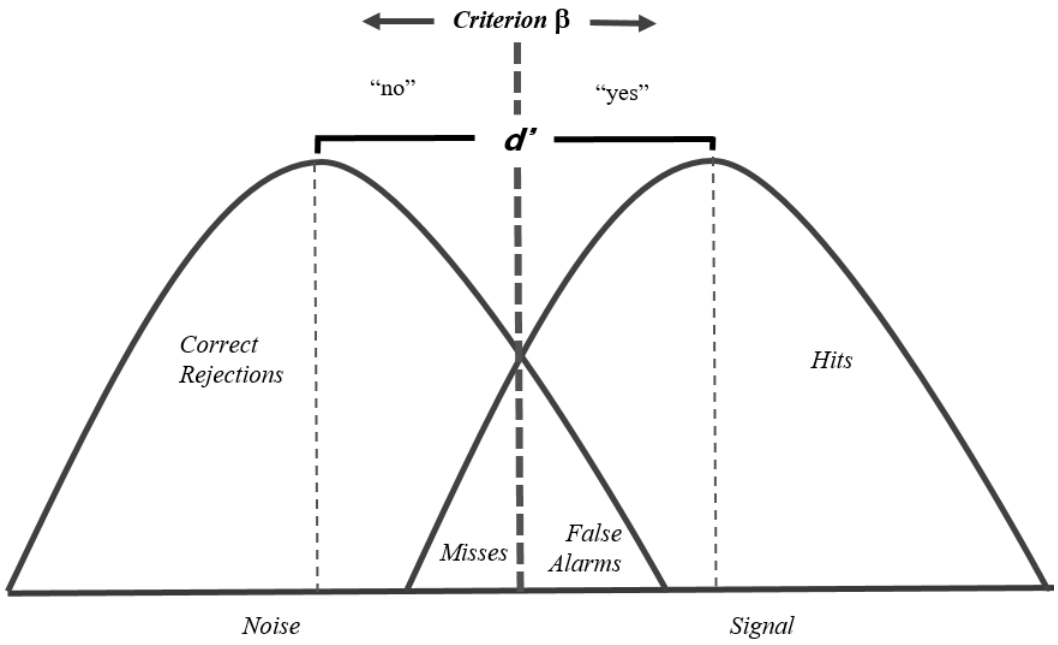
Figure 1. 1. *A signal detection matrix reflecting four possible behavioural outcomes of a signal detection task*



Sensitivity (also referred to as *d’*) reflects the standardized distance between signal trial distributions and noise trial distributions hence, larger *d’* values would indicate higher accuracy in discriminating between noise and signal trials (Stanislaw & Todorov, 1999; Macmillan & Creelman, 2004). Response bias on the other hand reflects the general tendency to respond towards detecting a signal on noise trials or vice versa. This process towards an affinity to a

"yes" or "no" answer depends on the value that a decision variable achieves during each trial (Stanislaw & Todorov, 1999). This means that if a threshold is met at a given trial the participant will answer "yes" otherwise the answer selected will be "no". The value that defines the threshold is called criterion (also referred to as *criterion*β) which reflects the minimum degree of internal confidence needed for the participant to reach a decision (see Figure 1.2.). Depending on the costs and benefits of the decision process the stance of the criterion can favour a liberal decision-making process (looking at Figure 1.2. the β dotted line would be placed towards the left) or a conservative decision-making process (the β dotted line in Figure 1.2. would be placed towards the right; Macmillan & Creelman, 2004; Kantowitz et al., 2014).

Figure 1. 2 *Distribution of signal and noise trials with a balanced criterion* β *and large distance between mean distributions reflecting a high sensitivity (d').*



As mentioned in previous sections, evolutionary psychologists argue that a biased decision-making process is an adaptive evolutionary strategy as making a false assumption in specific uncertain situations minimizes the risk of a greater error (Haselton et al.,2013). For example, misperceiving the shadow of a cat as one of a person (false alarm) while walking at

night through an alley prepares the body for a fight or flight response which, the other way around (miss response – perceiving the shadow of a person as one of a cat), would be more costly (possibility of getting hurt). This type of mechanism seems to have been around in humans and other organisms for millions of years and it is thought that biased information processing facilitates faster responses and thus makes use of fewer cognitive resources (Nesse, 2005). Nonetheless, a rigid tendency to misperceive trivial stimuli as dangerous can become dysfunctional in everyday life, particularly concerning social stimuli. Psychiatric diagnoses such as social anxiety and post-traumatic stress disorder are some examples of conditions in which biases towards misperceiving neutral situations as threatening have been observed, which in turn leads to avoidant or hypervigilant maladaptive behaviours (Stevens & Jonavic, 2018; Chen et al.,2020). In the case of paranoia, the evidence is mixed, with some studies showing a general deficit in discriminating social stimuli (Huang et al., 2011) whereas other studies reveal the contrary (i.e., bias towards threatening social stimuli; Green &Philips, 2004). Few studies on paranoia have been carried out from a signal detection perspective, and as mentioned in the previous section, fewer have done so looking at the trustworthiness of unfamiliar faces.

1.4. Conclusion

This chapter tried to document the enormous progress in the understanding of persecutory delusions and paranoid beliefs that has been achieved since the early days of psychiatry. This progress has accelerated dramatically since the final decade of the last century, largely because of the pursuit of systematic psychological research programmes. This chapter has also reviewed recent psychological models and rather than focus on the disagreements between these models, a summary of what is now agreed upon in terms of the processes and mechanisms of paranoid thinking will be stated:

First, the idea that paranoia should be considered a phenomenon in its own right, and one that encompasses a continuum between persecutory delusions and ordinary suspiciousness, has been extraordinarily fruitful. Although questions remain about this continuum, many findings from psychological research have been replicated both with patients suffering from persecutory delusions and non-patients scoring high on measures of paranoid belief.

Second, there is consistent evidence that paranoia is associated with specific psychological-emotional processes: an exaggerated self-serving attributional bias, negative self-schematic processes, and insecure attachment styles. In addition, the most extreme form of paranoid belief – persecutory delusions – seems to be associated with more general cognitive biases in the ability to integrate information and test hypotheses.

Third, it is now widely agreed that adverse experiences, especially in childhood, play an important causal role in psychotic experiences and this seems to be true for paranoia. In particular, disrupted early relationships with caregivers, exposure to harsh environments, and experiences of victimization seem to be important risk factors for the later development of paranoid traits.

Fourth, an approach that integrates socio-cognitive models of information processing (explicit vs implicit) with evolutionary insights seems to provide an appropriate framework for understanding and researching core aspects of paranoid beliefs such as social perception and meta-cognitive processes.

Fifth, the role of trust processes (i.e., interpersonal trust and trustworthiness) in paranoia has not been extensively studied and requires further exploration. .

Sixth, the implementation of signal detection theory seems to be suitable for the empirical study of the trustworthiness of unfamiliar faces in the context of paranoia research.

These findings provide the basis for developing an integrative clinical and scientific framework for answering research questions that would help improve psychological models of paranoid beliefs as well as potential clinical interventions.

1.5. Aims of the current thesis

Based on the review of this chapter, this thesis will focus on answering a neglected question by the clinical and the research community: What is the relationship between paranoia and trust processes (i.e., interpersonal, trustworthiness) in clinical and non-clinical populations?

For this, we propose a model that takes into account a continuum approach to psychopathology which incorporates the theoretical accounts of the defensive and cognitive model of paranoia (implicit and explicit self-esteem) but also adding the role of early attachment disruptions. Moreover, we study interpersonal trust as well as trustworthiness from a socio-cognitive perspective by exploring "self" and "other" relational schemas using behavioural, psychophysiological and self-report instruments. This model would ideally be tested using epidemiological, experimental, and clinical research designs.

Thus, the structure of the current thesis will be as follows:

Chapter II (Study 1): The second chapter of this thesis aims to explore in three large international representative samples which of several different trust predictors (institutional, interpersonal, and perceptual) are specifically associated with paranoid beliefs in contrast to a similar construct characterized by distrust, conspiracy mentality (i.e., tendency to believe and divulge conspiracy theories).

Chapter III (Study 2): This chapter aims to create a psychological model of paranoia by including bias towards mistrust as a potential mechanism mediating between insecure

attachment styles and paranoid beliefs when also controlling for negative self-esteem in a large UK representative sample.

Chapter IV (Study 3): By employing a between-within-subjects experimental design with an online convenience sample, this chapter will aim to test the effect of a novel evaluative conditioning experimental paradigm by activating relational schemas (i.e., self, other) and its effect on paranoid-related constructs such as attachment, self-esteem and trustworthiness judgements.

Chapter V (Study 4): The fifth chapter aims to explore the psychophysiological mechanisms of trustworthiness judgements in clinically paranoid and non-clinical samples by implementing event-related potential (ERP) techniques employing mobile encephalogram (EEG) equipment. This proof-of-concept study will also examine the association between psychological and behavioural (paranoid traits, attachment, trust judgments) variables with specific psychophysiological markers related to face perception.

Chapter VI: This chapter will discuss the findings of this PhD thesis as well as its limitations. Moreover, it will reflect on directions for future research in this area as well as the theoretical, methodological and clinical implication of this specific research topic.

Chapter II – First Empirical Study

Paranoid beliefs and conspiracy mentality are associated with different forms of mistrust: A three nation study

Contributions

Anton P. Martinez (conceptualization, formal analysis, methodology, writing – original draft)

Mark Shevlin (statistical conceptualization, supervision, writing, review and editing)

Carmen Valiente and Philip Hyland (writing, review, and editing)

Richard P. Bentall (supervision, methodology, writing– original draft, review, and editing.)

The following chapter has been peer-reviewed and published at Frontiers in Psychology:

Martinez, A. P., Shevlin, M., Valiente, C., Hyland, P., & Bentall, R. P. (2022). Paranoid beliefs and conspiracy mentality are associated with different forms of mistrust: A three-nation study. *Frontiers in Psychology*, 6186.

Findings from this chapter have been presented as part of a symposium at the European Conference on Schizophrenia Research (ECSR):

Martinez, A. P., and Bentall, R.P. (2021). Paranoid beliefs and conspiracy mentality are associated with different forms of mistrust: A three-nation study. *ECSR*, Berlin, Germany (Online).

Abstract

Paranoia and conspiracy are terms typically used interchangeably. However, although the underlying content of these types of beliefs might be similar (e.g., seeing others as powerful and threatening), recent research suggests that these constructs differ in important ways. One important feature shared by both constructs is excessive mistrust but this aspect might play different roles in each belief system. In this study, we explored the strength of associations of different trust predictors (i.e., trust in institutions, trust in sources of information, perceptual trust, and interpersonal trust) between conspiracy mentality and paranoid beliefs. We tested this association in a large representative multinational sample (UK n=2025; Spain n= 1951; and Ireland n= 1041). Confirmatory factor analysis supported a two-factor model of conspiracy and paranoid beliefs in each nation sample. Path and equality of constraints analysis revealed that paranoia was more strongly associated with perceptual mistrust (bias towards mistrusting unfamiliar faces) whereas conspiracy was more strongly associated with mistrust in political institutions. Although interpersonal mistrust and trust in social sources of information were associated significantly with conspiracy, their association with paranoid beliefs was stronger. These findings clarify the role of different trust processes in both belief systems. Limitations of this study are discussed.

Keywords: *Paranoid beliefs, conspiracy mentality, institutional trust, trustworthiness, interpersonal trust.*

2.1 Introduction

The terms ‘conspiracy theorist’ and ‘paranoid’ are often used interchangeably when referring to people who are suspicious of other people’s intentions, doubt the veracity of historical events, or who think that important governmental decisions are part of secret plots orchestrated by powerful others. For example, American historian Richard Hofstadter described the ‘paranoid style’ of American politics as a sense of heated exaggeration, suspiciousness and conspiratorial fantasy, although emphasizing that he was not using the term ‘paranoid’ in a clinical sense (Hofstadter, 1965). Some clinical definitions also appear to conflate the two concepts. For example, in the ICD-10 Classification of Mental and Behavioral Disorders, one of the characteristics of paranoid personality disorder was preoccupation with unsubstantiated ‘conspiratorial’ explanations of events both immediate to the patient and in the world at large (World Health Organisation, 1993), although the concept was dropped from the later, 11th edition (World Health Organization, 2018). However, most definitions of paranoia and conspiracy theories point to different conceptualizations of these constructs. Whereas the former refers to unfounded beliefs that involve intentional harm to the self from others (Bentall et al., 2001; Freeman & Garety, 2000), the latter is usually defined as an explanation for significant social and political events that involves secret plots by powerful and malevolent others (Douglas et al., 2017). Thus, although both constructs attribute events to the presence of threatening agents, and while there is consistent evidence that the two belief systems are modestly correlated (Imhoff & Lamberty, 2018; Alsuhibani et al., 2022), the locus of vulnerability for each appears to be different (i.e., the individual in the case of paranoia and society in general in the case of conspiracy theories; Greenburgh & Raihani, 2022; Greenburgh et al., 2022).

It is estimated that nearly a third (26.7%) of the general population are convinced that there is a conspiracy behind many world events (Freeman & Bentall, 2017). It has also been

observed that believing in a specific conspiracy theory is often associated with belief in many others (Goertzel, 1994; Swami et al., 2011), suggesting that conspiracy ideation can be considered a trait-like predisposition which is sometimes referred to as conspiracy mentality (Bruder et al., 2013). On the other hand, paranoid beliefs are not exclusive to clinical populations as they are experienced by at least 10-15% of the general population (Freeman, 2007) suggesting that they lie on a continuum from less severe to more dysfunctional forms (Bebbington, McBride, Steel, Kuipers, Radovanovic, et al., 2013; Elahi et al., 2017). Hence, researchers have tried to explore the psychological precursors of both types of beliefs by conducting studies in non-clinical populations. Factors such as negative self-esteem, disrupted attachment experiences, as well as various cognitive biases (i.e., jumping to conclusions or external locus of control) have been associated with the development and maintenance of paranoia (Bentall et al., 2014). Equally, narcissism, exaggerated positive view of the self, specific cognitive biases (i.e., confirmatory bias or illusory correlations), and poorer analytical reasoning have all been associated with conspiracy mentality (Douglas et al., 2017; Goreis & Voracek, 2019).

Although conspiracy mentality and paranoid beliefs therefore appear to be associated with different specific psychological factors, both constructs are thought to involve excessive mistrust (Freeman & Bentall, 2017). Trust is considered to be a fundamental aspect of everyday social interactions (Simpson, 2007). By accepting being vulnerable to the actions of another party we hold positive expectations regarding the other party's intentions and behaviours despite the uncertainty of what the outcome of that dyadic relationship will be (Lewicki et al., 2006; Lewicki & Brinsfield, 2012).These expectations are not restricted to interpersonal interactions but can extend to various social systems and large institutions (e.g., companies, banks, governmental agencies), or any circumstances in which we perceive another party as having control over our resources and life options (Hatzakis, 2009; Simpson, 2007). Thus, trust

appears to be a key component in the way we build our social relationships as well as in social behaviours we engage in, ranging from voting for a political party, reading a specific newspaper or getting vaccinated.

The association between mistrust and paranoia is well established in the literature, the former being a subcomponent of the paranoia spectrum usually present in non-clinical populations (Bebbington, McBride, Steel, Kuipers, Radovanovic, et al., 2013; Bell & O'Driscoll, 2018). However theorizing and research about mistrust in paranoia usually focuses on interpersonal mistrust (Barreto Carvalho et al., 2017; Furnham & Crump, 2015; Wickham et al., 2014) and by extension to the untrustworthiness of unfamiliar faces (Kirk et al., 2013b; Abbott et al., 2018; Martinez et al., 2020). On the other hand, conspiracy beliefs are usually linked to mistrust relating to society at large (Van Prooijen et al., 2021). For example, several studies have reported an association between conspiracy mentality and institutional mistrust, particularly in respect to political institutions (Kim & Kim, 2021; Mari et al., 2021). Nonetheless, a relationship between conspiracy mentality and lack of trust in other people has also been reported (Abalakina-Paap et al., 1999; Goertzel, 1994). One recent study reported that interpersonal mistrust was associated with both conspiracy mentality and paranoia, with the association between interpersonal mistrust and paranoia being stronger (Imhoff & Lamberty, 2018). This distinction between interpersonal and institutional mistrust is in line with the view that paranoia reflects a threat to the 'self' whereas, in the case of conspiracy mentality, the threat is orientated to society more generally (Imhoff & Lamberty, 2018; Van Prooijen et al., 2021).

To the best of our knowledge, no study to date has explored the role of specific trust processes (i.e., interpersonal, institutional, trust in sources of information, trust perceptions) in relation to both conspiracy and paranoid beliefs using large representative samples. The current study aimed to expand our understanding of the relationships between conspiracy mentality

and paranoid beliefs by examining how these beliefs co-varied with specific types of mistrust in representative population samples from three nations: the United Kingdom, Ireland, and Spain. The data was collected as part of a multinational study of the psychological impact of the COVID pandemic, during the earliest stages of the emergency. Following the work of Imhoff and Lamberty (2018) and Alsuhibani et al. (2022), we expected that paranoid beliefs and conspiracy mentality would form two distinct but correlated factors in all three nation samples. We employed measures of a wide range of forms of trust (i.e., interpersonal, institutional, trust in sources of information, perceptual trust) and looked at the specific contributions of each to paranoia and conspiracy mentality. Thus, we hypothesized that conspiracy mentality would be specifically associated with institutional mistrust, in particular with political institutions, and with mistrust in traditional media as well as with mistrust in institutional sources of information. On the other hand, following the findings of Martinez et al. (2020), we expected paranoid beliefs to be specifically associated with a tendency to judge face stimuli as untrustworthy (i.e., bias or perceptual mistrust). With regards to interpersonal mistrust, we predicted that this form of trust would be associated with both conspiracy mentality and paranoid beliefs however, we expected a stronger association with the latter[1].

2.2. Material and methods

2.2.1. *Participants/Procedure*

This study was based on data collected in the first wave of parallel surveys conducted in the United Kingdom (UK: *N*=2025), Ireland (*N* = 1041), and Spain (*N* = 1951) as part of the COVID-19 Psychological Research Consortium (C19PRC)[2] designed to monitor multiple indicators of psychosocial health during the early stages of the COVID-19 pandemic. Adult participants aged 18 years and over were recruited by the survey company Qualtrics in the UK

[1] Hypotheses of this study have not been pre-registered.

[2] https://www.sheffield.ac.uk/psychology-consortium-covid19

and Ireland, and by SONDEA in Spain. In each country quota sampling methods were used to reach a representative sample stratified by age, sex, household income, and geographical distribution within the countries (for more methodological information about–the C19PRC please see McBride et al., 2021; Spikol et al., 2021 and visit https://osf.io/2y45r for UK, Irish and Spanish surveys respectively). Participants responded to questionnaires and other measures presented on the Qualtrics survey platform and measures were comparable in all three countries by design. Information regarding age, sex, ethnicity, and dates of data collection for each country is presented in Table 2.1. Specific information about whether participants had received a mental health diagnosis was not collected in this study.

Table 2. 1. *Sample characteristics mean (M) standard deviation (SD) and percentages of sex distribution and ethnicity.*

Country	Age		Sex		Ethnicity	Data Collection Dates	Announcement of 1st Lockdown
	M	*SD*	M %	F %			
UK	45.44	15.90	48.2	51.8	85.5% White British	March 23rd - March 28th 2020	23rd of March 2020
Ireland	44.97	15.76	48.2	51.8	74.8% White Irish	March 30th - April 5th 2020	27th of March 2020
Spain	45.13	12.81	52.8	47.2	93% Spanish	April 8th - April 10th 2020	14th of March 2020

Ethical approval was granted by The University of Sheffield (Ref: 033759), the Social Research Ethics Committee at Maynooth University (Ref: SRESC-2020-2402202), and The Complutense University of Madrid (Ref: 2019/20-034) for the UK, Irish and Spanish samples, respectively. All participants were presented with an information page which detailed the

purpose of the study, and confidentiality of their data (under GDPR guidelines), and their right to withdraw at any time.

2.2.2. *Measurements*

The Revised Paranoia and Deservedness Scale (PADS-R; Melo et al., 2009). Paranoid beliefs were assessed by rating the agreement with five items of the PADS-R persecutory subscale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Items involved statements such as *''I'm often suspicious of other people's intentions towards me*'' and *''People will almost certainly lie to me''*. This scale has been previously validated in clinical and general population samples and its internal reliability in this study was very good across three countries (UK, α=.86; Spain, α=.84; Ireland, α=.83).

The Conspiracy Mentality Scale (CMS; Imhoff & Bruder, 2014) was used as a measurement of conspiracy mentality in which participants have to rate how likely based on their opinion each statement is true from 0% (*certainly not*) to 100% (*Certainly*). This scale included five statements such as *"Events which superficially seem to lack a connection are often the result of secret activities"* and *"Many important things happen in the world, which the public is never informed about"*. The internal reliability of this scale across the three countries was very good (UK, α=.85; Spain, α=.83; Ireland, α=.84).

Institutional trust. Participants had to indicate to which extent they trusted the following institutions (1) Parliament (UK), Dáil Éireann (Irish Parliament), Congreso de diputados (Spanish Congress); (2) The government; (3) The police (UK), An Garda Síochána (Irish Police), La policía (Spanish Police) ; (4) The legal system; (5) Political parties; (6) Scientists; (7) Doctors and other health professionals. Responses ranged from 1 (*do not trust at all*) to 5 (*completely trust*). Items 1, 2 and 5 were combined and used as a measure of *Trust in political institutions* (UK, α=.87; Ireland, α=.88; Spain α=.84) whereas items 3 and 4 were used as a

measurement of *Trust in legal institutions* (UK, α=.82; Ireland, α=.78; Spain α=.68) and with items 6 and 7 used as an indicator of *Trust in scientific institutions* (UK, α=.82; Ireland, α=.78; Spain α=.68).

Trust in sources of information. As with *Institutional trust*, participants were requested to indicate how much they trusted information from each of the following sources: (1) Newspapers, (2) Television; (3) Radio, (4) Internet websites, (5) Social media, (6) Doctors, (7) Other healthcare professionals, (8) Government agencies, and (9) Family or friends. Responses were recorded on a Likert scale ranging from 1 (*do not trust at all*) to 5 (*completely trust*). Items 1 to 3 were combined and used as a measurement of *Trust in traditional sources of information* (UK, α=.78; Ireland, α=.83; Spain α=.86) whereas items 4, 5 and 9 were used as an indicator of *Trust in informal sources of information* (UK, α=.68; Ireland, α=.66; Spain α=.65) and items 6 to 8 were used as *Trust in institutional sources of information* (UK, α=.85; Ireland, α=.81; Spain α=.69).

Facial trust detection task (FTDT; Oosterhof &Todorov, 2008) .To measure perceptual trust, data-driven computer-generated face stimuli were obtained from the University of Chicago Perception and Judgement Lab database[3]. These faces have been previously validated in terms of apparent trustworthiness ranging from 1 (not at all trustworthy) to 9 (extremely trustworthy). From this database 6 faces calibrated as more trustworthy (+ 3 and +2 SD) and 6 calibrated as less trustworthy (- 3 and -2 SD) were selected and presented in random order. Participants were asked to indicate if they trusted each face. Responses were recorded in a binary way (Yes/No) to allow us to calculate the signal detection outcomes of bias (tendency to judge a trustworthy face as untrustworthy or vice-versa) and sensitivity (perceiver's accuracy in discriminating trustworthy faces from untrustworthy ones). Computation of signal detection outcomes was

[3] https://tlab.uchicago.edu/

based on Stanislaw and Todorov (1999) calculations (equation 1 for sensitivity and 7 for bias)[4]. Further details of this test and method of scoring are available in Martinez et al. (2020).

General interpersonal trust. Participants were asked to indicate how much they agreed with the following statement "*Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people*?" on a 5 point Likert scale ranging from 1 (*Need to be very careful)* to 5 (*Most people can be trusted)*. This item has been adapted from the European Values Survey (Inglehart et al., 2000).

Neighborhood trust. Respondents were asked to rate how comfortable they were when taking the following actions (1) *"Asking a neighbour to keep a set of keys to your home for emergencies"* and (2) *"Asking a neighbour to collect a few shopping essentials for you, if you were ill and at home on your own"*. The sum of both ratings scored on a 4-point Likert scale (from 1 '*very uncomfortable*' to 4 *'very comfortable'*) was used as a measurement of neighborhood trust. Internal reliabilities were good across three countries (UK, α=.84; Ireland, α=.83; Spain α=.82) These questions were taken from the UK Community Life Survey (Cabinet Office, 2015).

2.2.3. *Statistical Analysis*

Our analyses aimed to (1) test the latent structure of paranoia and conspiracy mentality, (2) determine if the latent structure is invariant across countries, (3) determine if there are significant cross-country differences in mean levels of paranoia and conspiracy mentality, and (4) estimate the associations between trust variables and paranoia and conspiracy mentality, testing which associations are specific to one of these belief systems or the other.

[4] For more information about signal detection calculations see: Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers*, 31(1), 137-149.

The latent structure of paranoia and conspiracy was tested using confirmatory factor analysis, conducted on each of the national samples (UK, Spain, and Ireland), to compare two alternative models: Model 1, in which all the paranoia and conspiracy items loaded on a single factor (testing if conspiracy and paranoia represent the same underlying construct) and Model 2 in which paranoia and conspiracy were treated as separate but correlated factors. Once the best fitting model was determined, the cross-country invariance of the latent structure was tested using a multi-group model. Three, increasingly restrictive, levels of measurement invariance was tested; configural invariance which tests for the same factor structure, metric invariance where the factor loadings were constrained to be equal across the countries, and scalar invariance that tests for the equality of intercepts. Configural and metric invariance was evidenced if the fit of the models were acceptable, and the difference in fit between them was negligible. A criterion of -.01 change in CFI and changes in RMSEA of .015 and SRMR of .030 have been suggested as criteria to evaluate measurement invariance (Putnick & Bornstein, 2016). To test for scalar invariance (differences in intercepts) a Multiple Indicator – Multiple Cause (MIMIC) model was specified to test for Differential Item Functioning (DIF). Scalar invariance would be evidenced by the degree of DIF. Hence, a MIMIC model was specified using dummy coded nation variables (with UK nation as reference) as predictors of the paranoia and conspiracy mentality latent variables to determine if there were significant country differences in paranoia and conspiracy mentality. Then, a baseline model was defined where each direct path between the observed items and dummy-coded country variables were constrained to zero. DIF is the presence of 'significant' direct effects from the dummy coded variables to the observed items. As proposed by Kaplan (1989) a combination of modification indices (MI) and standardized expected parameter change (SEPCs), with values higher than 10 and greater than 0.20 respectively, were used to determine which direct effects should be freely estimated in the model. Thus, the path with the greatest MI/SEPC was freely estimated in the

model and then the model was re-estimated until there were no MI and SEPC values higher than 10 and 0.20.

To estimate the associations between trust variables and paranoia and conspiracy, all of the trust variables were added to the free estimated model as predictors of the latent variables. Given our large sample size and to assist with the interpretation of practical significance, we reported semi-partial correlations (*sr*) as they reflect the specific effect of each predictor variable on the dependent variable (Abdi, 2007; Dudgeon, 2016). Cohen's (1992) criteria was used to interpret the magnitude of the effect, with *sr* values of ≥ 0.50, ≥ 0.30, ≥ 0.10, ≤ 0.09, considered large, medium, small and trivial respectively. To test for specificity, equality constraints were tested using Wald tests, to determine if regression coefficients were significantly different.

All analyses were carried out in R 4.0.4 using the *lavaan* package *cfa* function (Rosseel, 2012) for conducting confirmatory factor analyses and for assessing measurement invariance (configural and metric), the *modindices* function for assessing MI and SEPC values, the *sem* function for calculating the regression models and the *lavTestWald* function for calculating equality of constraints between regression coefficients. The *fastDummies* package *dummy cols* function (Kaplan, 2020) was used for coding nation as dummy variables whereas the *ppcor* package *spcor* function (Kim, 2015) was used for calculating *sr* between predictor and outcome variables.

We report seven goodness of fit indices: the chi-square test; the Comparative Fit Index (CFI; Bentler, 1990); Tucker–Lewis Index (TLI; Tucker & Lewis, 1973); Root Mean Square Error of Approximation (RMSEA; MacCallum et al., 1996); Standardized Root Mean Squared Residual (SRMR; Hu & Bentler, 1999); the Bayesian Information Criterion (BIC; Schwarz, 1978) and Akaike Information Criterion (AIC; Akaike, 1987). Non-significant chi-square

values, CFI and TLI values above .90, RMSEA and SRMR values smaller than .08, and lower AIC and BIC values were considered indicators of good model fit.

2.3. Results

Fit indices for the confirmatory factor analysis model are shown in Table 2.2., supporting a two-factor model over a one-factor model in each of the national samples. Nonetheless, while most of the fit indices are within the cut-off criteria, it should be noted that the RMSEA values vary between .08 and .10 indicating that model fit is neither good nor bad (Hu & Bentler, 1999); hence these values should be interpreted with caution.

Measurement invariance results shown in Table 2.3. confirmed that the two-factor model was supported across the three national samples. Standardized paths from dummy-coded country variables to paranoia revealed significant differences in the factor means for UK and Spain ($\beta = -.388$, $p < .001$) and for UK and Ireland ($\beta = -.103$, $p = .01$). Likewise, standardised regression coefficients between dummy-coded country variables and conspiracy mentality were significant for UK and Spain ($\beta = .613$, $p < .001$) and for UK and Ireland ($\beta = .147$, $p < .001$).

Results shown in Table S2.1. reflect mean latent variable differences between nations regarding conspiracy mentality (higher in participants from Spain in comparison to participants from UK and Ireland) and paranoia (higher in participants from the UK in comparison to participants from Spain and Ireland). These values are in line with previously reported research with the same measures (Melo et al., 2009; Al-Suhibani et al., 2022; Bruder et al., 2013; Đorđević et al., 2021; for more information see Table S2.2.).

Table 2. 2. *Model fit Paranoia and Conspiracy one and two factor model*

Samples	Model	χ^2	df	*p*	CFI	TLI	AIC	BIC	RMSEA	SRMR
UK	1	4807.312	35	<.001	.496	.352	76292.781	76405.048	.259	.215
	2	751.458†	34	<.001	.924†	.900†	72238.927†	72325.807†	.102†	.055†
Spanish	1	4603.076	35	<.001	.455	.299	70563.720	70675.242	.259	.227
	2	705.211†	34	<.001	.920†	.894†	66667.854†	66784.952†	.101†	.057†
Irish	1	3851.702	35	<.001	.504	.362	36161.009	36258.493	.236	.196
	2	320.385†	34	<.001	.925†	.900†	34558.715†	34661.073†	.093†	.050†

Note † indicates better model fit. *CFI (Comparative Fit Index), TLI (Tucker-Lewis Index), AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion),RMSEA (Root Mean Square Error Approximation), SRMR (Standardized Root Mean Squared Residual). Model 1 (items load onto one factor) Model 2 (items loads onto two different factors).*

Results from DIF analysis revealed the greatest MI/SEPC values corresponded to the direct path between the dummy-coded variable representative of Ireland and the third item of the conspiracy mentality questionnaire (…*government agencies closely monitor all citizens*; MI = 428.063, SEPC = .234). The model was run again with this path freely estimated revealing that the path between the dummy-coded variable representative of Ireland and the fourth item of the conspiracy mentality questionnaire exhibited the largest MI/SEPC values (…*events which superficially seem to lack a connection are often the result of secret activities*, MI = 375.601, SEPC = .194). Once the model was run again with this path freely estimated there were no longer large MI/SEPC suggestive of adding new free parameters to the model. The two paths freely estimated in the final model were statistically significant (item 3, $\beta = .1768$, $p < .001$; item 4, $\beta = .1197$, $p < .001$) nonetheless R-squared difference before and after the inclusion of these freely estimated paths was small accounting for 6% of the variance for item 3 (from .429 to .487) and 1.3% for item 4 (.716 to .729). Fit statistics for these DIF models are presented in Table S2.3.

Table 2. 3. *Fit of measurement invariance (configural and metric) of two factor model*

	χ^2	df	CFI	TLI	AIC	BIC	RMSEA	SRMR	Δ χ^2	Δ df	ΔCFI	ΔRMSEA	ΔSRMR
Configural	1689***	102	.923	.898	167807	168409	.100	.050	-	-	-	-	-
Metric	1893***	118	.914	.902	167978.9	168477.3	.098	.058	204***	16	.009	.002	.008

Note. *p<.05, **p<.01,***p<.001. *CFI (Comparative Fit Index), TLI (Tucker-Lewis Index), AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion),RMSEA (Root Mean Square Error Approximation), SRMR (Standardized Root Mean Squared Residual). Δ indicates fit indices and degrees of freedom difference between Configural and Metric models.*

Bivariate and semi-partial correlation coefficients, standardized regression coefficients, and Wald test statistics are shown in Table 2.4. A significant, small to medium effect was found between mistrust in political institutions and conspiracy mentality (*sr*= -.19) whereas the effect between mistrust in political institutions and paranoia, although significant, was very small (*sr*= -.06). This difference was reflected in a significant Wald test showing that the association between mistrust in political institutions and conspiracy mentality was stronger than the association of the same predictor with paranoia. Examining the associations between the rest of the institutional trust predictors (scientific and legal) and paranoia and conspiracy mentality, we found significant but very small effects (*sr*< .08).

Regarding the associations between trust in sources of information and the two belief systems, significant but small positive effects were found between trust in informal sources of information and conspiracy mentality (*sr* =.11) and also paranoia (*sr* =.16). However, a significant Wald test revealed that the association between trust in informal sources of information and paranoia was stronger than the association between the same trust predictor and conspiracy mentality. On the other hand, the associations between trust in other sources of information (traditional media, institutional sources) and paranoia and conspiracy mentality, although significant, were very small (*sr* < .09).In the case of perceptual trust, a significant relationship was found between a bias towards judging face stimuli as untrustworthy and paranoia (*sr* = -.12), but not conspiracy mentality.

Table 2. 4. *Bivariate correlations, semi-partial correlations, standardized regression coefficients and Wald tests of equality of constraints from multivariate regression model predicting paranoia and conspiracy beliefs for whole sample controlling for nation.*

Predictor Variables		Conspiracy			Paranoia			Wald
		r	*sr*	β (se)	*r*	*sr*	β (se)	
Institutional Trust	Political	-.34***	-.19***	-.204(.010) ***	-.04*	.06***	.059 (004) ***	154.564***
	Scientific	-.007	.07***	.071(.016) ***	-.24***	-.06***	-.066(.007)***	29.704***
	Legal	-.20***	-.05***	-.081(.015) ***	-.19***	-.05***	-.074(.007)***	5.909*
Trust Sources	Media	-.12***	-.05***	-.083(.014) ***	-.03*	-.07***	-.094(.006) ***	4.544*
	Informal	.13***	.11***	.196 (.015) ***	.08***	.16***	.221(.007) ***	25.587***
	Institutional	-.09***	-.05***	-.094(.014) ***	-.16***	-.08***	-.104(.006) ***	6.546*
Perceptual Trust	Bias	-.09***	-.007	-.025(.027)	-.25**	-.12***	-.138(.012) ***	5.712*
	Sensitivity	.01	.01	.015(.022)	-.004	.01	.015(.010)	0.475
Interpersonal Trust	General	-.19***	-.12***	-.160 (.024) ***	-.35***	-.22***	-.272(.011) ***	5.393*
	Neighbourhood	-.05***	.05***	.057(.012) ***	-.16***	-.06***	-.088(.006)***	41.7615***
Nation (UK)	D1 (Spain)	.26***	.12***	-.164 (.061) ***	-.13***	-.14***	-.227(.028) ***	189.361***
	D2 (Ireland)	-.07***	.06***	-.055(.067) ***	.04**	-.01*	-.040(.028) **	4.582*
R^2	.239				.257			

Note. *p<.05, **p<.01,***p<.001. Bolded values represent practical significant coefficients based on semi partial correlations (*sr*).

However, no significant associations were found between both belief systems and the sensitivity of trust/mistrust judgments.

When considering interpersonal trust, a significant but small negative association was found with conspiracy mentality (*sr* = - .12) whereas a higher significant negative association was found with paranoia (*sr* = - .22), indicating low levels of interpersonal trust in relation to both types of beliefs; a significant Wald test indicated that the latter association was stronger than the former. Finally, regarding neighbourhood trust, a significant positive association was found with conspiracy mentality (people higher in conspiracy mentality trusted their neighbours more) but a significant negative association with paranoia was also found. The Wald test revealed that this difference was significant but both effects were very small (*sr* = .05; *sr* = - .06, respectively).

2.4. Discussion

In this study, we first examined if paranoid beliefs and conspiracy mentality were two separate but correlated phenomena and found that, as expected, a two-factor model was superior to a single-factor model in three large representative nation samples. These findings are in line with those of Imhoff and Lamberty (2018) and Alsuhibani et al (2022). However, whereas several studies have reported moderate to high associations between conspiracy mentality and paranoid ideation (Barron et al., 2014; Brotherton & Eser, 2015; Bruder et al., 2013; Cichocka et al., 2016; Darwin et al., 2011; Grzesiak-Feldman & Ejsmont, 2008) the correlation we observed (*r*=.11) between these constructs was much smaller. Thus, while the two-factor model provided optimal fit, the correlation between the factors strongly points towards two, distinct factors. In this context, we note that our sample was much larger, more international, and more representative of the participating nations than any hitherto study (most of the aforementioned studies used student or convenience samples). Moreover, some of these earlier studies measured paranoia within the context of schizotypal traits which are usually

regarded as an expression of a latent psychopathological entity (e.g., schizophrenia spectrum disorder; American Psychiatric Association, 2013). Although the authors of these articles do not regard conspiracy theories as a reflection of an underlying psychopathology, they consider that certain psychopathological traits might facilitate the belief in conspiracy theories (Barron et al., 2014; Darwin et al., 2011; Swami et al., 2013). Contrary to this view, authors such as the philosopher Quassim Cassam (2019) argue that studying conspiracy theories from an individual differences perspective fails to address one of the most important features of these theories, which is that they are often politically motivated. From this perspective, conspiracy theories can be thought of as ideologies (i.e., set of ideas and beliefs) that structure the understanding of the political world, and thus considering them as a trait of an underlying psychopathology underestimates the social harm that can cause (Cassam, 2019). In future research, it would be useful to compare how paranoia and conspiracy theories relate to political psychology variables, for example authoritarianism, collective mistrust, mistrust to specific outgroups, mistrust to political figures and authorities, which we anticipate would be associated with the latter and not the former.

As a second aim, we explored the association between different types of trust variables and conspiracy mentality and paranoid beliefs. Our expectations regarding conspiracy mentality were partially met as we found an association between conspiracist thinking and mistrust in political institutions whereas associations with trust in scientific or legal institutions were trivial. Moreover, whereas conspiracy mentality was related to trust in informal sources of information (i.e., friends, social media) associations with trust in traditional and institutional sources of information were very weak. On the other hand, our expectations in relation to paranoia were supported by our findings, as higher paranoid beliefs were associated with a bias towards judging unfamiliar faces as untrustworthy. Furthermore, interpersonal mistrust was associated with both conspiracy and paranoia, with Wald tests revealing that this relationship

was stronger for the latter. Finally, and surprisingly, we found that paranoia, more than having a conspiracy mentality, was associated with trust in informal sources of information.

Some of our findings appear to be in line with previous research whereas others seem less consistent, particularly in relation to conspiracy mentality. Authors such as Pierre (2020) have suggested that conspiracy theories are the consequence of epistemic mistrust, which is defined as mistrust of knowledge from authoritative sources (e.g., government). When disregarding information from well-established institutions, conspiracy theorists look for alternative explanations of events (Abalakina-Paap et al., 1999; Meuer & Imhoff, 2021; Pierre, 2020). Following this line of thought, authors such as Hartman et al (2021) argue that circumstances related to global uncertainty, for example the Covid-19 pandemic, are particularly likely to give rise to conspiracy theories, although different specific conspiracy theories may be associated with different specific factors. For example, various studies have shown that distrust in governments is usually associated with general conspiracy theories (Goertzel, 1994; Mari et al., 2021; Pierre, 2020) while distrust in scientific or other institutions tends to be associated with specific conspiracies, for example about COVID-19 or HIV (Ball et al., 2013; De Coninck et al., 2021; Hartman et al., 2021). Our findings are consistent with this previous research, as mistrust in political institutions (i.e., government, political parties and parliament) was associated with conspiracy mentality but the relationship between conspiracist thinking and mistrust in scientific and legal institutions was trivial. Although some studies have found that conspiracy ideation is related with mistrust in traditional media (Freeman et al., 2020; van der Linden et al., 2020) others have reported no association at all (De Coninck et al., 2021). Our results are aligned with those of those of De Coninck et al (2021) who found that conspiracy mentality was associated with relying on social media and personal contacts when gathering information about COVID-19.

Conspiracy theories are generally transmitted through social networks, online blogs and social media use (De Coninck et al., 2021; Enders et al., 2021; Mari et al., 2021; Parsons et al., 1999; Stempel et al., 2007) and it is therefore unsurprising that someone with a conspiracy mentality would be more likely to trust their own sources of information. However, paranoia often occurs in the context of loneliness (Alsubibani et al, 2022), social isolation (Butter et al., 2017), and insecure attachment (Wickham et al., 2014) so a positive association with trust in informal sources of information is surprising in the case of this kind of belief. Possibly, people with high paranoia tend to trust their own personal sources of information given the potential anxiety elicited by face to face contact with others outside of family and friends circles. More studies are needed that includes measurements of social anxiety, social isolation, attachment and size of social networks in order to test these alternative explanations to our findings. In our study, the associations between both conspiracy mentality and paranoia and mistrust in official sources of information was very small. Previously, one multi-study article has reported moderate to strong correlations between paranoia and this kind of mistrust (e.g., government, mainstream media, scientists; van der Linden et al., 2020). The authors of this article argue that distrust in formal sources of information and paranoia are mechanism that explain the association between political conservatism and conspiracy thinking, supporting Hofstadter's (1962) idea of the paranoid style in American politics (van der Linden et al., 2020). However, it has been previously argued that Hofstadter misunderstood the nature of paranoia, and that what he was really referring to was conspiracist thinking (Alsuhibani et al., 2022).

Interpersonal mistrust was also associated with conspiracy mentality but not as strongly as with paranoia, a finding which is aligned with previous studies of conspiracy theories (Abalakina-Paap et al., 1999; Goertzel, 1994;Green and Douglas, 2018; Meuer & Imhoff, 2021) and paranoia (Axelrod et al., 1997; Furnham & Crump, 2015; Greenaway et al., 2019; Kramer, 1994; Kong, 2017; Murphy et al., 2012; Wickham et al., 2014). By contrast, a bias

towards judging unfamiliar faces as untrustworthy was specifically related to paranoid beliefs, replicating the finding of Martinez et al (2020). In dangerous and uncertain contexts, a tendency to classify social cues as untrustworthy would be an evolutionary adaptive strategy, as it would enable individuals to avoid the highly costly consequences of underestimating threat (Haselton & Buss, 2000; Haselton & Nettle, 2006). This bias is consistent with current models of paranoia, which assume that the expectation of harm from others is its core feature (Hooker et al., 2011; Kirk et al.,2013; Martinez et al., 2020). Consistent with this account, studies using game theory paradigms have found that paranoia is associated with an increased tendency to make attributions of harmful intent (Greenburgh et al., 2019; Barnby et al., 2020), leading researchers to argue that people with paranoid beliefs are prone to make these kinds of attributions when they experience marked uncertainty about the world (Reed et al., 2020; Lebert et al., 2021). Conversely, a recent study by Meuer and Imhoff, (2021) failed to provide evidence that conspiracy theories were related to the detection of social threat. Thus, our findings support the view that paranoia reflects self-relevant concerns as opposed to conspiracy theories which involves societal ones (Imhoff & Lamberty, 2018).

2.4.1. *Limitations*

This study has some limitations. Firstly, our design is cross-sectional, which limits our ability to make causal claims. Future studies could benefit from employing experimental or longitudinal designs when studying causal relationships between trust processes and paranoid and conspiracy mentality. Moreover, the use of computer-generated face stimuli for measuring trustworthiness of unfamiliar faces might lack ecological validity and generalizability, as the face stimuli were male, bald and Caucasian. This lack of ecological validity can also be applied to the self-report instruments used to measure interpersonal trust, as they cannot tap into the social-interactive nature of interpersonal trust processes. Thus, future studies would benefit from including socially interactive measures such as game theory paradigms or virtual reality

scenarios. Finally, this data was collected during the very beginning of the first national lockdown during the first wave of the COVID-19 pandemic (see Table 2.1.). Since paranoid and conspiracy beliefs are thought to be heightened during threating and uncertain circumstances (Freeman, 2007; van Prooijen & Jostmann, 2013) it is possible that these beliefs were affected by the global impact the pandemic had at a societal and personal level.

We found that mean levels of conspiracy and paranoid beliefs differed significantly between the three nation samples. Paranoid levels were higher in UK and Ireland than those reported in Spain. Conversely, the UK and Ireland reported lower levels of conspiracy endorsement than Spain (see Table S2.1.). These differences might be due to various factors, such as degree of restriction enforcement by governmental institutions, perception of personal threat caused by the virus and habituation to the situation. The Spanish population had been quarantined for almost three weeks when the participants from that country completed the survey. Thus, it is possible that, in Spain at that time, the virus was not as threatening to individuals as COVID restrictions were, fuelling feelings of discontent towards the government. In contrast, UK and Irish participants completed the survey shortly after lockdown was announced and thus the threat from the virus might have been associated with a high level of interpersonal vulnerability. Another interpretation of these findings could be in terms of cultural differences between individualistic (UK, Ireland) and collectivist (Spain) countries. Whereas the former might orientate external threats to the self the latter might focus those threats to society, resulting in higher conspiracy mentality in Spain and higher paranoid beliefs in Ireland and UK. However, this interpretation is highly speculative, further research would be required to test it, and future studies should therefore consider including instruments measuring cultural variables

2.4.2. *Conclusions*

This study employed large representative samples from three different countries allowing us to conduct high-powered statistical tests as well as to generalize our results to the general population. Our findings show that paranoid beliefs and conspiracy mentality are two related but separated constructs and that this relationship between the two constructs did not differ between countries. Moreover, we found that conspiracy mentality and paranoid beliefs have shared and different trust predictors, with mistrust in political institutions being specifically associated with conspiracy mentality whereas a bias towards mistrust was uniquely associated with paranoid beliefs. Interpersonal mistrust and trust in informal sources of information were related with both paranoia and conspiracy mentality although the association of both trust predictors was stronger with paranoid beliefs. Our findings clarify the role of different trust mechanism in the two belief systems. Whereas conspiracy beliefs are conceptualized as ideologies which serves a political function as a response to social vulnerability (Abalakina-Paap et al., 1999; Miller, et al.,2016; Cassam, 2019) paranoid beliefs reflects a cognitive structure that underlies personal vulnerability in which others are regarded as threatening. These findings may potentially lead to a better conceptualization of paranoia, which will lead to more accurate assessment and targeted interventions in clinical settings for example, by focusing specifically on interpersonal and perceptual mistrust. Further research is needed in order to replicate these findings and to establish the causal processes that are responsible for the associations reported here.

Chapter III - Second Empirical Study

Mistrust and negative self-esteem: Two paths from attachment styles to paranoia

Contributions

Anton P. Martinez (conceptualization, formal analysis, methodology, writing – original draft)

Maximilian Agostini (formal analysis, methodology, writing – original draft)

Azzam Al-Suhibani (conceptualization, funding acquisition, writing – review & editing)

Richard P. Bentall (conceptualization, supervision, – writing original draft, review & editing)

The following chapter has been peer-reviewed and published at Psychology and Psychotherapy: Theory, Research and Practice

Martinez, A. P., Agostini, M., Al-Suhibani, A., & Bentall, R. P. (2021). Mistrust and negative self-esteem: Two paths from attachment styles to paranoia. *Psychology and Psychotherapy: Theory, Research and Practice*, *94*(3), 391-406.

Findings from this chapter have been presented as an open paper at the European Association of Behavioural and Cognitive Therapies (EABCT):

Martinez, A. P. (2021). Mistrust and negative self-esteem: Two paths from attachment styles to paranoia. *EABCT*, Belfast, Northern Ireland.

Abstract:

Objectives: Paranoia is known to be associated with insecure attachment, with negative self-esteem as a mediator, but this pathway is insufficient to explain the paranoid individual's beliefs about malevolent others. Mistrust is a likely candidate as it is a core feature of paranoid thinking also associated with insecure attachment styles. In this study, we tested whether mistrust - operationalized as judgments about the trustworthiness of unfamiliar faces - constitutes a second pathway from insecure attachment to paranoia. *Methods*: A nationally representative British sample of 1508 participants aged 18- 86, 50.8% female, recruited through the survey company Qualtrics, completed measurements of attachment style, negative self-esteem, and paranoid beliefs. Usable data was obtained from 1121 participants. Participants were asked to make trustworthiness judgments about computer-generated faces and their outcomes were analyzed by conducting signal detection analysis, which provided measures of bias (the tendency to assume untrustworthiness in conditions of uncertainty) and sensitivity (accuracy in distinguish between trustworthy and untrustworthy faces). The design of the study was cross-sectional. *Results*: Results using structural equation modelling revealed a good model fit (*RMSEA* = 0.071, 95%CI: 0.067 - 0.075, *SRMR* = .045, CFI = .93, TLI = .92). We observed indirect effects through bias towards mistrust both for the relationship between attachment anxiety and avoidance (β = 0.003, *95%CI:* 0.001 - 0.005, p <.001) as well as attachment anxiety and paranoia (β = 0.003, *95% CI* 0.002 - 0.006, p < .001). We observed an indirect effect through negative self-esteem only for the relationship between attachment anxiety and paranoia (β = 0.064, *95% CI:* 0.053 - 0.077, p <.001). Trust judgments and negative self-esteem were not associated with each other. *Conclusions*: We find that a bias towards mistrust is associated with greater paranoia. We also find indirect effects through bias towards mistrust between attachment styles and paranoia. Finally, we reaffirm the strong indirect effects

through negative self-esteem between attachment anxiety and paranoia. Limitations of the study are discussed.

Keywords: *Signal detection, attachment, paranoia, self-esteem, mistrust, response bias*

Practitioner Points

- When working with individuals suffering from paranoia, clinicians should consider not only explicit, deliberative cognitive processes of the kind addressed in cognitive behaviour therapy (e.g. cognitive restructuring) but also the way in which their patients make perceptual judgments (e.g. their immediate reactions on encountering new people) by for example employing bias modification training.
- Assessment and clinical interventions for people should consider the role of trust judgments and the way in which they combine with low self-esteem to provoke paranoid beliefs.
- Psychological interventions targeting paranoid beliefs should focus on both attachment anxiety and attachment avoidance.

3.1. Introduction

Paranoid delusions are the most common symptom of psychosis (Bentall et al., 2001) and can be defined as unfounded beliefs characterized by a high degree of conviction, preoccupation and distress in which the core theme includes intentional harm to the person who is holding the belief (American Psychiatric Association, 2013). However, less severe forms of paranoid beliefs are also experienced by at least 10-15% of the general population (Freeman, 2007a) suggesting that clinical paranoia lies on a continuum with subtle subclinical forms (Bebbington et al., 2013; Elahi et al., 2017). These lesser forms of paranoid beliefs, although held with less conviction in comparison to clinical paranoid delusions, are still associated with distress, social isolation and feelings of powerlessness (Freeman et al., 2005). By understanding the precursors of paranoid beliefs in the general population, we may therefore gain an understanding of mechanisms that may be responsible for more severe forms in clinical groups.

Consistent with the hypothesis that attachment processes may play a role in the development of paranoid beliefs (Bentall & Fernyhough, 2008), many studies have reported a strong association between insecure attachment styles and paranoia in both clinical and non-clinical samples (Carr et al., 2018; Gumley et al., 2014; Pickering et al., 2008; Ringer et al., 2014; Sitko et al., 2014; Wickham et al., 2015). These styles can be conceptualised as internalised representations of relationships with primary caregivers that take the form of working models of the self and others which in turn guide interpersonal behaviour (Bowlby, 1982; Shaver & Mikulincer, 2005). In adulthood, attachment styles can be secure or insecure, the former reflecting confidence in the availability of attachment figures, the latter reflecting the contrary. Two underlying dimensions of attachment security/insecurity in adults are anxiety and avoidance (Mikulincer, 1995). Attachment avoidance is associated with insecurity about other's intentions, preference for emotional distance, and a negative view of others; attachment

anxiety reflects negative self-image, fear of rejection and excessive need of approval (Berry et al., 2007; Mikulincer, 1995).

Several cross-sectional studies have reported that the association between insecure attachment styles – particularly attachment anxiety – and paranoia is mediated by negative self-esteem (Pickering et al., 2008; Ringer et al., 2014; Wickham et al., 2015). In longitudinal (experience sampling) studies, paranoid symptoms are predicted by fluctuations in self-esteem (Thewissen et al., 2008) and attachment-related cognitions (Sitko et al., 2016). Self-esteem involves the evaluation of attributes of the self (Hahn & Gawronski, 2015) which are influenced by internal working models about the self and others (i.e., attachment styles; Sitko et al., 2016). Hence, negative internal working models lead to negative evaluations about the self which in turn lead to feelings of vulnerability and the anticipation of social threats (Bentall & Fernyhough, 2008). However, based on the observation that the mediating role of self-esteem is also present in the association between attachment-anxiety and depression symptoms (Lee & Hankin, 2009; Roberts et al., 1996), the relationship between insecure attachment and paranoia would seem to require the involvement of additional factors. Mistrust is a likely candidate for this relationship, as it is a subcomponent of the paranoia spectrum often present in sub-clinical populations (Bebbington et al., 2013; Bell & O'Driscoll, 2018) and insecure attachment styles are associated with reduced interpersonal trust (Fett et al., 2016; Mikulincer, 1995, 1998).

3.1.1. *The role of trustworthiness judgments in paranoid beliefs*

During everyday life, most people meet numerous persons and judgments of trustworthiness have to be made rapidly without effortful deliberation (Sutcliffe et al., 2012). Trustworthiness judgements are a dominant mode of appraisal when encountering unfamiliar faces (Oosterhof & Todorov, 2008). Healthy individuals typically make judgments of trust very quickly – within a few hundred milliseconds – and consistently rate some unfamiliar faces as

less trustworthy than others (Todorov et al., 2009). However, studies of trustworthiness judgments in patients with a schizophrenia diagnosis (for which paranoia is a prominent symptom) have shown inconsistent findings. On the one hand, in one study participants diagnosed with schizophrenia judged unfamiliar face images to be more trustworthy than controls (Baas et al., 2008), potentially as a consequence of reduced social cognitive abilities (Green et al., 2008). Conversely, in another study, after being negatively primed with threat-related images, participants with a schizophrenia diagnosis judged face stimuli as more untrustworthy in comparison with non-clinical controls (Hooker et al., 2011); however this difference was not present in neutral priming conditions. Using computer generated face stimuli (Todorov et al., 2013) another study found that participants with high paranoia ideation judged unfamiliar faces as less trustworthy than those with low paranoia but that this was true for faces previously calibrated to appear trustworthy, untrustworthy or neutral, suggesting a general bias towards mistrust (Kirk et al., 2013).

3.1.2. *Signal detection theory and trustworthiness*

Signal detection theory (SDT; Swets, Dawes, & Monahan, 2000) offers a useful framework for analysing behavioural decision-making outcomes when judging trustworthiness of faces that have been previously selected to appear trustworthy or untrustworthy. There are four possible outcomes: hits (in this case, identifying a trustworthy face when the trustworthy face is present); false alarms (identifying a trustworthy face when the non-trustworthy face is present); correct rejections (identifying a non-trustworthy face when the non-trustworthy face is present); and misses (identifying a non-trustworthy face when the trustworthy face is present). These decision outcomes can be operationalised into two components (Stanislaw & Todorov, 1999): (1) *response bias* which reflects the general tendency to respond "Yes" (e.g., trustworthy) vs. "No" (e.g., non-trustworthy) and (2) *sensitivity* which mirrors the perceiver's

ability/ accuracy to discriminate between a target and a non-target stimulus (e.g., make correct vs. incorrect decisions; Lynn & Barrett, 2014).

While decreased sensitivity might indicate a deficit in information processing (Frith, 1979), response biases are sensitive to the differential costs and benefits of the different response outcomes (Correll et al., 2002; Haselton & Buss, 2000) and are hypothesized to be evolutionary adaptive strategies for minimizing the more costly type of error in uncertain and complex environments - for example perceiving trustworthy faces as non-trustworthy (Nettle & Haselton, 2006). Therefore response bias may suggest a tendency towards a liberal criterion for detecting threat or a strict criterion for detecting non-threat (Haselton & Buss, 2000). Consistent with this account, studies have shown paranoid participants adopted a liberal response bias for detecting angry faces after viewing anxiety evoking pictures, although not when not anxious (Westermann & Lincoln, 2010). Similarly, in another study participants diagnosed with schizophrenia showed a liberal response criterion when recognising fear and sad emotions in comparison to controls (Tsoi et al., 2008).

3.1.3. *Aims of the current study*

This study has two main aims: First, we will implement signal detection theory to test the association between judgments of mistrust and paranoia in non-clinical participants; we hypothesize that a response bias towards judging faces as untrustworthy, but not sensitivity to trustworthiness cues, will be associated with paranoia. Second, we aim to test whether there is an indirect effect of attachment on paranoia through mistrust bias; given that self-esteem involves the evaluation of the self (Nugent & Thomas, 1993) whereas judgements of trustworthiness involve the evaluation of others (Burns & Conchie, 2015), we hypothesize that this indirect effect will be separate and independent of the already established indirect effect through negative self-esteem. Moreover, we hypothesise that mistrust will be particularly

associated with attachment avoidance since this style involves a negative working model of others (Fett et al., 2016).

3.2. Methods

3.2.1. *Participants*

We recruited a nationally representative sample of 1,508 British participants for a multipurpose survey, age range 18 – 86 (M=47.8, SD= 17.2), 50.7% female, through the survey company Qualtrics. Participants were stratified on basis of the Office for National Statistics data by age, sex and household income[5]. Regarding ethnicity, 89% of the sample was white British/Irish whereas the remaining 11% consisted of white non-British/Irish (2.2%), Indian (1.8%), Pakistani (1.3%), Chinese (0.9%), Other-Asian (0.9%), Afro-Caribbean (0.7%), African (0.7%), Bangladeshi (0.7%) and other ethnic groups (1.8%).

An additional 344 participants were removed due to incomplete survey responses or completing the survey implausibly quickly (our pre-defined cut-off criteria based on pilot work and recommended by the survey company was 12 minutes). Differences regarding most demographics were non-significant between completers and non-completers, however age was significantly higher in the non-completers group, although this effect was small (Table S3.2.).

3.2.2. *Procedure*

After completing informed consent, participants filled out a number of measures regarding self-esteem, attachment styles and paranoid beliefs before engaging in the facial trust detection task.

[5]https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/bulletins/householddisposableincomeandinequality/financialyearending2016

3.2.3. *Measures*

The revised Paranoia and Deservedness Scale (PaDS –R) was designed on the basis of psychometric analyses of the original scale (Melo et al., 2009b) in a large sample of non-clinical individuals and patients with psychosis (Elahi et al., 2017). In line with recent findings (Bebbington, McBride, Steel, Kuipers, Radovanovič, et al., 2013), the 8 item scale measures the four elements of paranoia (two items per element): interpersonal sensitivity (e.g. "*My friends often tell me to relax and stop worrying about being deceived or harmed*"), mistrust (e.g. "*You should only trust yourself*"), fear of persecution (e.g. "*I believe that some people want to hurt me deliberately*"), and ideas of reference (e.g. "*Sometimes I think there are hidden insults in things that other people say or do*"). Items were answered on a 5-point scale ranging from *"Strongly disagree"* to *"Strongly agree".* Scale reliability was good (α = 0.87) and responses were normally distributed.

The Relationship Questionnaire (RQ) was used to assess attachment style (Bartholomew & Horowitz, 1991). Participants read four vignettes describing secure, fearful, preoccupied, and dismissing prototypical styles and had to choose the one that describes them best. They were then asked to rate each vignette "*according to how well or poorly each description corresponds to [their] general relationship style*" on 7 point scales ranging from *"Disagree strongly"* to *"Agree strongly".* Scores on the four scales were used to compute higher order measures of attachment anxiety (negative model of self) by subtracting the sum of secure and dismissing items from the sum of preoccupied and fearful items, and attachment avoidance (negative model of other) by subtracting the sum of secure and preoccupied items from the sum of dismissing and fearful items. Thus, the formula can be summed up as follow: model of self = (secure + dismissing) - (preoccupied + fearful), model of other = (secure + preoccupied) - (dismissing + fearful), where higher scores indicates the presence of each type of attachment style.

The short version of the Self-esteem rating scale (SERS; Lecomte, Corbière, & Laisné, 2006) is a 20-item scale, designed to assess self-esteem independently of mood. It consists of 10 positive statements about the self, e.g. "*I feel good about myself*" and 10 negative statements about the self, e.g. "*I feel that others do things much better than I do*". Participants rated each statement from 1, "*never*", to 7, "*always*". For both positive self-esteem (α = 0.94) and negative self-esteem (α = 0.94) scale reliability was good. However, because negative self-esteem rather than positive self-esteem has been found to be a strong predictor of paranoia in previous studies (Bentall et al., 2008) only negative self-esteem is considered in this study.

Facial trust detection task was based on the trustworthiness dataset (25 identities; Oosterhof & Todorov, 2008). This dataset contains computer-generated faces created using FaceGen 3.1 and obtained from the Princeton Social Perception Lab database[6]. The database includes identities manipulated on different traits (attractiveness, competence, dominance, extroversion, likeability, threat, and trustworthiness). From this data set, 10 bald Caucasian male computer-generated faces (5 prior rated as trustworthy and 5 prior rated as untrustworthy; see Figure 3.1.) were randomly selected by using the website www.Random.org. Participants were presented with each face followed by a fixation cross and were asked: "*How much would you trust this person*". Answers were given on a 7-point Likert scale (1 = *"I would not trust this person at all*" to 7 = *"I would trust this person completely"*). Reliabilities were good for the overall scale (α = .94), the trustworthy items (α = .93), and the untrustworthy items (α = .93). In the sample as a whole, the trustworthy faces were rated as more trustworthy (M = 20.10, SD = 5.58) than the untrustworthy faces (M = 16.49, SD = 5.62, t = 30.56, p < .001, showing that the faces were clearly discriminable). However, mean ratings for the two types of faces were correlated (r = .66, p < .001) suggesting individual differences in trust judgments.

[6] http://tlab.princeton.edu/databases/secretdatabaseportal/

3.2.4. *Signal Detection Analysis*

In order to analyse facial trust detection task outcomes using a signal detection paradigm, participant's responses were recoded as binary outcomes (a total of 15,080 trials for all participants) by recoding responses from 1 to 3 as a *NO* outcome (5,603 trials for all participants), responses from 5 to 7 as a *YES* outcome (3,524 trials for all participants), and responses of 4 as null outcomes (5,953 trials for all participants), so that positive values reflect bias towards trustworthiness. The amount of null responses resulted in a loss of 387 participants for which we were unable to calculate signal detection variables due to the inability of distinguishing yes and no responses. This meant our final sample size was reduced to 1121 participants.

Figure 3. 1. *Left image example of a trustworthy computer-generated face. Right image example of an untrustworthy computer-generated face (Images obtained from the Princeton Social Perception Lab database).*

Signal detection outcomes were analysed based on Stanislaw & Todorov's (1999) calculations using equation 7 for response bias: $c = \frac{\Phi^{-1}(\mathrm{H})+\Phi^{-1}(\mathrm{F})}{2}$ where Φ^{-1} ("inverse phi") function converts hit (H) rates (dividing the number of hits by total number of signal trials) and false-alarm (F) rates (dividing the number of false-alarms by total number of noise trials) into z

scores. This measure evaluates whether people have a bias towards pressing the "trust" or "not trust" button. Sensitivity calculations were based on Stanislaw and Todorov's equation 1: $d' = \Phi^{-1}(H) - \Phi^{-1}(F)$ with adjustments for potential assumption violations[7] (see Stanislaw & Todorov, 1999 for details). The sensitivity measure evaluates whether people are accurate in identifying whether faces are trustworthy or non-trustworthy; hence, it is a measure of sensitivity to trustworthiness cues. As we were also interested in specific errors on judgment that trustworthy faces were non-trustworthy (misses) and that non-trustworthy faces were trustworthy (false alarms), these were also separately recorded and reported. Although sensitivity and response bias measures are superior measures, false alarms and misses are inherently easier to interpret.

3.2.5. *Statistical Analysis*

Analyses were carried out in R 1.1.463, using CAR package *lm* function for linear regression and *cor* for correlations (Fox & Weisberg, 2019; Hlavac, 2018; R Core Team, 2018; RStudio Team, 2016). Our first aim was to assess if signal detection variables (sensitivity and bias) correlated with paranoia. To test our second, indirect effects hypothesis, we implemented a structural equation model in AMOS 25.0.0. For this model, attachment avoidance and attachment anxiety were modelled as independent variables, negative self-esteem and trust judgment response bias were included as indirect effect/ mediator variables, and paranoia was the outcome variable. We modelled negative self-esteem and paranoia as latent constructs and trust and attachment as observed variables since paranoia and self-esteem were measured through items tapping into their respective latent constructs. Attachment styles and signal detection variables were measured directly either through vignettes or the face rating task and were therefore modelled as observed variables (see Figure 3.2). Following suggestions of Kline

[7] *Extreme values (0s and 1s) were adjusted following using the approach of Stanislaw and Todorov, (1999; see p.144) in which rates of 0s are replaced with 0.5/n, and rates of 1s are replaced with (n-0.5)/n, where n is the number of signal or noise trials.*

(2015) we report five goodness of fit indices: the Chi-Square test; Root Mean Square Error of Approximation (RMSEA; MacCallum et al, 1996); Standardized Root Mean Squared Residual (SRMR, note: inflated with large sample sizes; L. Hu & Bentler, 1999); the Comparative Fit Index (CFI), and Tucker Lewis Index (TLI). We also report bootstrap bias-corrected confidence intervals (based on 5,000 bootstrap samples) to avoid problems of non-normal data.

3.3. Results

3.3.1. *Association between paranoia and trustworthiness judgments*

Correlational analyses showed that there were only trivial associations between self-esteem and the face judgment measures (r=-.09). Higher levels of paranoia were positively related to misses (i.e., judging a face as untrustworthy when the target is trustworthy). Higher levels of paranoia were also negatively associated with false alarms (i.e., judging a face stimulus as trustworthy when it is non-trustworthy). Using signal detection measures, there was a stronger association between paranoia and response bias (r = -.20) than between paranoia and sensitivity (r = -.10; z (1508) = 3.10, p =.002; Lee & Preacher, 2013) although, contrary to expectation, the latter association was significant.

Table 3. 1. *Bivariate correlations between main variables with (M) means and (SD) standard deviations*

Variables	*M*	*SD*	1	2	3	4	5	6	7	8
1. Paranoia	2.78	1.00	-	.41**	.14**	.74**	-.10**	-.20**	-.13**	.21**
2. Attachment Anxiety	-.87	3.50		-	-.17**	.39**	-.03	-.13**	-.10**	.12**
3. Attachment Avoidance	.85	3.57			-	.05	-.04	-.14**	-.11**	.13**
4. Negative Self-Esteem	3.35	1.43				-	-.09**	-.09**	-.03	.11**
5. Sensitivity	.67	.22					-	.19**	-.38**	-.61**
6. Response Bias	-.31	1.15						-	.83**	-.88**
7. False Alarms	.026	.33							-	-.47**
8. Misses	.44	.40								-

These findings were confirmed using regression analyses. Sensitivity and response bias (both centred) predicted a significant proportion of variance in paranoia, R^2_{adj} = .04, F(2, 1118) = 26.30, p < .001. Higher sensitivity, *b* = -.30 95%-CI [-0.57, -0.04], t(1118) = -2.25, p = .02, and a stronger response bias towards trustworthiness, *b* = -.16 95%-CI [-0.21, -0.11], t(1118) = -6.36, p < .001, predicted lower paranoia scores. It is important to note that, because sensitivity and bias are measured on different scales, this finding does not imply a greater effect for sensitivity. When standardized for comparison, we see that the response bias, *b* = -.19, 95%-CI [-0.25, -0.13], t(1118) = -6.36, p < .001, is in fact a stronger predictor of paranoia than sensitivity, *b* = -.07, 95%-CI [-0.13, -0.009], t(1118) = 2.25, p =.02.[8]

3.3.2. *Indirect effects analysis*

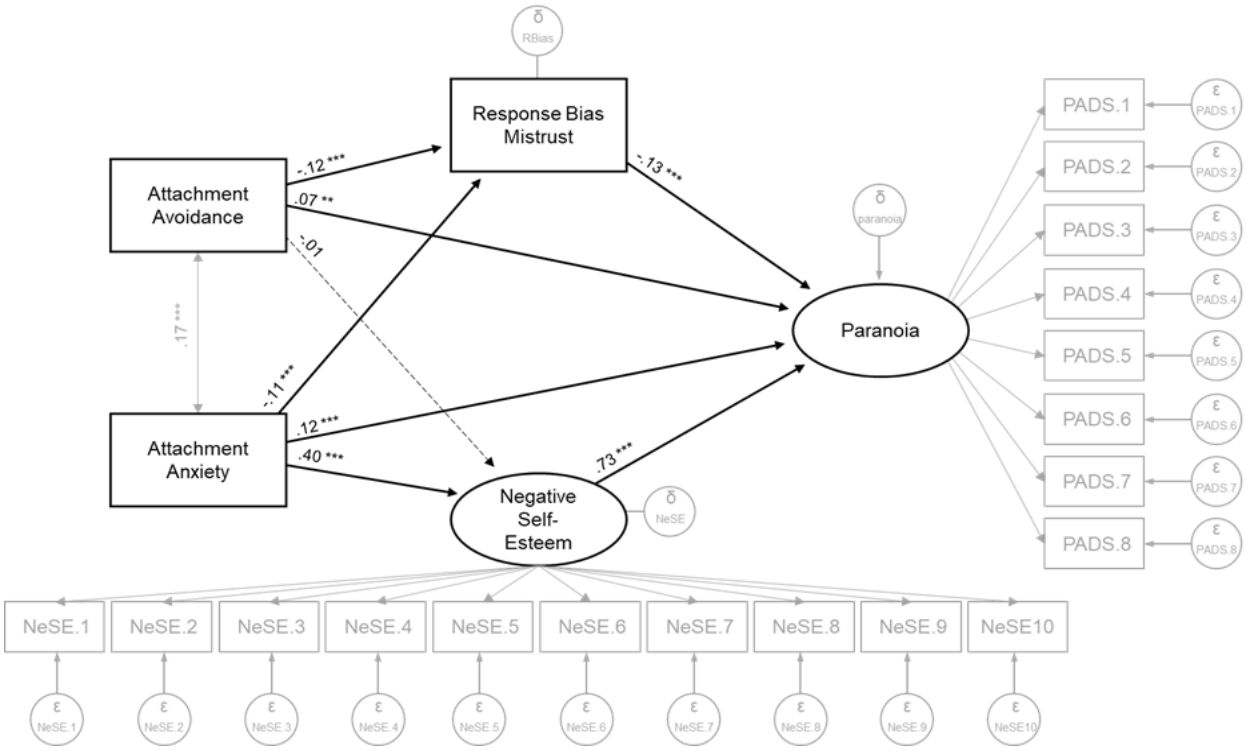
The model chi-squared test was statistically significant (χ^2 (183, *N* = 1121) = 1209.93, *p* < .001), which was to be expected given the large sample size (Kenny, 2015). The other recommended fit indexes suggested good model fit, with the absolute fit measures *RMSEA* = 0.071, 95%-CI 0.067 - 0.075, and *SRMR* = .045 being smaller than the recommended .08. The *CFI* = .93 and *TLI* = .92 were above the .9 rule of thumb.

We observed significant direct effects for all of our paths other than the path from attachment avoidance to negative self-esteem, β = -0.01, *95% CI* -0.07, - 0.05, *p* = .77 (Figure 3.2.). Importantly, as in previous work (Wickham, Stiko, and Bentall, 2015), the path from attachment anxiety to negative self-esteem, β = 0.40, *95% CI* 0.35 - 0.45, *p* <.001, and the path from negative self-esteem to paranoia, β = 0.73, *95% CI* 0.68 - 0.77, *p* <.001 were the strongest.

[8] Standardization does not always allow for direct comparison of effects (King, 1986). However, as the estimate of the response bias when considering its confidence interval falls further away from zero, we can be more confident that in the long run its true estimate may fall more often outside of 0. Overall, the results hint towards response bias being a stronger predictor.

In terms of indirect/ mediation effects, we observed a strong indirect effect between attachment anxiety on paranoia through negative self-esteem, β = 0.064, *95% CI* 0.053 - 0.077, p <.001, and a weaker indirect effect of attachment anxiety on paranoia through a response bias towards mistrust, β = 0.003, *95%CI* 0.001 - 0.005, p <.001. For the indirect effect of attachment avoidance on paranoia, we found no effect through negative self-esteem, β = -0.002, *95% CI* -0.011 - 0.008, p =. 78, and a weak effect through response bias, β = 0.003, *95% CI* 0.002 - 0.006, p < .001. $^{**}p < 0.01$, $^{***}p < 0.001$.

Figure 3. 2. *Full Mediation Model between attachment styles, response bias towards mistrust, negative self-esteem and paranoia traits. All estimates are standardized. Level of significance* $^{**}p < 0.01$, $^{***}p < 0.001$



To discriminate whether the indirect effect went through the response bias or sensitivity measure, we first added sensitivity as a third mediator. This undermined model fit considerably. We then replaced response bias with sensitivity. None of the direct effects from attachment styles to sensitivity, nor the direct effect from sensitivity to paranoia, nor any of the indirect effects through sensivity were significant; all ps > .15 and $\beta < .03$ (also see appendix B Figure

S3.1.). This suggests that the indirect effect from attachment styles on paranoia goes through response bias but not sensitivity.

3.4. Discussion

In this study, our first aim was to assess whether paranoid traits are associated with judgments of untrustworthiness operationalised as signal detection outcomes (bias and sensitivity), hypothesizing an association with bias but not sensitivity. Unexpectedly, we observed that paranoid traits were associated with reduced sensitivity, suggesting a reduced ability to detect subtle facial cues that signal trustworthiness. Nonetheless, in our subsequent regression analyses response bias was the stronger and more stable predictor. This finding suggests a tendency towards a liberal criterion for initiating a threat-related response (e.g., perceiving a trustworthy face as untrustworthy; Correll, Park, Judd, & Wittenbrink, 2002; Haselton & Buss, 2000) and is consistent with other evidence that people with high paranoia traits show an increased tendency to anticipate social threat (Bentall et al., 2009). Previous studies have addressed the association between judgments of faces and paranoia in clinical and non-clinical populations revealing inconsistent findings either by suggesting that patients diagnosed with schizophrenia might have difficulties discriminating facial stimuli (Baas, Van't Wout, Aleman, & Kahn, 2008) or proposing that participants with high paranoia traits show a bias towards mistrust (Kirk, Gilmour, Dudley, & Riby, 2013) . However, none of these studies assessed mistrust outcomes from a signal detection perspective. Thus, to the best of our knowledge, this is the first study that analyses judgments of trust outcomes using a signal detection framework.

As our second aim, we wanted to expand our understanding of the psychological mechanisms that mediate between insecure attachment and paranoia (Bantall & Fernyhough, 2008; Pickering et al., 2008; Wickham et al., 2015). For this purpose, we considered bias towards mistrust as a second mediator that is independent of negative self-esteem. In line with

our second hypothesis, we found indirect effects of attachment anxiety on paranoid beliefs through both bias towards mistrust as well as negative self-esteem. As stated in our third hypothesis, the indirect effect from attachment avoidance on paranoia went only through mistrust. Attachment avoidance reflects negative view of others whereas attachment anxiety reflects a negative view of oneself (Fett et al., 2016) and, hence, it is perhaps unsurprising that the mediating effect for self-esteem was only found in the case of anxiety. When we substituted sensitivity for response bias, the model did not hold. Together, these findings suggest that response bias plays a larger role than sensitivity in explaining the association between insecure attachment styles and paranoia, and that mistrust is an additional component of paranoia that is independent of self-esteem.

Several authors suggest that dysfunctional attachment styles, as a result of repeated experiences of victimization, are likely to heighten negative self-esteem and thus contribute to the feeling of being vulnerable to the actions of powerful others (Bentall & Fernyhough, 2008; Freeman, 2007a). Moreover, childhood attachment disruption experiences may limit the availability of secure attachment figures, leading to feelings of mistrust of others (Mikulincer, 1995; Sitko et al., 2016). Consistent with these accounts, empirical research supports an association between disrupted early attachment relationships and paranoia (Bentall et al., 2012a; Varese et al., 2012) and suggests that insecure attachment helps to explain this association (Sitko et al., 2014). Given that trust judgments dominate initial evaluations of new people, we expected that, when considered from a decision-making perspective, the feelings of mistrust created by an insecure attachment style would be manifest in a bias towards assuming that novel faces are untrustworthy. Although the design of our study was cross-sectional, our results point to how this attachment-based model can be expanded to account for the negative beliefs about the self and the intentions of others that are the key feature of paranoid thinking.

Future research would benefit by employing experimental and longitudinal designs to establish causality whilst incorporating measures of childhood adversity, attachment and mistrust.

3.4.1. *Limitations*

We acknowledge a number of limitations of this study. In terms of the facial trust detection ask, our mass online testing allowed us to present only ten faces to participants, which may have limited the precision of our signal detection measures (Essien et al., 2017); in future research it will be useful to employ more trials. A second limitation was that we only used male Caucasian faces. This decision was made in the light of evidence that there is a bias to classify bald, hairless faces as males (Todorov et al., 2013) thus, it is possible that our findings may not extend to females faces as well as faces from different ethnic backgrounds. Moreover the use of computer-generated faces as stimuli might have limited the ecological validity of our study, and it would be useful to replicate our findings with, for example, video-recorded images of real people. A third limitation is the large number of participants who had to be excluded because they consistently indicated "4: neutral" on the facial trust detection task, which may have reflected failure to engage with it. Our cross-sectional data allows only limited capability to make causal statements and our findings should be seen as consistent with an attachment-based developmental pathway rather than proving the existence of such a pathway. Finally, the effect sizes for the mistrust pathway were smaller than the effect sizes for the self-esteem pathway. It would be tempting, but in our view premature, to assume that the self-esteem pathway is more important. Small effect sizes between signal detection operationalisations and self-report scales are common in research (Mekawi & Bresin, 2015), and these kinds of effects can have large societal implications at the population level (see, e.g. Mekawi & Bresin, 2015). Our small effects may be also be a reflection of our methods. We would expect larger effects in a controlled lab environment and in clinical samples. This being said, our large and highly

representative sample, allows us to generalize to the UK population and allowed us to conduct high-powered statistical tests.

3.4.2. *Conclusions*

To the best of our knowledge, this is the first study to examine the relationship between mistrust operationalized as signal detection outcomes and paranoid beliefs. Our results revealed that participants with high paranoia traits show a bias towards mistrust when judging unfamiliar faces and that this process was also associated with insecure attachment styles. Moreover, our findings also revealed an indirect mediating effect of negative self-esteem between attachment anxiety and paranoia but not between attachment avoidance and paranoia. Although future research is needed to replicate these findings and to establish the direction of these associations, these findings should encourage clinicians to consider the role of mistrust in clients who are experiencing paranoia and to develop interventions for these patients that specifically target insecure attachment and trust judgments in combination with already established interventions for self-esteem.

Chapter IV – Third Empirical Study

Manipulating relational schemas to explore psychological processes associated with paranoid beliefs: an online experimental study

Contributions

Anton P. Martinez (conceptualization, formal analysis, methodology, writing – original draft)

Elizabeth Milne (conceptualization, methodology, supervision, review)

Georgina Rowse (conceptualization, methodology, supervision, review)

Richard P. Bentall (supervision, methodology, writing– original draft, review, and editing.)

Findings from this chapter have been presented as an open paper at the European Association of Behavioural and Cognitive Therapies (EABCT):

Martinez, A. P., and Trucharte, A. (2022). Effect of a classical conditioning intervention on social processes associated with paranoid beliefs: evidence from two experimental studies. *EABCT*, Barcelona, Spain.

Abstract

Background: Information about the self and others is organized in cognitive-affective structures that influence and guide interpersonal behaviour. These structures are referred to as relational schemas and are thought to be influenced by early interpersonal experiences with significant others leading to secure or insecure attachment patterns as adults. When insecure, these patterns appear to contribute to paranoid interpretations about the intentions of others by indirect pathways such as negative self-esteem and a bias towards untrustworthiness. Experimental studies employing classical conditioning (CC) interventions have been successful in manipulating these schemas, finding significant effects on various psychological outcomes such as attachment styles, implicit self-esteem, and paranoid beliefs. However, no study to date has explored these effects on trustworthiness judgements. *Objective*: This study aims to replicate the findings from previous experiments by also testing the effect of manipulating relational schemas on trustworthiness evaluations. *Methods*: A convenience online sample of 307 participants completed a series of tasks and questionnaires measuring attachment styles, explicit and implicit self-esteem, paranoia, and trustworthiness evaluations before and after a brief CC intervention, which involved being randomly allocated to three conditions. In each of these conditions, information about the self was always paired with either positive face stimuli (proximity-seeking condition), negative face stimuli (self-threat condition), or neutral face stimuli (control condition). *Results*: This study failed to replicate findings as previously reported in published experiments, only finding a marginally significant effect on attachment styles on the proximity-seeking CC condition. Moreover, no effect was found regarding trustworthiness judgements. *Discussion*: Limitations such as the online nature of the study and methodological aspects are discussed.

Keywords: *Relational schemas, trustworthiness, attachment styles, paranoia, classical conditioning*

4.1. Introduction

Social psychology researchers conceptualize relational schemas as cognitive-affective structures that, based on experiences of social interactions, organize patterns of interpersonal behaviour and guide social information processing (Baldwin, 1992). These cognitive structures can be categorized into self and other schemas that reflect beliefs and expectations about oneself and other people respectively (Brunson et al., 2015). It is thought that the content of these schemas is organized in associative networks that can be activated when primed with related information (Baldwin, 2005). For instance, studies have shown that participants with negative views of the self who have been presented with self-flaw cues display an attentional vigilance bias toward social rejection stimuli (Ravary & Baldwin, 2018). Behavioural psychologists argue that self-negative information becomes more accessible due to recurrent paring of aversive experiences with social stimuli (i.e., self, other) creating an association that facilitates fearful responses when related cues are presented (Lissek et al., 2008). For example, being repeatedly victimized by peers during adolescence can lead to a self-concept of worthlessness and a view of others as powerful which can elicit hypervigilant behaviours as an adult when faced with similar social situations (Bentall et al., 2012; Stapinski et al., 2014).

The construct of relational schemas is similar to the concept of working models proposed by attachment theorists. However, while the former reflects general knowledge structures about the self, others, and expectations of interpersonal experiences, the latter describes schemas that are activated in attachment-related situations (Brunson et al., 2015; Dewitte & De Houwer, 2011). Hence, attachment styles function as a system that is triggered when faced with a threatening environment with the goal of achieving a general sense of safety (Baldwin & Kay, 2003).The availability of a secure base will depend on the ability of the primary caregiver (i.e., attachment figure) to successfully meet the emotional needs of the child. If those needs are met, the child is likely to develop a secure attachment style as an adult

while, if they are not, an insecure style is more probable (Mikulincer, 1995). People with an insecure-anxious attachment style tend to manifest an excessive longing for closeness as a consequence of their attachment figures being inconsistent in early developmental stages. Conversely, a person with an insecure-avoidant style will deactivate their attachment system, exhibiting extreme distancing from close relationships, reflecting an irresponsive attachment figure during childhood (Dewitte et al., 2007; Dewitte & De Houwer, 2011). Furthermore, these insecure styles are characterized by specific relational schemas. Whereas anxious styles are characterized by having a negative self-schema and a positive other-schema, avoidant styles display the opposite, a positive self-schema and negative other-schema (Dewitte and De Houwer, 2011). Therefore, attachment-working models serve to regulate-distress in threatening interpersonal situations and are, in turn, influenced by specific relational schemas (Shaver & Mikulincer, 2005).

Perceptions of social threat and feelings of interpersonal distress stem from a primal human need that dates from early evolutionary times, the need to belong and feel accepted (Baumeister & Leary, 1995). Given that living in groups was crucial for survival, being ostracized or rejected would have diminished the chances to seek food, defend oneself or reproduce, leading to certain death and preventing genes from being passed to further generations (Fiske and Taylor, 2017). Within this framework, social and personality psychologists have argued that self-esteem serves as a *sociometer* that monitors the degree of one's acceptance and connection with others and directs screening for environmental signals related to socio-evaluative concerns (Howell et al., 2019; Leary, 2005). For example, experimental studies have shown that inducing feelings of acceptance results in increased levels of state self-esteem (Blackhart et al., 2009). Conversely, low-trait self-esteem seems to enhance sensitivity to social cues whereas high-trait self-esteem appears to mitigate the effects of negative social evaluations (Howell et al., 2019). Thus, self-esteem can be conceptualized

as a general sense of personal worth that facilitates access to relational schemas representing expectations about social approval or disapproval (Baldwin & Kay, 2003).

Early attachment experiences involve warmth and nurturing provided by attachment figures and these initial interpersonal relations threfore set the basis for high or low levels of self-esteem (Hart et al., 2005). Being regarded as competent and worthy of affection by significant others during childhood and adolescence can have enduring psychological effects on a person's sense of worth throughout their lifespan (Sroufe, 2005). For example, clinical researchers have found that the association between insecure attachment styles (particularly anxious) and depressive symptoms is explained by negative self-esteem (Lee & Hankin, 2009b; Roberts et al., 1996). Similarly, the same mediation pathway has been shown to explain the association between insecure attachment and paranoid beliefs (Humphrey et al., 2021; Sood et al., 2022). Moreover, a recent study revealed that although negative self-esteem explained the relationship between anxious attachment and paranoia, a bias towards mistrust was found to explain the association between both attachment styles (avoidant and anxious) and paranoid beliefs (Martinez et al., 2020). Trust is regarded as a core component in relational schemas, particularly in early attachment relations, as it involves positive expectations that significant others would be available in fulfilling one's emotional needs (Mikulincer, 1998). Hence, by having secure relational schemas one can form automatic impressions that others are trustworthy, easing social interactions with unknown individuals (Todorov, 2008). Conversely, feeling vulnerable and having negative expectations about interpersonal relations can lead to rapid mistrust judgements of unfamiliar faces, enabling avoidance of potentially dangerous strangers but at the same time facilitating hostile interpretations of other people's intentions (i.e., paranoia).

Several studies have tried to manipulate relational schemas in order to test whether there is an effect in the abovementioned psychological processes (i.e., self-esteem, attachment,

paranoia). For example, using a classical conditioning paradigm Baccus et al. (2004) found that participants whose self-relevant information (e.g., name, date of birth) was paired with smiling faces exhibited higher implicit (but not explicit) self-esteem than those participants whose self-relevant information was paired with random facial expressions. Espinosa et al. (2018) replicated this finding in a student sample with subclinical levels of paranoia with the added effect of showing that the intervention also lowered positive subclinical symptoms (i.e., unusual experiences) although paranoid levels were unaffected. However, using the same paradigm but with an added negative condition (self-relevant information paired with angry faces) Trucharte (2022)[9] found that student participants in that group reported higher levels of state paranoia. Conversely, participants in a positive condition (self-relevant information paired with happy faces) reported a reduction in interpersonal sensitivity as well as in state anxious and avoidant attachment insecurity. To summarise, it seems that by carrying out different manipulations with regards to relational schemas, various mechanisms seems to be triggered. Warm associations with the self seem to activate positive relational schemas and a sense of security whereas hostile associations elicit negative relational schemas and hypervigilant states.

To date few studies have employed associative or priming interventions to test the effect of relational schemas on trustworthiness judgements. One study found that clinical participants with high paranoia levels in comparison to non-clinical controls rated neutral faces as more untrustworthy when primed with negatively valenced images (Hooker et al., 2011). Although no relational schema primes were used, the study provided evidence that trustworthiness judgements in clinical samples can be influenced by negative emotional states. Moreover, although explicit self-esteem seems to be unaffected by a positive associative intervention (Baccus et al., 2004; Espinosa et al., 2018) the effect of negative conditions has not been

[9] It should be noted that the cited results are part of a doctoral dissertation and have not been peer-reviewed or published yet.

studied. Some authors have theorised that, in self-threatening situations, paranoid individuals might adopt a defensive attitude as reflected in a discrepancy between implicit and explicit self-esteem (Bentall et al., 2001; Jordan et al., 2003). According to this kind of defensive model, it might be predicted that pairing self-relevant information with threatening face images will elicit an increase in explicit state self-esteem but, at the same time, a reduction in implicit self-esteem.

The current study aims to replicate Baccus, Espinosa and Trucharte's findings and to extend them to consider the role of mistrust and paranoia in the light of Martnez et al.'s (2020) findings. We employed an experimental design in which participants were randomly assigned to either a self-threat, proximity-seeking, or control condition. In the self-threat condition, self-relevant information was paired with threatening faces whereas, in the proximity-seeking condition, self-relevant information was paired with non-threatening (i.e., likeable) faces. A control condition involved pairing self-relevant information with random threatening, neutral and likeable faces. The study explored the effect of these conditions on trustworthiness judgements operationalized by an affective priming task, using relational schemas as primes (self and other relevant information) and previously validated trustworthy and untrustworthy faces as targets. Trustworthiness judgements were analysed using signal detection analysis to calculate bias scores (i.e., the tendency to judge an untrustworthy face as trustworthy or vice versa). Finally, both implicit and explicit self-esteem were also measured as outcome variables to explore the effect of the self-threat intervention on these variables.

Based on the aforementioned design we expected the following results:

1. In comparison to the control condition, participants in the self-threat condition would report higher state explicit self-esteem levels but lower implicit self-esteem levels while

in the proximity-seeking condition, an increase in implicit self-esteem levels would be found but state explicit state self-esteem levels would not change.

2. An increase in state paranoia would be evident in the self-threat condition in comparison to the control and proximity-seeking conditions.

3. After the intervention, in comparison to the control condition, participants in the proximity-seeking condition would report lower levels of state attachment insecurity (i.e., avoidant and anxious) whereas higher levels were expected in the self-threat condition.

4. Finally, following the classical conditioning manipulation, in comparison to the control condition, participants in the self-threat condition would show a bias towards mistrust following a self-relevant prime but not when an other-relevant prime is presented. Conversely, participants in the proximity-seeking condition would report a bias towards trustworthiness when presented with a self-relevant prime but not following an other-relevant prime.

4.2. Methods

4.2.1 *Participants*

Participants were recruited via social media platforms (e.g., Twitter) as well as the volunteer staff list from the University of Sheffield and were offered a £5 AMAZON voucher once they have completed the study. The ages of the participants ranged from 18 to 73 (*M*age=28.9, *SD*= 7.7), and the sample was characterized by being majority male (51.9%), highly educated (59.4% graduates), and employed (62%). From 307 respondents, 266 were finally selected after excluding participants who did not pass more than 50% (3) of attention checks (e.g., "*Please select option number six*") as well as those who were considered extreme outliers of survey completion time based on Mahalanobis D (Curran, 2016). Excluded

participants did not differ from those who were included in any of the demographics or psychological variables (p >.05).

4.2.2. *Materials*

The revised Paranoia and Deservedness Scale (PaDS –R; Elahi et al., 2017): The PaDS is a paranoia trait measure validated in clinical and non-clinical populations that consists of 10 items which are answered on a 5-point scale from 0 (*"Certainly False"*) to 4 (*"Certainly True"*) with total scores ranging from 0 to 40. This instrument in turn has two scales, a persecution one that measures paranoid ideation and a deservedness one that assesses the degree to which respondents feel they deserve what is described in each persecution item. For this study, only the persecution scale was used reflecting good reliability (α = 0.77).

The Relationship Questionnaire (RQ; Bartholomew & Horowitz, 1991). The RQ is a self-report instrument that describes in four short paragraphs secure, fearful, preoccupied, and dismissing attachment patterns. By reading each description participants rate how well or poorly each vignette defines their corresponding relationship pattern on a 7-point scale ranging from 1(*"Disagree strongly"*) to 7 (*"Agree strongly"*). The scoring of each scale serves to compute measures of insecure styles such as attachment anxiety (negative model of self) and attachment avoidance (negative model of other)[10]. Negative scores will indicate the presence of insecure models for each attachment representation whereas positive scores will reflect the opposite. Reliability analysis using Chronbach's α cannot be calculated because there is only one item per attachment type , but good psychometric properties such as test-retest, construct, convergent and divergent validity for this scale have been established (Wongpakaran et al., 2021).

[10] Each style can be calculated as follows: model of self = (secure + dismissing) - (preoccupied + fearful); model of other = (secure + pre-occupied) - (dismissing + fearful)

Rosenberg Self-Esteem Scale (RSES; Rosenberg,1965): The positive subscale comprises 10 items designed to measure trait positive self-esteem by asking participants to rate on a 4-point Likert scale each statement from 1 (*"Strongly Agree"*) to 4 (*"Strongly Disagree"*). Total scores range between 10 to 40 and reliability analyses revealed acceptable levels (α = 0.63).

Name letter preference task (NLT; Nuttin, 1985): The NLT is an implicit self-esteem measure that consists of asking participants to rate their liking of the letters of the alphabet from 1 (*"Not at all"*) to 10 (*"A lot"*). The rationale behind this task lies in the premise that people with high implicit self-esteem tend to rate more positively the letters of their own names over other letters of the alphabet. A recommended algorithm for calculating implicit self-esteem scores is the ipsatized double-correction algorithm that controls for differences in the likeability of the different letters as well as the frequency of more generally used letters (LeBel & Bertram, 2009). The NLT has shown good levels of internal validity (α = 0.83) and test re-test reliability (Krause et al., 2011).

State Self-Esteem Scale (SSES; Heatherton & Polivy, 1991): The SSES is a 21-item scale designed to assess momentary states of self-esteem. Responses are provided on a 5-point Likert scale (1 = not at all, 2 = a little bit, 3 = somewhat, 4 = very much, and 5 = extremely). This scale has shown good internal reliability (α=.75 to α=.80) and good construct validity with total scores varying between 10 and 105.

State Adult Attachment Measure (SAAM; Gillath et al., 2009): is a scale developed to assess temporary states of insecure (anxious, avoidant) and secure attachment in response to experimental manipulations. The scale includes 21 items where participants have to rate the extent to which they agree or disagree with different statements based on how they currently feel from 0 (*"Disagree strongly"*) to 7(*"Agree strongly"*). For this study, we only used the insecure attachment subscales (7 items per subscale) with total scores ranging from zero to 49.

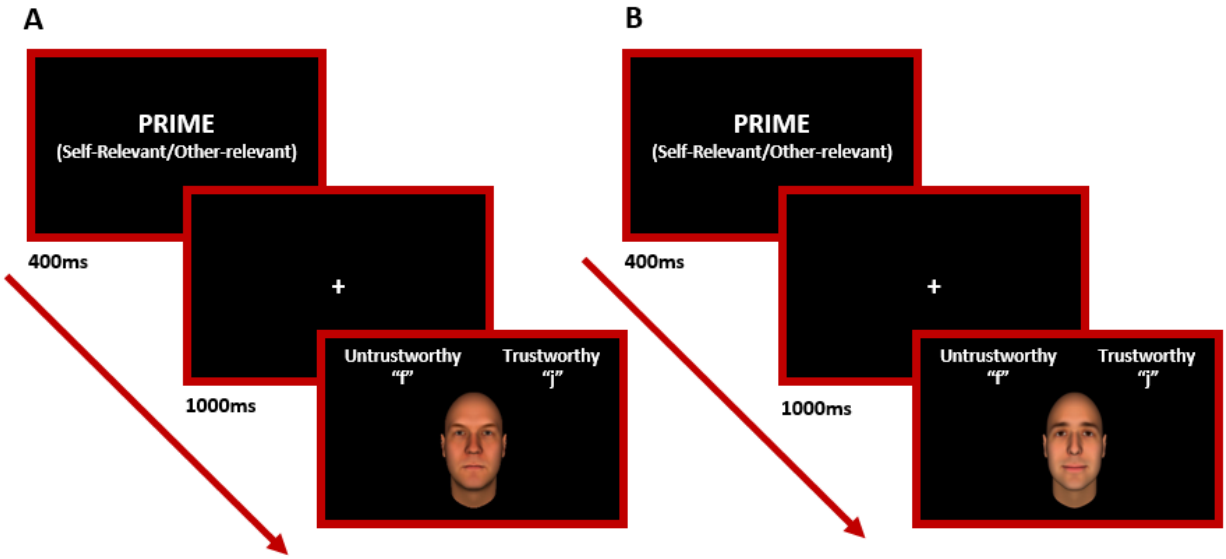
The SAAM has shown good internal reliability (α =.83 to .87), discriminant, convergent, and criterion validity.

State Paranoia Checklist (SPC; Schlier et al., 2005): is an 18-item scale designed to assess mild persecutory ideas. This version of the SPC has been adapted to measure state paranoid beliefs by asking participants to what extent each item applies to them "at the moment". Answers are provided on a Likert scale that ranges from 0 (not at all) to 10 (very much) with total scores varying from 0 to 180. The SPC has excellent internal reliability (α =.96) and good convergent validity (Lincoln et al., 2013).

Affective priming task (APT; Fazio et al., 1986): The APT is an evaluation task in which participants categorise a target stimuli in a binary way (e.g., positive or negative image) after being primed with a valenced stimulus (e.g., positive or negative word). The rationale behind this task is that if the prime triggers the same response as the target the response is facilitated, reflected in lower error rates. However, if the prime and target are incongruent the response becomes conflicted leading to a higher error rate. In this study, an adaptation of this task was conducted by presenting self-relevant and other-relevant information as primes and trustworthy and untrustworthy face stimuli as targets. For this, computer-generated faces from the Princeton Social Perception Lab dataset (Oosterhof & Todorov, 2008) previously calibrated in the trustworthiness dimension as either more trustworthy (+3 and +2 SD) or less trustworthy (-3 and –2 SD) were used as target stimuli (Figure S4.3.). Word stimuli (personal information and non-personal information) provided for the conditioning paradigm were used as prime stimuli. The structure of the task comprised 144 trials of which two blocks of 72 trials involved trustworthy and untrustworthy presentations of targets respectively. Moreover, of those 72 trials, 36 trials included "other" primes whereas the remaining 36 included "self" primes. Finally, each block of 36 trials encompassed 12 trials for three different types of face ethnicity (White, Black, Southeast Asian). For each trial, the prime word was displayed for 400ms

followed by a central fixation point (+) that was presented for 1000ms and the target image (e.g., trustworthy or untrustworthy face) appeared immediately afterward and remained on the screen until participants made a response by pressing the "j" (trustworthy) or "f" (untrustworthy) key (see Figure 4.1.). One hundred and forty-four trials in total were randomized for each participant to control for an order effect. Participants completed six practice trials with neutral words before commencing with the actual task.

Figure 4. 1. *Flow trial diagram for untrustworthy (A) and trustworthy (B) trials.*



4.2.3. *Procedure*

To assess the causal effect of self-threat and proximity-seeking on trustworthiness judgments, implicit self-esteem, and state measures (attachment, paranoia, and self-esteem), the study followed an experimental between/within-subjects design. Participants were randomly assigned either to the experimental conditions (self-threat, proximity-seeking) or the control condition (neutral) and completed outcome measurements before and after the experimental manipulation. Participants who took part in the study had to read the participant information sheet and consent form to take part in the online study. Once they agreed on participating, they were asked to answer self-report questionnaires regarding trait measures

(attachment styles, paranoia, and self-esteem) and state measures (self-esteem, paranoia, attachment). They also were asked to complete affective priming tasks on trustworthiness judgments and implicit self-esteem to establish a baseline before manipulation. This pre-measurement baseline phase was established three days before the intervention to control for any type of carryover effects that could influence the experimental manipulation. This period between pre and post-manipulation was based on Dewitte & De Houwer's (2011) findings in which participants were measured three days before being primed with specific attachment schemas. For this purpse, participants were asked to enter their email addresses so they could be reminded[11] to participate in the second part of the study and receive their £5 AMAZON voucher once the whole study was completed.

The classical conditioning intervention was based on the one implemented by Baccus et al. (2004), which involves asking participants to provide self-relevant information (i.e., first name/nickname, last name, the month and day of birth, personal pronouns me/mine). This information was collected at baseline which was also used for the affective priming task. These words were matched with control words such as names, surnames, personal pronouns, and months and days different from the information provided by the participants (see Table S4.3.). Participants were randomized to either the experimental (self-threat, proximity-seeking) or control conditions using the balanced randomization mode provided in the Gorilla online experiment builder. For completing the intervention, they were informed that a word would appear randomly in one of the quadrants on the computer screen and they were instructed to click on the word as quickly as possible, using the mouse. In addition, they were told that when they did so an image would be displayed briefly (for 500 ms) in that quadrant preceded by a fixation cross (for 250 ms, Figure 4.2.). This procedure was repeated for 252 trials. Self-

[11] This was done by using the Gorilla Experiment builder delay node which automatically sent an email with a link to complete the second part of the experiment 72 hs after participants finished the first part of the study.

relevant words (self-relevant information) and other-relevant words (control words) were presented in a preprogrammed pseudorandom order. In the control condition, once the participant clicked the word stimuli, a random selection of threatening (84 times), non-threatening (84 times), and neutral (84 times) photographs of faces followed both self-relevant and non-self-relevant words. In the experimental self-threat condition, self-relevant words were always paired with an image of a threatening face (126 times) whereas non-self-relevant words were paired with non-threatening (42 times), threatening (42 times), and neutral faces (42 times). In the experimental proximity-seeking condition, self-relevant words were always paired with non-threatening (i.e., likeable) face images (126 times), and non-self-relevant words were paired with neutral (42 times), non-threatening (42 times), and threatening face images (42 times). Face images were also downloaded from the Princeton Social Perception Lab dataset which has been previously validated in the social threatening dimension calibrated as more threatening (+3 and +2 SD) as well as less threatening (-3 and –2 SD; Figure S4.3.). These face stimuli also represented different ethnicities (White, Black, Southeast Asian) which were equally presented in each condition. The experiment was programmed in a way that participants' self-relevant information did not overlap with the other-relevant information. The task was self-paced with forced responses, so participants could not proceed to the following trial unless they clicked the quadrant in which the word appeared to minimize careless responses.

After the classical conditioning intervention, participants were asked to complete again the state measurements (attachment, self-esteem, and paranoia) as well as the affective priming and implicit self-esteem task administered in the pre-measurement phase. The order in which the measurements were presented, at both pre as well as post-time points, were randomized using *Latin square* mode provided by Gorilla online experiment builder. Ethical approval was

granted by the Department of Psychology Research Ethics Committee at The University of Sheffield (Ref: 041111).

4.2.4. *Statistical Analysis*

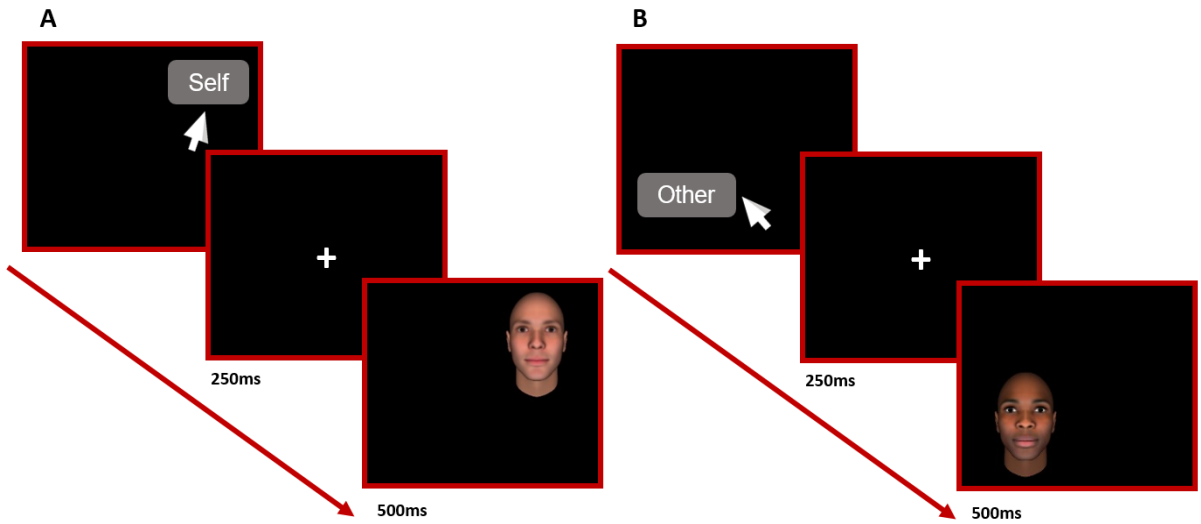
Chi-square tests as well as univariate ANOVA were conducted to compare differences at baseline regarding demographics along with psychological variables between groups. For the trustworthiness outcome, a 2 (time) x 3 (ethnicity) x 3 (conditions) repeated measures ANOVA was conducted to control for the effect of ethnicity on trustworthiness judgements. For the rest of the outcomes, a 2 (time) x 3 (conditions) repeated measures ANOVA was conducted. All analyses were conducted in SPSS v28.

For trustworthiness judgements, signal detection analysis was conducted to calculate the bias (β) parameter using the formula 7 reported by Stanislaw and Todorov (1999). In this context, a positive value would indicate a tendency to judge an untrustworthy face as trustworthy whereas a negative value would reflect the opposite. This outcome was used to compute a prime index for the affective priming task by subtracting the bias scores of the "self" priming condition from the "other" priming condition (Pime Index = βSelf - βOther). A positive prime index would indicate a positive bias when primed with self-relevant information in comparison to being primed with other-relevant information. Conversely, a negative prime index would reflect a negative bias when primed with self-relevant information in comparison to being primed with other-relevant information (Wentura & Degner, 2010).

A statistical power analysis was performed for sample size estimation, based on data from a pilot study (N=90). The effect size (ES) in this study for behavioural measures was η_p^2 =.03, considered to be small using Cohen's (1988) criteria. With an alpha = .05 and power = 0.85, the projected sample size needed with this effect size (GPower 3.1) is approximately n = 75 for the simplest between/within-group comparison. Thus, our proposed sample size of

n=307 was more than adequate for the main objective of this study and also allowed for expected attrition and our additional objectives of controlling for possible factors/subgroup analysis.

Figure 4. 2. *Flow trial diagram for self (A) and other (B) classical conditioning trials*



4.3. Results

No significant differences were found when comparing demographic and trait psychological variables between the experimental and control groups (Table 4.1.).

For the rest of the mixed repeated measures ANOVA, normality and homogeneity of variance assumptions were checked by using visual inspections (Q-Q plots and histograms, see Figure S4.1. and S4.2.) and Levene's test respectively. Levene's tests revealed non-significant results for all variables (Table S4.1.) meaning that error variances were equal across groups.

Regarding normality, all variables seem to display normal distributions of their residuals except for the state paranoia variables in which their distribution seemed to be moderately negatively skewed (-.72). Transformations of non-normal distributions as well as

outliers corrections were based on the recommendations detailed in Tabachnick and Fidell (2013)[12]. Sphericity assumption was check for the first analysis as ethnicity had more than two within-subject levels (White, Black, and Southeast Asian) unlike time which only had two (pre and post measures). Mauchly's test of sphericity revealed non-significant results [χ^2 (2) =.99. *p*=.20] indicating that this assumption was not violated.

Table 4. 1. *Descriptive statistics*

Variables	Condition			Total	*p*
	Self-Threat	Proximity-Seeking	Control	*N=266*	
	N=93	*N=81*	*N=92*		
Demographics					
Mean Age	29.1 (6.9)	28.5 (8.4)	29.0 (7.7)	28.9 (7.7)	.85
Gender	48.3% (*M*)	56.7% (*M*)	51.1% (*M*)	51.9% (*M*)	.53
Education	58.1% (*HE*)	59.2% (*HE*)	60.7% (*HE*)	59.4% (*HE*)	.79
Employment	65.6% (*E*)	58% (*E*)	61.9% (*E*)	62% (*E*)	.55
Psychological Traits					
Paranoia	30.1 (6.1)	28.3 (6.5)	28.3 (6.7)	28.9 (6.4)	.10
Self-Esteem	24.2 (3.7)	23.5 (4.5)	23.7 (4.0)	23.8 (4.1)	.48
Attachment Avoidance	-.20 (2.9)	.05 (3.7)	.17 (3.1)	.003 (3.2)	.72
Attachment Anxiety	.50 (3.9)	1.1 (4.0)	.45 (4.0)	.66 (3.7)	.50

Note. *SD in brackets for numerical variables. M= Male, HE= Higher Education (Undergraduate Degree or higher), E= Employed*

With respect to trustworthiness judgments, the main effect of ethnicity on the Prime Index (β) [*F* (2, 526) = 2.03, *p*=.13, η_p^2 = .01], as well as the interaction effect between condition, time, and ethnicity [F (2, 526) = 1.88, *p*=.11, η_p^2 = .01], were non-significant. This would indicate that the ethnicity of the stimuli did not influence trustworthiness judgements

[12] NEWX = SQRT (K-X). Where SQRT refers to squared root transformation and K-X is the reflected variable in which each score has been subtracted from the largest score plus 1 in the distribution.

regardless of the condition participants were in or the time at which the task was completed. When looking at the interaction between condition and time, no significant main effect of time [F (1, 263) = .16, p=.70, η_p^2 = .001], nor significant interaction between time and condition [F (2, 263) = .006, p=.97, η_p^2 = .00] were found.

Concerning state paranoia, the analysis yielded non-significant results for the effect of time[F (1, 263) = 2.00, p=.16, η_p^2 = .008], and the interaction effect between time and condition [F (2, 263) = .1.06, p=.35, η_p^2 = .008].

Regarding implicit self-esteem as measured by the name letter task, no significant effect of time was found [F (1, 263) = 3.29, p=.07, η_p^2 = .012], nor interaction effect between time and condition [F (2, 263) = .23, p=.80, η_p^2 = .002]. The same non-significant results were found on state self-esteem for the effect of time [F (1, 263) = .3.01, p=.08, η_p^2 = .01], and the interaction effect between time and condition [F (2, 263) = .460, p=.63, η_p^2 = .003].

Table 4. 2. *Pairwise Comparisons*

	State Anxious Attachment							
Condition	(I) Time	(J) Time	M_{diff} (I – J)	*SE*	Sig. [b]	η_p^2	95% CI [LB / UB]	
Control	1	2	.97	.54	.07	.01	-.09	2.03
Proximity-Seeking	**1**	**2**	**2.23**	**.320**	**<.001**	**.05**	**1.10**	**3.36**
Self-Threat	1	2	.33	.53	.53	.001	-.72	1.39

Note. Based on estimated marginal means [b]. Adjustment for multiple comparisons: Bonferroni.

A significant main effect of time was found for state attachment avoidant style [F (1, 263) = 4.27, p=.04, η_p^2 = .02] revealing a reduction of attachment-avoidant scores from pre (M_{T1}=26.98) to post (M_{T2}=26.35) experimental manipulation. Nonetheless, this change was not moderated by experimental conditions as the interaction effect between time and condition was not significant [F (2, 263) = .61, p=.54, η_p^2 = .005]. In the case of state attachment anxiety a main effect of time was also found [F (1, 263) = 13.80, p<.001, η_p^2 = .05] revealing a significant decrease of state attachment-anxiety levels from pre (M_{T1}=30.28) to post (M_{T2}=29.10) intervention. Moreover, results revealed a marginally significant interaction effect between time and condition [F (2, 263) = 3.00, p=.05, η_p^2 = .02]. Pairwise comparisons with Bonferroni corrections showed that participants in the Proximity-Seeking condition reported lower levels of insecure attachment levels post-intervention (M_{T1}=30.90, M_{T2}=28.66) in comparison to the control and Self-Threat condition in which no changes were found (Table 4.2.).

4.4. Discussion

This study aimed to explore the effect of relational schemas on attachment, self-esteem, trustworthiness judgements, and paranoia processes by pairing self and other-relevant information with likeable, threatening, and neutral face stimuli. For our first hypothesis we expected that, compared to the control condition, participants in the proximity-seeking condition would exhibit higher implicit self-esteem but no change in explicit state self-esteem. For participants in the self-threat condition, we predicted they would report a discrepancy between higher state self-esteem and lower implicit self-esteem reflecting a defensive response to threat. In contrast to what Baccus et al. (2004) and Espinosa et al. (2018) found, implicit self-esteem levels in the proximity-seeking condition did not increase after the intervention. Likewise, a discrepancy between implicit and explicit self-esteem was not found in the self-threat condition after the experimental manipulation. Moreover, in our second hypothesis, we

did not find higher state paranoia levels in the self-threat group after the intervention, thereby not replicating Trucharte's (2022) findings. Nonetheless, our third hypothesis was partially met as a decrease in state avoidant attachment levels was found regardless of the effect of a particular intervention whereas lower state anxious attachment levels were marginally explained by the proximity-seeking condition. In our last hypothesis, we stated that the condition in which participants were would influence trustworthiness judgements. Relative to the control condition, we expected that in the self-threat condition, participants would show a biased response towards mistrust when a self-relevant prime preceded the targets and the opposite (a bias towards trust when primed with self-relevant cues) in the proximity-seeking condition. Results did not support this hypothesis, as when or in which condition participants completed the task did not affect their trustworthiness judgements.

Findings from the first and second hypotheses mirror failed replications that, to some extent, can be explained by different factors. First, we conducted the study online due to the COVID-19 pandemic whereas the studies on which we based the replication were conducted in-person. Although online studies pose clear advantages over lab-based experiments such as rapid recruitment and relatively low cost for researchers, many other aspects may compromise the reliability of online responses. For example, not being able to check if the information participants provide is accurate (e.g., incorrect demographics), or if they are trying to profit from the study (e.g., taking part in the study more than once), or not being motivated enough (e.g., careless responses). Given this, several aspects were taken into account when designing the study to ensure the methodological quality, such as randomization procedures, implementing attentional checks, controlling for carry-over effects, and financial reward upon completion of the whole study. Nonetheless, accounting for these potential sources of unreliable responses is probably not enough to ensure the completion of the intervention in a

controlled and quiet environment, possibly compromising the effectiveness of the experimental procedure.

A second factor that differentiates the current study and the ones upon which we based our replication is the use of face stimuli for the classical conditioning intervention. Whereas the original studies employed face stimuli expressing happy, neutral, and angry expressions from the Karolinska Directed Emotional Faces database, we used computer-generated faces on the threatening–likeable dimension from the Princeton Social Perception Lab database. The decision of employing the latter database was due to certain advantages over face photographs of human actors. For example, controlling for specific features of facial expressions such as different variations of the emotion displayed as well as individual differences exhibited by the actors in their expressions (Said & Todorov, 2011). Although controlling for these aspects could be beneficial for psychophysiological studies, the use of computerized faces might seem more unnatural in contrast to the faces of human actors and thus might not be ideal for activating relational schemas. However, the use of this dataset seemed to have had an effect, albeit small, on state-anxious attachment. Given that an insecure-anxious attachment style is characterized by having a negative self-schema and a positive other-schema, it might not be surprising that pairing likeable faces with self-relevant information would elicit a general sense of security. Another explanation for this finding is that state attachment style is more sensitive to change in comparison to other constructs such as implicit self-esteem and thus it is more easily activated using computerized face stimuli. Future research should focus on the effect that different types of face stimuli (i.e., computer-generated; human faces) can have on eliciting relational schemas and its effect on various related psychological variables (i.e., self-esteem, attachment, paranoia, trustworthiness).

Finally, trustworthiness judgements while primed with relational schemas were not affected by the classical conditioning intervention. Implementing an evaluative (or classical)

conditioning intervention involves establishing an association between an unconditioned stimulus (*US*: self-relevant word) with a conditioned stimulus (*CS*: threatening/likeable face). Thus, we operationalized trustworthiness judgements using an evaluative (affective) priming task so that the learning effect of the conditioning intervention would be reflected in a facilitated response when a prime (*US*) preceded a target (*CS*). One possible explanation for this null effect could be that the valence of the *CS* was not high enough for the *US* to trigger a conditioned response (Hofmann et al., 2010). Furthermore, a bias towards mistrust measured as the outcome of an affective priming paradigm at baseline did not reveal significant correlations with paranoid traits nor with attachment styles (Table S4.2.) probably indicating weak convergent validity. Previous studies using standard facial recognition tasks with the same stimuli have found significant associations between a bias toward mistrust and paranoia as well as with insecure attachment styles in large international representative samples (Martinez et al., 2020, 2022). Explicit and implicit emotion recognition tasks are designed to measure different psychological processes. Whereas the former focuses on the direct processing of emotions, the latter aims to capture nuanced responses elicited by contextual information. (Kliemann et al., 2013). Thus, this could lead to differences in observed associations between the same construct measured differently and an outcome. Hence, the fact we did not employ a representative sample in this study or used weak primes in our affective priming task may have led to an unreliable measure of trustworthiness judgements.

4.4.1. *Limitations*

As mentioned in the abovementioned paragraphs, this study is not exempt from limitations. First, the use of an online convenience sample does not reflect the characteristics of a representative general population sample. Although the gender distribution was not distal from the one of the general population, the sample in this study was mainly highly educated, young, and employed and hence our results are not generalizable. Moreover, online recruitment

was done via social media, and may have attracted careless responders, rather than via well-known survey platforms (e.g., Prolific, MTurk) which tend to guarantee quality respondents. Lastly, although the face stimuli dataset used for both the trustworthiness task as well as for the classical conditioning intervention included three different types of ethnicity the faces were primarily male thus limiting the generalizability of the results. Though the ethnicity of the faces was controlled for, revealing that did not have an effect on trustworthiness judgements, the ethnicity of the participants was not collected thus constraining our interpretations of this finding.

4.4.2. *Conclusions*

This study aimed to experimentally manipulate relational schema to see its effect on different psychological constructs that tap onto social cognitive processes that are thought to be underlying mechanisms of psychological symptoms-traits such as paranoia. Previous studies employed classical conditioning interventions to elicit such processes however, they did not consider trustworthiness judgements, a core component of attachment and paranoid beliefs. Due to contextual circumstances such as the COVID-19 pandemic, this research was conducted online and thus faced several limitations. Nonetheless, significant and null findings of this research can lead to a number of conclusions. First, the same procedures should be followed when replicating experimental paradigms, in the case of the current study real face stimuli as well as in-person data collection may have had an impact on the results. Second, it is possible that, in order to generate a significant effect on trustworthiness ratings operationalized as an affective priming task, the valence of the *CS* should be powerful enough for the *US* to elicit a response. Thirdly, when measured explicitly or implicitly, trustworthiness judgements might tap into different psychological processes leading to different observed associations between the same construct and other variables. Lastly, insecure state anxious attachment seems to be more sensitive to change meaning that, by activating positive self-relational schemas, insecure

attachment state levels are reduced leading to an overall feeling of security. However, given the abovementioned limitations as well as the marginal effect found, these findings and conclusions should be interpreted with caution. Future research should replicate these findings by comparing online versus lab-based data collection as well as by implementing different face stimuli datasets to explore the effect of human faces versus computer-generated ones.

Chapter IV – Fourth Empirical Study

Psychophysiological underpinnings of trustworthiness judgements in clinically paranoid and non-clinical samples: a proof-of-concept study

Contributions

Anton P. Martinez (conceptualization, formal analysis, methodology, writing – original draft)

Elizabeth Milne (conceptualization, supervision, analysis, methodology, review, and editing.)

Almudena Trucharte (conceptualization, methodology)

Stephen Kellet (conceptualization, methodology)

Richard P. Bentall (conceptualization, supervision, methodology, review, and editing.)

Abstract

Background: Facial information is used in everyday life to infer the mental and emotional states of others. Based on emotional facial cues individuals tend to make fast social attributes that prepare them to adopt approach or avoidant responses. Trustworthiness is the most common attribute that people make when encountering strangers. Evidence has found that people with paranoid traits have a general tendency in mistrusting unfamiliar faces and fMRI (functional Magnetic Resonance Imaging) studies have shown increased amygdala activation when patients with paranoid schizophrenia judged untrustworthy face stimuli. Although fMRI studies provide highly accurate information regarding brain structures when processing information, no study to date has explored temporal dynamics in trustworthiness evaluations in paranoid patients using ERP (Event Related Potential) techniques. *Objectives*: This proof-of-concept study aims to explore the temporal dynamics of trustworthiness evaluations in participants with paranoid schizophrenia spectrum conditions and non-clinical controls. *Methods*: Nine clinical and 15 non-clinical participants completed a series of questionnaires and a trustworthiness face recognition task while their neural activity was recorded through EEG (Electroencephalogram) equipment. *Results*: Participants in the clinical group showed greater bias towards mistrust as well as high levels of interpersonal mistrust. Clinical participants revealed enhanced amplitudes of the left P100 when processing untrustworthy faces and reduced right N170 regardless of the trustworthiness of the stimuli in comparison to non-clinical controls. *Discussion*: Clinically paranoid participants appear to allocate cognitive resources to negatively valenced stimuli at very early stages of face processing, revealing also differences in amplitudes when encoding face stimuli in comparison to controls. Limitations of the study such as small sample size and methodological aspect are discussed.

Keywords: *Persecutory delusions, trustworthiness, bias, EEG, ERP, P100, N170, face processing, interpersonal trust, signal detection*

5.1. Introduction

In our everyday life we constantly make social judgements about other people. Using facial information, we can infer the mental and emotional states of others in fractions of seconds. Moreover, important social outcomes may be determined by these kinds of judgments. For example, voting behaviour seem to be predicted by the fast processing of social attributes (e.g., competence) based on the physical appearance of the prospective candidates, particularly among uninformed voters (Todorov & Oh, 2021). Some studies have shown that people can form first impressions of unfamiliar faces when displayed for only 100ms, with lengthier exposures strengthening their initial decisions (Willis & Todorov, 2006). Scientists have therefore argued that face processing is a highly adaptive skill shaped by evolutionary mechanisms, since identifying different emotions accurately and rapidly allowed our ancestors to recognise their own kin and detect potential threats, facilitating both cooperative and defensive strategies (Fiske & Taylor, 2017). First impressions are thus formed on a valence dimension, based on the positivity or negativity of facial cues, eliciting approach or avoidant responses that enable individuals to engage or distance themselves from others (Todorov, 2008). Thus, the processing of human faces involves complex global visual systems that prepare an individual for social interactions, and which have an impact on the development and organization of the mind throughout the lifespan (Bigelow et al., 2021).

Trustworthiness is considered one of the more general and stable social attributes that people infer when encountering unfamiliar faces (Oosterhof & Todorov, 2008; Todorov & Oh, 2021). Interpersonal trust involves forming positive expectations about the intentions of other people while accepting personal vulnerability to their actions, and is a process that is important for enabling prosocial interactions (Lewicki et al., 2006). Unfamiliar faces that are rated as trustworthy tend to be associated with attributes such as warmth and likeability whereas untrustworthy faces are associated with threatening and erratic traits (Todorov & Oh, 2021).

Evidence from neuroimaging research has revealed a strong association between trustworthiness judgements and the activation of the amygdala, particularly when evaluating untrustworthy face stimuli (Santos et al., 2016). The amygdala is a brain region involved in the processing of emotional information and is thought to play a key role in detecting salient stimuli to prepare the organism for potential danger (Öhman, 2005; Whalen, 2007). Thus, from an evolutionary point of view, it seems more adaptive for the organism to be sensitive to untrustworthiness cues than trustworthiness cues to avoid the more costly error of being vulnerable to the actions of an untrustworthy individual (Schaller, 2008).

5.1.1. *The importance of trust in paranoia*

Paranoid delusions (one the most prevalent symptoms of schizophrenia; Moutoussis et al., 2007), concern the negative core belief that others are intentionally trying to harm the individual, which in turn is reflected in defensive attitudes such as hypervigilance and suspiciousness (Bentall et al., 2009). Within this context, mistrust is considered a central feature of paranoid beliefs (Bell & O'Driscoll, 2018) with epidemiological evidence indicating that a bias towards mistrust face evaluation mediates between insecure attachment and paranoid traits (Martinez et al., 2020). Early interpersonal experiences with primary caregivers influence the development of attachment styles leading to secure or insecure relational patterns that guide interpersonal behaviour (Bowlby, 1982). The upbringing of individuals with secure relational patterns tends to be characterized by caring and nourishing environments. On the other hand, insecure individuals generally report inattentive and distant attachment figures throughout childhood leading to negative expectations about the reliability of significant others as adults (Shaver & Mikulincer, 2005). Since attachment styles guide social information processing, it would not be surprising that these patterns would affect automatic face evaluations. Thus, insecure styles may be characterized by a tendency towards untrustworthiness, which in turn culminates in paranoid interpretations of social interactions.

Studies carried out in clinical populations using fMRI (functional Magnetic Resonance Imaging) have revealed that the association between amygdala activation and untrustworthiness face evaluations is stronger in clinically paranoid participants when compared to non-paranoid clinical and non-clinical samples (Pinkham et al., 2015). Moreover, evidence from experimental manipulations that have been designed to induce paranoia in non-clinical samples has shown a greater association between amygdala activation and untrustworthy judgements in comparison to control participants who have not received a paranoia induction, providing further evidence of the role of amygdala hyperactivity in paranoid beliefs (Pinkham et al., 2022).

5.1.2. *Exploring trustworthiness judgments using EEG*

Although neuroimaging studies using fMRI provide information about the structure and function of brain areas involved in face evaluation, they are not able to provide insight into the temporal dynamics of neural activity when processing face information. Electroencephalographic (EEG) recording of neural activity is considered a reliable measure for capturing momentary changes in brain activation (Key et al., 2005). By acquiring the signal of synchronised post-synaptic activation of neuron populations in response to a stimulus in a specific time-window, researchers can explore the polarity and latency of the EEG signal associated with that event (Gantiva et al., 2020). This technique is referred to as event-related potential (ERP) and it provides informative components about specific cognitive processes and their association with brain activity (Key et al., 2005). Several ERP components concerning face processing have been identified. For example, early components such as the P100 consist of a positive polarity with the signal reaching its peak roughly around 100ms after stimulus onset with larger amplitudes reflecting visual selective attention and processing of lower level characteristic of stimuli (e.g., luminance, shape) at occipital regions (Smith et al., 2013). A key component when evaluating face stimuli is the N170, which peaks negatively around 130-

200ms after stimulus onset usually at lateral occipital, temporal and parietal sites. This component is thought to reflect the structural encoding of faces, particularly on the right hemisphere, and has been found to be sensitive to emotional expressions, with angry, fearful and happy faces generating greater amplitudes respectively in comparison to neutral ones (Hinojosa et al., 2015). Finally, a positive deflection approximately 300-400ms post-stimulus onset at central sites is defined as Late Positive Potential (LPP) which has been found to be related to attentional processes towards emotional stimuli, being enhanced when evaluating angry and fearful faces (Smith et al., 2013).

Considering that people with a diagnosis of schizophrenia tend to report difficulties in social cognitive skills (Green et al., 2015), numerous studies have explored the neural processing of faces in this clinical population. For example, a meta-analysis of 21 studies showed that N170 amplitudes were significantly smaller in the schizophrenia sample in comparison with non-clinical controls regardless of whether the faces were neutral or emotionally valenced (McCleery et al., 2015). Similarly, another meta-analytic study revealed that participants with a schizophrenia diagnosis exhibited reduced P100 amplitudes in comparison to non-clinical controls (Earls et al., 2016). However, in this case, the emotional valence of the stimuli did moderate the results indicating that the effect was stronger particularly when viewing happy and neutral faces. Finally, meta-analytic findings exploring differences between schizophrenia and control samples on LPP revealed a significant, albeit small, effect. Clinical participants exhibited a reduction in LPP amplitude when processing negative valenced faces but no effect was found for neutral or happy stimuli (Castro et al., 2019). Together, evidence from these studies suggests that individuals with schizophrenia might have a general deficit when encoding face stimuli. This reduced ERP amplitude appears to be greater for neutral and positive stimuli at very early processing stages and for negative stimuli during the latter stages. Nonetheless, although these findings are generalizable for

individuals with a diagnosis of schizophrenia, no study has explored these effects from a symptom-specific perspective.

Given the heterogeneous symptom presentation in schizophrenia spectrum disorders and their association with different underlying mechanisms, conducting studies from a symptom-based approach would be suitable for a better understanding of psychopathological conditions (Pinkham et al., 2016). For example, evidence from clinical studies has established clear differences in behavioural as well as on neuroimaging measures between paranoid and non-paranoid schizophrenia patients when processing social stimuli (Phillips et al., 1999; Pinkham et al., 2015, 2016). To the best of our knowledge, no studies have explored the temporal dynamics of trustworthiness evaluation in paranoid-clinical samples, nor in samples with a schizophrenia diagnosis. On the other hand, findings from non-clinical research looking at differences between trustworthy and untrustworthy evaluations of faces revealed enhanced P100, N170 and LPP amplitudes when behavioural judgements were congruent with the valence of the stimuli (e.g., trustworthy faces judged as trustworthy; Marzi et al., 2014). However, these effects did not remain when only looking at the influence of face valence, suggesting the involvement of decision-making processes in the early and late stages of trustworthiness evaluations. Conversely, another study focusing on the latter stages of face processing, found that valence did modulate LPP amplitude, which was larger when processing untrustworthy faces, stressing the role of structural face properties in trustworthy evaluations. Moreover, LPP predicted behavioural ratings of trustworthiness with larger amplitudes being associated with untrustworthy appraisals of faces, pointing to the role of attentional allocation processes when evaluating negatively valenced stimuli (Yang et al., 2011). Taken together, these findings appear to suggest an interplay between top-down and bottom-up processes, pointing towards a more nuanced understanding of trustworthiness appraisals. Nonetheless,

more studies are required to establish a clearer picture of the underlying mechanisms of trust and mistrust processes in non-clinical as well as clinical samples.

5.1.3. *Purpose of the present research*

This study used an ERP paradigm to explore the neural dynamics of trustworthiness face evaluations in clinically paranoid and non-clinical samples, and to study the associations between these judgments and behavioural, clinical and psychological variables. Based on the above-reviewed evidence and design we expect the following outcomes:

1. Levels of mistrust operationalized as behavioural outcomes as well as with self-report instruments will be significantly higher in clinical than non-clinical participants.

2. Participants in the clinical group will exhibit smaller P100 amplitudes in comparison to the non-clinical group and this effect will be moderated by valence, being particularly decreased in the trustworthy condition.

3. Reduced N170 amplitude will be evident for the clinical group however, this effect will not be modulated by the trustworthiness of the faces.

4. The control group will reflect an enhanced LPP amplitude in comparison to the control group, this effect will be moderated by the negative valence of the stimuli, untrustworthy faces.

5. Analyses will also be carried out to investigate whether the amplitude of ERP components related to different stages of face processing are associated with psychological variables such as attachment, mistrust, and paranoid traits.

5.2. Methods

5.2.1. *Participants*

Inclusion and exclusion criteria for clinical participants are detailed in Table 5.1. Inclusion criteria for non-clinical participants involved being 18 years old or over, fluent in English, and having no history of complex mental conditions. Recruitment of clinical participants was done through adult mental health inpatient and outpatient services in the English National Health Service (NHS).

Table 5. 1. *Inclusion and exclusion criteria for clinical participants*

Inclusion Criteria	Exclusion Criteria
1. Adult service users 18 years old or over who are currently inpatients or in receipt of treatment in the community.	1. Under 18 years old
2. Being clinically diagnosed with a schizophrenia spectrum disorder or other psychotic disorders (delusional disorder, schizoaffective disorder, schizophreniform disorder, and brief psychotic episode) as determined and evaluated by the clinical team.	2. Does not meet the diagnosis of Schizophrenia Spectrum Disorder (or other psychotic disorders)
3. Current or past experiences of persecutory or paranoid delusions (checked using the R-GPTS, 2021).	3. Participants with organic impairments or learning disabilities
4. Being fluent in the English language	4. Not fluent in the English language
5. Having the capacity to give informed consent as determined by their clinical team.	5. Being unable to provide informed consent

We used clinically validated scales of paranoia, depression, and general anxiety to confirm referrals from clinicians as well as the absence of severe mental health symptoms in the non-clinical condition. Out of 14 clinical participants who expressed an initial interest to their respective clinicians in the taking part in the study, only 9 agreed to participate and completed the experiment. Of those 9 participants, 5 had a diagnosis of paranoid schizophrenia

whereas the remaining 4 were diagnosed with schizoaffective disorder. From the total of clinical participants, 7 were recruited from inpatients wards while 2 were recruited from outpatient services. In the case of non-clinical participants, 15 agreed and took part in the study and were recruited through adverts and word of mouth.

5.2.2. *Instruments*

The Revised Green Paranoid Thought Scale (R-GPTS; Freeman et al., 2021). The R-GPTS is an 18-item scale designed to assess the severity of paranoid thoughts by evaluating persecutory and self-referential symptoms. Respondents need to indicate the degree to which certain thoughts and feelings may apply to them throughout the last month on a 5-point Likert scale (from 0 = *"Not at all"* to 3 = *"Totally"*). This scale has been validated in non-clinical, subclinical, and clinical samples providing clear cut-off scores to identify participants across the paranoia spectrum. The validation of this scale has revealed excellent psychometric properties and in this sample, internal reliability in this sample was very good (α=.96).

Patient Health Questionnaire – 9 (PHQ-9; Kroenke et al., 2002). The PHQ-9 is a self-report measure in which participants are asked to rate how often they experience, over the last two weeks, nine depressive symptoms on a 4-point Likert scale (0 = *"Not at all"* to 3 = *"Nearly every day"*. This scale allows to screen, monitor and inform the clinical diagnosis of depressive disorders and is being considered a valid and reliable tool of depression severity (α=.81).

Generalized Anxiety Disorder 7-item Scale (GAD-7; Spitzer et al., 2006). This scale requests participants to indicate the frequency in which a series of problems had an impact on them over the past 2 weeks on a four-point Likert scale from (0) *"Not at all"* to (3) *"Nearly every day"*. Higher scores indicates higher levels of generalised anxiety. The reliability of this scale in this sample was good (α=.82).

The Multi-theme delusional inventory (MTDI, Martinez et al., in prep*).* An adapted version of the Paranoid Deservedness Scale (PADS-R; Melo et al., 2009) was used to measure paranoia traits in both non-clinical and clinical populations in order to capture these beliefs as a continuum rather than focusing only on the severity of paranoid symptoms as measured by the R-GPTS. This scale revealed excellent internal reliability (α=.92).

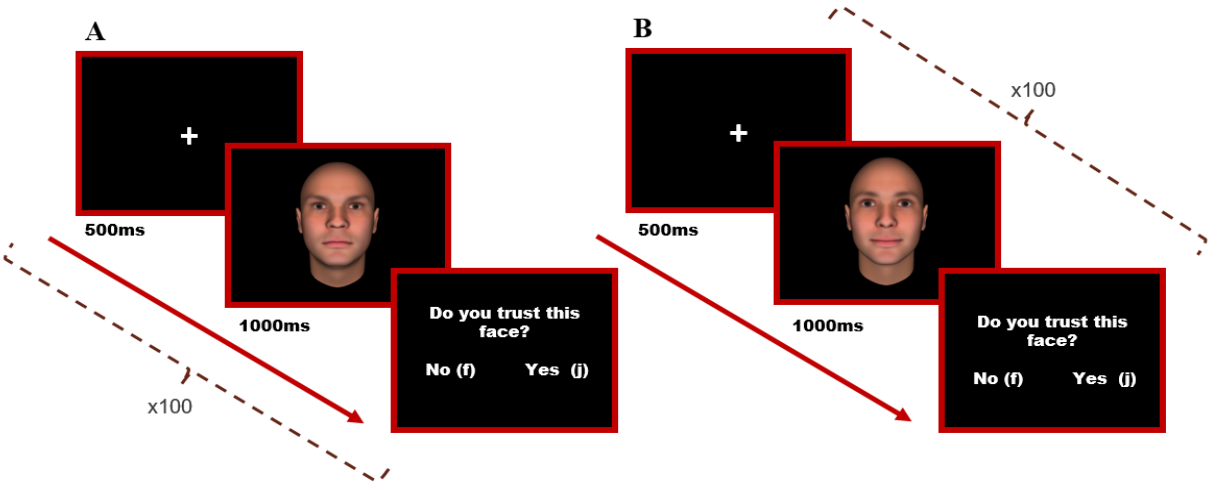
General interpersonal mistrust. Respondents are asked to indicate how much they agreed with the following statement *"Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?"* on a 5-point Likert scale ranging from 1 (*"Most people can be trusted"*) to 5 (*"Need to be very careful"*). This item has been adapted from the European Values Survey (Inglehart et al., 2000).

The Relationship Questionnaire (RQ; Bartholomew & Horowitz, 1991). To measure different attachment working models participants need to rate four vignettes describing secure, fearful, preoccupied, and dismissing prototypical patterns by indicating how adequately each description characterizes their own relationship style on a 7-point Likert scale (from 1= *"Disagree strongly"* to 7= *"Agree strongly"*). Based on these ratings, scores measures of attachment anxiety (negative model of self) and attachment avoidance (negative model of other) are computed with higher scores indicating the presence of secure styles.

Facial trust detection task (FTDT; Oosterhof & Todorov, 2008). To measure trustworthiness judgements we used data-driven computer-generated faces previously validated in the trustworthiness dimension obtained from the University of Chicago Perception and Judgement Lab database. From this database, a block of 20 Caucasian faces calibrated as more trustworthy (+ 3) and a block of 20 Caucasian faces calibrated as less trustworthy (- 3 SD) were selected and presented in random order. Each trial began with a central fixation cross that lasted for 500ms which was followed by the face stimulus that was presented for 1000ms at the centre of

the screen. Participants were asked to indicate their trustworthiness rating of the target stimulus immediately after it was displayed by pressing the *"J"* keyboard if they trusted it or the *"F"* keyboard if they did not. To improve the noise-to-signal ratio each participant completed each block five times leading to a total number of 100 trials per face category condition (see Figure 5.1.). Given the role of decisional processes in face evaluation, we computed trustworthiness ratings by implementing signal detection analysis based on Stanislaw and Todorov's (1999) calculations using equations 1 and 7 to calculate sensitivity and response bias parameters. Positive scores would indicate a tendency to judge an untrustworthy face as trustworthy in the case of bias, and higher accuracy in discriminating between a trustworthy and an untrustworthy face stimulus in the case of sensitivity.

Figure 5. 1. *Trial structure diagram for untrustworthy (A) and trustworthy (B) condition*



5.2.3. *Procedure*

For the clinical group, multi-disciplinary NHS clinical staff (i.e., consultant psychiatrists, consultant clinical psychologists, and mental health nurses) identified potential participants, discussed with them the nature of the study, and asked them about their willingness to participate. Individuals who were interested in participating met with the researcher who discussed information about the project such as what the study entailed (i.e., questionnaires, tasks, and EEG recording), potential risks (facing potentially sensitive

material), the reward for taking part (£10 Love2Shop voucher), and their right to withdraw without any consequences. After discussing the study and expressing interest in participating, the experimenter provided the participant information sheet followed by the consent form. Once written consent was provided, participants were asked to complete a series of questionnaires presented on a laptop using the online platform QUALTRICS. As described above, these questionnaires covered demographics, measures of trust, psychological measures (RQ), and mental health measures (PHQ-9, GAD-7, R-GPTS). Subsequently, the experimenter proceeded to put the EEG cap on the participant's head and asked them to wait a few moments while the electrodes were being calibrated. Once the EEG equipment was ready to start recording neural activity the participant started to perform the facial recognition task presented in the experimental software Open Sesame v4.0. Finally, participants stared at a fixation cross for one minute and then closed their eyes for another minute while their neural activity was being recorded at rest. Resting state data will not be analysed further in this study, but was obtained to allow for further analyses that may be carried out in the future. Upon completion, a debrief sheet about the study was provided to the participant as well as their £10 Love2Shop voucher. Throughout the study, participants had the chance to ask questions to the researcher at any point. Non-clinical control participants underwent the same process as clinical participants with the difference that they were recruited through the community by flyers or by word-of-mouth. Additionally, non-clinical participants met the researcher and engaged in the study in a psychological lab at The University of Sheffield, while clinical participants met the experimenter and participated at agreed NHS locations depending on the services they were attending. In total, the study lasted approximately 50 min for non-clinical participants and around 1 hour and 30 minutes for clinical participants. This difference was mainly attributed to the clinical participants taking more time to complete self-report measures which at times

were assisted by the experimenter. A Research Ethics Committee from the NHS Health Research Authority ethically approved this research study (Ref: 22/NW/0354).

5.2.4. *EEG data recording and pre-processing*

We recorded neural activity from 24 Ag/AgCl electrodes using a saline-based net for eego^{TM} amplifier system (ANT-Neuro) as part of a mobile EEG equipment ensuring that impedance was below 50kΩ. The electrical signal was amplified at a sampling rate of 500Hz and underwent filtering with a band-pass of 1-30 Hz. Noisy channels identified through visual inspection were removed and interpolated. Forehead ground and reference electrodes were used for common average referencing. Stimulus-locked epochs comprised 200ms before and 1000ms after stimulus onset and were extracted for each face condition. The signal recorded 200ms prior to stimulus presentation was baseline corrected for each epoch and artifacts greater than 40μV were automatically rejected. To carry out ERP analyses we selected the average signal from the P100, N170, and LPP components recorded at occipital (O1, O2), parietal (P3, P9, P4, P10), and central (C3, C4) sites respectively based on findings from Marzi et al. (2014). Then, we quantified each ERP component as the average voltage in a specific latency range, which encompassed between 110-130ms for P100, 130-220ms for N170, and 300-500ms for LPP. From the total of trials in both face conditions, 72% were accepted after artefact rejection across all participants in the clinical group whereas 80% were accepted for the non-clinical group. One parietal and two occipital channels were interpolated in the non-clinical participants whereas five parietal and one occipital channels were interpolated in the clinical participants. One non-clinical participant was not included in the ERP analysis due to the presence of very noisy data reflected in a very large number channels and trials removal making that data unusable.

5.2.5. *Statistical analysis*

Pre-processing of EEG and ERP data was carried out using the MATLAB-EEGLAB software toolbox whereas statistical analyses were conducted in SPSS v28.To analyse if there were any differences between the groups in demographics as well as in clinical symptoms chi-square as well as independent sample t-tests were conducted. To explore the effect of group and face category on the mean amplitude of the P100, N170, and LPP components we conducted a 2 (clinical, non-clinical) x 2(trustworthy, untrustworthy) repeated measures ANOVA. Given our small sample size and the exploratory nature of this study, we adopted a liberal approach to maintain statistical power by trying to keep as many degrees of freedom as possible. Thus, lateralization was not included as a factor in the repeated measures ANOVA but rather we conducted separate analyses for each location (left and right). Similarly, to explore the association between the mean amplitude of the ERP components and psychological variables, zero-order bivariate correlations were analysed in both samples combined.

A statistical power analysis was performed for sample size estimation, based on data from a meta-analysis (McCleery et al., 2015) following a similar design as proposed in our study. The weighted mean effect size (ES) in this study was considered to be medium (Hedge's g = .64) using Cohen's (1988) criteria. With an alpha = .05 and power = 0.80, the projected sample size needed with this effect size (GPower 3.1) is approximately N = 26 for a between/within-group comparison.

5.3. Results

Visual inspection using QQ-plots and histograms as well as Levene’s test for homogeneity of variance revealed that residuals of the main variables were normally distributed and that error variances were equal across conditions (Table S5.1; Figure S5.1.; Figure S5.2.).

5.3.1. *Sample characteristics*

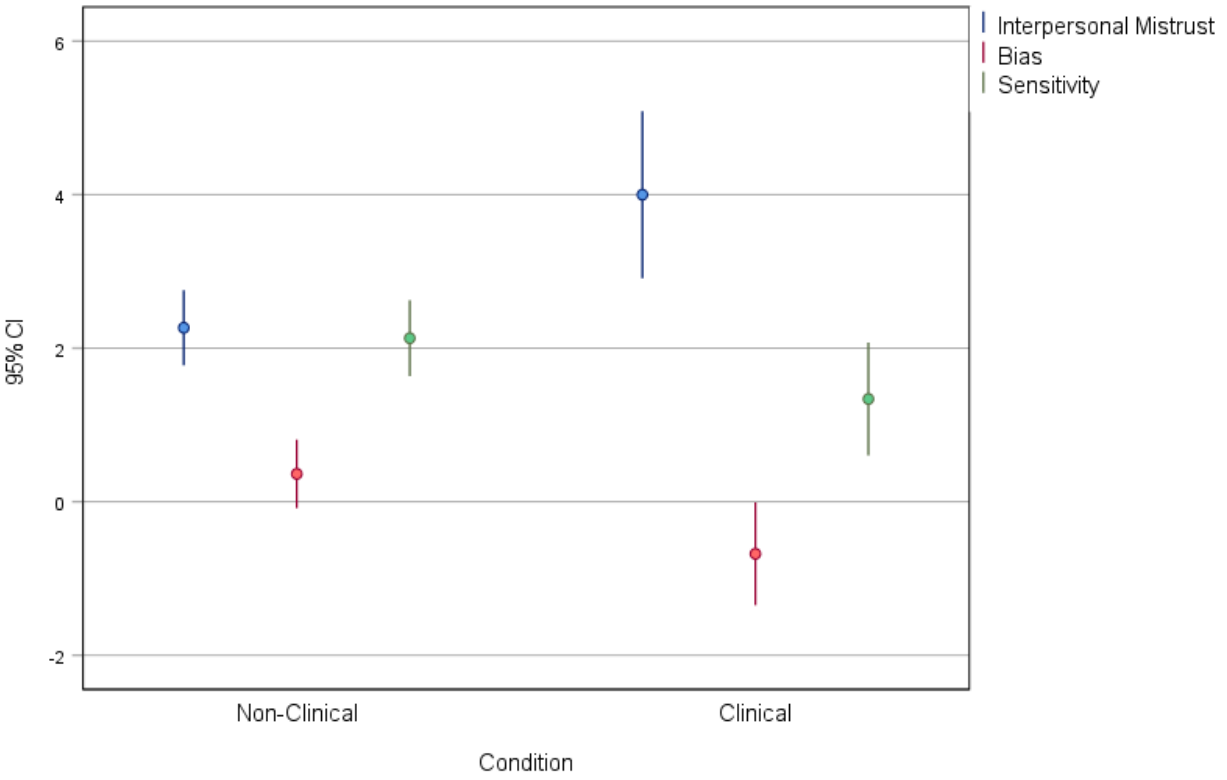
Descriptive statistics are presented in Table 5.2. Regarding demographics, non-clinical participants were significantly younger and more highly educated than clinical participants. However, no differences regarding gender or ethnicity were found between the groups with the majority of the sample being male and white. Concerning mental health symptoms, the non-clinical group reported significantly lower levels of anxiety and depressive symptoms in comparison to the clinical group. Moreover, ideas of reference and persecution scores were significantly higher in the clinical group, indicating moderately severe levels of paranoid symptoms (Freeman et al., 2021) confirming clinicians referrals. Equally, paranoia traits were significantly higher in the clinical group. All clinical participants were under psychopharmacological treatment when they took part in this study. As mentioned in the statistical analysis section, due to our small sample size and to preserve statistical power we decided only to include age as a covariate when analysing electrophysiological variables given its relevant association with ERP components related to face processing (Gao et al., 2009).

5.3.2. *Behavioural outcomes*

When considering trustworthiness ratings operationalized as signal detection outcomes as well as self-report mistrust, independent t-tests revealed significant effects of interpersonal mistrust [$t(22) = -3.71$, $p<.01$, $d_= .1.57$], bias [$t(22) = 2.98$, $p<.01$, $d_= 1.25$], and a marginally significant effect of sensitivity [$t(22) = 2.05$, $p=.05$, $d_= .86$]. These results reflected a marked tendency towards rating the faces as more untrustworthy in the clinical participants ($M= -.68$, $SD= .87$) in comparison to the non-clinical group ($M= .36$, $SD= .80$). Likewise, higher interpersonal mistrust was evident in the clinical ($M= 4.0$, $SD= .1.41$) than in the non-clinical group ($M= 2.27$, $SD= .88$). Although marginal, results concerning sensitivity indicated that participants in the non-clinical group ($M= 2.13$, $SD= .89$) had a higher degree of accuracy in

discriminating trustworthy from untrustworthy faces in comparison to the clinical group (*M*= 1.33, *SD*= .96; Figure 5.2.)

Figure 5. 2. *Error bar graphs showing mean and 95%CI for each of the trust variables*



5.3.3. *ERP results*

Repeated measures ANOVA revealed no significant effects of face condition [*F* (1, 20) = .25 *p*=.35, η_p^2 = .04] nor an effect for group [*F* (1, 20) = 1.94, *p*=.18, η_p^2 = .09] when considering the left P100 component. However, a significant interaction effect was revealed between face condition and group [*F* (1, 20) = 4.87, *p*=.04, η_p^2 = .20] with Bonferroni post-hoc pairwise comparisons indicating that in the clinical group, the P100 amplitude was significantly larger when evaluating untrustworthy faces (*M*= -.76, *SD*= .80) than trustworthy ones (*M*=.19, *SD*= .81) in comparison to the non-clinical group (Table 5.3.). Concerning the right P100 signal, findings revealed no significant effects for face condition [*F* (1, 20) = .74, *p*=.40, η_p^2 = .04], group [*F* (1, 20) = .79, *p*=.44, η_p^2 = .03], nor for the interaction between face condition and group [*F* (1, 20) = .53, *p*=.47, η_p^2 = .02].

Table 5. 2. *Descriptive statistics.*

Variables	Group			*p*
	Clinical *N=9*	Non-Clinical *N=15*	Total *N=24*	
Demographics				
Mean Age	41 (2.8)	30.3 (2.5)	28.9 (7.7)	<.01
Gender	89% (*M*)	73% (*M*)	79% (*M*)	.36
Education	22% (*HE*)	100% (*HE*)	70.8% (*HE*)	<.01
Ethnicity (White)	67% (*W*)	87% (*W*)	79% (*W*)	.24
Clinical Symptoms				
Ideas of Persecution	15 (13.2)	1.2 (2.9)	6.4 (10.6)	.01
Ideas of Reference	14.9 (10.2)	5.2 (4.9)	8.9 (8.6)	.02
Paranoid Traits	29.2 (8.3)	16.3 (5.5)	21.2 (9.1)	<.001
Depression	20.8 (5.1)	12.2 (2.5)	15.4 (5.6)	<.001
Generalised Anxiety	15.4 (5.7)	11.3 (3.4)	12.9 (4.7)	.04
Clinical Characteristics				
Mean Medication Dosing (mg)^{CPZ}	459.3 (200.9)	-	-	-
Illness duration (years)	13 (5.4)	-	-	-

Note. *SD in brackets for numerical variables. M= Male, HE= Higher Education (Undergraduate Degree or higher), W=White British/non-British, CPZ= Chlorpromazine equivalent*

A significant effect of face condition on the left N170 component [F (1, 20) = 5.05, p=.04, η_p^2 = .20] was found, with a higher amplitude in the trustworthy (M=1.35, SD= 1.8) than the untrustworthy (M=1.23, SD= 1.9) condition. However, a significant interaction effect between face conditions and the covariate age [F (1, 20) = 5.54, p=.03, η_p^2 = .22] was revealed, suggesting that differences in face conditions were influenced by age. There was no significant interaction effect between face and group conditions [F (1, 20) = 2.5, p=.12, η_p^2 = .11] nor an effect for the group variable [F (1, 20) = 1.52, p=.23, η_p^2 = .07]. Regarding the right N170

component, no significant effect of face condition [F (1, 20) = .26, p=.59, η_p^2 = .01], nor face condition and group interaction [F (1, 20) = .09, p=.77, η_p^2 =.004] was found. However, a marginally significant effect of group was found [F (1, 20) = 4.33, p=.05, η_p^2 = .18] revealing that the amplitude was enhanced in the non-clinical group (M=2.26, SD= .44) in comparison to the clinical group (M=.44, SD= .59).

Table 5. 3. *Pairwise Comparisons*

	P100 (Left)							
Condition	(I) Trustworthy	(J) Untrustworthy	M_{diff} (I – J)	SE	Sig. [b]	η_p^2	95% CI [LB / UB]	
Non-Clinical	1	2	-.06	.23	.79	.004	-.56	.43
Clinical	**1**	**2**	**.95**	**.32**	**<.01**	**.31**	**.28**	**1.62**

Note. Based on estimated marginal means [b]. Adjustment for multiple comparisons: Bonferroni.

Regarding, the left LPP component, no significant findings were evident for face condition [F (1, 20) = 1.23, p=.26, η_p^2 = .06], group [(1, 20) = .10, p=.76, η_p^2 =.005], nor interaction effect between face condition and group [F (1, 20) = .11, p=.74, η_p^2 = .005]. Similarly, results from the right LPP component were non-significant for the effect of face condition [F (1, 20) = .33, p=.60, η_p^2 = .02], group [F (1, 20) = 1.67, p=.21, η_p^2 = .08], nor for the interaction effect of face condition and group [F (1, 20) = .22, p=.65, η_p^2 = .01]. Grand averages of each ERP component stratified for each group and each face condition is shown in Figure 5.3.

5.3.4. *Correlation analysis*

Exploratory bivariate correlation analysis between the amplitude of significant ERP components and psychological variables was conducted in both groups combined. Given that paranoia traits scores were better distributed in the combined sample than paranoid symptomatology (Figure S5.3.) we decided to employ the traits measure for this correlation analysis. Negative significant moderate associations between paranoid traits and the left P100 component in untrustworthy conditions (r=-.46, p<.05), the left N170 component in trustworthiness conditions (r=-.53, p<.01), and the right N170 component in untrustworthiness conditions (r=-.41, p<.05) were found. Similarly, a negative significant association was also found between the left N170 component in trustworthiness conditions and interpersonal mistrust (r=-.44, p<.05; Table 5.4.; Figure S5.3.).

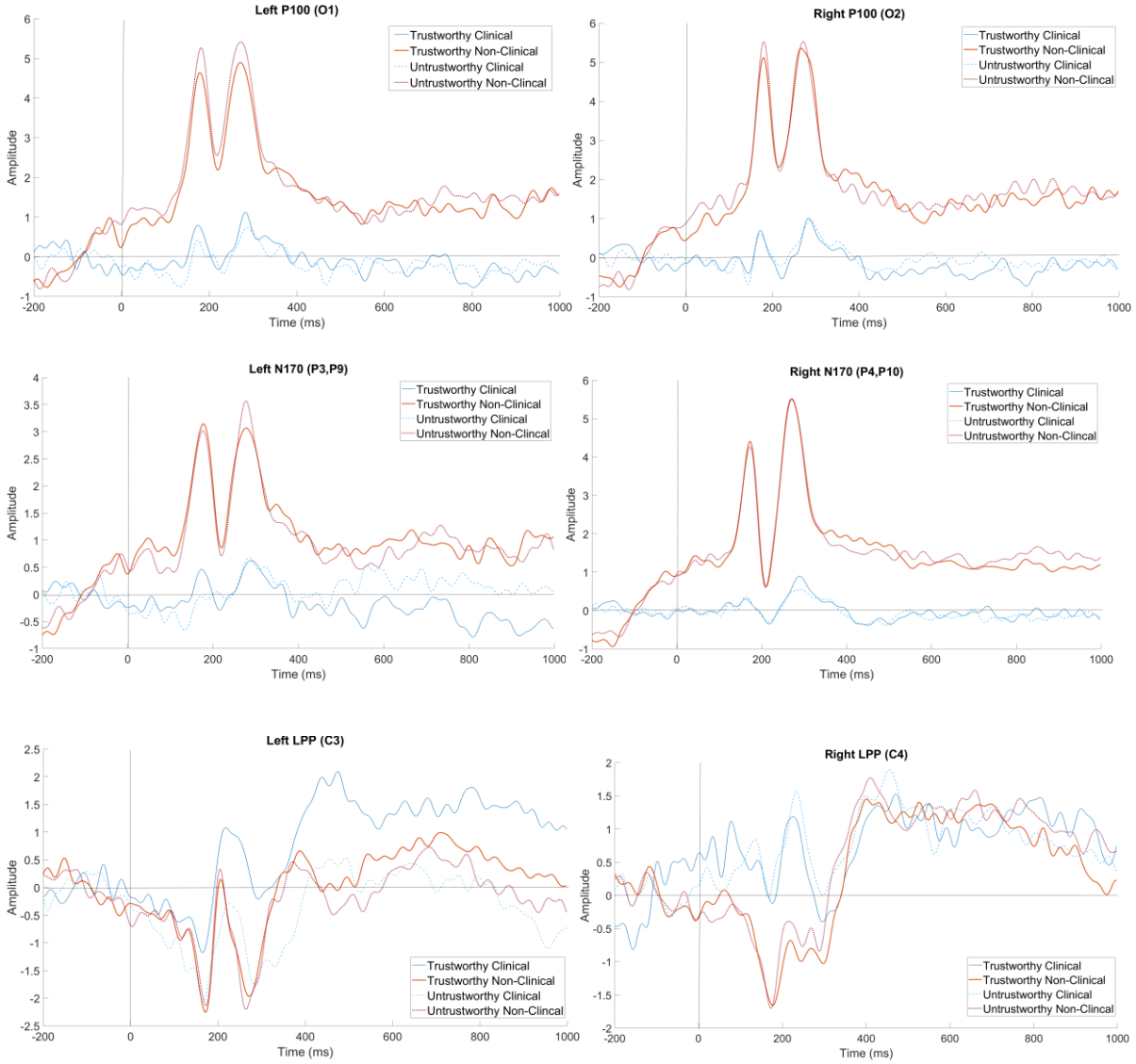
Table 5. 4. *Bivariate correlation between psychological variables and ERP components (combined sample).*

Variables	1	2	3	4	5	6	7	8	9	10	11	12
1. Paranoid Traits	1	-.59**	-.35	.66**	.32	-.32	-.32	-.46*	-.53**	-.34	-.36	-.41*
2. Bias		1	.13	-.70**	-.09	-.42*	.22	.25	.33	.32	.30	.24
3. Sensitivity			1	-.20	.51*	.19	.07	.21	.25	.10	.14	.26
4. Interpersonal Mistrust				1	.18	-.49*	-.29	-.38	-.44*	-.34	-.39	-.40
5. Attachment Anxiety					1	-.03	-.04	-.06	.07	.32	.07	-.05
6. Attachment Avoidance						1	.16	.21	.12	.27	.30	.18
7. P100 Left Trustworthy							1	.93**	.75**	.73**	.74**	.74**
8. P100 Left Untrustworthy								1	.80**	.77**	.76**	.77**
9. N170 Left Trustworthy									1	.83**	.74**	.75**
10.N170 Left Untrustworthy										1	.81**	.73**
11.N170 Right Trustworthy											1	.89**
12.N170 Right Untrustworthy												1

*Correlations are significant at the .05 level

*Correlations are significant at the .01 level

Figure 5. 3. *Grand averages for trustworthy and untrustworthy conditions for each group and ERP component*



5.4. Discussion

This study aimed to explore the underlying psychophysiological mechanisms of trustworthy evaluations of unfamiliar faces in a clinically paranoid sample in comparison to a non-clinical control. For our first hypothesis, we expected that the clinical sample would be characterized by higher levels of mistrust measured by self-report as well as behavioural measures. This was confirmed as clinical participants reported elevated scores of general interpersonal mistrust than non-clinical controls. Likewise, clinical participants exhibited a strong tendency to judge face stimuli as untrustworthy and, to a lesser degree, were less accurate in discriminating a trustworthy face from an untrustworthy one. Mistrust is considered

a core feature of paranoia (Bell & O'Driscoll, 2018) and findings from large international representative non-clinical samples have found that a mistrust bias is specifically associated with paranoid beliefs (Martinez et al., 2022). Although differences in trustworthiness ratings have also been found between participants with persecutory delusions and non-clinical controls (Pinkham et al., 2008), no previous study has done so by operationalizing mistrust from a signal detection perspective. Thus, these findings appear to provide a more nuanced understanding of trustworthiness processes in clinical populations by indicating that clinically paranoid participants have a more conservative criterion when deciding whether or not to trust an unfamiliar face.

Regarding psychophysiological analysis, we first predicted that the amplitude for the P100 component would be significantly smaller in the clinical group and that this difference would be modulated by the valence of the face stimuli. This finding was fully met in the left location of the electrode and is in line with meta-analytic findings from Earls et al. (2016) with the only difference being that those authors found that the effect was modulated by happy and neutral faces only. Conversely, our findings revealed an enhanced amplitude in the clinical group when evaluating untrustworthy face stimuli in contrast to trustworthy ones. It is possible that our lack of a neutral condition and the fact that we only included a paranoid clinical sample in our study might reflect our different results regarding the effect of face valence. The P100 component is considered an index of early visual selective attention processing with studies suggesting its involvement in fear conditioning (Pizzagalli et al., 2003). Thus, since paranoid participants are characterized by hypervigilance states these findings could reflect the neural sensitivity of paranoid patients in allocating attentional resources toward potentially dangerous stimuli.

Concerning our second hypothesis, a greater amplitude was expected for the non-clinical group regardless of face valence. This hypothesis was also fully met in the right

hemisphere, with significantly different amplitudes between both groups. Nonetheless, in the left location, an effect of face valence and age was found, meaning that a difference in amplitudes between trustworthy and untrustworthy faces was explained by individual differences related to age. Several studies have shown right-lateralization asymmetries when analyzing the N170 component (Hinojosa et al., 2015) revealing stronger amplitudes in the right hemisphere of the brain compared to the left. Moreover, a study by Marzi et al. (2014), on which we based our experimental paradigm, only found an effect of face conditions on the right N170 component when trials were congruent with the responses. Moreover, by visually inspecting the ERP plots of our study one can appreciate a cleaner signal in the right component in comparison to the left one. Hence, the left parietal electrodes might have been more sensitive to noise in comparison to the right ones. Nonetheless, our findings seem to be aligned with meta-analytic evidence of a lesser N170 amplitude evident in clinical populations in comparison to non-clinical controls reflecting a possible deficit of paranoid participants when encoding faces.

For our last electrophysiological hypothesis, we predicted a significantly enhanced LPP amplitude in clinical participants and that this effect would be modulated by the negative valence of the face stimuli. This hypothesis was not confirmed, as no significant differences were found between groups or face conditions. Evidence from a meta-analysis found a small effect regarding this component between clinical and non-clinical samples (Castro et al., 2019). Hence, the absence of an effect in our study is not surprising given our small sample size and possibly the lack of power to detect a small effect. Moreover, similar to the left N170 component, right and left LPP signals appear to be less clear for the central electrodes, being probably more sensitive to external noise.

Finally, when exploring bivariate associations between significant ERP components and psychological variables, negative moderate associations were found between the left P100

and paranoid traits. Neuroimaging studies have reported that when processing negative valenced faces, the left side of the amygdala activity appears to be more enhanced than the right side (Fusar-Poli et al., 2009). Thus, it is not surprising that the allocation of attentional resources to untrustworthy stimuli reflected in a nuanced P100 amplitude would be associated with higher paranoid traits, particularly on the left side. Likewise, paranoid traits and interpersonal mistrust were negatively associated with the left N170 component, although under trustworthy conditions. Additionally, higher paranoid traits correlated with lesser right N170 amplitudes when evaluating untrustworthy faces. Hence, deficits in the structural encoding of trustworthy and untrustworthy faces may be influenced by hypervigilant and defensive attitudes toward others. Insecure attachment styles did not correlate significantly with any of the electrophysiological variables nor with paranoid symptomatology. Nonetheless, a negative moderate association between attachment avoidant and trustworthiness bias, as well as a positive association between attachment anxiety and sensitivity, were found. This is in line with findings in non-clinical samples showing that holding negative views of others, which is a characteristics of avoidant attachment, can lead to a biased tendency in judging trustworthy faces as untrustworthy (Martinez et al., 2022). Conversely, a positive view of the self and others (reflected in positive scores on the Relationship Questionnaire) can lead to improved discrimination of trustworthy from untrustworthy unfamiliar faces. Given our small sample size and the exploratory nature of this study, these findings should be interpreted with caution as the strength of these associations, particularly with paranoia measures, is probably led by high scores reported in the clinical group.

5.4.1. *Limitations*

This study had several limitations that should be considered when interpreting the above-presented results. First, the sample size of this study was relatively small for conducting between-within-subjects analyses. The projected sample size needed for detecting a medium

effect following the design outlined in this study was 26, indicating that our sample of 23 for analysing the main variables was underpowered. Nonetheless, considering the challenges of recruiting participants with severe mental health conditions, the obtained results still hold practical significance, as evident from the large effect sizes and significant or marginally significant findings. Second, following our sample size limitation both groups were mainly male and white, limiting their representativeness of the general as well as clinical populations, particularly regarding gender and ethnicity. Moreover, the non-clinical participants were highly educated and considerably younger than those from the clinical group. Although age was included in all main analyses as a covariate, differences found in the outcome variables might be influenced by levels of education. However, as mentioned in previous sections, a liberal statistical approach was adopted when conducting the respective analyses as preserving statistical power was a priority in this study. Third, the stimuli used for the experimental task were restricted to bald, male, and Caucasian faces which limits the generalizability of the faces to other genders or ethnicities. Moreover, while the validated stimuli used in this study offer several advantages for psychophysiological research (i.e., control over stimulus characteristics and individual differences; Oosterhof & Todorov, 2008) the avatar-like appearance of the faces may impact the ecological validity of the task. Another methodological limitation entails the absence of a neutral condition in our experiment, restricting the possibility to compare neural activity when evaluating valenced as opposed to neutral faces. This decision was made to mitigate potential fatigue effects, particularly among clinical participants when completing the face recognition task. Furthermore, given that the P100 component is also associated with the perception of other types of stimuli (e.g., objects), the lack of a non-face condition in our design limits our conclusions. Additionally, while clinical participants completed the experiment in various locations depending on the service they were attending, non-clinical participants completed the task in controlled psychological lab environments where extraneous variables

such as ambient lighting and noise were kept constant. As a result, findings from this study may have been influenced by extraneous sources of variability that were better controlled in the non-clinical compared to the clinical group. Also, a 20-30% of total trials across participants were removed due to the presence of artifacts (e.g., eyeblink) reflected in a lower signal-to-noise ratio, this would suggest that a design with more trials is preferable. Lastly, all clinical participants were in receipt of pharmacotherapy at the time of the experiment hence, differences found on the EEG signal between groups could be influenced by the effect of the antipsychotic medication.

5.4.2. *Conclusions*

The purpose of this proof of concept study was to test the psychophysiological underpinnings of trustworthiness evaluations by collecting electrophysiological data using mobile EEG equipment from a clinical sample with severe mental health conditions. Findings from our study demonstrate the feasibility of conducting such investigations highlighting the importance of designing studies that account for the specific characteristics and context of the clinical population to study. Overall, our results appear to be consistent with the existing scientific literature regarding the temporal dynamics of face evaluation suggesting that participants with a schizophrenia diagnosis exhibit differences in encoding face stimuli at early stages, as reflected in smaller amplitudes. Moreover, our findings suggest that the early stages of processing faces, particularly when evaluating stimuli signalling untrustworthiness, may play a role in paranoid traits. However, future research with larger sample sizes and the inclusion of a non-paranoid clinical sample, should be conducted to further explore the influence of clinical as well as psychological variables on face processing and examine the interplay between psychophysiological as well as behavioural aspects of trustworthiness. Taken together, these findings contribute to our understanding of the psychophysiological processes underlying trustworthiness evaluations in clinical and non-clinical populations and

emphasize the importance of considering the specific characteristics of individuals with severe mental health conditions when designing psychophysiological research studies.

Chapter VI -General Discussion

Contributions

Anton P. Martinez (conceptualization, writing – original draft)

Elizabeth Milne (supervision)

Richard P. Bentall (supervision, review, and editing)

Summary

This chapter aims to summarise the evidence presented across the four empirical studies in this thesis. Moreover, the theoretical, methodological, and clinical implications of these findings are discussed. Following these implications, this chapter reflects on the strengths and limitations of the studies presented in this thesis as well as future directions regarding the research of mistrust and paranoia. Finally, an overall conclusion of this thesis is provided.

6.1. Thesis main findings

The current thesis aimed to explore the role of trust processes and their relationship with paranoid beliefs in clinical and non-clinical populations. Moreover, we intended to adopt a socio-cognitive and evolutionary approach by implementing signal detection theory as well as a relational framework for the study of mistrust. Through epidemiological, experimental, and clinical studies we tried to expand current conceptualizations of paranoia contributing to the already established defensive and cognitive psychological models.

The first empirical study explored whether different forms of trust were particularly associated with either paranoid beliefs or conspiracy mentality. To achieve that, data on different measures of trust such as institutional trust, trust in different sources of information, interpersonal trust, and trustworthiness judgements were collected from three large international representative samples (Spain, the UK, and the Republic of Ireland). Our findings revealed that a bias towards evaluating face stimuli as untrustworthy was specifically associated with paranoia whereas mistrust in political institutions was solely related to conspiracy mentality. Moreover, while both constructs were associated with interpersonal mistrust this relationship seemed particularly stronger for paranoia than for conspiracy mentality.

Given the specific relation between mistrust bias and paranoia, the second empirical study aimed to further explore the role of trustworthiness in paranoid beliefs. For this purpose, this study examined whether trust judgements operationalized as signal detection outcomes (i.e., bias and sensitivity) were associated with insecure attachment styles, negative self-esteem, and paranoia in a large UK non-clinical representative sample. Additionally, a model was tested to explore the indirect effect of bias and sensitivity between insecure attachment styles and paranoia while controlling for the mediational pathway of self-esteem. Results revealed significant associations between paranoid beliefs and response bias as well as

sensitivity, although the association for the former was greater. Moreover, response bias was also related to both insecure anxious, and avoidant styles but not negative self-esteem. Additionally, a significant indirect pathway involving mistrust bias was found between both insecure attachment styles and paranoia, although negative self-esteem only appeared to mediate between anxious attachment and paranoia. A model where an indirect effect was added for sensitivity resulted in non-significant results, strengthening the notion that a tendency to mistrust unfamiliar faces, rather than difficulties in discriminating between them, seems to better explain the relationship between insecure relational patterns and paranoia.

The third empirical study intended to further explore the potential causal role of relational schemas since insecure attachment styles appear to be a precursor of a tendency toward untrustworthiness leading to paranoid beliefs. For this purpose, an experiment was conducted with an online convenience sample. By implementing an evaluative classical conditioning intervention pairing participant's personal information with likeable, threatening, and neutral faces, no effects were found on mistrust judgements operationalized as affective priming outcomes. Moreover, the intervention affected neither paranoia nor explicit or implicit self-esteem or attachment avoidance. Only a marginal effect was evident for attachment anxiety, being lower for participants whose personal information was consistently paired with positive face stimuli.

Finally, the fourth empirical study intended to test the feasibility of conducting a psychophysiological study by measuring the brain activity of non-clinical and clinically paranoid participants while evaluating the trustworthiness of face stimuli. Findings from this last study revealed, in an exploratory manner, that clinical participants appear to allocate attentional resources to untrustworthy stimuli at very early stages of face processing in comparison with non-clinical controls. Additionally, clinical participants showed marked differences when encoding faces regardless of the valence of the stimuli in comparison to

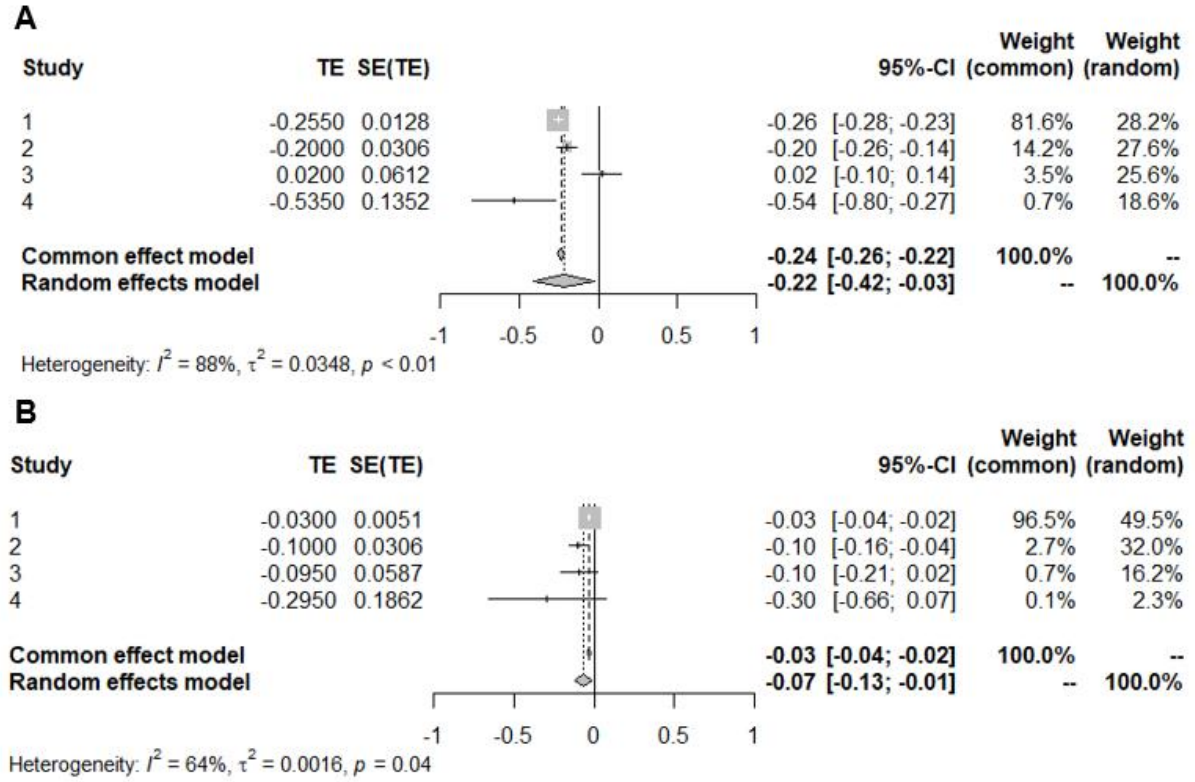
controls, as the neural activity was significantly greater for the latter. These neural signals appear to be influenced by higher paranoid traits as well as interpersonal mistrust in both clinical and non-clinical samples. When examining psychological variables, only a bias toward mistrust was associated with greater paranoid traits, whereas insecure attachment styles and sensitivity were not.

The main findings of this thesis suggest a consistent association between paranoid traits and trustworthiness judgements operationalized as signal detection outcomes. A mini correlation meta-analysis integrating the Pearson coefficients of each empirical study revealed a small-to-moderate association between paranoia and mistrust bias [r=-.22, 95%CI (-.42,-.03)], and, while trivial, a significant negative association with sensitivity [r=-.07, 95%CI (-.13,-.01); Figure 6.1.][13]. Although the heterogeneity between studies was high due to differences in sample size, the populations of the studies, and methodological designs, associations between measures of paranoid traits and responses to face stimuli were consistent. Thus, albeit interpreted with caution, these findings seem to support the notion that perceptual biases concerning the untrustworthiness of unfamiliar faces rather than a difficulty in accurately discerning between valenced stimuli contribute to paranoid interpretations Conversely, the role of insecure attachment style was less consistent throughout the empirical studies, particularly when trying to explore it from an experimental perspective in empirical study 3. Nonetheless, findings from empirical study 2 and associations between attachment avoidance and mistrust bias in study 4, inform us about the possible importance of the relationships between insecure attachment styles and paranoid beliefs. Interpersonal mistrust was not a key variable when answering most of the main aims of each empirical study in this thesis however, it was measured in empirical studies 1,2, and 4. Just by looking at bivariate associations between those studies, one can appreciate moderate to large associations between

[13] The mini meta-analysis was performed using R metafor package (Viechtbauer, 2010)

general interpersonal mistrust and paranoid traits and to a lesser degree with insecure attachment styles (for associations in empirical study 2 see Table S3.1.).

Figure 6. 1. *Forest plot of the effect sizes (Pearson's r) for studies measuring the association between paranoid traits and Bias (A) as well as Sensitivity (B). Error bars indicate 95% confidence intervals of the effect.*



6.2. Theoretical implications

6.2.1. *Individual versus collective vulnerability*

Findings from the first empirical study contribute to the notion that conspiracy mentality and paranoid beliefs in non-clinical populations are two different constructs that, to a lesser degree, are associated with each other. Moreover, expected vulnerability to the potential actions from other parties concerned political institutions in the case of conspiracy mentality and the perception of unfamiliar faces in the case of paranoia. Although the general belief that others should not be trusted was related to both constructs, this was particularly associated with paranoid beliefs. Our results are in line with meta-analytic findings showing

that conspiracy beliefs were associated with mistrust in government but less so with general people, while the opposite was evident with paranoid traits (Imhoff & Lamberty, 2018). Social and clinical psychologists have argued that, while individuals with a conspiracy mentality tend to feel vulnerable to powerful groups, people with paranoia have a general tendency to perceive potential danger from other people irrespective of their social status (Freeman & Bentall, 2017; Imhoff & Lamberty, 2018). Hence, this perceived threat appears to be directed towards the self in the case of paranoia and towards society at large for conspiracists. Studies employing large samples have shown that, while both belief systems are associated with poor locus of control and loneliness, paranoia is characterized by negative self-esteem and insecure attachment whereas conspiracy mentality is related to positive self-esteem and narcissism (Alsuhibani et al., 2022). Thus, the uncertainty generated by negative expectations of the actions of others, either as part of a selective group or individually, appears to activate different psychological mechanisms that can lead to conspiracist or paranoid beliefs. Therefore, the finding that a bias towards perceiving unfamiliar face stimuli as untrustworthy is specifically associated with paranoid traits adds to the current conceptualizations of paranoia which in turn helps to differentiate between these two types of belief systems.

6.2.2. *The explanatory role of mistrust in paranoid beliefs*

Evidence from the second empirical study proposes mistrust bias as a possible mediator in the well-established relationship between insecure attachment styles and paranoid beliefs (Murphy et al., 2020) in non-clinical populations while controlling for negative self-esteem. These results are aligned with meta-analytic findings revealing moderate to large associations between paranoia and both self and other-negative schemas (Humphrey et al., 2021). Moreover, studies have found that negative beliefs about the self and others mediated the effect between insecure attachment styles and paranoia (Sood et al., 2022) arguing that early attachment disruptions can lead to paranoid interpretations through personal vulnerability and

perceived danger from others. However, most of these studies have conceptualized negative other-schemas as a general belief rather than a specific construct. Fewer studies have explored mistrust as an explanatory role between adverse experiences and paranoid thoughts in non-clinical populations. For example, studies looking at the association between paranoia and adverse childhood memories found that interpersonal mistrust mediated that relationship (Murphy et al., 2012). Moreover, epidemiological findings have revealed that lack of trust was a mediator between environmental risk factors, such as social deprivation, and paranoid beliefs (Wickham et al., 2014) suggesting that not only psychological but also social adversities contributes to paranoid experiences through mistrust. Additionally, studies looking at psycho-social variables have found that interpersonal trust explains the negative relationship between stronger national group identification and lower paranoid cognitions (Greenaway et al., 2019). Finally, a study conducted in a clinical sample with persecutory delusions found that distrust-based competition mediated the association between state paranoia and uncooperative behaviours (Ellett et al., 2023). Altogether, the above-mentioned evidence would indicate that mistrust can be considered an underlying mechanism that could explain the association between personal as well as social adversities and paranoid beliefs. Thus, our findings expand this understanding by including bias toward untrustworthiness, a feature of mistrust that has not previously been regarded as a potential mechanism.

6.2.3. *Psychological models*

Current psychological models of paranoia have considered insecure attachment styles as a precursor of this type of belief. However, as discussed in the previous section, the potential mediational role of mistrust between those constructs is less clear. On the other hand, negative self-esteem appears to be a consistent mediator between insecure attachment and paranoia, and its role has been further developed in the paranoia literature. For example, the defensive model of paranoia conceptualized self-esteem as a discrepancy between an “ideal” and “actual” self,

which would lead to a self-serving bias to reduce negative feelings generated by that discrepancy (Bentall et al., 2001). Instead, the direct model of paranoia argues that negative self-esteem directly reflect feelings of vulnerability concerning others, leading to paranoid interpretations about those negative emotional states (Freeman et al., 2005). Hence, discrepancies between implicit and explicit self-esteem are not necessarily considered in this model.

Through an evaluative conditioning intervention, the third empirical study intended to elicit different trustworthiness evaluations as well as implicit and explicit self-esteem responses as a result of experimentally manipulating relational schemas. By triggering negative relational schemas one would expect a tendency toward judging face stimuli as untrustworthy, revealing the precursor role of relational schemas in mistrust bias. Moreover, whereas negative implicit and positive explicit self-esteem levels after a negative activation of relational schemas would support the defence model, both negative explicit and implicit self-esteem scores would provide evidence for the direct model of paranoia. Nonetheless, findings from this experiment did not support the role of mistrust bias nor the defence or direct model of paranoia. However, this could have been either due to the lack of sufficiently strong interpersonal stimuli required to elicit a negative state in the participants or because the intervention was not sensitive enough to have an effect. Previous studies using the same classical conditioning intervention found effects on implicit self-esteem when employing the positive intervention, or higher levels of state paranoia when triggering negative relational schemas (Baccus et al., 2004; Espinosa et al., 2018; Trucharte, 2022). However, none of them included all measures presented in empirical study 3 to reflect the different models of paranoia.

Although researching self-esteem was not the main focus of this thesis, when measured in empirical studies 2 and 3, moderate to strong associations were found with paranoid traits. Nonetheless, while implicit self-esteem was not significantly correlated with paranoid traits, it

was moderately associated with state paranoia before the classical conditioning intervention (Table S4.2.). Thus, while negative explicit self-esteem might feed paranoid cognitions through feelings of vulnerability, implicit self-esteem might have a stronger influence on momentary paranoid states than on more global and stable traits. An important aspect to consider in this thesis is that in both empirical studies 2 and 3 self-esteem was not (or only trivially) associated with mistrust bias and interpersonal mistrust (Table S3.1.), suggesting that both mistrust and negative self-esteem may contribute to paranoid beliefs through different pathways.

6.2.4. *Face perception and mistrust*

Findings from the fourth empirical study revealed that clinically paranoid participants appear to be more hypervigilant to untrustworthy stimuli at the initial phases of face processing and presented encoding differences when processing faces in comparison to non-clinical controls. These results are in line with evidence from fMRI studies showing an association between amygdala hyperactivation and persecutory symptoms as well as paranoid states (Pinkham et al., 2022; Pinkham et al., 2015). Moreover, these findings are also aligned with evidence from non-clinical samples showing amygdala activation when processing untrustworthy stimuli (Todorov et al., 2008). Additionally, results from empirical study 4 replicated meta-analytic findings showing significantly reduced right N170 amplitudes in patients with a schizophrenia diagnosis in comparison to non-clinical participants (McCleery et al., 2015). Furthermore, the fact that the valence of face stimuli did not moderate the effect of the N170 amplitudes is consistent with findings in non-clinical samples using the same stimuli dataset that was employed in empirical study 4 (Marzi et al., 2014). Nonetheless, when looking at later components of face processing, no significant differences were found between the clinical and non-clinical groups in contrast to previous findings (Castro et al., 2019). However, this discrepancy may suggest either that more statistical power is needed to detect the same effect or that paranoia is linked to rapid-bottom-up processes when regarding face

evaluation. Some authors argue that face valence underlies a general principle when making a first impression that leads to approach or avoidance responses (Chen & Bargh, 1999; Todorov et al., 2008). Paranoid beliefs are characterized by hypervigilant states towards potentially dangerous social stimuli. Therefore, it might not be surprising that clinically paranoid participants would early detect negative valenced stimuli reflected in enhanced P100 amplitudes. This heightened sensitivity to negative stimuli would lead to a general perception of untrustworthiness which in turn would result in a decreased allocation of cognitive resources to the encoding of faces reflected in lower N170 amplitudes and biased face judgements. It is worth noting that these explanations are highly speculative and that further research is needed to confirm these hypotheses. Nonetheless, findings from empirical study 4 appear to contribute to the understanding of face evaluation in paranoid participants when attributing trustworthiness judgments to face stimuli by employing temporal dynamics techniques to the study of these processes.

6.3. Methodological implications

This thesis is characterized by employing different methods for studying paranoia and its association with mistrust. Hence, several methodological aspects should be discussed. First, the use of the face stimuli dataset in this thesis allowed us to compare the effects between the four empirical studies as well as with empirical findings from other experiments. Second, conducting online experimental studies without the use of reliable participant recruitment platforms can lead to unreliable results, thus the implementation of attentional checks and analysis of survey completion time can be convenient for the removal of careless respondents. Conversely, using appropriate online experimental software can provide an array of tools such as proper randomization techniques as well as features for designing longitudinal studies improving the quality of the experiment. Nevertheless, trying to elicit negative or positive psychological states without controlling environmental variables can introduce a lot of external

error such as variations in the size of devices, ambient noise, lighting and temperature which could impact the experimental manipulation. Thus, conducting experiments in adequate psychological lab environments might be preferable when trying to activate relational schemas. Finally, unlike traditionally used lab-based EEG equipment that involves a higher number of electrodes, (varying from 32-128) we used mobile EEG equipment with 24 electrodes. Although the former can provide greater brain coverage and higher sensitivity, the conductance between the scalp and the electrode is achieved by using gel. Conversely, the mobile EEG equipment used in empirical study 4 used a saline solution which was less invasive and less time-consuming than gel-based EEG systems. Our findings, are not that far from those reported in the previously cited meta-analysis, indicating that the mobile EEG equipment provided reliable results. Moreover, given the characteristics of the clinical sample, implementing less-time consuming and less-invasive procedures was prioritized. Thus, the application of mobile EEG equipment when testing clinical samples with schizophrenia diagnosis for conducting ERP experiments is feasible.

6.4. Clinical implications

Recent meta-analytic evidence regarding the efficacy of Cognitive Behavioural Therapy for psychosis (CBTp) has found modest effects for the treatment of delusions in patients with schizophrenia spectrum disorders (Turner et al., 2020). Moreover, these effects have not been disseminated concerning persecutory beliefs. Given that each delusional system answers to different mechanisms, general CBTp techniques might not be sufficient for achieving efficacious treatment outcomes. A series of key causal factors such as worry, negative beliefs about the self, interpersonal sensitivity, sleep disturbances, abnormal internal experiences, and reasoning biases, have been proposed to be taken into account when treating patients with persecutory delusions (Freeman & Garety, 2014). Thus, several interventions have been adapted to target these mechanisms such as mindfulness, meta-cognitive training,

and CBT for insomnia and worry (Ellett et al., 2020; Freeman et al., 2017; Garety et al., 2021; Moritz et al., 2022). However, most of these clinical approaches seem to focus on cognitive mechanisms which are characterized by being slow and deliberative, thus clinical research aiming at other psychological factors such as mistrust and attachment appears to be scarce.

Findings from this thesis suggest that other pathways such as early attachment disruptions and mistrust judgements might also contribute through fast and implicit processes to paranoid thinking in clinical and non-clinical populations. Within this context, clinical interventions focused on relational components such as schema therapy and imagery-focused therapy have been adapted to work with persecutory delusions (Taylor et al., 2019, 2020). Moreover, although not tailored to persecutory delusions, other types of interventions have been designed to improve emotional facial recognition in patients with schizophrenia showing promising results regarding social cognition outcomes (Lahera et al., 2021). Therefore, considering the relationship between higher paranoid beliefs and a tendency to mistrust unfamiliar faces, which appears to be influenced by insecure attachment styles, it may be worthwhile to develop interventions that target bias modification training and focus on relational aspects. Such interventions may have the potential to reduce paranoid cognitions by addressing and mitigating these biases.

6.5. Strengths and limitations

This thesis is not exempt from limitations but also strengths. First, the design of the first two empirical studies was cross-sectional and thus restricted the possibility of making causal inferences regarding the predictor, mediator and outcome variables in the models. Moreover, the implementation of a small set of trials when assessing trustworthiness judgements might have affected the precision of the construct being measured. Additionally, the fact that a considerable proportion of respondents were removed in empirical study 3 might limit the representativity of the sample. Even though no differences were found in most demographic

characteristics between participants who were removed and those who completed the task, the former group was significantly older, although this difference was small (Table S3.2.). Nonetheless, the samples employed for these studies were large, international, and representative of the general population enabling the generalization of results to the population at large as well as the implementation of high-powered statistical tests.

Second, a common limitation across the four empirical chapters was the use of the Princeton Social Perception Lab database as stimuli for measuring trustworthiness. This database poses two main limitations. First, the fact that the faces look avatar-like and thus not real restricts the ecological validity of the task. Second, the faces are male, Caucasian (in empirical studies 1,2, and 4), and young. Consequently, this would limit the generalization of gender, ethnicity, and age thus not capturing the full range of trustworthiness variations presented in the real world. However, this database offers several advantages for researching face perception. For instance, this database has been computer-generated through AI (Artificial Intelligence) which means that allows for controlling and manipulating facial features that cannot be done with standard pictures of actors posing an emotion. Also, by manipulating these features, several degrees of trustworthiness can be accurately calibrated across faces. Therefore, it allows better control for external sources of variability that cannot be accounted for when using real-face stimuli, which is highly beneficial for conducting psychophysiological research. Moreover, this database has been validated using large samples, across time, and neural correlates have been established associated with trustworthiness perception (Todorov et al., 2008; Todorov & Oh, 2021). Lastly, this database is freely available, making it accessible to researchers across the world who are interested in conducting or replicating studies in face perception or social cognitive neuroscience. However, new research is emerging in this field, and due to the advances in AI technology newer face stimuli datasets are being computer-generated while keeping a real-face-like appearance (Walker et al., 2018). Another potential

limitation regarding ecological validity when measuring trust processes was the lack of social interactive paradigms such as game theory tasks, which have been proven useful to study interpersonal processes in clinically paranoid populations (Ellet et al., 2023).

A third general limitation involves the use of convenience samples for both the empirical study 3 as well as the non-clinical sample of empirical study 4. Although gender distribution in empirical study 3 was not far from what is reported in the general population, other demographics were not. Moreover, in empirical study 4, control participants were highly educated and considerably younger than the clinical group whereas both groups were predominantly male and white. Given the characteristics of the studies, employing unrepresentative samples tends to be a common limitation, particularly when conducting neurophysiological research. Furthermore, when conducting online studies, the use of representative samples through online survey platforms can be costly and, unless supported by considerable funding, samples may be unrepresentative in terms of key demographic variables. Nevertheless, the sample size in empirical study 3 was large enough to allow between-within statistical tests, and while unrepresentative, empirical study 4 includes a hard-to-recruit clinical sample as well as objective psychophysiological measures.

Finally, as mentioned in the previous paragraph, the sample size for the clinical group was small. Moreover, aspects such as medication and higher severity of depression and anxiety in comparison to the non-clinical group are hard to disentangle when interpreting the results. Given that the underlying mechanisms of psychopharmacological medication target neurotransmitters in cortical and subcortical regions, it is possible that effects observed in ERP signals might be influenced by antipsychotic dosage. Moreover, since almost half of the clinical participants had a schizoaffective disorder diagnosis, higher depression scores are not unsurprising in the clinical group. Additionally, no information regarding the presence of other positive symptoms, such as hallucinations, was collected from the clinical sample, restricting

interpretations about the specific role of paranoia in trustworthiness judgements. Lastly, although 14 patients were initially interested in taking part, only 9 ended up participating. Of those 5 initially interested, two were outpatients whereas three were inpatients. Data collection from participants who suffer from severe mental conditions is challenging, particularly when they are hospitalized in inpatient wards as the mental stability of the patients is fragile. While showing interest in taking part in a study one day, the next day they may not feel well or motivated enough. Thus, data collected from clinical participants in these conditions is highly valuable and is not only informative quantitatively but also qualitatively.

6.6. Future directions

Future studies researching trust and paranoia from a relational approach using different methods can benefit from acknowledging the strengths and limitations discussed in the previous section. For example, longitudinal studies in representative samples can employ ecological momentary assessment tools to monitor and assess whether trust processes precede paranoid thoughts or vice-versa. Moreover, methodological studies exploring the reliability and validity of avatar-like and real-face-like stimuli and their relationship with paranoid beliefs can provide information about which type of face stimuli better captures trustworthiness judgements. Additionally, although the evaluative conditioning intervention employed in empirical study 3 affected attachment anxiety, other forms of experimental methods should be employed to test the effect of relational schemas on trust. Experimental studies have been successful in priming insecure and secure attachment styles by using imagery techniques, finding an effect on state paranoia (Sood et al., 2022; Sood & Newman-Taylor, 2020). Thus, to explore the causal role of attachment style on trustworthiness judgements, studies could design experiments implementing this kind of priming technique. Finally, given the high degree of co-occurrence of symptoms in participants with a schizophrenia diagnosis, studies investigating the association between the psychophysiological underpinnings of mistrust and

paranoia should employ designs with clinically paranoid, non-clinical, and non-paranoid clinical groups. Alternatively, other quasi-experimental designs can be applied by having non-clinical groups with high paranoid traits and low paranoid traits or by using experimental methods to induce paranoid states as implemented by Pinkham et al. (2022).

It is worth noting that the COVID-19 pandemic of 2020-2022 had an impact on this thesis, resulting in modifications to the original PhD plan. Therefore, two out of the three studies had to be conducted online to accommodate the circumstances caused by the pandemic.

6.7. Overall conclusions

This thesis has highlighted the role of mistrust in paranoid beliefs by operationalizing it as signal detection outcomes from a relational and evolutionary approach. Moreover, epidemiological, experimental, and psychophysiological methods have been implemented to explore the relationship between these two variables in large international representatives, convenience, and clinical samples. Given the social nature of paranoid beliefs and since trust is a key component in social interactions, findings from this thesis contribute to current conceptualizations of paranoia. These contributions should inform potential new psychological interventions in clinical and non-clinical populations. Nonetheless, these findings should be interpreted with caution as these studies had several limitations. Future studies should benefit from acknowledging the strengths and limitations of this thesis and expanding the understanding of mistrust and paranoid beliefs.

APPENDIX A

Supplementary material S2

Table S2. 1. *Mean differences between nations. Pairwise comparison with Bonferroni correction for paranoia and conspiracy*.

Variables	Nation			Pairwise Comparisons	Mean difference 95%CI [L/U]	*p*
	UK Mean (SD)	*Spain* Mean (SD)	*Ireland* Mean (SD)			
Paranoia	7.43 (5.77)	5.96(4.43)	7.26(4.87)	UK-Spain UK-Ireland Ireland-Spain	-1.47 [-1.86/-1.07] -.017 [-.06/.03] 1.30 [.82/1.78]	<.001 1.00 <.001
Conspiracy	35.17 (9.17)	40.69 (9.20)	36.37 (9.26)	UK-Spain UK-Ireland Ireland-Spain	5.52[4.81/6.24] 1.20[.32/2.07] -4.32[-5.19/-3.45]	<.001 <.001 <.001

Table S2. 2. *Mean and standard deviations (SD) for the combined sample as well as same statistics reported in other studies*

Variables	*Combined Sample* Mean (SD)	*Melo et al (2009)* Mean (SD)	*Al-Suhibani et al (2022)* Mean (SD)	*Bruder et al (2013)* Mean (SD)	*Đorđević et al (2021)* Mean (SD)
Paranoia	1.37 (1.02)	1.18 (.078)	1.69 (0.95)	-	-
Conspiracy	7.51 (1.90)	-	6.91 (2.11)	6.48 (2.40)	7.88 (2.26)

Note. All values have been averaged based on the number of total items that the authors have used in order to make the results easier to interpret.

Table S2. 3. *Fit Statistics for MIMIC model*

Model	χ^2	**df**	***p***	**CFI**	**TLI**	**RMSEA**	**SRMR**
Baseline constrained	2449.638	50	<.001	.898	.867	.098	.057
Ireland -> CMQ Item3	1990.389	49	<.001	.918	.891	.089	.048
Ireland -> CMQ Item4	1595.736	48	<.001	.934	.911	.080	.042

CMQ (Conspiracy Mentality Questionnaire)

Table S2. 4. *Correlation Table with main variables Study 2*

Variables	1	2	3	4	5	6	7	8	9	10	11	12
1.Trust in Political Institutions	1	.49**	.16**	.29**	.05**	.25**	**.11****	.004	.17**	.15**	-.03*	-.34**
2.Trust in Legal Institutions		1	.47**	.22**	.11**	.30**	**.17****	.04**	.24**	.18**	-.18**	-.20**
3.Trust in Scientific Institutions			1	.13**	.14**	.41**	**.18****	.06**	.22**	.12**	-.24**	-.003
4.Trust in Traditional Media				1	.45**	.41**	**.10****	.02	.03*	.15**	-.03*	-.12**
5.Trust in Informal Sources					1	.37**	.02	.01	.005	.005	.07**	.13**
6.Trust in Institutional Sources						1	**.18****	.05**	.15**	.15**	-.16**	-.09**
7.Bias							1	.06**	**.26****	**.15****	**-.25****	-.09**
8.Sensitivity								1	-.012	.003	-.008	.018
9.Interpersonal Trust									1	**.17****	**-.35****	**-.19****
10.Neighbourhood Trust										1	**-.16****	-.05**
11.Paranoia											1	.11**
12.Conspiracy Mentality												1

APPENDIX B

Supplementary material S3

Figure S3. 1. *The only significant indirect effect was attachment anxiety → negative self-esteem → paranoia, β = 0.065, 95% CI 0.055 - 0.077, p <.001. None of the effects went through the sensitivity measure.*

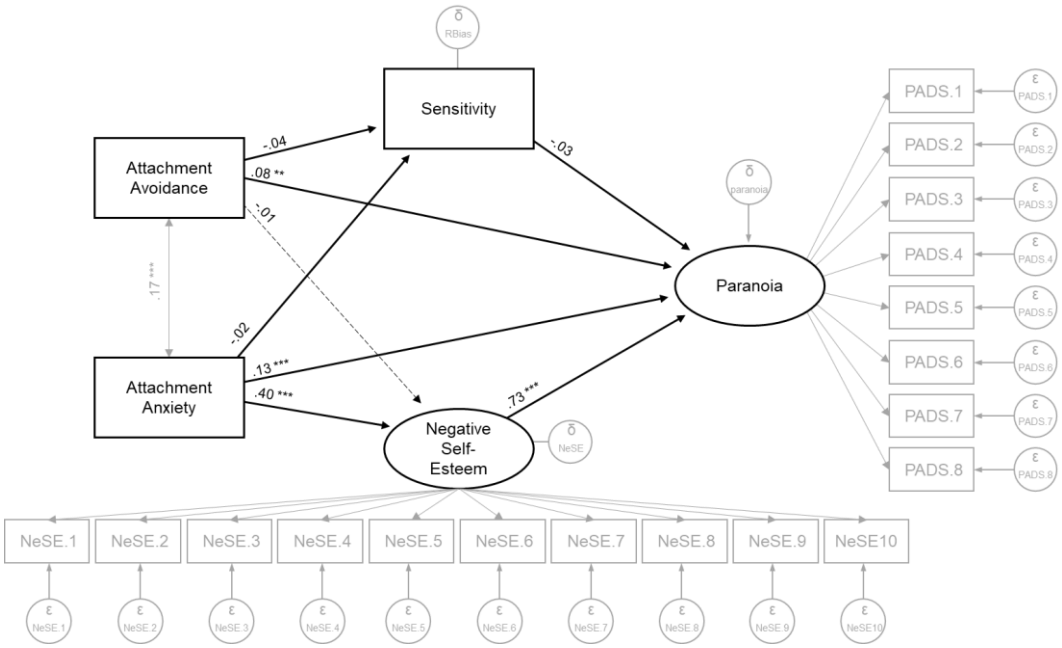


Table S3. 1. *Differences between completers and non-completers.*

	t	**df**	***d***	M_{diff}	***p***
Age	5.71	1517	.33	5.67	<.001
			χ^2	**df**	***p***
Gender			2.55	1	.11
Ethnicity			10.92	10	.36

Added measurement in Table S3.1

General Trust Scale (GTS; Yamagishi and Yamagishi, 1994). Respondents were asked to read six statements measuring general beliefs about honesty and trustworthiness and indicate their agreement on a 5-point Likert scale (from 1= *"Strongly Disagree"* to 5= *"Strongly Agree"*). This scale revealed good internal reliability (α=.88).

Table S3. 2. *Bivariate correlations between main variables with means and standard deviations*

Variables	*M*	*SD*	1	2	3	4	5	6	7
1. Paranoia	2.78	1.00	-	.41**	.14**	.74**	-.10**	**-.20****	**-.29****
2. Attachment Anxiety	-.87	3.50		-	-.17**	.39**	-.03	-.13**	**.17****
3. Attachment Avoidance	.85	3.57			-	.05	-.04	-.14**	**.22****
4. Negative Self-Esteem	3.35	1.43				-	-.09**	-.09**	.14**
5. Sensitivity	.67	.22					-	.19**	.06*
6. Response Bias	-.31	1.15						-	**.30****
7. Interpersonal Trust	21.10	4.23							-

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

APPENDIX C

Supplementary material S4

Figure S4. 1. *QQ-plots of residuals for Dependent Variables at different time points*

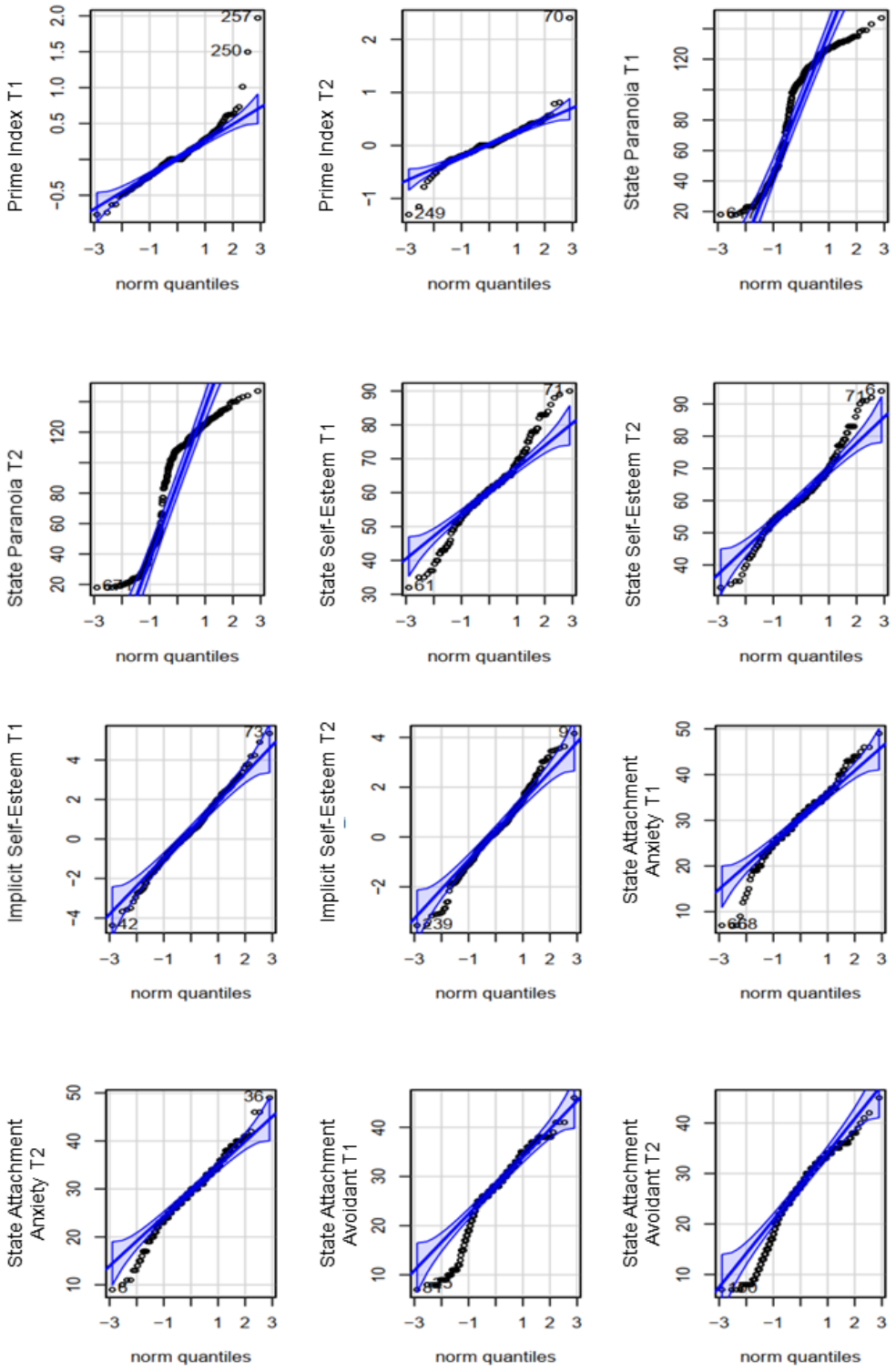


Figure S4. 2. *Histograms of residuals for Dependent Variables at different time points*

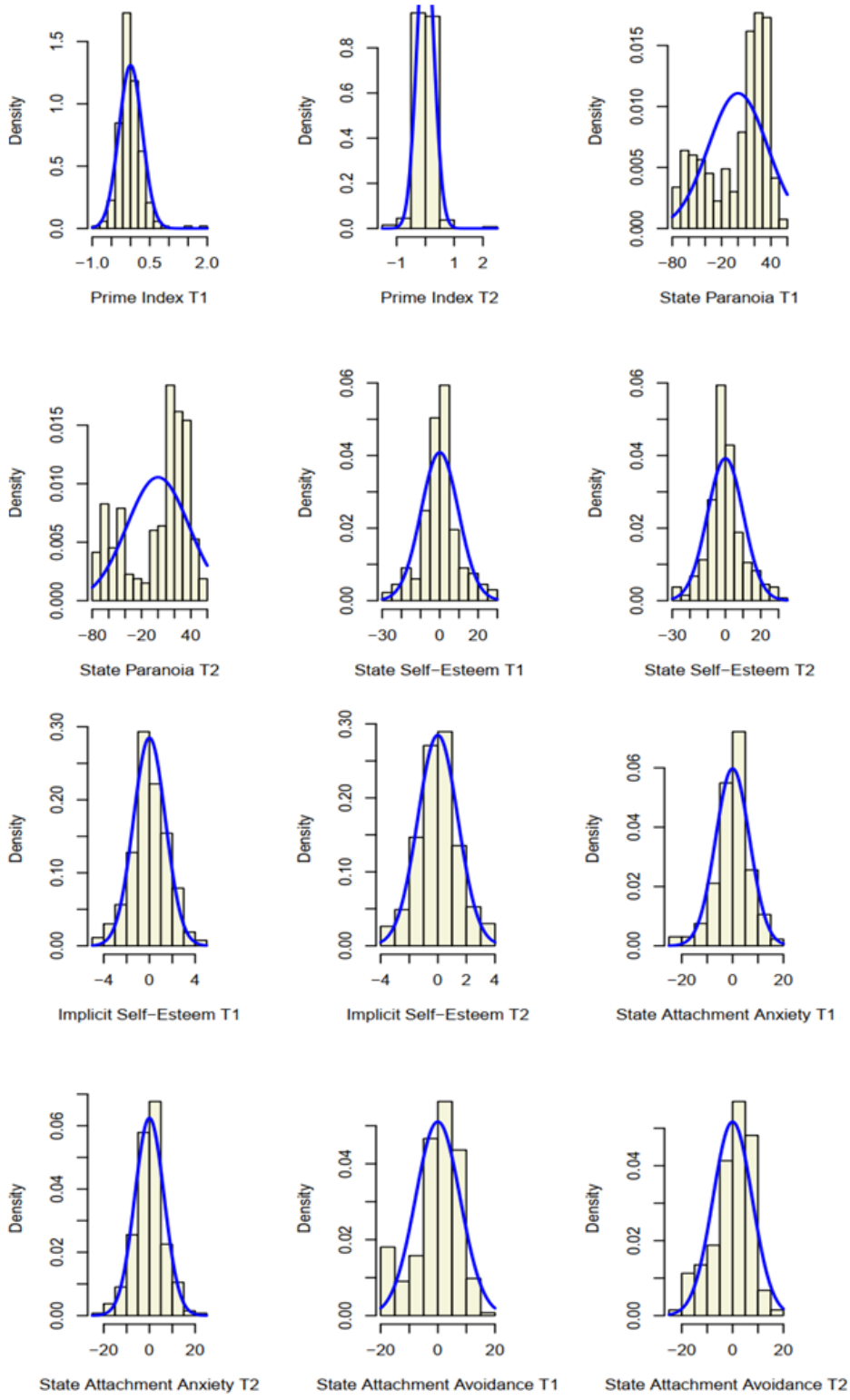


Table S4 1. *Levene's Test of Equality of Error Variances*[a]

	Levene Statistic	df1	df2	Sig.
Prime Index (T1)	.91	2	263	.40
Prime Index (T2)	.94	2	263	.39
State Paranoia (T1)	.75	2	263	.47
State Paranoia (T2)	.50	2	263	.61
State Self-Esteem (T1)	1.35	2	263	.26
State Self-Esteem (T2)	1.54	2	263	.22
Implicit Self-Esteem (T1)	2.27	2	263	.10
Implicit Self-Esteem (T2)	.14	2	263	.87
State Attachment Anxiety (T1)	1.21.	2	263	.30
State Attachment Anxiety (T2)	1.61	2	263	.20
State Attachment Avoidance (T1)	.27	2	263	.76
State Attachment Avoidance (T2)	1.80	2	263	.17

[a]Levene's values based on the means

Table S4 2. *Bivariate correlation between trait and state psychological variables measured at baseline.*

Variables	*M*	*SD*	1	2	3	4	5	6	7	8	9	10
1. Paranoia Trait	28.96	6.49	1	.55**	-.31**	-.10	-.11	.02	.27**	.66**	.79**	-.51**
2. Negative Self-Esteem	23.80	4.08		1	-.40**	-.12	.09	.05	.11	.47**	.43**	-.64**
3. Attachment Anxiety	.66	3.86			1	.07	.01	-.02	-.25**	-.16**	-.19**	.33**
4. Attachment Avoidant	.004	3.26				1	.03	.02	.06	-.25**	-.03	.11
5. Implicit Self- Esteem (T1)	.42	1.56					1	-.02	-.14*	-.18**	-.24**	-.14*
6. Prime Index - **β** (T1)	.03	.30						1	-.04	.02	.01	.01
7. State Attachment Anxiety (T1)	30.25	6.69							1	.16**	.25**	-.20**
8. State Attachment Avoidant (T1)	27.01	7.87								1	.68**	-.41**
9. State Paranoia (T1)	93.18	36.06									1	-.32**
10. State Self-Esteem (T1)	60.63	9.77										1

*Note.*** Correlation is significant at the 0.01 level (2-tailed), * is significant at the 0.05 level (2-tailed)

Table S4 3. *Other-relevant information for classical conditioning task.*

Names	Surnames	Dates	Cities	Pronouns
Marvin	*Dankworth*	*February 29th*	*Berlin*	*He*
Siobhann				*She*
Ludovic	*McQoid*	*December 21st*	*Brussels*	*His*
Tatania				*Her*

Figure S4. 3. *Examples of face stimuli in each dimension for each ethnicity*

APPENDIX D

Supplementary Material Study 4

Figure S5. 1. *QQ-plots for the main variables.*

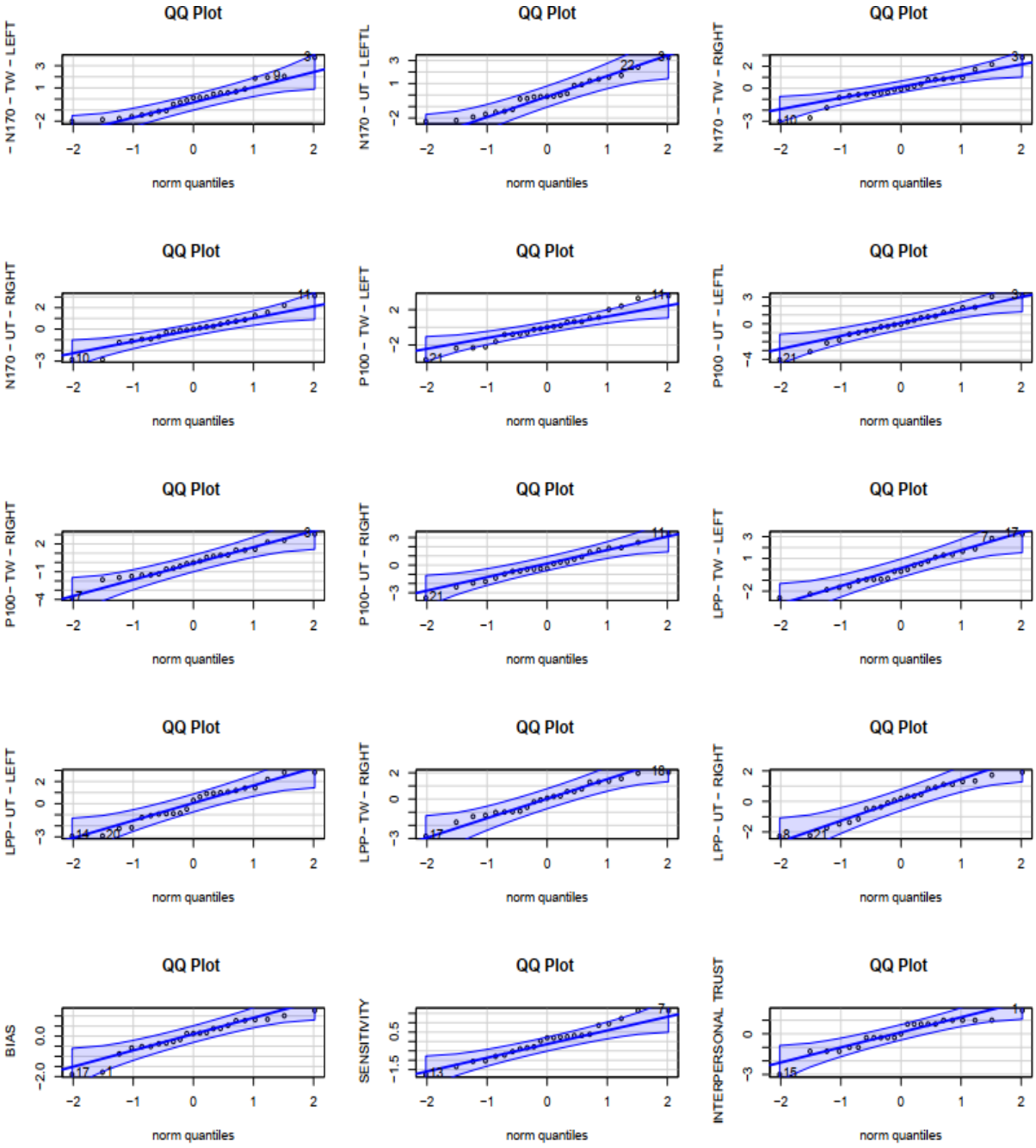


Figure S5. 2. *Histograms of residuals for main variables*

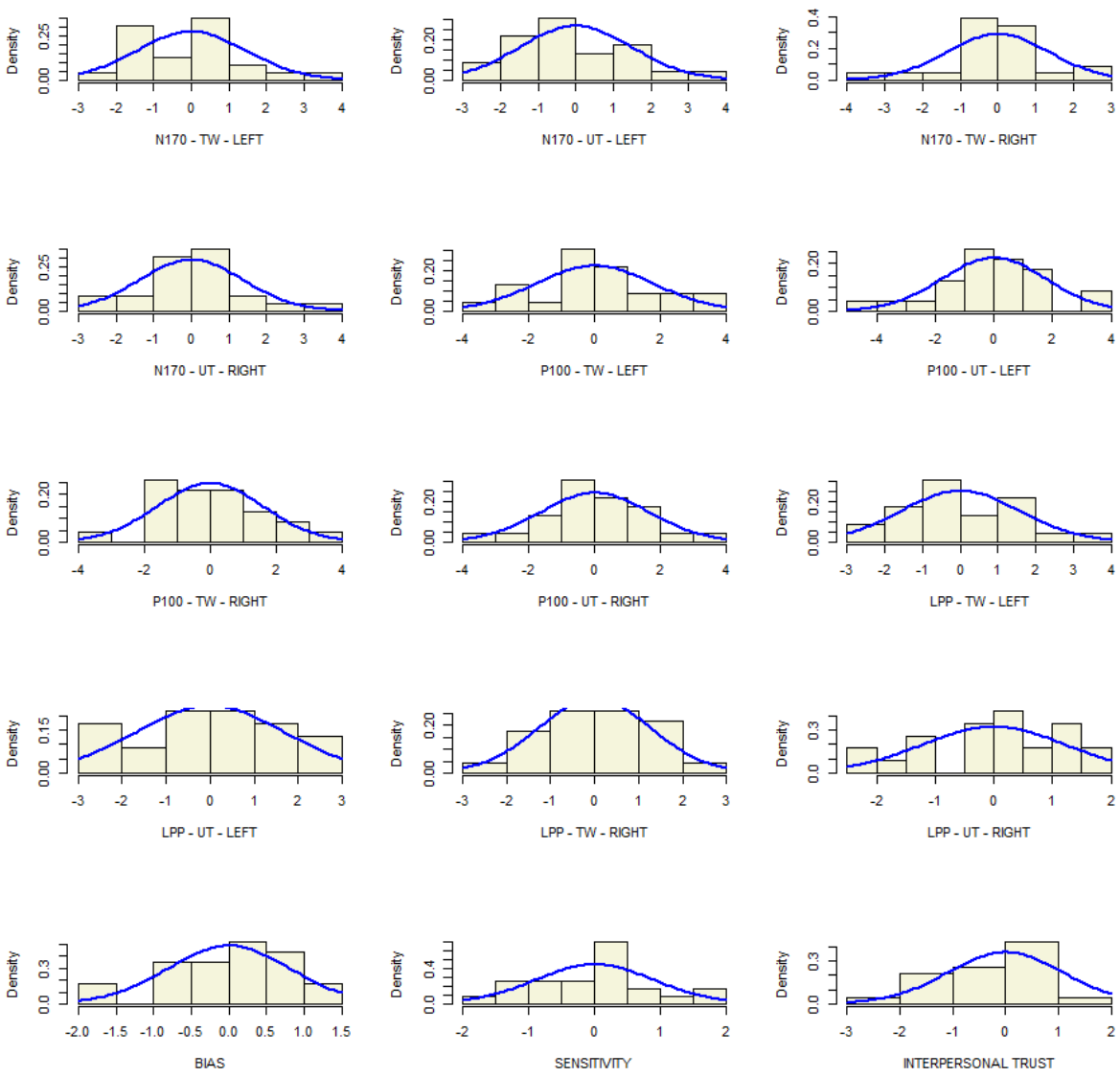


Figure S5. 3. *Histograms of frequencies regarding paranoid traits, ideas of reference and persecution symptoms*

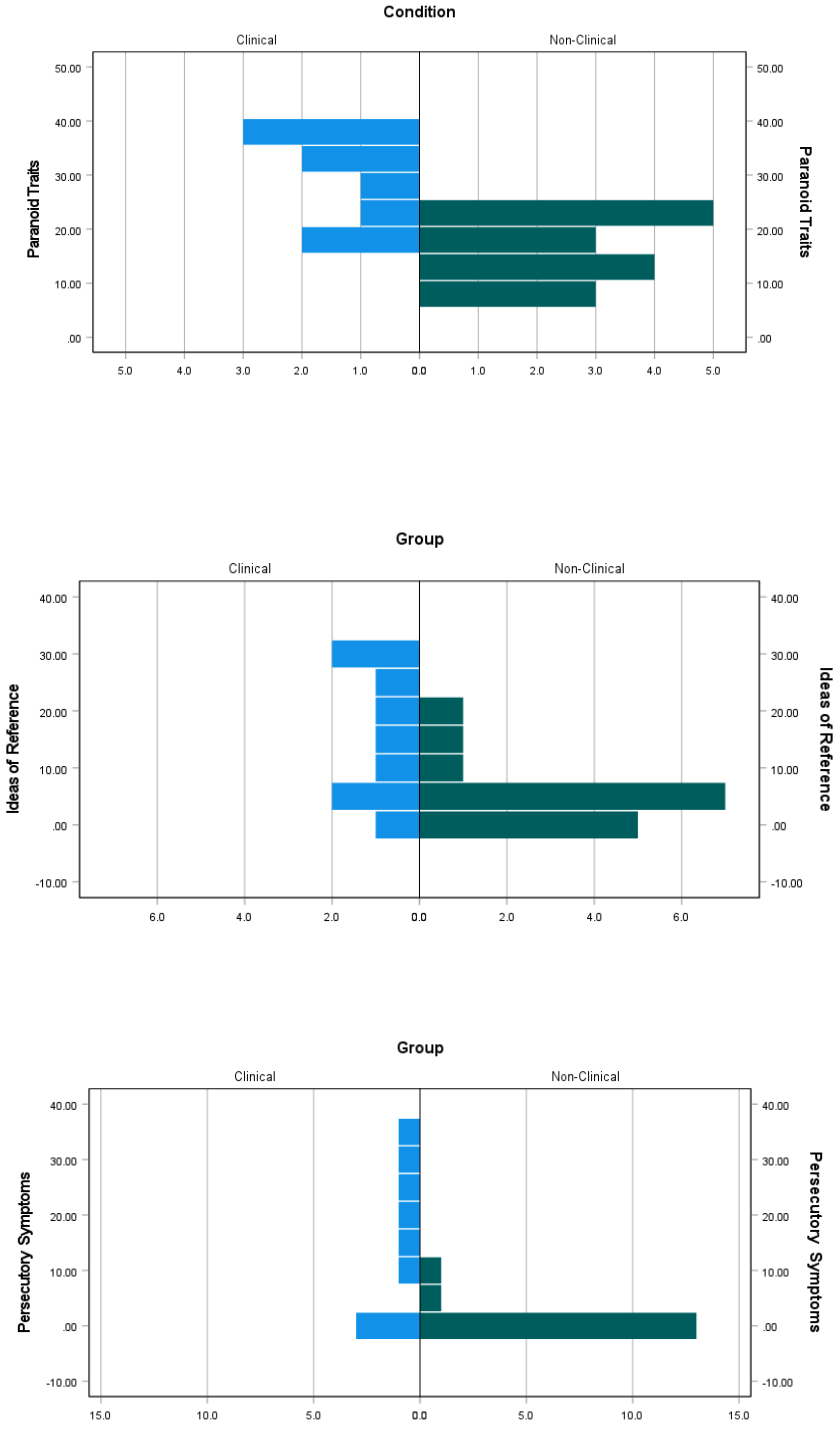


Figure S5. 4. *Scatterplots for correlation between paranoia as well as interpersonal mistrust and significant ERP components*

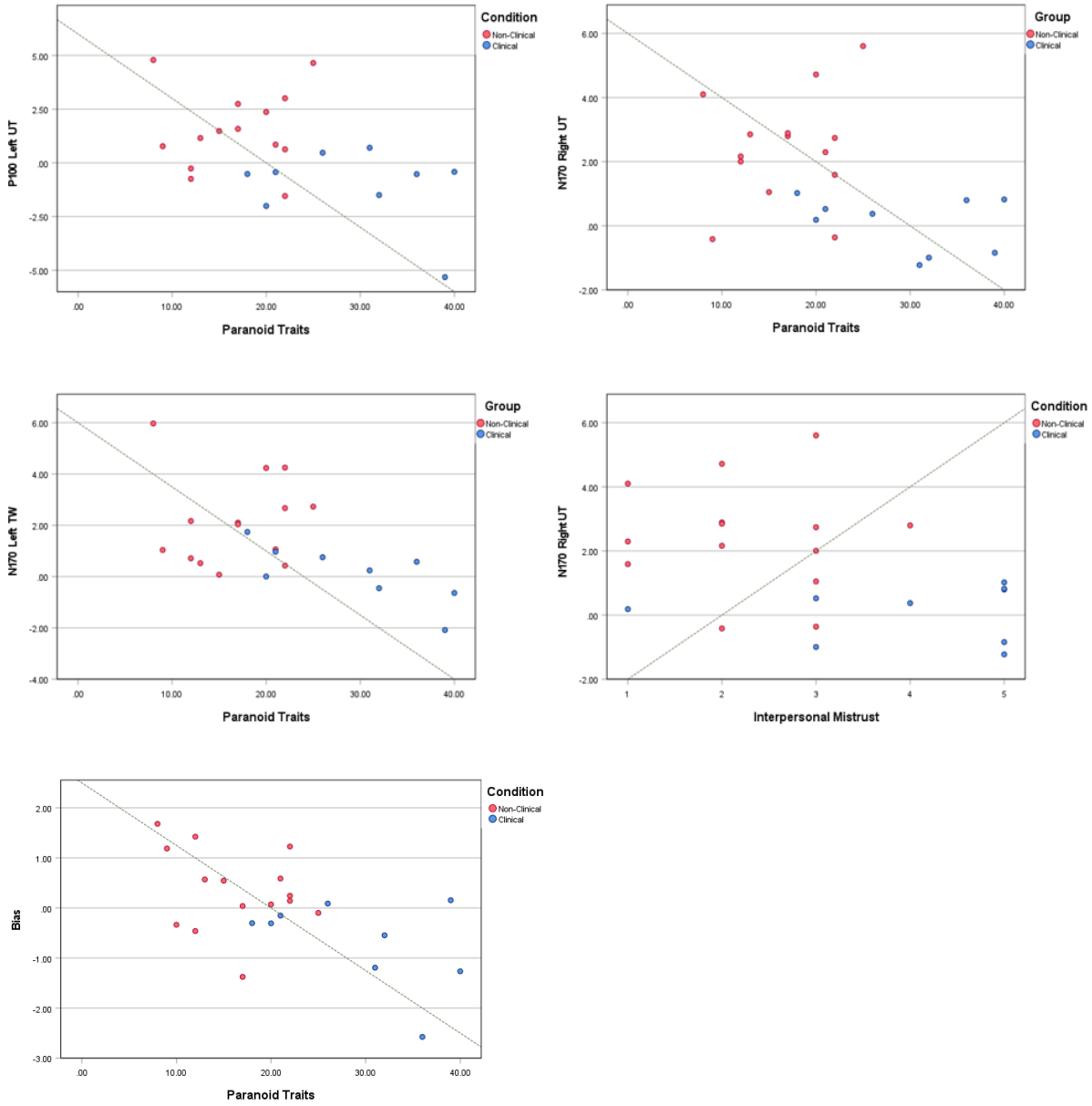


Table S5. 1. Levene's Test of Equality of Error Variances[a]

	Levene Statistic	df1	df2	Sig.
Ideas of Persecution	25.16	1	22	<.001
Ideas of Reference	9.50	1	22	<.01
Depression	7.16	1	22	.01
Generalised Anxiety	3.46	1	22	.08
Paranoid Traits	3.59	1	22	.07
Bias	.05	1	22	.82
Sensitivity	.13	1	22	.72
Interpersonal Trust	1.88	1	22	.18
P100 (Right- Trust)	.01	1	22	.93
P100 (Right- Mistrust)	.41	1	22	.53
P100 (Left – Trust)	.13	1	22	.78
P100 (Left – Mistrust)	.08	1	22	.72
N170 (Right- Trust)	3.11	1	22	.09
N170 (Right- Mistrust)	1.91	1	22	.18
N170 (Left – Trust)	.87	1	22	.36
N170 (Left – Mistrust)	.83	1	22	.37
LPP (Right- Trust)	.13	1	22	.72
LPP (Right- Mistrust)	.37	1	22	.53
LPP (Left – Trust)	.15	1	22	.70
LPP (Left – Mistrust)	.01	1	22	.95

[a]Levene’s values based on the means

REFERENCES

Abalakina-Paap, M., Stephan, W. G., Craig, T., and Gregory, W. L. (1999). Beliefs in conspiracies. Polit. Psychol. 20, 637–647. doi: 10.1111/0162-895X.00160.

Abbott, J., Middlemiss, M., Bruce, V., Smailes, D., and Dudley, R. (2018). The effect of arousal and eye gaze direction on trust evaluations of stranger's faces: A potential pathway to paranoid thinking. J. Behav. Ther. Exp. Psychiatry 60, 29–36. doi: 10.1016/j.jbtep.2018.02.007.

Abdi, H. (2007). "Part (Semi Partial) and Partial Regression Coefficients," in Encyclopedia of Measurement and Statistics, ed. N. Salkind (Sage), 736–740.

Akaike, H. (1987). Factor analysis and AIC. Psychometrika 52, 317–332. doi: 10.1007/BF02294359.

Alsuhibani, A., Shevlin, M., Freeman, D., Sheaves, B., & Bentall, R. P. (2022). Why conspiracy theorists are not always paranoid: Conspiracy theories and paranoia form separate factors with distinct psychological predictors. PLoS ONE, 17(4 April), 1–23. https://doi.org/10.1371/journal.pone.0259053

Alsuhibani, A., Shevlin, M., Freeman, D., Sheaves, B., and Bentall, R. P. (2022). Why conspiracy theorists are not always paranoid: Conspiracy theories and paranoia form separate factors with distinct psychological predictors. PLoS One 17, 1–23. doi: 10.1371/journal.pone.0259053.

American Psychiatric Association. (1980). Diagnostic and statistical manual of mental disorders (3rd ed.). Washington, DC: Author. https://doi.org/10.1176/ajp.145.10.1301

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington D.C. https://doi.org/10.1176/appi.books.9780890425596

Axelrod, S. R., Widiger, T. A., Trull, T. J., and Corbitt, E. M. (1997). Relations of five-factor model antagonism facets with personality disorder symptomatology. J. Pers. Assess. 69, 297–313. doi: 10.1207/s15327752jpa6902_4.

Baas, D., Van't Wout, M., Aleman, A., & Kahn, R. S. (2008). Social judgement in clinically stable patients with schizophrenia and healthy relatives: Behavioural evidence of social brain dysfunction. Psychological Medicine, 38(5), 747–754. https://doi.org/10.1017/S0033291707001729

Baccus, J. R., Baldwin, M. W., & Packer, D. J. (2004). Increasing implicit self-esteem through classical conditioning. Psychological Science, 15(7), 498–502. https://doi.org/10.1111/j.0956-7976.2004.00708.x

Bach, B., & First, M. B. (2018). Application of the ICD-11 classification of personality disorders. BMC Psychiatry, 18(1), 1–14. https://doi.org/10.1186/s12888-018-1908-3

Baldwin, M. W. (1992). Baldwin - 1992 - Relational Schemas and the Processing of Social Information.pdf. Psychol. Bull. 112, 461–484.

Baldwin, M. W. ed. (2005). Interpersonal Cognition. New York City, NY, US: The Guilford Press.

Baldwin, M. W., and Kay, A. C. (2003). Adult attachment and the inhibition of rejection. J. Soc. Clin. Psychol. 22, 275–293. doi: 10.1521/jscp.22.3.275.22890.

Ball, K., Lawson, W., and Alim, T. (2013). Medical Mistrust, Conspiracy Beliefs & HIV-Related Behavior Among African Americans. J. Psychol. Behav. Sci. 1, 1–7.

Barnby, J. M., Deeley, Q., Robinson, O., Raihani, N., Bell, V., & Mehta, M. A. (2020). Paranoia, sensitization and social inference: Findings from two large-scale, multi-round behavioural experiments. Royal Society open science, 7(3), 191525.

Barreto Carvalho, C., Sousa, M., da Motta, C., Pinto-Gouveia, J., Caldeira, S. N., Peixoto, E. B., … Fenigstein, A. (2017). Paranoia in the general population: A revised version of the General Paranoia Scale for adults. Clinical Psychologist, 21(2), 125–134. https://doi.org/10.1111/cp.12065

Barron, D., Morgan, K., Towell, T., Altemeyer, B., and Swami, V. (2014). Associations between schizotypy and belief in conspiracist ideation. Pers. Individ. Dif. 70, 156–159. doi: 10.1016/j.paid.2014.06.040.

Bartholomew, K., & Horowitz, L. M. (1991). Romantic Love Conceptualized as an Attachment Process. Journal of Personality and Social Psychology, 61(2), 226–244. http://www.ncbi.nlm.nih.gov/pubmed/1920064

Baumeister, R. F., & Leary, M. R. (1995). The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation. Psychological Bulletin, 117(3), 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Bebbington, P. E., McBride, O., Steel, C., Kuipers, E., Radovanovič, M., Brugha, T., … Freeman, D. (2013). The structure of paranoia in the general population. British Journal of Psychiatry, 202(6), 419–427. https://doi.org/10.1192/bjp.bp.112.119032

Bell, V., & O'Driscoll, C. (2018). The network structure of paranoia in the general population. Social Psychiatry and Psychiatric Epidemiology, 53(7), 737–744. https://doi.org/10.1007/s00127-018-1487-0

Bentall, R. P. (2003). Madness explained: Psychosis and human nature. Penguin.

Bentall, R. P. (2009). Doctoring the mind: Why psychiatric treatments fail. Penguin.

Bentall, R. P. (2021). The role of early life experience in psychosis. In C. A. Tamminga, E. I. Ivleva, U. Reininghaus, & J. van Os (Eds.), Psychotic disorders: Comprehensive conceptualization and treatments (pp. 406-414). Oxford University Press. Owen

Bentall, R. P., & Fernyhough, C. (2008). Social predictors of psychotic experiences: Specificity and psychological mechanisms. Schizophrenia Bulletin, 34(6), 1012–1020. https://doi.org/10.1093/schbul/sbn103

Bentall, R. P., Corcoran, R., Howard, R., Blackwood, N., & Kinderman, P. (2001). Persecutory Delusions: a Review and Theoretical Integration. Clinical Psychology Review, 21(8), 1143–1192. https://doi.org/10.1016/S0272-7358(01)00106-4

Bentall, R. P., de Sousa, P., Varese, F., Wickham, S., Sitko, K., Haarmans, M., et al. (2014). From adversity to psychosis: Pathways and mechanisms from specific adversities to specific symptoms. Soc. Psychiatry Psychiatr. Epidemiol. Int. J. Res. Soc. Genet. Epidemiol. Ment. Heal. Serv. 49, 1011–1022. doi: http://dx.doi.org/10.1007/s00127-014-0914-0.

Bentall, R. P., Kinderman, P., & Kaney, S. (1994). The self, attributional processes and abnormal beliefs: Towards a model of persecutory delusions. Behaviour Research and Therapy, 32(3), 331–341. https://doi.org/10.1016/0005-7967(94)90131-7

Bentall, R. P., Rouse, G., Kinderman, P., Blackwood, N., Howard, R., Moore, R., … Corcoran, R. (2008). Paranoid delusions in schizophrenia spectrum disorders and depression: The transdiagnostic role of expectations of negative events and negative self-esteem. Journal of Nervous and Mental Disease, 196(5), 375–383. https://doi.org/10.1097/NMD.0b013e31817108db

Bentall, R. P., Rowse, G., Shryane, N., Kinderman, P., Howard, R., Blackwood, N., … Corcoran, R. (2009). The cognitive and affective structure of paranoid delusions: A transdiagnostic investigation of patients with schizophrenia spectrum disorders and depression. Archives of

General Psychiatry, 66(3), 236–247. https://doi.org/http://dx.doi.org/10.1001/archgenpsychiatry.2009.1

Bentall, R. P., Wickham, S., Shevlin, M., & Varese, F. (2012). Do specific early-life adversities lead to specific symptoms of psychosis? a study from the 2007 the adult psychiatric morbidity survey. Schizophrenia Bulletin, 38(4), 734–740. https://doi.org/10.1093/schbul/sbs049

Bentler, P. M. (1990). Comparative fit indices in structural equation models. Psychol. Bull. 107, 238－246.

Berry, K., Wearden, A., & Barrowclough, C. (2007). Adult attachment styles and psychosis: An investigation of associations between general attachment styles and attachment relationships with specific others. Social Psychiatry and Psychiatric Epidemiology, 42(12), 972–976. DOI 10.1007/s00127-007-0261-5

Bigelow, F. J., Clark, G. M., Lum, J. A. G., & Enticott, P. G. (2021). The development of neural responses to emotional faces: A review of evidence from event-related potentials during early and middle childhood. Developmental Cognitive Neuroscience, 51(June), 100992. https://doi.org/10.1016/j.dcn.2021.100992

Blackhart, G. C., Nelson, B. C., Knowles, M. L., and Baumeister, R. F. (2009). Rejection elicits emotional reactions but neither causes immediate distress nor lowers self-esteem: A meta-analytic review of 192 studies on social exclusion. Personal. Soc. Psychol. Rev. 13, 269–309. doi: 10.1177/1088868309346065.

Blanchard, J. J., Andrea, A., Orth, R. D., Savage, C., & Bennett, M. E. (2020). Sleep disturbance and sleep-related impairment in psychotic disorders are related to both positive and negative symptoms. Psychiatry Research, 286(February), 112857. https://doi.org/10.1016/j.psychres.2020.112857

Blashfield, R. K., & Blashfield, R. K. (1984). The Neo-Kraepelinian Movement in American Psychiatry. The Classification of Psychopathology: Neo-Kraepelinian and Quantitative Approaches, 25-58. DOI: 10.1007/978-1-4613-2665-6_2

Borsboom, D., & Cramer, A. O. J. (2013). Network Analysis: An Integrative Approach to the Structure of Psychopathology. Annual Review of Clinical Psychology, 9(1), 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Bosqui, T. J., Hoy, K., & Shannon, C. (2014). A systematic review and meta-analysis of the ethnic density effect in psychotic disorders. Social Psychiatry and Psychiatric Epidemiology, 49(4), 519–529. DOI 10.1007/s00127-013-0773-0

Bowlby, J. (1969). Attachment and loss: Vol 1 - Attachment. Hogarth Press.

Bowlby, J. (1982). Attachment and loss: Vol.1. Attachment: Vol. I (2nd ed). Basic Books.

Brotherton, R., and Eser, S. (2015). Bored to fears: Boredom proneness, paranoia, and conspiracy theories. Pers. Individ. Dif. 80, 1–5. doi: http://dx.doi.org/10.1016/j.paid.2015.02.011.

Brown, P, Waite, F., & Freeman, D. (2019). “Twisting the lion’s tail”: Manipulationist tests of causation for psychological mechanisms in the occurrence of delusions and hallucinations. Clinical Psychology Review, 68, 25–37. https://doi.org/https://dx.doi.org/10.1016/j.cpr.2018.12.003

Brown, Poppy, Waite, F., & Freeman, D. (2020). Parenting behaviour and paranoia: a network analysis and results from the National Comorbidity Survey-Adolescents (NCS-A). Social Psychiatry and Psychiatric Epidemiology. https://doi.org/https://dx.doi.org/10.1007/s00127-020-01933-6

Brown, T. A., & Barlow, D. H. (2005). Dimensional versus categorical classification of mental disorders in the fifth edition of the diagnostic and statistical manual of mental disorders and

beyond: Comment on the special section. Journal of Abnormal Psychology, 114(4), 551–556. DOI: 10.1037/0021-843X.114.4.551

Bruder, M., Haffke, P., Neave, N., Nouripanah, N., and Imhoff, R. (2013). Measuring Individual Differences in Generic Beliefs in Conspiracy Theories Across Cultures: Conspiracy Mentality Questionnaire. Front. Psychol. 4. doi: 10.3389/fpsyg.2013.00225.

Brunson, J. A., Acitelli, L. K., and Sharp, C. (2015). Exploring the dynamic nature of "us": An integrative approach to relationship cognition. Bull. Menninger Clin. 79, 131–165. doi: 10.1521/bumc.2015.79.2.131.

Buck, B. E., Pinkham, A. E., Harvey, P. D., & Penn, D. L. (2016). Revisiting the validity of measures of social cognitive bias in schizophrenia: Additional results from the Social Cognition Psychometric Evaluation (SCOPE) study. The British Journal of Clinical Psychology, 55(4), 441–454. DOI:10.1111/bjc.12113

Burns, C., & Conchie, S. (2015). Measuring Implicit Trust and Automatic Attitude. In Handbook of research methods on trust. (pp. 239–248). Edward Elgar Publishing.

Butter, S., Murphy, J., Shevlin, M., and Houston, J. (2017). Social isolation and psychosis-like experiences: a UK general population analysis. Psychosis 9, 291–300. doi: 10.1080/17522439.2017.1349829.

Cabinet Office (2015). Cabinet Office. Available at: Community Life Survey Technical Report 2019/20.

Camerer, C. F. (2003). Behavioural studies of strategic thinking in games. Trends in Cognitive Sciences, 7(5), 225–231. https://doi.org/10.1016/S1364-6613(03)00094-9

Carr, S. C., Hardy, A., & Fornells-Ambrojo, M. (2018). Relationship between attachment style and symptom severity across the psychosis spectrum: A meta-analysis. Clinical Psychology Review, 59(November 2016), 145–158. https://doi.org/10.1016/j.cpr.2017.12.001

Castro, M. K., Bailey, D. H., Zinger, J. F., & Martin, E. A. (2019). Late electrophysiological potentials and emotion in schizophrenia: A meta-analytic review. Schizophrenia Research, 211, 21–31. https://doi.org/10.1016/j.schres.2019.07.013

Černis, E., Bird, J. C., Molodynski, A., Ehlers, A., & Freeman, D. (2021). Cognitive appraisals of dissociation in psychosis: A new brief measure. Behavioural and Cognitive Psychotherapy, 49(4), 472–484. doi:10.1017/S1352465820000958

Chapman, L. J., & Chapman, J. P. (1988). The genesis of delusions. In T. F. Oltmanns & B. A. Maher (Eds.), Delusional beliefs (pp. 167-183). John Wiley.

Cheesman, R., Selzam, S., Ronald, A., Dale, P. S., McAdams, T. A., Eley, T. C., & Plomin, R. (2017). Childhood behaviour problems show the greatest gap between DNA-based and twin heritability. Translational Psychiatry, 7(12). DOI 10.1038/s41398-017-0046-x

Chen, M., & Bargh, J. . (1999). Immediate behavioral predispositions to approach or avoid the stimulus. Personality and Social Psychology Bulletin, 25(2), 215–224.

Cichocka, A., Marchlewska, M., and de Zavala, A. G. (2016). Does Self-Love or Self-Hate Predict Conspiracy Beliefs? Narcissism, Self-Esteem, and the Endorsement of Conspiracy Theories. Soc. Psychol. Personal. Sci. 7, 157–166. doi: 10.1177/1948550615616170.

Cohen, J. (1992). A Power Primerr. Psychol. Bull. 112, 155–159. doi: 10.1037/0033-2909.112.1.155.

Combs, D. R., Adams, S. D., Michael, C. O., Penn, D. L., Basso, M. R., & Gouvier, W. D. (2006). The conviction of delusional beliefs scale: Reliability and validity. Schizophrenia Research, 86(1–3), 80–88. https://doi.org/10.1016/j.schres.2006.06.023

Combs, D. R., Penn, D. L., Cassisi, J., Michael, C., Wood, T., Wanner, J., & Adams, S. (2006). Perceived racism as a predictor of paranoia among African Americans. Journal of Black Psychology, 32(1), 87–104. DOI: 10.1177/0095798405283175

Conrad, K. (1958/2012). Beginning schizophrenia: Attempt for a Gestalt-analysis of delusion. In M. R. Broome, R. Harland, G. S. Owen, & A. Stringaris (Eds.), The Maudsley reader in phenomenological psychiatry (pp. 176-193). Cambridge University Press.

Cooper, A. F., & Curry, A. R. (1976). The pathology of deafness in the paranoid and affective psychoses of later life. Journal of Psychosomatic Research, 20(2), 97–105. https://doi.org/10.1016/0022-3999(76)90035-0

Corcoran, R., Mansfield, R., de Bezenac, C., Anderson, E., Overbury, K., & Marshall, G. (2018). Perceived neighbourhood affluence, mental health and wellbeing influence judgements of threat and trust on our streets: An urban walking study. PLoS ONE, 13(8). https://doi.org/10.1371/journal.pone.0202412

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: using ethnicity to disambiguate potentially threatening individuals. Journal of Personality and Social Psychology, 83(6), 1314–1329. https://doi.org/10.1037//0022-3514.83.6.1314

Couture, S. M., Penn, D. L., Addington, J., Woods, S. W., & Perkins, D. O. (2008). Assessment of social judgments and complex mental states in the early phases of psychosis. Schizophrenia Research, 100(1–3), 237–241. https://doi.org/10.1016/j.schres.2007.12.484

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. J. Exp. Soc. Psychol. 66, 4–19. doi: 10.1016/j.jesp.2015.07.006.

Dalenberg, C. J., Brand, B. L., Gleaves, D. H., Dorahy, M. J., Loewenstein, R. J., Cardeña, E., … Spiegel, D. (2012). Evaluation of the evidence for the trauma and fantasy models of dissociation. Psychological Bulletin, 138(3), 550–588. https://doi.org/10.1037/a0027447

Darwin, H., Neave, N., and Holmes, J. (2011). Belief in conspiracy theories. The role of paranormal belief, paranoid ideation and schizotypy. Pers. Individ. Dif. 50, 1289–1293. doi: 10.1016/j.paid.2011.02.027.

Davies, C., Segre, G., Estradé, A., Radua, J., De Micheli, A., Provenzani, U., … Fusar-Poli, P. (2020). Prenatal and perinatal risk and protective factors for psychosis: a systematic review and meta-analysis. The Lancet Psychiatry, 7(5), 399–410. https://doi.org/10.1016/S2215-0366(20)30057-2

De Bont, P. A. J. M., Van Den Berg, D. P. G., Van Der Vleugel, B. M., De Roos, C., De Jongh, A., Van Der Gaag, M., & Van Minnen, A. M. (2016). Prolonged exposure and EMDR for PTSD v. a PTSD waiting-list condition: Effects on symptoms of psychosis, depression and social functioning in patients with chronic psychotic disorders. Psychological Medicine, 46(11), 2411–2421. https://doi.org/10.1017/S0033291716001094

De Coninck, D., Frissen, T., Matthijs, K., d'Haenens, L., Lits, G., Champagne-Poirier, O., et al. (2021). Beliefs in Conspiracy Theories and Misinformation About COVID-19: Comparative Perspectives on the Role of Anxiety, Depression and Exposure to and Trust in Information Sources. Front. Psychol. 12, 1–13. doi: 10.3389/fpsyg.2021.646394.

DeWall, C. N., & Bushman, B. J. (2011). Social acceptance and rejection: The sweet and the bitter. Current Directions in Psychological Science, 20(4), 256–260. https://doi.org/10.1177/0963721411417545

Dewitte, M., and De Houwer, J. (2011). Attachment-style Differences in the Appraisal of the Attachment Figure. Eur. J. Pers. 25, 173–183. doi: 10.1002/per.

Dewitte, M., De Houwer, J., Buysse, A., & Koster, E. H. W. (2008). Proximity seeking in adult attachment: Examining the role of automatic approach-avoidance tendencies. British Journal of Social Psychology, 47(4), 557–573. https://doi.org/10.1348/014466607X265148

Dewitte, M., Koster, E. H. W., De Houwer, J., and Buysse, A. (2007). Attentive processing of threat and adult attachment: A dot-probe study. Behav. Res. Ther. 45, 1307–1317. doi: 10.1016/j.brat.2006.11.004.

Đorđević, J. M., Žeželj, I., & Đurić, Ž. (2021). Beyond general political attitudes: Conspiracy mentality as a global belief system predicts endorsement of international and local conspiracy theories. Journal of Social and Political Psychology, 9(1), 144-158.

Douglas, K. M., Sutton, R. M., and Cichocka, A. (2017). The psychology of conspiracy theories. Curr. Dir. Psychol. Sci. 26, 538–542. doi: 10.1177/0963721417718261.

Dowbiggin, I. (2000). Delusional diagnosis? The history of paranoia as a disease concept in the modern era. History of Psychiatry, 11(41), 37–69. https://doi.org/10.1177/0957154x0001104103

Dudgeon, P. (2016). A Comparative Investigation of Confidence Intervals for IndependentVariables in Linear Regression. Multivariate Behav. Res. 51, 139–153. doi: 10.1080/00273171.2015.1121372.

Dudley, R. E. J., John, C. H., Young, A. W., & Over, D. E. (1997). The effect of self-referent material on the reasoning of people with delusions. British Journal of Clinical Psychology, 36(4), 575–584.doi:10.1111/j.2044-8260.1997.tb01262.x

Earls, H. A., Curran, T., & Mittal, V. (2016). Deficits in early stages of face processing in schizophrenia: A systematic review of the P100 component. Schizophrenia Bulletin, 42(2), 519–527. https://doi.org/10.1093/schbul/sbv096

Eichner, C., & Berna, F. (2016). Acceptance and efficacy of metacognitive training (mct) on positive symptoms and delusions in patients with schizophrenia: A meta-analysis taking into account important moderators. Schizophrenia Bulletin, 42(4), 952–962. https://doi.org/10.1093/schbul/sbv225

Elahi, A., McIntyre, J. C., Hampson, C., Bodycote, H. J., Sitko, K., White, R. G., & Bentall, R. P. (2018). Home is where you hang your hat: Host town identity, but not hometown identity, protects against mental health symptoms associated with financial stress. Journal of Social and Clinical Psychology, 37(3), 159–181. https://doi.org/10.1521/jscp.2018.37.3.159

Ellett, L., Allen-Crooks, R., Stevens, A., Wildschut, T., & Chadwick, P. (2013). A paradigm for the study of paranoia in the general population: the Prisoner's Dilemma Game. Cognition & Emotion, 27(1), 53–62. https://doi.org/https://dx.doi.org/10.1080/02699931.2012.689757

Ellett, L., Freeman, D., & Garety, P. (2008). The psychological effect of an urban environment on individuals with persecutory delusions: The Camberwell walk study. Schizophrenia Research, 99(1–3), 77–84. https://doi.org/http://dx.doi.org/10.1016/j.schres.2007.10.027

Ellett, L., Tarant, E., Kouimtsidis, C., Kingston, J., Vivarelli, L., Mendis, J., & Chadwick, P. (2020). Group mindfulness-based therapy for persecutory delusions: A pilot randomised controlled trial. Schizophrenia Research. https://doi.org/https://dx.doi.org/10.1016/j.schres.2020.04.023

Ellett, L., Wildschut, T., & Chadwick, P. (2023). Schizophrenia and Increased Distrust-Based Competitiveness in Interpersonal Interactions: A Serial Process Model. Schizophrenia Bulletin, 1–7. https://doi.org/10.1093/schbul/sbad021

Ellis, H. D., Young, A. W., Quayle, A. H., & De Pauw, K. W. (1997). Reduced autonomic responses to faces in Capgras delusion. Proceedings of the Royal Society B: Biological Sciences, 264(1384), 1085–1092.http://dx.doi.org/10.1098/rspb.1997.0150

Enders, A. M., Uscinski, J. E., Seelig, M. I., Klofstad, C. A., Wuchty, S., Funchion, J. R., et al. (2021). The Relationship Between Social Media Use and Beliefs in Conspiracy Theories and Misinformation. Polit. Behav. doi: 10.1007/s11109-021-09734-6.

Espinosa, R., Valiente, C., Varese, F., & Bentall, R. P. (2018). Can We Ameliorate Psychotic Symptoms by Improving Implicit Self-Esteem? A Proof-of-Concept Experience Sampling Study of an Evaluative Classical Conditioning Intervention. Journal of Nervous and Mental Disease, 206(9), 699–704. https://doi.org/10.1097/NMD.0000000000000858

Espinosa, R., Valiente, C., Varese, F., and Bentall, R. P. (2018). Can We Ameliorate Psychotic Symptoms by Improving Implicit Self-Esteem? A Proof-of-Concept Experience Sampling Study of an Evaluative Classical Conditioning Intervention. J. Nerv. Ment. Dis. 206, 699–704. doi: 10.1097/NMD.0000000000000858.

Essien, I., Stelter, M., Kalbe, F., Koehler, A., Mangels, J., & Meliß, S. (2017). The shooter bias: Replicating the classic effect and introducing a novel paradigm. Journal of Experimental Social Psychology, 70, 41–47. https://doi.org/10.1016/j.jesp.2016.12.009

Faris, R. E. L., & Dunham, H. W. (1939). Mental disorders in urban areas. Chicago University Press.

Fenigstein, A., & Vanable, P. A. (1992). Paranoia and self-consciousness. Journal of Personality and Social Psychology, 62(1), 129–138. https://doi.org/http://dx.doi.org/10.1037/0022-3514.62.1.129

Fett, A. K. J., Shergill, S. S., Korver-Nieberg, N., Yakub, F., Gromann, P. M., & Krabbendam, L. (2016). Learning to trust: Trust and attachment in early psychosis. Psychological Medicine, 46(7), 1437–1447. https://doi.org/10.1017/S0033291716000015

Fett, A.-K. J., Shergill, S. S., Joyce, D. W., Riedl, A., Strobel, M., Gromann, P. M., & Krabbendam, L. (2012). To trust or not to trust: the dynamics of social interaction in psychosis. Brain : A

Journal of Neurology, 135(Pt 3), 976–984. https://doi.org/https://dx.doi.org/10.1093/brain/awr359

Forkert, A., Brown, P., Freeman, D., & Waite, F. (2021). A compassionate imagery intervention for patients with persecutory delusions. Behavioural and Cognitive Psychotherapy, 1–13. doi:10.1017/S1352465821000229

Fornells-Ambrojo, M., & Garety, P. A. (2005). Bad me paranoia in early psychosis: A relatively rare phenomenon. British Journal of Clinical Psychology, 44(4), 521–528. https://doi.org/10.1348/014466505X35326

Fox, J., & Weisberg, S. (2019). An (R) Companion to Applied Regression. Thousand Oaks, CA: Sage.

Freeman, D. (2007). Suspicious minds: The psychology of persecutory delusions. Clin. Psychol. Rev. 27, 425–457. doi: 10.1016/j.cpr.2006.10.004.

Freeman, D. (2008). The assessment of persecutory ideation. In D. Freeman, R. Bentall, & P. Garety (Eds.), Persecutory delusions: Assessment, theory, treatment (pp. 23-52). Oxford University Press.

Freeman, D. (2016). Persecutory delusions: a cognitive perspective on understanding and treatment. The Lancet Psychiatry, 3(7), 685–692. https://doi.org/10.1016/S2215-0366(16)00066-3

Freeman, D., & Bentall, R. P. (2017). The concomitants of conspiracy concerns. Social Psychiatry and Psychiatric Epidemiology, 52(5), 595–604. https://doi.org/https://dx.doi.org/10.1007/s00127-017-1354-4

Freeman, D., & Garety, P. (2014). Advances in understanding and treating persecutory delusions: A review. Social Psychiatry and Psychiatric Epidemiology: The International Journal for

Research in Social and Genetic Epidemiology and Mental Health Services, 49(8), 1179–1189. https://doi.org/http://dx.doi.org/10.1007/s00127-014-0928-7

Freeman, D., and Bentall, R. P. (2017). The concomitants of conspiracy concerns. Soc. Psychiatry Psychiatr. Epidemiol. 52, 595–604. doi: https://dx.doi.org/10.1007/s00127-017-1354-4.

Freeman, D., and Garety, P. A. (2000). Comments on the content of persecutory delusions: Does the definition need clarification? Br. J. Clin. Psychol. 39, 407–414. doi: 10.1348/014466500163400.

Freeman, D., Dunn, G., Murray, R. M., Evans, N., Lister, R., Antley, A., … Morrison, P. D. (2015). How cannabis causes paranoia: using the intravenous administration of 9-tetrahydrocannabinol (THC) to identify key cognitive mechanisms leading to paranoia. Schizophrenia Bulletin, 41(2), 391–399. https://doi.org/https://dx.doi.org/10.1093/schbul/sbu098

Freeman, D., Garety, P. A., Bebbington, P. E., Smith, B., Rollinson, R., Fowler, D., … Dunn, G. (2005). Psychological investigation of the structure of paranoia. The British Journal of Psychiatry, 186(5). https://doi.org/10.1192/bjp.186.5.427

Freeman, D., Garety, P. A., Bebbington, P. E., Smith, B., Rollinson, R., Fowler, D., Kuipers, E., Ray, K., & Dunn, G. (2005). Psychological investigation of the structure of paranoia. The British Journal of Psychiatry, 186(5). https://doi.org/10.1192/bjp.186.5.427

Freeman, D., Garety, P. A., Kuipers, E., Fowler, D., & Bebbington, P. E. (2002). A cognitive model of persecutory delusions. British Journal of Clinical Psychology, 41(4), 331–347. https://doi.org/10.1348/014466502760387461

Freeman, D., Garety, P. A., Kuipers, E., Fowler, D., Bebbington, P. E., & Dunn, G. (2007). Acting on persecutory delusions: The importance of safety seeking. Behaviour Research and Therapy, 45(1), 89–99. https://doi.org/10.1016/j.brat.2006.01.014

Freeman, D., Loe, B. S., Kingdon, D., Startup, H., Molodynski, A., Rosebrock, L., Brown, P., Sheaves, B., Waite, F., & Bird, J. C. (2021). The revised Green et al., Paranoid Thoughts Scale (R-GPTS): Psychometric properties, severity ranges, and clinical cut-offs. Psychological Medicine, 51(2), 244–253. https://doi.org/10.1017/S0033291719003155

Freeman, D., Pugh, K., Green, C., Valmaggia, L., Dunn, G., & Garety, P. (2007b). A measure of state persecutory ideation for experimental studies. Journal of Nervous and Mental Disease, 195(9), 781–784. https://doi.org/10.1097/NMD.0b013e318145a0a9

Freeman, D., Pugh, K., Vorontsova, N., & Southgate, L. (2009). Insomnia and paranoia. Schizophrenia Research, 108(1–3), 280–284. https://doi.org/http://dx.doi.org/10.1016/j.schres.2008.12.001

Freeman, D., Sheaves, B., Goodwin, G. M., Yu, L. M., Nickless, A., Harrison, P. J., … Espie, C. A. (2017). The effects of improving sleep on mental health (OASIS): a randomised controlled trial with mediation analysis. The Lancet Psychiatry, 4(10), 749–758. https://doi.org/10.1016/S2215-0366(17)30328-0

Freeman, D., Sheaves, B., Goodwin, G. M., Yu, L.-M., Nickless, A., Harrison, P. J., Emsley, R., Luik, A. I., Foster, R. G., Wadekar, V., Hinds, C., Gumley, A., Jones, R., Lightman, S., Jones, S., Bentall, R., Kinderman, P., Rowse, G., Brugha, T., … Espie, C. A. (2017). The effects of improving sleep on mental health (OASIS): a randomised controlled trial with mediation analysis. The Lancet. Psychiatry, 4(10), 749–758. https://doi.org/https://dx.doi.org/10.1016/S2215-0366(17)30328-0

Freeman, D., Startup, H., Dunn, G., Černis, E., Wingham, G., Pugh, K., … Kingdon, D. (2013). The interaction of affective with psychotic processes: A test of the effects of worrying on working memory, jumping to conclusions, and anomalies of experience in patients with persecutory delusions. Journal of Psychiatric Research, 47(12), 1837–1842. https://doi.org/10.1016/j.jpsychires.2013.06.016

Freeman, D., Waite, F., Rosebrock, L., Petit, A., Causier, C., East, A., et al. (2020). Coronavirus Conspiracy Beliefs, Mistrust, and Compliance with Government Guidelines in England. Psychol. Med. doi: 10.1017/S0033291720001890.

Freeman, D., Waite, F., Startup, H., Myers, E., Lister, R., McInerney, J., … Yu Felicity. (2015). Efficacy of cognitive behavioural therapy for sleep improvement in patients with persistent delusions and hallucinations (BEST): A prospective, assessor-blind, randomised controlled pilot trial. The Lancet Psychiatry, 2(11), 975–983. https://doi.org/http://dx.doi.org/10.1016/S2215-0366%2815%2900314-4

Freud, S. (1911/1950). Psychoanalytic notes upon an autobiographical account of a case of paranoia (Dementia Paranoides). In Collected Papers (Vol. III, pp. 387-466). Hogarth Press.

Frith, C. D. (1979). Consciousness, information processing and schizophrenia. British Journal of Psychiatry, 134(3), 225–235. https://doi.org/10.1192/bjp.134.3.225

Furnham, A., & Crump, J. (2015). A big five facet analysis of a Paranoid Personality Disorder: The validity of the HDS Sceptical scale of subclinical paranoia. Journal of Individual Differences, 36(4), 199–204. https://doi.org/10.1027/1614-0001/a000174

Gantiva, C., Sotaquirá, M., Araujo, A., & Cuervo, P. (2020). Cortical processing of human and emoji faces: an ERP analysis. Behaviour and Information Technology, 39(8), 935–943. https://doi.org/10.1080/0144929X.2019.1632933

Gao, L. E. I., Xu, J., Zhang, B., Zhao, L. U. N., Harel, A., & Bentin, S. (2009). Aging effects on early-stage face perception : An ERP study. 46, 970–983. https://doi.org/10.1111/j.1469-8986.2009.00853.x

Garety, P. A, & Freeman, D. (2013). The past and future of delusions research: From the inexplicable to the treatable. The British Journal of Psychiatry, 203(5), 327–333. https://doi.org/http://dx.doi.org/10.1192/bjp.bp.113.126953

Garety, P., Ward, T., Emsley, R., Greenwood, K., Freeman, D., Fowler, D., Kuipers, E., Bebbington, P., Rus-Calafell, M., McGourty, A., Sacadura, C., Collett, N., James, K., & Hardy, A. (2021). Effects of SlowMo, a Blended Digital Therapy Targeting Reasoning, on Paranoia among People with Psychosis: A Randomized Clinical Trial. JAMA Psychiatry, 78(7), 714–725. https://doi.org/10.1001/jamapsychiatry.2021.0326

Garety, P.A, Hemsley, D. ., & Wessely, S. (1991). Reasoning in Deluded Schizophrenic and Paranoid Patients. The Journal of Nervous and Mental Disease, 179(4), 194–201.

Garety, Philippa A., & Freeman, D. (1999). Cognitive approaches to delusions: A critical review of theories and evidence. British Journal of Clinical Psychology, 38(2), 113–154. https://doi.org/10.1348/014466599162700

George A. Keepers, M.D. , ,, Laura J. Fochtmann, M.D., M.B.I. , ,, Joan M. Anzia, M.D. ,, Sheldon Benjamin, M.D. ,, Jeffrey M. Lyness, M.D. ,, Ramin Mojtabai, M.D. ,, Mark Servis, M.D. ,, Art Walaszek, M.D. ,, Peter Buckley, M.D. ,, Mark F. Lenzenweger, Ph.D. ,, Alexander S. Young, M.D., M.S.H.S. ,, Amanda Degenhardt, M.D. ,, & Seung-Hee Hong. (2020). The American Psychiatric Association Practice Guideline for the Treatment of Patients With Schizophrenia. American Journal of Psychiatry, 177(9), 868-872. https://doi.org/10.1176/appi.ajp.2020.177901

Gillath, O., Hart, J., Noftle, E. E., and Stockdale, G. D. (2009). Development and validation of a state adult attachment measure (SAAM). J. Res. Pers. doi: 10.1016/j.jrp.2008.12.009.

Goertzel, T. (1994). Belief in conspiracy theories. Politic.Psychol.731-742.

Goreis, A., and Voracek, M. (2019). A systematic review and meta-analysis of psychological research on conspiracy beliefs: Field characteristics, measurement instruments, and associations with personality traits. Front. Psychol. 10, 1–13. doi: 10.3389/fpsyg.2019.00205.

Green, C. E. L., Freeman, D., Kuipers, E., Bebbington, P., Fowler, D., Dunn, G., & Garety P. A.; ORCID: http://orcid.org/0000-0002-5637-1340, P. A. A. I.-O. http://orcid. org/Garet. (2008). Measuring ideas of persecution and social reference: The Green et al. Paranoid Thought Scales (GPTS). Psychological Medicine, 38(1), 101–111. https://doi.org/http://dx.doi.org/10.1017/S0033291707001638

Green, M. F., Horan, W. P., & Lee, J. (2015). Social cognition in schizophrenia. Nature Reviews Neuroscience, 16(10), 620–631. https://doi.org/10.1038/nrn4005

Green, M. F., Penn, D. L., Bentall, R., Carpenter, W. T., Gaebel, W., Gur, R. C., … Heinssen, R. (2008). Social Cognition in Schizophrenia: An NIMH Workshop on Definitions, Assessment, and Research Opportunities. Schizophrenia Bulletin, 34(6), 1211–1220. https://doi.org/10.1093/schbul/sbm145

Green, R., and Douglas, K. (2018). Anxious attachment and belief in conspiracy theories. Pers. Individ. Dif. 125, 30–37. doi: http://dx.doi.org/10.1016/j.paid.2017.12.023.

Greenaway, K. H., Haslam, S. A., & Bingley, W. (2019). Are “they” out to get me? A social identity model of paranoia. Group Processes and Intergroup Relations, 22(7), 984–1001. https://doi.org/10.1177/1368430218793190

Greenburgh, A. G., Liefgreen, A., Bell, V., & Raihani, N. (2022). Factors affecting conspiracy theory endorsement in paranoia. Royal Society Open Science, 9(1), 211555.

Greenburgh, A., Bell, V., & Raihani, N. (2019). Paranoia and conspiracy: group cohesion increases harmful intent attribution in the Trust Game. PeerJ, 7, e7403.

Gromann, P. M., Heslenfeld, D. J., Fett, A.-K., Joyce, D. W., Shergill, S. S., & Krabbendam, L. (2013). Trust versus paranoia: abnormal response to social reward in psychotic illness. Brain : A Journal of Neurology, 136(Pt 6), 1968–1975. https://doi.org/https://dx.doi.org/10.1093/brain/awt076

Grzesiak-Feldman, M., and Ejsmont, A. (2008). Paranoia and conspiracy thinking of jews, arabs, germans, and russians in a polish sample. Psychol. Rep. 102, 884–886. doi: 10.2466/PR0.102.3.884-886.

Gumley, A. I., Taylor, H. E. F., Schwannauer, M., & MacBeth, A. (2014). A systematic review of attachment and psychosis: Measurement, construct validity and outcomes. Acta Psychiatrica Scandinavica, 129(4), 257–274. https://doi.org/10.1111/acps.12172

Gummerum, M., Hanoch, Y., & Keller, M. (2008). When child development meets economic game theory: An interdisciplinary approach to investigating social development. Human Development, 51(4), 235–261. DOI: 10.1159/000151494

Guze, S. B. (1989). Biological psychiatry: Is there any other kind. Psychological Medicine, 19(2), 315–323. https://doi.org/10.1017/S0033291700012356

Hahn, A., & Gawronski, B. (2015). Implicit Social Cognition. International Encyclopedia of the Social & Behavioral Sciences: Second Edition, 4, 714–720. https://doi.org/10.1016/B978-0-08-097086-8.24066-X

Hart, J., Shaver, P. R., and Goldenberg, J. L. (2005). Attachment, self-esteem, worldviews, and terror management: Evidence for a tripartite security system. J. Pers. Soc. Psychol. 88, 999–1013. doi: 10.1037/0022-3514.88.6.999.

Hartman, T. K., Marshall, M., Stocks, T. V. A., McKay, R., Bennett, K., Butter, S., et al. (2021). Different Conspiracy Theories Have Different Psychological and Social Determinants: Comparison of Three Theories About the Origins of the COVID-19 Virus in a Representative Sample of the UK Population. Front. Polit. Sci. 3, 1–17. doi: 10.3389/fpos.2021.642510.

Haselton, M. G., & Buss, D. M. (2000). Error Management Theory: A New Perspective on Biases in Cross-Sex Mind Reading (pp. 1–11). pp. 1–11.

Haselton, M. G., & Nettle, D. (2006). The Paranoid Optimist: An Integrative Evolutionary Model of Cognitive Biases Correspondence should be sent to. Personality and Social Psychology Review, 10(1), 47–66.

Hatzakis, T. (2009). Towards a framework of trust attribution styles. Br. J. Manag. 20, 448–460. doi: http://dx.doi.org/10.1111/j.1467-8551.2008.00596.x.

Haut, K. M., & MacDonald, A. W. (2010). Persecutory delusions and the perception of trustworthiness in unfamiliar faces in schizophrenia. Psychiatry Research, 178(3), 456–460. https://doi.org/10.1016/j.psychres.2010.04.015

Heatherton, T. F., and Polivy, J. (1991). Development and Validation of a Scale for Measuring State Self-Esteem. J. Pers. Soc. Psychol. 60, 895–910. doi: 10.1037/0022-3514.60.6.895.

Hemsley, D. R., & Garety, P. A. (1986). The formation of maintenance of delusions: A Bayesian analysis. British Journal of Psychiatry, 149(JULY), 51–56. https://doi.org/10.1192/bjp.149.1.51

Hemsley, D. R., & Zawada, S. L. (1976). “Filtering” and the cognitive deficit in schizophrenia. British Journal of Psychiatry, 128(5), 456–461. https://doi.org/10.1192/bjp.128.5.456

Hillmann, T. E., Ascone, L., Kempkensteffen, J., & Lincoln, T. M. (2017). Scanning to conclusions? Visual attention to neutral faces under stress in individuals with and without subclinical paranoia. Journal of Behavior Therapy and Experimental Psychiatry, 56, 137–143. https://doi.org/10.1016/j.jbtep.2016.08.013

Hinojosa, J. A., Mercado, F., & Carretié, L. (2015). N170 sensitivity to facial expression: A meta-analysis. Neuroscience and Biobehavioral Reviews, 55, 498–509. https://doi.org/10.1016/j.neubiorev.2015.06.002

Hlavac, M. (2018). Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.1. Bratislava, Slovakia: Central European Labour Studies Institute (CELSI).

Hoehne, K. A. (1988). Ernst Kretschmer’s multidimensional psychiatry. Journal of Personality Disorders, 2(1), 28–35.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., and Crombez, G. (2010). Evaluative Conditioning in Humans: A Meta-Analysis. Psychol. Bull. 136, 390–421. doi: 10.1037/a0018916.

Hofstadter, R. (1965). The paranoid style in American politics. Harper’s Mag., 77–86.

Hooker, C. I., Tully, L. M., Verosky, S. C., Fisher, M., Holland, C., and Vinogradov, S. (2011). Can I trust you? Negative affective priming influences social judgments in schizophrenia. J. Abnorm. Psychol. 120, 98–107. doi: https://dx.doi.org/10.1037/a0020630.

Howell, J. L., Sosa, N., and Osborn, H. J. (2019). Self-esteem as a monitor of fundamental psychological need satisfaction. Soc. Personal. Psychol. Compass 13, 1–16. doi: 10.1111/spc3.12492.

https://doi.org/https://dx.doi.org/10.1007/s00127-017-1354-4

Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct. Equ. Model. 6, 1–55. doi: 10.1080/10705519909540118.

Huddy, V., Brown, G. P., Boyd, T., & Wykes, T. (2014). An exploratory investigation of real-world reasoning in paranoia. Psychology and Psychotherapy, 87(1), 44–59. https://doi.org/https://dx.doi.org/10.1111/j.2044-8341.2012.02072.x

Humphrey, C., Bucci, S., Varese, F., Degnan, A., & Berry, K. (2021). Paranoia and negative schema about the self and others: A systematic review and meta-analysis. Clinical Psychology Review, 90(April), 102081. https://doi.org/10.1016/j.cpr.2021.102081

Humphrey, C., Bucci, S., Varese, F., Degnan, A., and Berry, K. (2021). Paranoia and negative schema about the self and others: A systematic review and meta-analysis. Clin. Psychol. Rev. 90, 102081. doi: 10.1016/j.cpr.2021.102081.

Imhoff, R., & Lamberty, P. (2018). How paranoid are conspiracy believers? Toward a more fine-grained understanding of the connect and disconnect between paranoia and belief in conspiracy theories. European Journal of Social Psychology, 48(7), 909–926. https://doi.org/10.1002/ejsp.2494

Imhoff, R., and Bruder, M. (2014). Speaking (un-)truth to power: Conspiracy mentality as a generalised political attitude. Eur. J. Pers. 28, 25–43. doi: http://dx.doi.org/10.1002/per.1930.

Imhoff, R., and Lamberty, P. (2018). How paranoid are conspiracy believers? Toward a more fine-grained understanding of the connect and disconnect between paranoia and belief in conspiracy theories. Eur. J. Soc. Psychol. 48, 909–926. doi: 10.1002/ejsp.2494.

Inglehart, R., Basanez, M., Diez-Medrano, J., Halman, L. and Luijkx, R., 2000. World values surveys and European values surveys, 1981-1984, 1990-1993, and 1995-1997. Ann Arbor-Michigan, Institute for Social Research, ICPSR version.

Janssen, I., Hanssen, M., Bak, M., Bijl, R. V., De Graaf, R., Vollenberg, W., McKenzie, K., & van Os, J. (2003). Discrimination and delusional ideation. British Journal of Psychiatry, 182, 71-76. https://doi.org/10.1192/bjp.182.1.71

Jetten, J., Haslam, S. A., Cruwys, T., Greenaway, K. H., Haslam, C., & Steffens, N. K. (2017). Advancing the social identity approach to health and well-being: Progressing the social cure research agenda. European Journal of Social Psychology, 47(7), 789-802. https://doi.org/https://doi.org/10.1002/ejsp.2333

Jordan, C. H., Spencer, S. J., Zanna, M. P., Hoshino-Browne, E., and Correll, J. (2003). Secure and Defensive High Self-Esteem. J. Pers. Soc. Psychol. 85, 969–978. doi: 10.1037/0022-3514.85.5.969.

Kahn-Greene, E. T., Killgore, D. B., Kamimori, G. H., Balkin, T. J., & Killgore, W. D. S. (2007). The effects of sleep deprivation on symptoms of psychopathology in healthy adults. Sleep Medicine, 8(3), 215–221. https://doi.org/10.1016/j.sleep.2006.08.007

Kaney, S., & Bentall, R. P. (1989). Persecutory delusions and attributional style. British Journal of Medical Psychology, 62(2), 191–198. https://doi.org/10.1111/j.2044-8341.1989.tb02826.x

Kaney, S., & Bentall, R. P. (1992). Persecutory Delusions and the Self-Serving Bias. Evidence from a Contingency Judgement Task. The Journal of Nervous and Mental Disease, 180(12), 773–780.

Kaplan, J. (2020). Package ‘ fastDummies .’

Kasanova, Z., Hajdúk, M., Thewissen, V., & Myin-Germeys, I. (2020). Temporal associations between sleep quality and paranoia across the paranoia continuum: An experience sampling study. Journal of Abnormal Psychology, 129(1), 122–130. https://doi.org/10.1037/abn0000453

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The positive and negative syndrome scale (panss) manual. Schizophrenia Bulletin, 13(2). https://doi.org/10.1016/0006-3223(91)91039-t

Kendler, K. (2018). The development of Kraepelin's mature diagnostic concepts of paranoia (die verrücktheit) and paranoid dementia praecox (dementia paranoides): A close reading of his textbooks from 1887 to 1899. JAMA psychiatry, 75(12), 1280-1288. https://doi.org/10.1001/jamapsychiatry.2018.2377

Kendler, K. S. (1988). Kraepelin and the diagnostic concept of paranoia. Comprehensive Psychiatry, 29(1), 4–11. https://doi.org/10.1016/0010-440X(88)90031-4

Kendler, K. S. (2015). A joint history of the nature of genetic variation and the nature of schizophrenia. Molecular Psychiatry, 20(1), 77–83. https://doi.org/10.1038/mp.2014.94

Kendler, Kenneth S. (1980). The nosologic validity of paranoia (simple delusional disorder): a review. Archives of General Psychiatry, 37(6), 699–706.

Kendler, Kenneth S. (1988). Kraepelin and the diagnostic concept of paranoia. Comprehensive Psychiatry. https://doi.org/10.1016/0010-440X(88)90031-4

Kesting, M. L., & Lincoln, T. M. (2013). The relevance of self-esteem and self-schemas to persecutory delusions: A systematic review. Comprehensive Psychiatry. https://doi.org/10.1016/j.comppsych.2013.03.002

Key, A. P. F., Dove, G. O., & Maguire, M. J. (2005). Linking brainwaves to the brain: An ERP primer. Developmental Neuropsychology, 27(2), 183–215. https://doi.org/10.1207/s15326942dn2702_1

Kim, S. (2015). ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. Commun. Stat. Appl. Methods 22, 665–674. doi: 10.5351/csam.2015.22.6.665.

Kim, S., and Kim, S. (2021). Searching for general model of conspiracy theories and its implication for public health policy: Analysis of the impacts of political, psychological, structural factors on conspiracy beliefs about the covid-19 pandemic. Int. J. Environ. Res. Public Health 18, 1–28. doi: 10.3390/ijerph18010266.

Kinderman, P., & Bentall, R. P. (1996). Self-discrepancies and persecutory delusions: Evidence for a model of paranoid ideation. Journal of Abnormal Psychology, 105(1), 106–113.

King, G. (1986). How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science. American Journal of Political Science, 30(3), 666. https://doi.org/10.2307/2111095

Kirk, H., Gilmour, A., Dudley, R., and Riby, D. (2013). Paranoid Ideation and Assessments of Trust. J. Exp. Psychopathol. 4, 360–367. doi: 10.5127/jep.027812.

Kliemann, D., Rosenblau, G., Bölte, S., Heekeren, H. R., and Dziobek, I. (2013). Face puzzle-two new video-based tasks for measuring explicit and implicit aspects of facial emotion recognition. Front. Psychol. 4, 1–13. doi: 10.3389/fpsyg.2013.00376.

Kong, D. T. (2017). Sojourners' Ineffective Sociocultural Adaptation: Paranoia as a Joint Function of Distrust toward Host Nationals and Neuroticism. Curr. Psychol. 36, 540–548. doi: 10.1007/s12144-016-9441-3.

Kramer, R. M. (1994). The sinister attribution error: Paranoid cognition and collective distrust in organizations. Motiv. Emot. 18, 199–230. doi: 10.1007/BF02249399.

Krause, S., Back, M. D., Egloff, B., and Schmukle, S. C. (2011). Reliability of Implicit Self-esteem Measures Revisited. Eur. J. Pers. 25, 239–251. doi: 10.1002/per.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2002). The PHQ-15: Validity of a new measure for evaluating the severity of somatic symptoms. Psychosomatic Medicine, 64(2), 258–266. https://doi.org/10.1097/00006842-200203000-00008

Lahera, G., Reboreda, A., Vallespí, A., Vidal, C., López, V., Aznar, A., Fernández, P., Escolar, M., Martínez-Alés, G., Rodriguez-Jimenez, R., Halverson, T., Frommann, N., Wölwer, W., & Penn, D. L. (2021). Social Cognition and Interaction Training (SCIT) versus Training in Affect Recognition (TAR) in patients with schizophrenia: A randomized controlled trial. Journal of Psychiatric Research, 142(May), 101–109. https://doi.org/10.1016/j.jpsychires.2021.07.029

Leary, M. R. (2005). Sociometer theory and the pursuit of relational value: Getting to the root of self-esteem. Eur. Rev. Soc. Psychol. 16, 75–111. doi: 10.1080/10463280540000007.

LeBel, E., and Bertram, G. (2009). How to Find What's in a Name: Scrutinizing the Optimality of FiveScoring Algorithms for the Name-letter Task. Eur. J. Pers. 23, 85–106. doi: 10.1002/per.

Lebert, L., Turkington, D., Freeston, M., & Dudley, R. (2021). Rumination, intolerance of uncertainty and paranoia in treatment resistant psychosis. Psychosis, 13(1), 65-70.

Lecomte, T., Corbière, M., & Laisné, F. (2006). Investigating self-esteem in individuals with schizophrenia: Relevance of the Self-Esteem Rating Scale-Short Form. Psychiatry Research. https://doi.org/10.1016/j.psychres.2005.08.019

Lee R. (2017). Mistrustful and Misunderstood A Review of Paranoid Personality Disorder. Current Behavioral Neuroscience Reports, 4(2), 151–165. https://doi.org/10.1007/s40473-017-0116-7.Mistrustful

Lee, A., and Hankin, B. L. (2009). Insecure attachment, dysfunctional attitudes, and low self-esteem predicting prospective symptoms of depression and anxiety during adolescence. J. Clin. Child Adolesc. Psychol. 38, 219–231. doi: 10.1080/15374410802698396.

Lee, I. A., & Preacher, K. J. (2013).Calculation for the test of the difference between two dependent correlations with no variable in common [Computer software]. Available from http://quantpsy.org/corrtest/corrtest3.htm

Lewicki, R. J., & Brinsfield, C. T. (2012). Measuring tust beliefs and behaviours. Handbook of Research Methods on Trust (F. Lyon, G. Möllering, & M. N. . Saunders, Eds.). Edward Elgar Publishing.

Lewicki, R. J., and Brinsfield, C. T. (2012). Measuring tust beliefs and behaviours. Handbook of Research Methods on Trust. , eds. F. Lyon, G. Möllering, and M. N. . Saunders Edward Elgar Publishing.

Lewicki, R. J., Tomlinson, E. C., & Gillespie, N. (2006). Models of Interpersonal Trust Development: Theoretical Approaches, Empirical Evidence, and Future Directions. Journal of Management, 32(6), 991–1022. https://doi.org/http://dx.doi.org/10.1177/0149206306294405

Lewis, A. (1970). Paranoia and paranoid: A historical perspective. Psychological Medicine, 1(1), 2–12. https://doi.org/10.1017/S0033291700039969

Lincoln, T. M., Hohenhaus, F., and Hartmann, M. (2013). Can paranoid thoughts be reduced by targeting negative emotions and self-esteem? An experimental investigation of a brief

compassion-focused intervention. Cognit. Ther. Res. 37, 390–402. doi: 10.1007/s10608-012-9470-7.

Lincoln, T. M., Salzmann, S., Ziegler, M., & Westermann, S. (2011). When does jumping-to-conclusions reach its peak? the interaction of vulnerability and situation-characteristics in social reasoning. Journal of Behavior Therapy and Experimental Psychiatry, 42(2), 185–191. https://doi.org/10.1016/j.jbtep.2010.09.005

Lissek, S., Levenson, J., Biggs, A. L., Johnson, L. L., Ameli, R., Pine, D. S., et al. (2008). Elevated fear conditioning to socially relevant unconditioned stimuli in social anxiety disorder. Am. J. Psychiatry 165, 124–132. doi: 10.1176/appi.ajp.2007.06091513.

Longden, E., Branitsky, A., Moskowitz, A., Berry, K., Bucci, S., & Varese, F. (2020). The Relationship Between Dissociation and Symptoms of Psychosis: A Meta-analysis. Schizophrenia Bulletin. https://doi.org/https://dx.doi.org/10.1093/schbul/sbaa037

Lynn, S. K., & Barrett, L. F. (2014). "Utilizing" Signal Detection Theory. Psychological Science, 25(9), 1663–1673. https://doi.org/10.1177/0956797614541991

MacCallum, R. C., Browne, M. W., and Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. Psychol. Methods 1, 130–149. doi: 10.1037/1082-989X.1.2.130.

Maher, B. (1999). Anomalous Experience in Everyday Life: Its Significance for Psychopathology. Monist, 82(4), 547–570. https://doi.org/10.5840/monist199982428

Maher, B. A. (1974). Delusional thinking and perceptual disorder. Journal of Individual Psychology, 30(1), 98–113.

Mari, S., Gil de Zúñiga, H., Suerdem, A., Hanke, K., Brown, G., Vilar, R., et al. (2021). Conspiracy Theories and Institutional Trust: Examining the Role of Uncertainty Avoidance and Active Social Media Use. Polit. Psychol. 0. doi: 10.1111/pops.12754.

Martinez, A. P., Agostini, M., Al-Suhibani, A., and Bentall, R. P. (2020). Mistrust and negative self-esteem: Two paths from attachment styles to paranoia. Psychol. Psychother. Theory, Res. Pract., 1–16. doi: 10.1111/papt.12314.

Martinez, A. P., Dorahy, M. J., Nesbit, A., Palmer, R., & Middleton, W. (2020). Delusional beliefs and their characteristics: A comparative study between dissociative identity disorder and schizophrenia spectrum disorders. Journal of Psychiatric Research, 131(July), 263–268. https://doi.org/10.1016/j.jpsychires.2020.09.015

Martinez, A. P., Shevlin, M., Valiente, C., Hyland, P., and Bentall, R. P. (2022). Paranoid beliefs and conspiracy mentality are associated with different forms of mistrust: A three-nation study. Front. Psychol. 13, 1–11. doi: 10.3389/fpsyg.2022.1023366.

Marzi, T., Righi, S., Ottonello, S., Cincotta, M., & Viggiano, M. P. (2014). Trust at first sight: Evidence from ERPs. Social Cognitive and Affective Neuroscience, 9(1), 63–72. https://doi.org/10.1093/scan/nss102

McBride, O., Murphy, J., Shevlin, M., Gibson-Miller, J., Hartman, T. K., Hyland, P., et al. (2021). Monitoring the psychological, social, and economic impact of the COVID-19 pandemic in the population: Context, design and conduct of the longitudinal COVID-19 psychological research consortium (C19PRC) study. Int. J. Methods Psychiatr. Res. 30, 1–16. doi: 10.1002/mpr.1861.

McCleery, A., Lee, J., Joshi, A., Wynn, J. K., Hellemann, G. S., & Green, M. F. (2015). Meta-analysis of face processing event-related potentials in schizophrenia. Biological Psychiatry, 77(2), 116–126. https://doi.org/10.1016/j.biopsych.2014.04.015

McElroy, E., McIntyre, J. C., Bentall, R. P., Wilson, T., Holt, K., Kullu, C., … Corcoran, R. (2019). Mental health, deprivation, and the neighbourhood social environment: a network analysis Authors: Clinical Psychological Science, 1–43.

McIntyre, J. C., Elahi, A., & Bentall, R. . (2016). Social identity and psychosis: Explaining elevated rates of psychosis in migrant populations. Social and Personality Psychology Compass, 10(11), 619–633.

McIntyre, J. C., Elahi, A., Barlow, F. ., White, R. ., & Bentall, R. . (2019). The relationship between ingroup identity and Paranoid ideation among people from African and African Caribbean backgrounds. Psychology and Psychotherapy: Theory, Research and Practice, 94(1), 16–32. https://doi.org/10.1111/papt.12261

McIntyre, J. C., Wickham, S., Barr, B., & Bentall, R. P. (2018). Social Identity and Psychosis: Associations and Psychological Mechanisms. Schizophrenia Bulletin, 44(3), 681–690. https://doi.org/https://dx.doi.org/10.1093/schbul/sbx110

McKay, R., Langdon, R., & Coltheart (2006). The Persecutory Ideation Questionnaire. Journal of Nervous and Mental Disease, 194(8), 628–631. https://doi.org/http://dx.doi.org/10.1097/01.nmd.0000231441.48007.a5

McLean, B. F., Mattiske, J. K., & Balzan, R. P. (2017). Association of the jumping to conclusions and evidence integration biases with delusions in psychosis: A detailed meta-analysis. Schizophrenia Bulletin, 43(2), 344–354. https://doi.org/10.1093/schbul/sbw056

Mehl, S., Werner, D., & Lincoln, T. M. (2015). Does Cognitive Behavior Therapy for psychosis (CBTp) show a sustainable effect on delusions? A meta-analysis. Frontiers in Psychology, 6(OCT). https://doi.org/10.3389/fpsyg.2015.01450

Mekawi, Y., & Bresin, K. (2015). Is the evidence from racial bias shooting task studies a smoking gun? Results from a meta-analysis. Journal of Experimental Social Psychology, 61(September 2015), 120–130. https://doi.org/10.1016/j.jesp.2015.08.002

Melo, S., Corcoran, R., Shryane, N., and Bentall, R. P. (2009). The persecution and deservedness scale. Psychol. Psychother. Theory, Res. Pract. 82, 247–260. doi: http://dx.doi.org/10.1348/147608308X398337.

Meuer, M., and Imhoff, R. (2021). Believing in hidden plots is associated with decreased behavioral trust: Conspiracy belief as greater sensitivity to social threat or insensitivity towards its absence? J. Exp. Soc. Psychol. 93, 104081. doi: 10.1016/j.jesp.2020.104081.

Mikulincer, M. (1995). Attachment style and the mental representation of the self. Journal of Personality and. Soc. Psychol. (Gott). 69, 1203–1215.

Mikulincer, M. (1998). Mikulincer, M. (1998). Attachment working models and the sense of trust: An exploration of interaction goals and affect regulation. Journal of Personality and Social Psychology, 74(5), 209–224.

Miller, J. M., Saunders, K. L., and Farhart, C. E. (2016). Conspiracy Endorsement as Motivated Reasoning: The Moderating Roles of Political Knowledge and Trust. Am. J. Pol. Sci. 60, 824–844. doi: 10.1111/ajps.12234.

Mirowsky, J., & Ross, C. E. (1983). Paranoia and the structure of powerlessness. American Sociological Review, 48(2), 228–239.

Mishara, A. L. (2010). Klaus Conrad (1905-1961): delusional mood, psychosis, and beginning schizophrenia. Schizophrenia Bulletin, 36(1), 9–13. https://doi.org/10.1093/schbul/sbp144

Moritz, S., Menon, M., Balzan, R., & Woodward, T. S. (2022). Metacognitive training for psychosis (MCT): past, present, and future. European Archives of Psychiatry and Clinical Neuroscience, 0123456789. https://doi.org/10.1007/s00406-022-01394-9

Moutoussis, M., Williams, J., Dayan, P., & Bentall, R. P. (2007). Persecutory delusions and the conditioned avoidance paradigm: Towards an integration of the psychology and biology of paranoia. Cognitive Neuropsychiatry, 12(6), 495–510. https://doi.org/10.1080/13546800701566686

Müller, H., Betz, L. T., & Bechdolf, A. (2021). A comprehensive meta-analysis of the self-serving bias in schizophrenia spectrum disorders compared to non-clinical subjects. Neuroscience and Biobehavioral Reviews, 120(May 2020), 542–549. https://doi.org/10.1016/j.neubiorev.2020.09.025

Murphy, J., Shevlin, M., Adamson, G., Cruddas, S., & Houston, J. (2012). Memories of childhood threat, fear of disclosure and paranoid ideation: A mediation analysis using a nonclinical sample. Journal of Aggression, Maltreatment & Trauma, 21(4), 459–476. https://doi.org/http://dx.doi.org/10.1080/10926771.2012.667521

Murphy, J., Shevlin, M., Adamson, G., Cruddas, S., and Houston, J. (2012). Memories of childhood threat, fear of disclosure and paranoid ideation: A mediation analysis using a nonclinical sample. J. Aggress. Maltreat. Trauma 21, 459–476. doi: http://dx.doi.org/10.1080/10926771.2012.667521.

Murphy, P., Bentall, R. P., Freeman, D., O'Rourke, S., & Hutton, P. (2018). The paranoia as defence model of persecutory delusions: a systematic review and meta-analysis. The Lancet. Psychiatry, 5(11), 913–929. https://doi.org/https://dx.doi.org/10.1016/S2215-0366(18)30339-0

Murphy, R., Goodall, K., & Woodrow, A. (2020). The relationship between attachment insecurity and experiences on the paranoia continuum: A meta-analysis. British Journal of Clinical Psychology, 59(3), 290–318. https://doi.org/10.1111/bjc.12247

Myers, E., Startup, H., & Freeman, D. (2011). Cognitive behavioural treatment of insomnia in individuals with persistent persecutory delusions: A pilot trial. Journal of Behavior Therapy and Experimental Psychiatry, 42(3), 330–336. https://doi.org/10.1016/j.jbtep.2011.02.004

National Institute for Health and Care Excellence. (2014). Psychosis and schizophrenia in adults: Prevention and management. NICE Guidelines [CG178].

Nettle, D., & Haselton, M. G. (2006). The Paranoid Optimist: An Integrative Evolutionary Model of Cognitive Biases. Personality and Social Psychology Review, 10(1), 47–66.

Nugent, W. R., & Thomas, J. W. (1993). Validation of a Clinical Measure of Self-Esteem. Research on Social Work Practice, 3(2), 191–207. https://doi.org/10.1177/104973159300300205

Öhman, A. (2005). The role of the amygdala in human fear: Automatic detection of threat. Psychoneuroendocrinology, 30(10), 953–958. https://doi.org/10.1016/j.psyneuen.2005.03.019

Oosterhof, N. N., and Todorov, A. (2008). The functional basis of face evaluation. Proc. Natl. Acad. Sci. 105, 11087–11092. doi: 10.1073/pnas.0805664105.

Owen, M. J. (2012). Implications of genetic findings for understanding schizophrenia. Schizophrenia Bulletin, 38(5), 904–907. https://doi.org/10.1093/schbul/sbs103

Parsons, S., Simmons, W., Shinhoster, F., and Kilburn, J. (1999). A test of the grapevine: An empirical examination of conspiracy theories among african americans. Sociol. Spectr. 19, 201–222. doi: 10.1080/027321799280235.

Phillips, M. L., Williams, L., Senior, C., Bullmore, E. T., Brammer, M. J., Andrew, C., Williams, S. C. R., & David, A. S. (1999). A differential neural response to threatening and non-

threatening negative facial expressions in paranoid and non-paranoid schizophrenics. Psychiatry Research - Neuroimaging, 92(1), 11–31. https://doi.org/10.1016/S0925-4927(99)00031-1

Pickering, L., Simpson, J., & Bentall, R. P. (2008). Insecure attachment predicts proneness to paranoia but not hallucinations. Personality and Individual Differences. https://doi.org/10.1016/j.paid.2007.11.016

Pierre, J. M. (2020). Mistrust and misinformation: A two-component, socio-epistemic model of belief in conspiracy theories. J. Soc. Polit. Psychol. 8, 617–641. doi: 10.5964/jspp.v8i2.1362.

Pilton, M., Varese, F., Berry, K., & Bucci, S. (2015). The relationship between dissociation and voices: A systematic literature review and meta-analysis. Clinical Psychology Review, 40, 138–155. https://doi.org/10.1016/j.cpr.2015.06.004

Pinkham, A. E., Harvey, P. D., & Penn, D. L. (2016). Paranoid individuals with schizophrenia show greater social cognitive bias and worse social functioning than non-paranoid individuals with schizophrenia. Schizophrenia Research: Cognition, 3, 33–38. https://doi.org/10.1016/j.scog.2015.11.002

Pinkham, A. E., Hopfinger, J. B., Pelphrey, K. A., Piven, J., & Penn, D. L. (2008). Neural bases for impaired social cognition in schizophrenia and autism spectrum disorders. Schizophrenia Research, 99(1–3), 164–175. https://doi.org/10.1016/j.schres.2007.10.024

Pinkham, A. E., Liu, P., Lu, H., Kriegsman, M., Simpson, C., & Tamminga, C. (2015). Amygdala hyperactivity at rest in paranoid individuals with schizophrenia. American Journal of Psychiatry, 172(8), 784–792. https://doi.org/10.1176/appi.ajp.2014.14081000

Pinkham, A., Bass, E., Klein, H., Springfield, C., Kent, J., & Aslan, S. (2022). Inducing Paranoia is Linked to Increased Resting Amygdala Cerebral Blood Flow in Healthy Individuals. Journal of Experimental Psychopathology, 13(3), 1–10. https://doi.org/10.1177/20438087221130693

Pizzagalli, D. A., Greischar, L. L., & Davidson, R. J. (2003). Spatio-temporal dynamics of brain mechanisms in aversive classical conditioning : high-density event-related potential and brain electrical tomography analyses. 41, 184–194.

Putnick, D. L., and Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. Dev. Rev. 41, 71–90. doi: 10.1016/j.dr.2016.06.004.

Qi, R., Palmier-Claus, J., Simpson, J., Varese, F., & Bentall, R. (2020). Sexual minority status and symptoms of psychosis: The role of bullying, discrimination, social support, and drug use - Findings from the Adult Psychiatric Morbidity Survey 2007. Psychology and Psychotherapy, 93(3), 503–519. https://doi.org/https://dx.doi.org/10.1111/papt.12242

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Raihani, N. J., & Bell, V. (2017). Paranoia and the social representation of others: a large-scale game theory approach. Scientific Reports, 7(1), 4544. https://doi.org/https://dx.doi.org/10.1038/s41598-017-04805-3

Rankin, P., Bentall, R., Hill, J., & Kinderman, P. (2005). Perceived relationships with parents and paranoid delusions: Comparisons of currently ill, remitted and normal participants. Psychopathology, 38(1), 16–25. DOI: 10.1159/000083966

Ravary, A., and Baldwin, M. W. (2018). Self-esteem vulnerabilities are associated with cued attentional biases toward rejection. Pers. Individ. Dif. 126, 44–51. doi: 10.1016/j.paid.2018.01.014.

Rawlings, D., & Freeman, J. L. (1996). A questionnaire for the measurement of paranoia/suspiciousness. The British Journal of Clinical Psychology, 35(3), 451–461.

Read, J., Fink, P. J., Rudegeair, T., Felitti, V., & Whitfield, C. L. (2008). Child maltreatment and psychosis: A return to a genuinely integrated bio-psycho-social model. Clinical Schizophrenia and Related Psychoses, 2(3), 235–254. https://doi.org/10.3371/CSRP.2.3.5

Reed, E. J., Uddenberg, S., Suthaharan, P., Mathys, C. D., Taylor, J. R., Groman, S. M., & Corlett, P. R. (2020). Paranoia as a deficit in non-social belief updating. Elife, 9, e56345.

Reeve, S., Emsley, R., Sheaves, B., & Freeman, D. (2018). Disrupting Sleep: The Effects of Sleep Loss on Psychotic Experiences Tested in an Experimental Study With Mediation Analysis. Schizophrenia Bulletin, 44(3), 662–671. https://doi.org/10.1093/schbul/sbx103

Renard, S. B., Huntjens, R. J. C., Lysaker, P. H., Moskowitz, A., Aleman, A., & Pijnenborg, G. H. M. (2017). Unique and overlapping symptoms in schizophrenia spectrum and dissociative disorders in relation to models of psychopathology: A systematic review. Schizophrenia Bulletin, 43(1), 108–121. https://doi.org/10.1093/schbul/sbw063

Rex B. Kline. (2015). Principles and Practice of Structural Equation Modeling. New York, NY: Guilford.

Ringer, J. M., Buchanan, E. E., Olesek, K., & Lysaker, P. H. (2014). Anxious and avoidant attachment styles and indicators of recovery in schizophrenia: Associations with self-esteem and hope. Psychology and Psychotherapy: Theory, Research and Practice, 87(2), 209–221. https://doi.org/10.1111/papt.12012

Roberts, J. E., Gotlib, I. H., and Kassel, J. D. (1996). Adult Attachment Security and Symptoms of Depression: The Mediating Roles of Dysfunctional Attitudes and Low Self-Esteem. J. Pers. Soc. Psychol. 70, 310–320. doi: 10.1016/S0272-6386(97)70025-8.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. J. Stat. Softw. 48. doi: 10.18637/jss.v048.i02.

RStudio Team. (2016). RStudio: Integrated development environment for R. Boston, MA: RStudio, Inc.

Said, C. P., and Todorov, A. (2011). A statistical model of facial attractiveness. Psychol. Sci. 22, 1183–1190. doi: 10.1177/0956797611419169.

Santos, S., Almeida, I., Oliveiros, B., & Castelo-Branco, M. (2016). The role of the amygdala in facial trustworthiness processing: A systematic review and meta-analyses of fMRI studies. PLoS ONE, 11(11), 1–28. https://doi.org/10.1371/journal.pone.0167276

Schaller, M. (2008). Evolutionary bases of first impressions. First Impressions., 15–34. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc6&NEWS=N&AN=2008-09248-001

Schlier, B., Moritz, S., & Lincoln, T. M. (2016). Measuring fluctuations in paranoia: Validity and psychometric properties of brief state versions of the Paranoia Checklist. Psychiatry Research. https://doi.org/10.1016/j.psychres.2016.05.002

Schreber, D. (1903/1955). Memoirs of my nervous illness (I. Macalpine & R. A. Hunter, Trans.). Dawsons.

Schwarz, G. (1978). Estimating the Dimension of a Model Author (s): Gideon Schwarz Reviewed work (s): Source : The Annals of Statistics , Vol . 6 , No . 2 (Mar ., 1978), pp . 461-464

Published by : Institute of Mathematical Statistics Stable URL : http://www.jstor.org/. Ann. Stat., 461–464.

Shaver, P. R., and Mikulincer, M. (2005). Attachment theory and research: Resurrection of the psychodynamic approach to personality. J. Res. Pers. 39, 22–45. doi: 10.1016/j.jrp.2004.09.002.

Shevlin, M., McAnee, G., Bentall, R. P., & Murphy, J. (2015). Specificity of association between adversities and the occurrence and co-occurrence paranoia and hallucinations: Evaluating the stability of childhood risk in an adverse adult environment. Psychosis: Psychological, Social and Integrative Approaches, 7(3), 206–216. https://doi.org/http://dx.doi.org/10.1080/17522439.2014.980308

Sieradzka, D., Power, R. A., Freeman, D., Cardno, A. G., Dudbridge, F., & Ronald, A. (2015). Heritability of Individual Psychotic Experiences Captured by Common Genetic Variants in a Community Sample of Adolescents. Behavior Genetics, 45(5), 493–502. https://doi.org/10.1007/s10519-015-9727-5

Simpson, J. A. (2007). Foundations of Interpersonal Trust. Social Psychology Handbook of Basic Principles. Second. , eds. A. W. Kruglanski and E. T. Higgins The Guilford Press doi: 10.4324/9781315147888.

Sitko, K., Bentall, R. P., Shevlin, M., O'Sullivan, N., & Sellwood, W. (2014). Associations between specific psychotic symptoms and specific childhood adversities are mediated by attachment styles: An analysis of the National Comorbidity Survey. Psychiatry Research, 217(3), 202–209. https://doi.org/10.1016/j.psychres.2014.03.019

Sitko, K., Varese, F., Sellwood, W., Hammond, A., & Bentall, R. (2016). The dynamics of attachment insecurity and paranoid thoughts: An experience sampling study. Psychiatry Research. https://doi.org/10.1016/j.psychres.2016.08.057

Smith, E., Weinberg, A., Moran, T., & Hajcak, G. (2013). Electrocortical responses to NIMSTIM facial expressions of emotion. International Journal of Psychophysiology, 88(1), 17–25. https://doi.org/10.1016/j.ijpsycho.2012.12.004

So, S. H. wai, Siu, N. Y. fan, Wong, H. lam, Chan, W., & Garety, P. A. (2016). 'Jumping to conclusions' data-gathering bias in psychosis and other psychiatric disorders — Two meta-analyses of comparisons between patients and healthy individuals. Clinical Psychology Review, 46, 151–167. https://doi.org/10.1016/j.cpr.2016.05.001

Sood, M., & Newman-Taylor, K. (2020). Cognitive Fusion Mediates the Impact of Attachment Imagery on Paranoia and Anxiety. Cognitive Therapy and Research, 44(6), 1150–1161. https://doi.org/10.1007/s10608-020-10127-y

Sood, M., Carnelley, K. B., & Newman-Taylor, K. (2022). How does insecure attachment lead to paranoia? A systematic critical review of cognitive, affective, and behavioural mechanisms. British Journal of Clinical Psychology, 1–35. https://doi.org/10.1111/bjc.12361

Sood, M., Carnelley, K. B., & Newman-Taylor, K. (2022a). Do emotion regulation strategies mediate the attachment–paranoia association? An experimental study of repeated attachment imagery priming and stress buffering. Psychology and Psychotherapy: Theory, Research and Practice, 95(3), 781–806. https://doi.org/10.1111/papt.12398

Sood, M., Carnelley, K. B., & Newman-Taylor, K. (2022b). How does insecure attachment lead to paranoia? A systematic critical review of cognitive, affective, and behavioural mechanisms. British Journal of Clinical Psychology, 61(3), 781–815. https://doi.org/10.1111/bjc.12361

Sood, M., Carnelley, K. B., and Newman-Taylor, K. (2022). How does insecure attachment lead to paranoia? A systematic critical review of cognitive, affective, and behavioural mechanisms. Br. J. Clin. Psychol. 61, 781–815. doi: 10.1111/bjc.12361.

Spikol, E., Mcbride, O., Butter, S., and Hyland, P. (2021). Tracking the Irish adult population during the first year of the COVID-19 pandemic : A methodological report of the COVID-19 psychological research consortium (C19PRC) study in Ireland. Acta Psychol.220, 103416

Spitzer, R. L., & Fleiss, J. L. (1974). A reanalysis of the reliability of psychiatric diagnosis. British Journal of Psychiatry, 125(10), 341–347. https://doi.org/10.1192/bjp.125.4.341

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. Archives of Internal Medicine, 166(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Sroufe, L. A. (2005). Attachment and development: A prospective, longitudinal study from birth to adulthood. Attach. Hum. Dev. 7, 349–367. doi: 10.1080/14616730500365928.

Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. Behav. Res. Methods, Instruments, Comput. 3, 37–149. doi: 10.3758/BF03207704.

Stapinski, L. A., Bowes, L., Wolke, D., Pearson, R. M., Mahedy, L., Button, K. S., et al. (2014). Peer victimization during adolescence and risk for anxiety disorders in adulthood: A prospective cohort study. Depress. Anxiety 31, 574–582. doi: 10.1002/da.22270.

Statham, V., Emerson, L.-M., & Rowse, G. (2019). A systematic review of self-report measures of paranoia. Psychological Assessment, 31(2), 139–158. https://doi.org/10.1037/pas0000645

Stefanis, N., Thewissen, V., Bakoula, C., van Os, J., & Myin-Germeys, I. (2006). Hearing impairment and psychosis: A replication in a cohort of young adults. Schizophrenia Research, 85(1–3), 266–272. https://doi.org/10.1016/j.schres.2006.03.036

Stempel, B. C., Hargrove, T., and Stempel, G. H. (2007). Media use, social structure, and belief in 9/11conspiracy theories. J&MC Quaterly 84, 353–372.

Strand, M. (2011). Towards an idea-centered, principle-base design to as creation approach support learning knowledge. Theory and Society, 40(3), 273–313. https://doi.org/10.1007/sl

Strauss, G. P., Lee, B. G., Waltz, J. A., Robinson, B. M., Brown, J. K., & Gold, J. M. (2012). Cognition-emotion interactions are modulated by working memory capacity in individuals with schizophrenia. Schizophrenia Research, 141(2–3), 257–261. https://doi.org/10.1016/j.schres.2012.08.010

Sullivan, P. F., Kendler, K. S., & Neale, M. C. (2003). Schizophrenia as a Complex Trait. Archives of General Psychiatry, 60(12), 1187. https://doi.org/10.1001/archpsyc.60.12.1187

Sutcliffe, A., Dunbar, R., Binder, J., & Arrow, H. (2012). Relationships and the social brain: Integrating psychological and evolutionary perspectives. British Journal of Psychology, 103(2), 149–168. https://doi.org/10.1111/j.2044-8295.2011.02074.x

Swami, V., Coles, R., Stieger, S., Pietschnig, J., Furnham, A., Rehim, S., et al. (2011). Conspiracist ideation in Britain and Austria: Evidence of a monological belief system and associations between individual psychological differences and real-world and fictitious conspiracy theories. Br. J. Psychol. 102, 443–463. doi: 10.1111/j.2044-8295.2010.02004.x.

Swami, V., Pietschnig, J., Tran, U. S., Nader, I. W., Stieger, S., and Voracek, M. (2013). Lunar Lies: The Impact of Informational Framing and Individual Differences in Shaping Conspiracist Beliefs About the Moon Landings. Appl. Cogn. Psychol. 27, 71–80. doi: 10.1002/acp.2873.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic criteria. Psychological Science in the Public Interest, 1(1), 1–26.

Taylor, C. D. J., Bee, P. E., Kelly, J., & Haddock, G. (2019). iMagery-focused psychological therapy for persecutory delusions in psychosis (iMAPS): A novel treatment approach. Cognitive and

Behavioral Practice, 26(3), 575–588. https://doi.org/http://dx.doi.org/10.1016/j.cbpra.2018.10.002

Taylor, C. D. J., Bee, P. E., Kelly, J., Emsley, R., & Haddock, G. (2020). iMAgery focused psychological therapy for persecutory delusions in PSychosis (iMAPS): a multiple baseline experimental case series. Behavioural and Cognitive Psychotherapy, 48(5), 530–545. https://doi.org/https://dx.doi.org/10.1017/S1352465820000168

Thewissen, V., Bentall, R. P., Lecomte, T., van Os, J., & Myin-Germeys, I. (2008). Fluctuations in Self-Esteem and Paranoia in the Context of Daily Life. Journal of Abnormal Psychology, 117(1), 143–153. https://doi.org/10.1037/0021-843X.117.1.143

Thewissen, V., Myin-Germeys, I., Bentall, R., De Graaf, R., Vollebergh, W., & Van Os, J. (2005). Hearing impairment and psychosis revisited. Schizophrenia Research, 76(1), 99–103. https://doi.org/10.1016/j.schres.2004.10.013

Thomas, J., Bentall, R. P., Hadden, L., & O'Hara L. (2017). Ethnic identity and paranoid thinking: Implicit out-group preference and language dominance predict paranoia in Emirati women. Journal of Behavior Therapy and Experimental Psychiatry, 56, 122–128. https://doi.org/http://dx.doi.org/10.1016/j.jbtep.2016.10.004

Tiernan, B., Tracey, R., & Shannon, C. (2014). Paranoia and self-concepts in psychosis: A systematic review of the literature. Psychiatry Research. https://doi.org/10.1016/j.psychres.2014.02.003

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

Todorov, A., & Oh, D. W. (2021). The structure and perceptual basis of social judgments from faces. In Advances in Experimental Social Psychology (1st ed., Vol. 63). Elsevier Inc. https://doi.org/10.1016/bs.aesp.2020.11.004

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: A model based approach. Social Cognitive and Affective Neuroscience, 3(2), 119–127. https://doi.org/10.1093/scan/nsn009

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. Emotion, 13(4), 724–738. https://doi.org/10.1037/a0032335

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating Faces on Trustworthiness After Minimal Time Exposure. Social Cognition, 27(6), 813–833. https://doi.org/10.1521/soco.2009.27.6.813

Trémeau, F., Antonius, D., Todorov, A., Rebani, Y., Ferrari, K., Lee, S. H., … Javitt, D. C. (2016). What can the study of first impressions tell us about attitudinal ambivalence and paranoia in schizophrenia? Psychiatry Research, 238, 86–92. https://doi.org/10.1016/j.psychres.2016.02.014

Trotta, A., Kang, J., Stahl, D., & Yiend, J. (2021). Interpretation Bias in Paranoia: A Systematic Review and Meta-Analysis. Clinical Psychological Science, 9(1), 3–23. https://doi.org/10.1177/2167702620951552

Trower, P., & Chadwick, P. (1995). Pathways to Defense of the Self: A Theory of Two Types of Paranoia. Clinical Psychology: Science and Practice, 2(3), 263–278. https://doi.org/10.1111/j.1468-2850.1995.tb00044.x

Trucharte, A. (2022). Social inference processes and attachment in the paranoia spectrum. Complutense University of Madrid.

Tso, I. F., Calwas, A. M., Chun, J., Mueller, S. A., Taylor, S. F., & Deldin, P. J. (2015). Altered attentional and perceptual processes as indexed by N170 during gaze perception in schizophrenia: Relationship with perceived threat and paranoid delusions. Journal of Abnormal Psychology, 124(3), 519–531. https://doi.org/10.1037/abn0000056

Tsoi, D. T., Lee, K. H., Khokhar, W. A., Mir, N. U., Swalli, J. S., Gee, K. A., … Woodruff, P. W. R. (2008). Is facial emotion recognition impairment in schizophrenia identical for different emotions? A signal detection analysis. Schizophrenia Research, 99(1–3), 263–269. https://doi.org/10.1016/j.schres.2007.11.006

Tucker, L. R., and Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. Psychometrika 38, 1–10.

Turner, D. T., Burger, S., Smit, F., Valmaggia, L. R., & van der Gaag, M. (2020). What constitutes sufficient evidence for case formulation-driven CBT for psychosis? Cumulative meta-analysis of the effect on hallucinations and delusions. Schizophrenia Bulletin, 46(5), 1072–1085. https://doi.org/10.1093/schbul/sbaa045

Udachina, A., Varese, F., Oorschot, M., Myin-Germeys, I., & Bentall, R. P. (2012). Dynamics of self-esteem in "poor-me “bad-me” paranoia. Journal of Nervous and Mental Disease, 200(9), 777–783. DOI: 10.1097/NMD.0b013e318266ba57

van der Linden, S., Panagopoulos, C., Azevedo, F., and Jost, J. T. (2020). The Paranoid Style in American Politics Revisited: An Ideological Asymmetry in Conspiratorial Thinking. Polit. Psychol. doi: 10.1111/pops.12681.

Van Os, J., & McGuffin, P. (2003). Can the social environment cause schizophrenia? British Journal of Psychiatry, 182(APR.), 291–292. https://doi.org/10.1192/bjp.182.4.291

van Prooijen, J. W., and Jostmann, N. B. (2013). Belief in conspiracy theories: The influence of uncertainty and perceived morality. Eur. J. Soc. Psychol. 43, 109–115. doi: 10.1002/ejsp.1922.

Van Prooijen, J. W., Spadaro, G., and Wang, H. (2021). Suspicion of institutions: How distrust and conspiracy theories deteriorate social relationships. Curr. Opin. Psychol. 43, 65–69. doi: 10.1016/j.copsyc.2021.06.013.

Varese, F., Smeets, F., Drukker, M., Lieverse, R., Lataster, T., Viechtbauer, W., … Bentall, R. P. (2012). Childhood adversities increase the risk of psychosis: A meta-analysis of patient-control, prospective-and cross-sectional cohort studies. Schizophrenia Bulletin, 38(4), 661–671. https://doi.org/10.1093/schbul/sbs050

Vassos, E., Pedersen, C. B., Murray, R. M., Collier, D. A., & Lewis, C. M. (2012). Meta-analysis of the association of urbanicity with schizophrenia. Schizophrenia Bulletin, 38(6), 1118–1123. https://doi.org/10.1093/schbul/sbs096

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of statistical software, 36(3), 1-48.

Walker, E. F., Diforio, D., & Walker, F. (1997). Schizophrenia: A Neural Diathesis-Stress Model there have been no integrative analy. Psychological Review, 104(4), 667–685.

Walker, M., Schönborn, S., Greifeneder, R., & Vetter, T. (2018). The basel face database: A validated set of photographs reflecting systematic differences in big two and big five personality dimensions. PLoS ONE, 13(3), 1–20. https://doi.org/10.1371/journal.pone.0193190

Wentura, D., and Degner, J. (2010). A practical guide to sequential priming and related tasks. Handb. implicit Soc. Cogn. Meas. theory, Appl., 95–116. Available at: http://www.google.cz/books?hl=en&lr=&id=8UiqBlGT7z4C&pgis=1.

Westermann, S., & Lincoln, T. M. (2010). Using signal detection theory to test the impact of negative emotion on sub-clinical paranoia. Journal of Behavior Therapy and Experimental Psychiatry, 41(2), 96–101. https://doi.org/10.1016/j.jbtep.2009.10.007

Whalen, P. J. (2007). The uncertainty of it all. Trends in Cognitive Sciences, 11(12), 499–500. https://doi.org/10.1016/j.tics.2007.08.016

Wickham, S., Shryane, N., Lyons, M., Dickins, T., and Bentall, R. (2014). Why does relative deprivation affect mental health? The role of justice, trust and social rank in psychological wellbeing and paranoid ideation. J. Public Ment. Health 13, 114–126. doi: 10.1108/JPMH-06-2013-0049.

Wickham, S., Sitko, K., & Bentall, R. P. (2015). Insecure attachment is associated with paranoia but not hallucinations in psychotic patients: The mediating role of negative self-esteem. Psychological Medicine, 45(7), 1495–1507. https://doi.org/10.1017/S0033291714002633

Wickham, S., Taylor, P., Shevlin, M., & Bentall, R. P. (2014). The impact of social deprivation on paranoia, hallucinations, mania and depression: The role of discrimination social support, stress and trust. PLoS ONE, 9(8), 1–9. https://doi.org/10.1371/journal.pone.0105140

Wickham, Sophie, & Bentall, R. (2016). Are specific early-life adversities associated with specific symptoms of psychosis?: A patient study considering just world beliefs as a mediator. Journal of Nervous and Mental Disease, 204(8), 606–613. https://doi.org/http://dx.doi.org/10.1097/NMD.0000000000000511

Wicks, S., Hjern, A., Gunnell, D., Lewis, G., & Dalman, C. (2005). Social adversity in childhood and the risk of developing psychosis: A national cohort study. American Journal of Psychiatry, 162(9), 1652–1657. https://doi.org/10.1176/appi.ajp.162.9.1652

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. Psychological Science, 17(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Winokur, G. (1977). Delusional Disorder (Paranoia). Comprehensive Psychiatry, 18(6), 511–521. https://doi.org/10.1176/ajp.121.10.1027-a

Wongpakaran, N., Demaranville, J., and Wongpakaran, T. (2021). Validation of the relationships questionnaire (Rq) against the experience of close relationship-revised questionnaire in a clinical psychiatric sample. Healthc. 9, 1–13. doi: 10.3390/healthcare9091174.

Woodward, T. S., Buchy, L., Moritz, S., & Liotti, M. (2007). A bias against disconfirmatory evidence is associated with delusion proneness in a nonclinical sample. Schizophrenia Bulletin, 33(4), 1023–1028. https://doi.org/10.1093/schbul/sbm013

Woodward, T. S., Moritz, S., Menon, M., & Klinge, R. (2008). Belief inflexibility in schizophrenia. Cognitive Neuropsychiatry, 13(3), 267–277. https://doi.org/10.1080/13546800802099033

World Health Organisation (1993). The ICD-10 Classification of Mental and Behavioural Disorders.

Yamagishi, T., and Yamagishi, M. (1994). Trust and Commitment in the United States and Japan. Motiv. Emot. 18, 130–166. doi: http:// dx.doi.org/10.1007/BF02249397.

Yang, D., Qi, S., Ding, C., & Song, Y. (2011). An ERP study on the time course of facial trustworthiness appraisal. Neuroscience Letters, 496(3), 147–151. https://doi.org/10.1016/j.neulet.2011.03.066

Young, H. F., & Bentall, R. P. (1997). Probabilistic reasoning in deluded, depressed and normal subjects: Effects of task difficulty and meaningful versus non-meaningful material. Psychological Medicine, 27(2), 455–465. https://doi.org/10.1017/S0033291796004540

Zavos, H. M. S., Freeman, D., Haworth, C. M. A., McGuire, P., Plomin, R., Cardno, A. G., & Ronald, A. (2014). Consistent etiology of severe, frequent psychotic experiences and milder, less frequent manifestations: A twin study of specific psychotic experiences in adolescence. JAMA Psychiatry, 71(9), 1049–1057. https://doi.org/10.1001/jamapsychiatry.2014.994