



Provided by the author(s) and University of Galway in accordance with publisher policies. Please cite the published version when available.

Title	Examining paranoia in the general population from a Contextual Behavioural Science (CBS) perspective: New insights from Relational Frame Theory (RFT) and the Implicit Relational Assessment Procedure (IRAP)
Author(s)	Stewart, Corinna
Publication Date	2018-01-26
Item record	<a href="http://hdl.handle.net/10379/7109">http://hdl.handle.net/10379/7109</a>

Downloaded 2024-04-25T10:16:43Z

Some rights reserved. For more information, please see the item record link above.



National University of Ireland, Galway



Examining paranoia in the general population from a Contextual Behavioural Science (CBS) perspective: New insights from Relational Frame Theory (RFT) and the Implicit Relational Assessment Procedure (IRAP)

Thesis submitted to the School of Psychology, College of Arts, Social Sciences & Celtic Studies, in fulfilment of the requirements for the degree of Doctor of Philosophy, National University of Ireland, Galway.

Corinna Stewart B.A. (Hons.)  
January 2018

Head of School: Dr. John Bogue

Thesis Supervisors: Dr. Ian Stewart & Dr. Yvonne Barnes-Holmes

## **Abstract**

The current thesis presents a first step towards examining paranoia from a functional-analytic perspective. The focus of the research was ‘self-beliefs’ as a wealth of evidence in the existing cognitive-clinical literature suggests that this is a key process in the development and persistence of paranoia. From the functional-analytic perspective, paranoia and related self-concepts are defined as patterns of behavior in context. The goal of this research involved elucidating these behaviors, determining how they influence each other, and identifying which contextual variables affect them. The work drew upon Relational Frame Theory (RFT), a functional-analytic theory of human language and cognition, and an RFT-based measure, the Implicit Relational Assessment Procedure (IRAP), to do so. Across five studies involving participants from the general (non-clinical) population, specific patterns of relational responding to the self that may be pertinent to paranoia were identified. Using an experimental approach involving threat-induction tasks, it was also demonstrated that paranoia and related responding to the self (e.g., as negative, vulnerable) and others (e.g., as trustworthy, devious) can be influenced by environmental factors. The IRAP was shown to be a useful measure in this regard, demonstrating predictive utility (Study 1), an ability to parse out patterns of responding relevant to high non-clinical paranoia (Study 4), and sensitivity to experimental manipulations (Studies 2, 3, and 5). Taken together, the findings from this research suggest that the functional-analytic perspective may compliment the cognitive-clinical approach to the study of paranoia and might also offer new and exciting avenues for research (e.g., novel procedures) in this domain.

## **Acknowledgements**

### **Ian Stewart**

Thank you for the creative freedom, unwavering support and invaluable guidance you have given me. I could not have asked for a better or more supportive mentor.

### **Yvonne Barnes-Holmes**

Thank you for your insight, guidance and advice over the years.

### **Sean Hughes**

I am so grateful to you for your endless encouragement, support and friendship.

### **NUI Galway Psychology department & postgrads**

Thank you for your support, warmth, and fun. Thanks to Declan Coogan, Olive O’Grady, and my GRC for their help over the years. I wish my wonderful fellow postgrads (past and present) all the best – thank you for your friendship, kindness, and banter. Special thank you to my G042 pals, Páraic Ó Súilleabháin, Amanda Sesker, Milou Fredrix and Sophi Arndt –

G042 4 LiF3!

### **commit and act**

Thank you, my international family, for encouraging, challenging, and always being there for me.

### **Matt**

Thank you for your love and support. You truly are a dreamboat!

### **Family & friends**

Thanks for everything! For your love and kindness, and for the help, laughs, patience and encouragement throughout this journey. I will treasure many fond memories of my time in Galway and 45, with Lisa, Teresa, Grainne, Michelle, and Sue.

## Table of Contents

<b>Abstract</b> .....	i
<b>Acknowledgements</b> .....	ii
<b>Table of Contents</b> .....	iii
<b>Publications</b> .....	v
<b>List of Figures</b> .....	vi
<b>List of Tables</b> .....	ix
<b>Table of Abbreviations</b> .....	xi
<b>List of Appendices</b> .....	xii
<b>Chapter 1: General Introduction</b> .....	1
1.1. Definition of paranoia .....	1
1.1.1. Prevalence, measurement and nature of paranoia within the general population .....	2
1.2. Conceptual, theoretical and methodological considerations .....	5
1.3. Contextual Behavioral Science .....	6
1.4. Relational Frame Theory .....	9
1.4.1. Properties of Arbitrarily Applicable Relational Responding (AARR) .....	11
1.4.2. The importance of AARR .....	14
1.4.3. AARR, psychological suffering and paranoia .....	15
1.5. Measuring AARR: The Implicit Relational Assessment Procedure (IRAP) .....	18
1.5.1. The utility of the IRAP in experimental-clinical research .....	21
1.6. Cognitive approaches to paranoia .....	23
1.6.1. Threat anticipation model of paranoia .....	24
1.6.2. Defensive model of paranoia .....	25
1.6.3. Self-esteem, self- and others-evaluations, and paranoia .....	26
1.6.4. Considerations regarding self-concepts in the context of paranoia .....	27
1.7. The current thesis .....	30
1.7.1. Thesis summary .....	32
1.7.2. Ethical considerations .....	33
<b>Chapter 2: The relationship between paranoia, self-reported self-esteem, and relational responding to the self on the IRAP</b> .....	
2.1. Introduction .....	35
2.2. Method .....	38

2.3. Results .....	42
2.4. Discussion .....	47

**Chapter 3: The effects of self-focused attention and task failure on state paranoia, self-reported self-esteem, and relational responding to the self on the IRAP**

3.1. Introduction .....	51
3.2. Method .....	54
3.3. Results .....	58
3.4. Discussion .....	65

**Chapter 4: The effects of social exclusion on state paranoia, self-reported self-esteem, and relational responding to the self on the IRAP**

4.1. Introduction .....	69
4.2. Method .....	72
4.3. Results .....	76
4.4. Discussion .....	83

**Chapter 5: Relational responding to the self (as ‘vulnerable’ versus ‘safe’) and others (as ‘devious’ versus ‘trustworthy’) on the IRAP in individuals with high non-clinical paranoia and controls**

5.1. Introduction .....	88
5.2. Method .....	91
5.3. Results .....	96
5.4. Discussion .....	109

**Chapter 6: The effect of social exclusion on state paranoia, self-reported safety and vulnerability, and relational responding to the self as ‘safe’ and ‘vulnerable’ on the IRAP**

6.1. Introduction .....	114
6.2. Method .....	117
6.3. Results .....	120
6.4. Discussion .....	135

**Chapter 7: General Discussion .....** 140

**References .....** 167

**Appendices .....** 191

## **Publications**

Sections of this thesis are currently published, in press, and under submission. The chapter number, chapter name, and the corresponding citations are listed below.

### **Chapter 1: General Introduction**

Stewart, C., Stewart, I., & Hughes, S. (2016). A functional account of (persecutory) delusions. *Journal of Contextual Behavioral Science*, 5, 235-246. DOI: 10.1016/j.jcbs.2016.09.002

### **Chapter 3: The effects of self-focused attention and task failure on state paranoia, self-reported self-esteem, and relational responding to the self on the IRAP**

Stewart, C., Rogers, F., Stewart, I., & Barnes-Holmes, Y. (*under review*). The effects of self-focused attention and task failure on state paranoia and explicit and implicit self-esteem. *The Psychological Record*.

### **Chapter 4: The effects of social exclusion on state paranoia, self-reported self-esteem, and relational responding to the self on the IRAP**

Stewart, C., Rogers, F., Pilch, M., Stewart, I., Barnes-Holmes, Y., & Westermann, S. (2017). Examining the impact of social exclusion on state paranoia and explicit and implicit self-esteem in a non-clinical sample. *Journal of Behavior Therapy & Experimental Psychiatry*, 57, 62-69. DOI: 10.1016/j.jbtep.2017.04.001

## List of Figures

- Figure 1.1. The Paranoia Hierarchy. Reprinted from "Psychological investigation of the structure of paranoia in a non-clinical population" by D. Freeman et al. (2005), *British Journal of Psychiatry*, 18, p. 433. Copyright 2005 by Royal College of Psychiatrists.
- Figure 1.2. Summary of the formation of a persecutory delusion according to the threat anticipation model. Reprinted from "A cognitive model of persecutory delusions" by D. Freeman et al. (2002), *British Journal of Clinical Psychology*, 41, 334. Copyright 2002 by the British Psychological Society.
- Figure 2.1. Examples of the four trial-types in the IRAP. Correct answers for responding on consistent and inconsistent blocks of trials are indicated by the arrows (which were not presented during the IRAP procedure).
- Figure 2.2. D-IRAP scores for each trial-type for low, mid-range, and high paranoia groups, with standard error bars (95% confidence intervals). Note, D-IRAP scores above zero indicate faster responding on consistent (i.e., faster 'Me-Positive' agreement and 'Me-Negative' disagreement) relative to inconsistent blocks of trials, whereas scores below zero indicate the opposite pattern of responding (i.e., faster responding on inconsistent trials). Asterisks indicate trial-types significantly different from zero.
- Figure 3.1. Mean D-IRAP scores for each IRAP trial-type by experimental condition, with standard error bars (95% confidence intervals): LSA + No Feedback; LSA + Failure feedback; HSA + No feedback; HSA + Failure feedback. Note, D-IRAP scores above zero indicate faster responding on consistent (i.e., faster 'Me-Positive' agreement and 'Me-Negative' disagreement) relative to inconsistent blocks of trials, whereas scores below zero indicate the opposite pattern of responding (i.e., faster responding on inconsistent trials). Asterisks denote trial-types significantly different from zero.
- Figure 4.1. Changes in mean RSES scores from baseline to post-Cyberball for Inclusion and Exclusion groups. Standard error bars represent 95% confidence intervals.
- Figure 4.2. D-IRAP scores for IRAP trial-types at baseline and post-Cyberball for the Inclusion and Exclusion groups, with standard error bars (95% confidence intervals). Note, D-IRAP scores above zero indicate faster agreement for 'Me-Positive' and 'Others-Positive' (relative to disagreement) and faster disagreement for 'Me-Negative' and 'Others-Negative' (relative to agreement), whereas scores below zero indicate the opposite pattern of responding (i.e., faster 'Me-Positive' and 'Others-Positive' disagreement and 'Me-Negative' and 'Others-Negative' agreement). Asterisks represent trial-types significantly different from zero.

Figure 5.1. D-IRAP scores for ‘Self’ IRAP trial-types for controls and participants with high non-clinical paranoia, with standard error bars (95% confidence intervals). Note, D-IRAP scores above zero indicate faster responding on consistent (i.e., faster ‘Me-Safe’ agreement and ‘Me-Vulnerable’ disagreement) relative to inconsistent blocks of trials, whereas scores below zero indicate the reverse pattern (i.e., faster responding on inconsistent trials). Asterisks indicate trial-types significantly different from zero.

Figure 5.2. D-IRAP scores for ‘Others’ IRAP trial-types for controls and participants with high non-clinical paranoia, with standard error bars (95% confidence intervals). Note, D-IRAP scores above zero indicate faster responding on consistent (i.e., faster ‘Others-Trustworthy’ agreement and ‘Others-Devious’ disagreement) relative to inconsistent blocks of trials, whereas scores below zero indicate the reverse pattern (i.e., faster responding on inconsistent trials). Asterisks indicate trial-types significantly different from zero.

Figure 6.1. Changes in mean self-reported ‘Me-Vulnerable’ (left-hand side) and ‘Me-Safe’ (right-hand side) ratings from baseline to post-Cyberball for Inclusion and Exclusion groups. Please note differences in scales due to differences in mean rating scores; that is, while the same response format was used for ‘Me-Vulnerable’ and ‘Me-Safe’ ratings, higher scores on ‘Me-Vulnerable’ indicate greater ‘vulnerability’ whereas higher scores on ‘Me-Safe’ indicate greater ‘safety’.

Figure 6.2. Changes in mean Negative-Self (left-hand side) and Positive-Self (right-hand side) scores on the BCSS from baseline to post-Cyberball for Inclusion and Exclusion groups. Please note differences in scales due to differences in mean rating scores; that is, while the same response format was used for Negative-Self and Positive-Self scales, higher scores on Negative-Self indicate greater ‘negativity’ towards the self whereas higher scores on Positive-Self indicate greater ‘positivity’ towards the self.

Figure 6.3. Changes in mean Negative-Others (left-hand side) and Positive-Others (right-hand side) scores on the BCSS from baseline to post-Cyberball for Inclusion and Exclusion groups. Please note differences in scales due to differences in mean rating scores; that is, while the same response format was used for Negative-Others and Positive-Others scales, higher scores on Negative-Others indicate greater ‘negativity’ towards others whereas higher scores on Positive-Others indicate greater ‘positivity’ towards others.

Figure 6.4. D-IRAP scores for IRAP trial-types at baseline (left-hand side) and post-Cyberball (right-hand side) for the Inclusion and Exclusion groups, with standard error bars (95% confidence intervals). Note, D-IRAP scores above zero indicate faster responding on consistent (i.e., faster ‘Me-Safe’ agreement and ‘Me-Vulnerable’ disagreement) relative to inconsistent blocks of trials, whereas scores below zero indicate the reverse pattern (i.e., faster responding on inconsistent trials). Asterisks represent trial-types significantly different from zero.

## List of Tables

Table 2.1. IRAP stimuli for Study 1.

Table 2.2. Ranges, Means, and SDs, for Paranoia Checklist total and subscales, RSES, and D-IRAP scores for each IRAP trial-type ( $N = 70$ ).

Table 2.3. Correlations between RSES, IRAP trial-types, and Paranoia Checklist total and subscale scores ( $N = 70$ ).

Table 2.4. Results from multiple regression analyses using age, gender, RSES scores and IRAP trial-types to predict total score on the Paranoia Checklist.

Table 3.1. Means and SDs for Paranoia Checklist total and subscales, state paranoia across the four time points, and RSES scores across the four threat-induction task conditions.

Table 3.2. Correlations between Paranoia Checklist total and subscales, state paranoia at Time 4, RSES, and IRAP trial-type scores ( $N = 66$ ).

Table 4.1. IRAP stimuli for Study 3.

Table 4.2. Correlations between measures of trait and state paranoia, RSES scores, and IRAP trial-types at baseline ( $N = 85$ ).

Table 4.3. Correlations between Paranoia Checklist total and subscale scores, state paranoia at post-Cyberball, and change scores (from baseline to post-Cyberball) for RSES and IRAP trial-types ( $N = 85$ ).

Table 5.1. Stimuli for ‘Self’ IRAP for Study 4 (and for Study 5).

Table 5.2. Stimuli for ‘Others’ IRAP for Study 4.

Table 5.3. Paranoia Checklist items endorsed by participants in the control ( $n = 33$ ) and high non-clinical paranoia ( $n = 30$ ) groups. Endorsement was defined as a Frequency rating (i.e., “How often have you had the thought?”) of  $\geq$  weekly (i.e., a score of  $\geq 3$ ). Percentages represent proportion of each group (control and high non-clinical paranoia) that rated each item with a Frequency score of  $\geq$  weekly.

Table 5.4. Means and SDs for ‘Self’ and ‘Others’ IRAP trial-types and self-report measures for participants high in non-clinical paranoia compared with controls ( $N = 63$ ).

Table 5.5. Correlations between Paranoia Checklist total and subscale scores, ‘Self’ and ‘Others’ IRAP trial-types, self-reported ‘Me-Safe’ and ‘Me-Vulnerable’ ratings, BCSS, RSES, and DASS scores ( $N = 63$ ). Note, 20 = Depression, 21 = Anxiety, and 22 = Stress scores (DASS).

Table 6.1. Correlations between measures of trait and state paranoia and IRAP trial-types, self-reported ‘Me-Safe’ and ‘Me-Vulnerable’ ratings, BCSS, and RSES scores at baseline ( $N = 81$ ).

Table 6.2. Correlations between trait paranoia, state paranoia change scores (from baseline to post-Cyberball) and change scores for Me-Safe, Me-Vulnerable, BCSS, RSES and IRAP trial-types ( $N = 81$ ).

## Table of Abbreviations

<b>Abbreviation</b>	<b>Term</b>
AARR	Arbitrarily applicable relational responding
BATs	Behavioral approach tasks
BCSS	Brief Core Schemas Scales
BIRR	Brief and immediate relational responding/responses
CBS	Contextual Behavioral Science
DASS	Depression, Anxiety and Stress Scales
DSM	Diagnostic and Statistics Manual of Mental Disorders
EERR	Extended and elaborate relational responding/responses
GNAT	Go/No-go Association Task
HSA	High self-awareness
IAT	Implicit Association Test
IRAP	Implicit Relational Assessment Procedure
LSA	Low self-awareness
MDML	Multidimensional Multilevel model
OC	Obsessive compulsive
OCD	Obsessive Compulsive Disorder
REC	Relational, Elaboration and Coherence model
RFT	Relational Frame Theory
RSES	Rosenberg Self-esteem Scale

## **List of Appendices**

Appendix A: The Paranoia Checklist (Freeman et al., 2005)

Appendix B: The state-adapted Paranoia Checklist (Westermann et al., 2012)

Appendix C: Rosenberg Self-esteem Scale (Rosenberg, 1965)

Appendix D: Cyberball Questionnaire (Williams et al., 2000; Williams, 2009)

Appendix E: Brief Core Schemas Scales (Fowler et al., 2006)

Appendix F: Depression, Anxiety and Stress Scale-21 (Lovibond & Lovibond, 1995)

Appendix G: Research Ethics Committee approval letter from NUI Galway with accompanying Participant Information Sheet, Consent Form, and Participant Debriefing Form

Appendix H: Research Ethics Committee approval letter from University College Dublin with accompanying Participant Information Sheet, Consent Form, and Participant Debriefing Form

## Chapter 1: General Introduction

### 1.1. Definition of paranoia

The overall aim of this thesis is to bring a functional-analytic perspective to the study of paranoia<sup>1</sup> within the general population. The definition of what constitutes a paranoid belief, or ‘persecutory delusion’ (the term used for clinical paranoia), has shifted across time and the evolution of this concept has been tightly linked to the syndromal approach adopted in the *Diagnostic and Statistical Manual of Mental Disorders (DSM)*. Within *DSM-V*, the broad category of ‘delusional beliefs’ (as applied to clinical contexts/populations) is defined as “fixed beliefs that are not amenable to change in light of conflicting evidence” (APA, 2013, p. 87). However, recent research indicates that paranoid beliefs constitute a more complex phenomenon. Freeman (2007) outlined a range of factors relevant when determining whether a persecutory belief is present, including the extent to which the belief involves (a) personal reference, (b) is unfounded (although for some it may reflect a kernel of truth that has been exaggerated or distorted), (c) is firmly held and resistant to change (even given contradictory evidence), (d) is preoccupying and possibly distressing, and (e) impairs functioning and/or leads to other negative life-outcomes (e.g., distress, aggression).

Freeman (2016) proposed that “the key psychological inroad into understanding persecutory delusions is to recognize them as threat beliefs” (p. 687). Freeman and Garety (2000) suggested that paranoid beliefs have two core defining properties: the individual believes that (1) harm is occurring or will occur, and (2) a perceived persecutor is deliberately trying to cause that harm. Evidence shows that paranoia is not exclusive to clinical populations. Rather there seems to be an exponential distribution of paranoid thoughts within the general population (Freeman et al., 2005; Bebbington et al., 2013). In other words, paranoid thoughts and ideas occur along a continuum, with normal experiences at one end

---

<sup>1</sup> The term ‘paranoia’ is adopted throughout the thesis as this research focuses on this phenomenon within the general (non-clinical) population. ‘Persecutory delusions’ typically refer to clinical levels of paranoia exhibited in individuals with a formal psychosis-related diagnosis (e.g., schizophrenia, schizoaffective disorder).

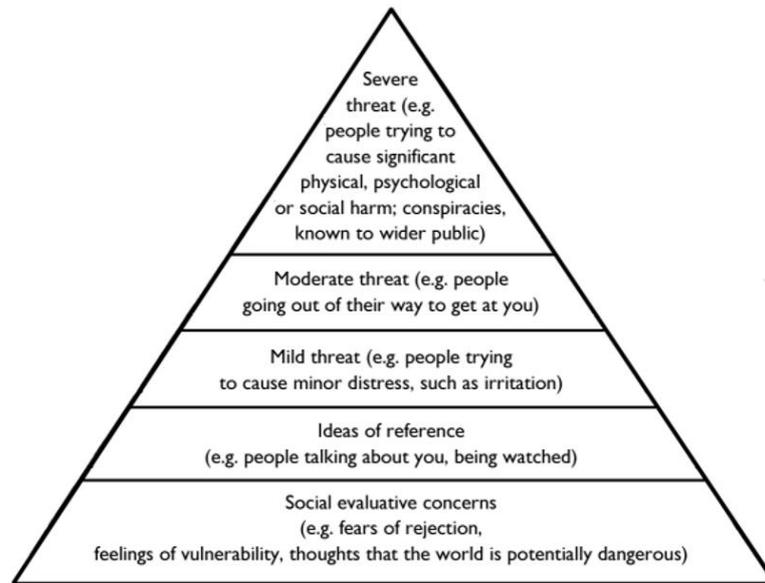
and persecutory delusions at the other, such that “many people have a few paranoid thoughts, and a few have many” (Freeman & Garety, 2014, p. 1179). Examining non-clinical paranoia within the general population may inform our understanding of clinical persecutory delusions and help to normalize these experiences. The current thesis explores paranoia within the general population.

### **1.1.1. Prevalence, measurement and nature of paranoia within the general population**

Recently researchers have investigated the prevalence and characteristics of delusions generally and paranoia specifically within the general (non-clinical) population (e.g., Bebbington et al., 2013; Freeman et al., 2005; Freeman et al., 2011; Green et al., 2008; Johns et al., 2004; Peters et al., 2004; Peters et al., 1999b). Earlier research of this type has mainly involved brief structured interviews (Psychosis Screening Questionnaire; Bebbington & Nayani, 1995) and/or self-report measures (e.g., the Paranoia Scale; Fenigstein & Venable, 1992), while more recent self-report measures adopt a dimensional approach (e.g., the Paranoia Checklist, Freeman et al., 2013; Peters et al. Delusions Inventory; Peters et al., 1999b; 2004). That is, participants rate items that refer to experiences that are commonplace (e.g., “People say negative things about me behind my back”) or rare (e.g., “There is a possible conspiracy against me”) in terms of the frequency of, or preoccupation with, these experiences, as well as levels of associated conviction and distress.

Based on the results of 15 studies, Freeman (2007) proposed that approximately 10 to 15 per cent of the general population regularly experience paranoid thoughts. Evidence indicates that endorsement of more unusual or rarer items (e.g., thoughts about conspiracy, threats of serious harm) is associated with endorsement of more common items (e.g., thoughts of negative comments being circulated about you; Bebbington et al., 2013; Freeman et al., 2005; Green et al., 2008). This suggests that there is a hierarchy of paranoia within the general population, whereby thoughts of threat appear to build on more familiar ideas (e.g.,

social evaluative concerns). Specifically, Freeman et al. (2005) proposed the Paranoia Hierarchy (depicted in Figure 1.1). At the base of this pyramid lie the most common type of suspicion, referred to as ‘social evaluative concerns’, which relate to themes of social anxiety (e.g., feelings of vulnerability, fears of rejection). These common suspicions inform worries about future threat and the intentions of others, known as ‘ideas of reference’ (e.g., people talking about you). Persecutory thoughts are hypothesized to emerge from these sensitivities and ideas of reference. These persecutory thoughts emerge at the first mild level of threat (e.g. people trying to cause minor distress such as irritation) and may progress to less common experiences that represent moderate threat (e.g. people going out of their way to get at you) and then progress even further to the least common but most severe perception of threat at the top of the pyramid (e.g. people trying to cause significant harm). As noted previously, Freeman et al. (2005) observed that individuals who endorsed the rarer and more severe items at the top of the pyramid also endorse the lower level more commonplace items, and that they do so with greater conviction and more distress relative to those who endorse the more common items only.



**Figure 1.1.** The Paranoia Hierarchy. Reprinted from “Psychological investigation of the structure of paranoia in a non-clinical population” by D. Freeman et al. (2005), *British Journal of Psychiatry*, 18, p. 433. Copyright 2005 by Royal College of Psychiatrists.

Higher levels of paranoia have also been associated with emotional and avoidant coping (considered maladaptive coping strategies; see Roger, Jarvis, & Najarian, 1993), engaging in social comparison, and greater use of submissive behaviors (Freeman et al., 2005). Hence, paranoid thoughts are not isolated events, but are accompanied by interpretations of these events. In other words, the impact of experiencing paranoid thoughts does not depend simply on their presence, but also on their interpretation. Put simply, “it’s not *what* you think, but *how* you think about it” (Peters et al., 2004, p. 1013). Indeed, studies involving participants from both clinical and non-clinical populations have reported overlap in the items endorsed by the two groups (e.g., Green et al., 2008; Peters et al., 1999b; 2004); however, the two were differentiated by their ratings of preoccupation, conviction, and distress. These findings suggest qualitative shifts in experience at the top end of the paranoia hierarchy and therein support the necessity of a multi-dimensional approach to determining whether or not an individual’s experience of paranoia is clinical and disabling.

## 1.2. Conceptual, theoretical and methodological considerations

To date, the study of paranoia has largely been conducted by cognitive-clinical researchers operating from a mechanistic, mentalistic philosophical tradition whereby paranoia and related phenomena (e.g., self-concepts, reasoning biases, etc.) are interpreted as mental mechanisms that influence behavior (e.g., Bentall, Kinderman, & Kaney, 1994; Bentall, Corcoran, Howard, Blackwood, & Kinderman, 2001; Freeman, Garety, Kuipers, Fowler, & Bebbington, 2002; Garety, Kuipers, Fowler, Freeman, & Bebbington, 2001; Kesting & Lincoln, 2013; Salvatore, Lysaker, Popolo, Procacci, Carcione, & Dimaggio, 2012). As such, the focus is almost exclusively on identifying the (mental) mechanisms involved and describing their interaction and influence. This thesis presents a first step towards investigating paranoia from an alternative functional-analytic perspective, and specifically using a contextual behavioral science approach (CBS; see Hayes, Barnes-Holmes, & Wilson, 2012). In order to provide context to the theoretical, conceptual and empirical work contained herein, we<sup>2</sup> begin with an overview of CBS, its philosophical assumptions (i.e. functional contextualism) and its distinction from the mechanistic approach. We then introduce a CBS theory of human language and cognition, known as Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001), and an RFT-based methodology called the Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes et al., 2006). We propose that recent advances in RFT research and the development of measures like the IRAP offer both a theoretical basis and a tool with which to examine psychological suffering generally, and paranoia specifically, from a functional-analytic approach. We then turn our attention to cognitive-clinical models of and research on paranoia and related variables (i.e., self-concepts) and argue that a functional-analytic perspective may

---

<sup>2</sup> For purely stylistic reasons the pronoun “we” is used in place of “I” even though the work presented here is the product of a single doctoral candidate.

offer new and supplementary avenues for research (e.g., novel procedures) and intervention in this domain.

### **1.3. Contextual Behavioral Science**

Contextual behavioral science (CBS; Hayes et al., 2012) is an empirical tradition rooted in functional contextualism, a philosophy of science that adopts a radically functional-analytic approach to truth and meaning (i.e., behavioral pragmatism; see Barnes-Holmes, 2000; Biglan & Hayes, 1997). Functional contextualism differs from mechanism in essential ways and these distinctions have important consequences for the assumptions, goals, and values of researchers, and by implication, the principles, theories, and methodologies that they draw upon (see Hughes, De Houwer, & Barnes-Holmes, 2016; Levin, Twohig, & Smith, 2016; Wilson, Whiteman, & Bordieri, 2013).

Mechanistically-oriented researchers operating at the mental level of analysis consider psychological events as “similar to a machine, composed of discrete parts that interact and are subject to specific operating conditions” (Hughes et al., 2016, p.7). Hence, the primary scientific goal involves identifying the mental mechanism(s) that mediate between input (environment) and output (behavior). The researcher’s role is to develop an account of mental representations and processes that mediate changes in behavior. The scientific value or “truth” of such accounts are based on the *correspondence* between the proposed mental mechanism(s) and ‘reality’. Hence, if this correspondence is high (i.e., achieves greater “truth”), then the theory (i.e., proposed mental mechanism) should yield better prediction of behavioral effects. Studying paranoia from a traditional mechanistic perspective, Freeman (2016) argued that the defining feature of a ‘causal’ factor in the development and/or maintenance of persecutory beliefs is that, when altered, it leads to a change in these beliefs. Freeman articulated the process through which mental mechanisms (e.g., self-concepts) may be deemed ‘causal’. Specifically, a given plausible mental mechanism is believed to influence

paranoia when: (1) it is present in a large proportion of individuals who experience paranoia, (2) it is potentially tractable or subject to control/influence, and (3) its causal role is tested by manipulation (e.g., in experimental or intervention studies). However, such an approach can be problematic as it involves *a priori* assumptions about the relation between mental constructs and behavior (De Houwer, 2011). That is, if the researcher's goal is to predict behavioral effects under certain environmental conditions (as is often the case for those working at the mental level of analysis), then any mental variable can be used to achieve prediction, so long as it reliably precedes that effect. However, if the behavioral effect can be *caused* by other mental constructs (in the absence of the assumed causal mental construct) then the behavioral effect does not allow firm conclusions to be drawn about the assumed mental construct. Furthermore, although cognitive models can be quite accurate predictors of psychological events, the assumption that one psychological event may cause or explain another, with limited reference to the impact of environmental or historical variables, means that researchers are left with little knowledge of the factors that gave rise to either psychological events or how to influence them given that mental constructs cannot be manipulated directly. Furthermore, this focus on identifying causal mental mechanisms or constructs might potentially limit knowledge of other factors that may be highly influential to behavior (e.g., histories of learning).

In contrast, CBS does not concern itself with the nature of 'reality', but with "successful working". From this perspective, scientific validity or "truth" is determined by examining whether an analysis leads to effective action and achieves its purpose or goal. Explicitly, the goal of CBS is prediction-*and-influence* of behavior with a limited number of behavioral principles and analytic concepts characterized by precision (a limited number of analytic concepts apply to a given case), scope (an analytic concept applies to a range of cases) and depth (analytic concepts cohere across various levels of analysis). CBS researchers

make no appeal to mental mechanisms in explicating behavior; rather, they focus on behavior “interacting in and with a context, considered both historically and situationally” (Hayes et al., 2012, p. 3). As noted above, this is primarily for pragmatic (and not ontological) reasons. That is, if the goal is to achieve prediction-and-influence, researchers must focus on elements that can be directly manipulated to influence behavior, making non-environmental or mental variables (e.g., thoughts, emotions) ‘poor choices’ as analytic end points because they cannot be directly manipulated (Levin et al., 2016). That is not to say that these variables are unimportant. Indeed, thoughts, emotions and related phenomena are particularly pertinent to psychological suffering, including paranoia (see Hayes, Strosahl, & Wilson, 1999; 2012). However, such variables are considered behaviors in themselves and may interact with other behaviors (e.g., paranoia), referred to as ‘behavior-behavior relations’. Neither behaviors nor behavior-behavior relations are directly manipulable and so a behavior-behavior relation cannot be considered a complete explanation of behavior in itself. Instead, such relations are themselves based on contextual factors (e.g., contingencies of reinforcement) that can be directly manipulated to achieve influence over behavior, seen within this approach as a fundamental analytic requirement (see Hayes & Brownstein, 1986). Nonetheless, while behavior-behavior relations are not sufficient in themselves to provide a complete explanation, they are part of complex human behavior including paranoia and thus, as suggested, they are still important within the context of a complete functional explanation and thus deserve attention within this context. Within this thesis, we explore such (behavior-behavior) relations and the influence of environmental contingencies on their interaction.

Within CBS, the unit of analysis is the ‘act-in-context’; behavior and its context are not fully separable, and history, circumstances, and consequences are aspects of the act itself in a functional sense. Furthermore, all events are deemed as behavioral functions rather than things that exist independently of behavior (see Barnes-Holmes, 2000). For instance, a

stranger may be defined as an eliciting stimulus for a response (e.g., fear) or may function as a reinforcing stimulus for a behavioral response (e.g., escape, avoidance). Thus, the event (e.g., a stranger) is defined in terms of its function within a given context. These functions are likely established via one's learning history and features of the current context. The history of learning may be relatively proximal (e.g., experiencing an unpleasant interpersonal event with a stranger the previous day) or more distal (e.g., traumatic childhood experiences involving strangers). Aspects of one's history of learning may interact with the current context whereby they influence, moderate, or increase the probability of certain responses within the current context (e.g., looking away following eye-contact with a stranger). Furthermore, the history of interactions that established the functions of strangers for one individual and the associated probability of this stimulus eliciting a response (e.g., escape, aggression) likely differ from the history that established the stimulus functions for another individual, which may help explain individual variability in responding to similar events in the environment. That is, it is important to appreciate how varying distal contexts or learning histories may produce variation in stimulus functions and associated behaviors for different individuals. However, the same behavioral principles (e.g., reinforcement, stimulus generalization) may be used to explicate the development of these varying stimulus functions and associated responses.

#### **1.4. Relational Frame Theory**

CBS is committed to developing basic functional-analytic accounts of complex behaviors and generates functional-analytic theory and models by using sets of interrelated behavioral principles to elucidate the history and situational determinants of a range of behaviors. An example of one such theory within CBS is Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001), which provides a functional-analytic theory of human language and cognition. The test of any behavioral theory is its coherence and utility

in serving as a coordinating account and in opening up new and important areas of research (Hayes, Blackledge, & Barnes-Holmes, 2001). RFT has provided a coherent and useful set of behavioral principles that allow symbolic behaviors to be predicted and influenced (see Dymond & Roche, 2013; Hughes & Barnes-Holmes, 2016a; 2016b; Stewart, 2016). Importantly, language is particularly critical to complex human behavior, including psychological suffering; hence a theory that can elucidate the processes involved in language is crucial for understanding, predicting-and-influencing such behavior.

RFT posits that the core unit underpinning language and cognition is a generalized operant called ‘arbitrarily applicable relational responding’, or ‘AARR’. Many species can respond based on the physical relationship between stimuli (e.g., picking something physically bigger or smaller than something else). However, humans uniquely learn to relate stimuli *regardless* of their physical properties and in ways never experienced or instructed in the past (i.e., to *derive* relations between stimuli). RFT refers to this behavior as AARR because it is based on contextual cues independent of the physical properties of the stimuli being related. AARR is defined in terms of three core properties: (i) *Mutual entailment* refers to the derived bi-directionality of stimulus relating such that if A is related to B in a specific context, then a relation from B to A is also entailed in that context (e.g., if  $A > B$  then  $B < A$ ); (ii) *Combinatorial entailment* refers to the combination of two or more relations. For example, if A is related to B and B is related to C, then a bidirectional relation between A and C is entailed (e.g., if  $A > B > C$ , then  $A > C$  and  $C < A$ ); (iii) *Transformation of stimulus functions* refers to the finding that when stimuli are related - and the stimulus functions (“psychological properties”) of one of those stimuli is modified in some way - the corresponding functions of related stimuli can change without additional training or instruction (e.g., if  $A < B < C$  and A is threatening, then C is more threatening than A).

Humans can learn to relate stimuli in many ways, from equivalence (e.g., A is the

same as B) and opposition (e.g., A is opposite to B; Barnes-Holmes, Barnes-Holmes, & Smeets, 2004), to comparison (e.g., A is bigger than B; Vitale, Barnes-Holmes, Barnes-Holmes & Cullen, 2008), temporality (A is before or after B; O’Hora et al., 2008), conditionality (e.g., If A then B; Augustson, Dougher, & Markham, 2000), hierarchy (e.g., A contains B; Gil, Luciano, Ruiz & Valdivia-Salas, 2012), analogy (e.g., A is to B as C is to D; Stewart, Barnes-Holmes, & Roche, 2004) and deixis (or perspective-taking; e.g., I am here and you are there; McHugh, Barnes-Holmes, & Barnes-Holmes, 2004a) (see Hughes & Barnes-Holmes, 2016a; Stewart & Roche, 2013; Törneke, 2010). The way in which people relate stimuli and the functions transformed through those relations is controlled by stimuli in the environment (i.e., ‘contextual cues’). Some (‘relational’) cues (e.g., the words “same as” or “bigger than”) specify how stimuli are related while other (‘functional’) cues specify the functions transformed through those relations. For example, in the phrase “Zim tastes like chocolate”, the relational cue ‘like’ specifies a sameness relation between the novel word (Zim) and chocolate while the functional cue ‘tastes’ specifies that the functions that should transform from chocolate to Zim are gustatory (and not visual or auditory). RFT posits that all relational responding is contextually controlled, as are the stimulus functions transformed by relational responding (i.e., the specific forms of the relations derived and functions transformed through these relations occur due to contextual cues). Furthermore, once a sophisticated repertoire of AARR is established, people have the capacity to elaborate entire ‘networks’ of stimulus relations quickly, to bring them under increasingly subtle forms of contextual control, and to transform stimulus functions through entire networks (Barnes-Holmes, O’Hora, Roche, Hayes, Bissett, & Lyddy, 2001).

#### **1.4.1. Properties of Arbitrarily Applicable Relational Responding (AARR)**

While mutual and combinatorial entailment and transformation of functions represent the core defining features of AARR, there are three dimensions of analysis that are worth

noting, particularly when considering the role of AARR in psychological suffering; namely complexity, derivation, and coherence (see Barnes-Holmes, Barnes-Holmes, Hussey, & Luciano, 2016)<sup>3</sup>. Firstly, the number and type of stimuli and ways in which they are related, functions that are transformed through those relations, and sources of contextual control that influence the probability of those responses can vary from high to low complexity. In the context of paranoia, individuals may develop ‘relational networks’ (which vary in complexity) that allow them to navigate their social world and which may influence responding to events in the (social) environment. These relational networks may alter the functions of specific events and the ability of such events to elicit particular responses or to function as effective reinforcers or punishers. For example, individuals who experience paranoia may formulate networks characterized by relations of coordination between the self and others and particular negative evaluations (e.g., “I am vulnerable”, “Others are devious”). Such networks may be established via a history of adverse interpersonal experiences (e.g., being betrayed or mistreated by someone). Once such events are experienced, AARR allows these events to be placed into contact with entirely unrelated events, which transforms the functions of those events as well (e.g., by transforming the functions of strangers from ‘neutral’ or ‘ambiguous’ to ‘threatening’) and may come to control the functions of many related events (e.g., novel interpersonal interactions with strangers) and to influence the probability of certain responses (e.g., approach or avoidance) in the presence of these events. Indeed, once someone can AARR, an aversive event (e.g., approaching a stranger) need not be directly experienced for someone to respond to it.

Secondly, the extent to which a particular pattern of relational responding has been derived (or ‘rehearsed’) previously can vary along a continuum from high derivation to low derivation, marked out as distance from the first point of derivation, which is invariably high.

---

<sup>3</sup>More recently, a fourth dimension of AARR has been suggested; namely relational flexibility, which refers to the extent to which a particular pattern of AARR may be modified by contextual variables (see Barnes-Holmes et al., 2016; Barnes-Holmes, Barnes-Holmes, Luciano, & McEntegart, *in press*).

For example, if an individual is informed that B is the same as A and that C is the same as B they will likely derive that A is the same as C. The first instance in which they derive the A-C relation could be considered a ‘highly derived’ response given that it is entirely derived from a limited set of prior learning experiences (i.e., it has not been derived before with those stimuli by that individual). However, given an ever-increasing number of opportunities to derive the relation between those same stimuli (i.e., A and C), across each of the following occasions of derivation the resulting response may come to be increasingly defined as a ‘low derivation’ response (Hughes, Barnes-Holmes, & Vahey, 2012). It is likely that relational responses that are low in derivation are experienced as ‘automatic’, especially if these responses are also low in relational complexity (e.g. “Others cannot be trusted”). The same may hold true for increasingly complex relational responses (sometimes referred to as ‘beliefs’). For instance, a belief derived for the first time may differ from a belief derived 100 or 1000 times. As these relations are increasingly derived, the responses they occasion (e.g., fear, avoidance) may come to be increasingly elicited ‘automatically’. We could apply the dimensions of complexity and derivation to Freeman’s (2016) definition of paranoia as ‘threat beliefs’. For instance, from an RFT perspective, ‘belief systems’ can be conceptualized as more or less complex relational networks that have been derived many times in the past.

A third dimension along which AARR can be analyzed is coherence. Coherence is a core assumption within RFT and refers to “the extent to which a particular pattern of relational responding yields relatively consistent consequences” (Hussey, Barnes-Holmes, & Barnes-Holmes, 2015, p. 13). RFT researchers propose that coherence is both a well-established repertoire and a potential reinforcer for AARR, in and of itself. As an operant, AARR can be shaped by contingencies of reinforcement or punishment. Early on in our development these contingencies are delivered by the socio-verbal community to ensure that

people AARR in an internally consistent or coherent manner (see Hayes, Fox, et al., 2001). Eventually coherence becomes a conditioned reinforcer for AARR because responding in coherent ways results in both socially mediated reinforcement (e.g., positive attention) and effective action (e.g., by being able to accurately, coherently, and contingently describe relations among events; see Wray, Dougher, Hamilton, & Guinther, 2012). Discriminating that one is deriving coherently (i.e., that one is “correct”) or that certain patterns of relational responding increase the likelihood of certain consequences (e.g., prevention of harm) may increase the probability that similar responses are emitted in future. The dimensions of derivation and coherence may covary whereby particular patterns of relational responding may be considered increasingly coherent with increased derivation opportunities. In the context of paranoia, as particular patterns of responding to the self (as ‘vulnerable’), others (as ‘devious’) and the (social) world (as ‘hostile’) are increasingly derived, these may be elaborated into relational networks that could be considered typical or characteristic of paranoia. Indeed, prominent cognitive-clinical models (e.g., Freeman et al., 2002; Bentall et al., 1994; 2001) and research on paranoia suggest that such patterns of responding to the self and others are pertinent to paranoia and are implicated in the development and persistence of paranoid beliefs (see Kesting & Lincoln, 2013). Furthermore, such networks are likely influenced or elicited by environmental contingencies (e.g., adverse interpersonal events), which reinforce these patterns of relational responding and their perceived coherence (e.g., “I was right. I am vulnerable. You can’t trust anyone.”).

#### **1.4.2. The importance of AARR**

CBS researchers are interested in AARR for several reasons, partly due to the advantages it affords language-able humans (in principle) and partly due to its utility for CBS researchers interested in examining and influencing complex human behavior. In relation to the former, learning how to relate stimuli, unconstrained by their physical properties,

unshackles humans from the physical world. Now (in principle) anything can be ‘symbolically’ related to anything else (see Barnes-Holmes, Hayes, Dymond, & O’Hora, 2001). Indeed, once a sophisticated repertoire of AARR is established, then the ability to arbitrarily relate stimuli in even just a few ways quickly allows people to respond in a complex manner (i.e., to generate relational networks under the control of contextual features specifying multiple relations and functions). Moreover, stimuli participating in derived relations can acquire entirely new functions or have their existing functions modified. This learned capacity to transform the functions of stimuli through derived relations appears to be the defining property of human language and cognition and responsible for its complexity, generativity, and semantic richness. Thus, AARR equips humans with “an extraordinarily efficient and generative means of interacting with the world around them” (Hughes & Barnes-Holmes, 2016b, pp. 184-185). In relation to the latter, AARR unlocks a functional-analytic account of a host of complex behaviors such as thinking, self-knowledge, problem-solving, and rule-following for CBS researchers interested in studying and affecting these phenomena. Furthermore, AARR provides useful units with which to conceptualize ‘private’ behavior in functional-analytic terms, thus opening up such behavior to prediction-and-influence (e.g., via manipulating contextual control to produce changes in AARR). In this way, even a small set of behavioral principles and analytic concepts may facilitate prediction-and-influence over an array of complex human behavior in a coherent and parsimonious manner (Hughes & Barnes-Holmes, 2016b).

#### **1.4.3. AARR, psychological suffering and paranoia**

AARR is quite simply a ‘game changer’: it enables humans to respond to and modify their own behavior in ways that many other organisms cannot. However, the bi-directionality that is characteristic of AARR can be a double-edged sword. On the one hand, derived stimulus relating can be extremely adaptive, allowing humans to link past with future events

as well as derive entirely new sets of affairs from a limited set of prior experiences. Yet, on the other hand, deriving relations between stimuli can also set the stage for many psychological problems, including paranoia (see Törneke, Luciano, & Valdivia Salas, 2008). For example, an individual who has had a traumatic experience (e.g., being betrayed) may find recalling this event highly aversive. This may be because the words used to describe this event participate in equivalence relations with the actual event, and as a result, acquire many of the functions (or psychological properties) of this event (via transformation of function). AARR can also put people into contact with events that might never, or have yet to, happen. This can be adaptive in some cases but maladaptive in others. For example, being told that a certain neighborhood is dangerous (e.g., crime is prevalent) places this locality in a relation of coordination with something aversive (e.g., being mugged). Hence, the neighborhood acquires stimulus functions (for the individual in that context) that it did not have prior to hearing this information. Consequently, the individual may avoid this neighborhood in future, despite not having directly contacted the stipulated consequences. Similarly, AARR can put people into contact with purely hypothetical events. For example, a potential future event involving being betrayed can be contacted via temporal relations (e.g., “If I trust someone again they will betray me”). Although this event has not, and may never, happen the individual may still experience distress via AARR. Indeed, a number of recent studies have modelled this experimentally, demonstrating that AARR may explain how people come to fear and avoid certain stimuli in the absence of prior experience or instruction (see, for example, Dymond, Schlund, Roche, & Whelan, 2014; Dymond, Schlund, Roche, De Houwer, & Freegard, 2012).

Increasingly complex instances of AARR (commonly referred to as ‘rules’) may also be central in the development and maintenance of paranoid beliefs. RFT defines ‘rules’ as complex stimulus relations involving “the coordination of behavior with a verbally specified

contingency that makes reference to antecedent, behavioral, and consequential events” (Hughes & Barnes-Holmes, 2016b, p. 195) and proposes that rules alter behavior through the appropriate transformation of functions that result from contact with the elements included in the rule. Let us consider paranoid beliefs in terms of ‘rules’. For example, individuals may formulate a network of related events characterized by fear and avoidance functions (e.g., “Approaching strangers is dangerous so I should avoid people I don’t know”). While these relational networks, or ‘rules’, may prove to be adaptive in some contexts, they may be maladaptive in others by rendering behavior less sensitive to direct environmental contingencies (e.g., Monestes, Villatte, Stewart, & Loas, 2014). For instance, rule-governed behavior may restrict the individual’s behavioral repertoire, while efforts to respond in ways that are coherent with those rules may increase their influence. For example, avoidance in the context of paranoia is likely negatively reinforced via short-term consequences of responding in-line with the rule (e.g., avoiding strangers will typically be followed by the absence of harm, reducing anxiety). However, engaging in such behaviors prevents contact with other contingencies providing evidence contrary to the rule (e.g., appetitive interpersonal events). As these paranoia-related ‘rules’ are increasingly derived, they may also increase in terms of complexity (e.g. as new events are experienced) and inflexibility, and the responses that they occasion (e.g., fear, avoidance) may come to be elicited in an increasingly ‘automatic’ manner. For instance, an individual who experiences paranoia may develop relational networks or ‘rules’ characterized by relations of coordination between the self and others and negative evaluations (e.g., as ‘vulnerable’ and ‘devious’ respectively) and conditionality between events such as approach by strangers and perceived aversive consequences (e.g., “If strangers approach me they will try to hurt me”). Accordingly, any attempt to help by others may readily be perceived as a threat (aversive), rather than being interpreted as support (that would typically be appetitive). Given an aversive or traumatic interpersonal history, AARR

potentially allows all other people, for example, to be coordinated, such that all strangers may be perceived as devious, threatening, etc., even though there is some direct experience of the contrary.

### **1.5. Measuring AARR: The Implicit Relational Assessment Procedure (IRAP)**

Over the past two decades, RFT has expanded rapidly into many areas of important concern, including psychological suffering (see Dymond & Roche, 2013; Stewart, 2016) and paranoia specifically (Stewart, Stewart, & Hughes, 2016). From this perspective, we define paranoia and related phenomena (e.g., self-concepts) in functional-analytic terms as behaviors, and specifically as AARR. We consider these patterns of behavior in context, and the goal of this program of PhD research involves determining how these behaviors influence each other and which contextual variables affect this. To this end, we utilize an RFT-based measure called the Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes et al., 2006) to examine paranoia and related behavior (e.g., relational responding towards the self and others) from a functional-analytic perspective. The IRAP is a computerized response latency procedure that allows researchers to target specific repertoires of relational responding that have been established by prior learning histories (such as those involved in psychological suffering), measure their strength, probability, and persistence, and capture AARR ‘in flight’ (see Barnes-Holmes et al., 2016; Hussey et al., 2015).

When it comes to measurement procedures, researchers operating from the functional-analytic perspective adopt different assumptions from their mechanistic counterparts and these have implications for the current research regarding interpreting and integrating the findings from the studies within the current thesis into the wider cognitive-clinical literature. Given that ‘cognitive’ researchers assume that mental mechanisms mediate between environment and behavior, behavioral outcomes obtained on measurement procedures are considered proxies for, or equivalent to, the presence of the assumed mental construct of

interest. Hence, the (measurement) context is considered separate from the mental processes that are hypothesized to cause behavior. For example, in the context of implicit cognition research, cognitive models conceptualize responses on implicit and self-report measures as reflecting structurally distinct mental representations or associations in memory (although see the associative-propositional evaluation, or APE, model; Gawronski & Bodenhausen, 2007a; 2007b). Therefore, the observed behavioral outcome on a measure like the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) is considered to reflect the assumed mental mechanism (i.e., associations in memory). In contrast, the Relational, Elaboration and Coherence (REC) model (Barnes-Holmes, Murphy, Barnes-Holmes, & Stewart, 2010; Hughes et al., 2012), an RFT interpretation of measurement procedures including the IRAP, proposes that the behavioral outcome obtained on a measurement procedure reflects an interaction between the individual's learning history with respect to the targeted relation(s) and the specific features of the context in which they are assessed. Thus, the measurement procedure – as a central feature of the current context – is considered an explanatory factor in understanding the observed behavior (Hughes et al., 2012). In other words, causation is examined in terms of the functional relation(s) between environment (including the measurement context) and behavior, without the need to appeal to mediating mental mechanisms. Although the IRAP is often referred to as a type of 'implicit' measure, this is a fundamental difference between the procedure and the interpretation of its output, relative to other more mechanistically-driven implicit procedures.

The current research involves both self-report and so-called 'implicit' measures. RFT researchers do not consider responses defined as 'explicit' (i.e., self-report) and 'implicit' as belonging to mutually exclusive classes (e.g., 'conscious' and 'unconscious'), nor do they contend that dual processes (e.g., associative or propositional) account for responding on these measures. Rather, they appeal to the single process of AARR and view responses on

both types of measures as a function of the individual's learning history with respect to the targeted stimuli and features of the current measurement context (see Barnes-Holmes et al., 2010; Hughes, Barnes-Holmes, & De Houwer, 2011; Hughes et al., 2012). According to the REC model, AARR can be carved into different patterns of behavior that vary in their relative levels of complexity and derivation; the degree to which different features of AARR will be made evident to the researcher will depend on "the specific constellation of conditions that characterize the context in which that behavior is measured" (Hughes et al., 2012, p. 25). Implicit measures are considered to capture brief and immediate relational responses (BIRRs), while their self-report counterparts capture extended and elaborated relational responding (EERRs) as a function of the assessment context. In everyday situations, BIRRs and EERRs likely interact with each other in a dynamic fashion. For instance, in an ambiguous interpersonal context an individual may emit a BIRR in response to a glance from a stranger (e.g., "They're watching me"). Subsequently emitted EERRs could either contradict this BIRR (e.g., "I don't have enough information to base this on") or converge with it (e.g., "They're out to get me"). From an empirical perspective, this has implications for examining the relationship between responses on implicit and self-report measures (e.g., whether they converge or diverge), and their predictive validity when it comes to related behavior (e.g., paranoia, avoidance, etc.)<sup>4</sup>. De Houwer (2011) argued that by studying potential moderators of behavioral effects (e.g., measurement context, stimuli presented),

---

<sup>4</sup> From a functional-analytic standpoint, however, these are considered behavior-behavior relations. Ultimately, we must identify the contingencies that will influence the probability of one class of responses (i.e., BIRRs or EERRs) predicting a second class of responses (e.g., avoidance). For example, in the context of an ambiguous interpersonal event (e.g., being approached by a stranger), particular features of the current context (e.g., this event occurring at night in an empty street) may increase the probability that one's BIRRs (e.g., "This person is dangerous", "They will try to hurt me") will influence subsequent behavior (e.g., attempting to escape or avoid the stranger). However, if the features of the current context differed such that the individual was approached by the same stranger but this time on a busy street during the daytime, the initial BIRRs may be followed by subsequent EERRs (e.g., "They're not going to attack me in front of all these people", "They could be lost and looking for directions"), increasing the probability of alternative subsequent behavior more aligned to EERRs than BIRRs occurring (e.g., approaching and offering to help the stranger).

more can be learned about the conditions under which certain elements in the environment influence behavior.

### **1.5.1. The utility of the IRAP in experimental-clinical research**

The IRAP procedure involves presenting specific relational terms (e.g., Similar, Opposite) as response options with respect to describing the relations among relevant label (e.g., ‘Me’ and ‘Others’) and target stimuli (e.g., ‘Positive’ and ‘Negative’ words). Participants are required to respond quickly and accurately in ways that are deemed either consistent (or ‘coherent’) or inconsistent with their pre-established history of responding to these stimuli. According to the REC model, trials may produce a BIRR before the participant actually presses a response key. By definition, the most probable immediate response (i.e., one that coheres with the individual’s learning history) will be emitted first most often, and thus any IRAP trial that requires a response that coordinates with that immediate response will be emitted relatively quickly. If, however, a trial requires a response that opposes the immediate relational response, then it may be emitted less quickly. The extent of the observed difference in response latency between these ‘consistent’ and ‘inconsistent’ trials is assumed to provide an index of the strength of the relational responses, or repertoires, being assessed (termed the “IRAP effect”).

There are currently over 50 published IRAP studies supporting its utility. There is also some evidence to suggest that the IRAP may offer some advantages over other so-called ‘implicit’ measures, such as the IAT (see Hughes et al., 2011; 2012). To illustrate, consider the measurement of implicit self-esteem (an important construct in the context of paranoia that we will return to later). The IAT assesses the relative strength of associations of two contrasted construct categories (e.g., ‘Me’ versus ‘Others’) with two contrasted attribute categories (e.g., ‘Positive’ versus ‘Negative’). Participants are required to respond as quickly as possible in accordance with two separate construct-attribute associations by mapping one

of each onto the same response keys (e.g., press ‘d’ for ‘Me’ and ‘Positive’, and ‘k’ for ‘Others’ and ‘Negative’). On ‘consistent’ blocks of trials, participants are required to categorize ‘Me’ with ‘Positive’ and ‘Others’ with ‘Negative’. On ‘inconsistent’ blocks, they must categorize ‘Me’ with ‘Negative’ and ‘Others’ with ‘Positive’. Faster responding on consistent (‘Me’ with ‘Positive’) than inconsistent (‘Me’ with ‘Negative’) blocks indicates positive implicit self-esteem. However, as the IAT juxtaposes self and others as contrast categories, this makes it impossible to explore self-based associations independently (Karpinksi, 2004; although see Pinter & Greenwald, 2005). Thus, the effect obtained does not specify whether the stronger self-positive association represents a ‘Me-Positive’ bias, an ‘Others-Negative’ bias, or some combination of the two. Also, the IAT does not capture the nature of an association (e.g., ‘Me-Positive’ could in principle be ‘Me’-same as-‘Positive’ or ‘Me’-opposite to-‘Positive’) and thus can only be deemed an *indirect* measure of potential underlying self-evaluations.

In contrast, the IRAP focuses on specific relations (e.g., coordination, opposition, distinction) between stimuli, rather than associations, and so stipulates the nature of these pairings. The format of IRAP trials is also different to that of the IAT. On each IRAP trial, one of two label stimuli (e.g., ‘Me’, ‘Others’) is presented top-screen with either a Positive (e.g., ‘good’, ‘friendly’) or Negative (e.g., ‘bad’, ‘rejected’) target stimulus presented center-screen. Participants must then choose one of two response options (e.g., ‘Similar’, ‘Opposite’) presented left and right bottom-screen. Hence, responding on consistent trials might involve selecting ‘Me-Positive-Similar’ while responding on inconsistent trials might involve selecting ‘Me-Positive-Opposite’. This format generates four individual trial-types which are considered four metrics of the individual relational responses that comprise the bias under scrutiny (e.g., a self-esteem bias; ‘Me-Positive’, ‘Me-Negative’, ‘Others-Positive’, ‘Others-Negative’), unlike the IAT’s single metric. Thus, the IRAP may afford greater

precision in terms of clarifying the observed pattern of relational responding. Furthermore, a rapidly expanding IRAP literature suggests that this measure can also capture more complex relational responses that are low in derivation and emitted quickly and accurately. For example, Remue, Hughes, De Houwer, and De Raedt (2014) found that while on an IAT, participants with self-reported dysphoria showed similar responding to controls, the groups diverged on the IRAP, with those with dysphoria showing a greater discrepancy between *actual* and *ideal* self-esteem (measured using two separate IRAPs). Such research highlights the utility of the IRAP in parsing out how specific self-concepts may be related to clinically-relevant phenomena.

Indeed, the IRAP has been increasingly used in experimental-clinical research and has provided some important insights therein (see Vahey, Nicholson, & Barnes-Holmes, 2015, for a recent meta-analysis). Much of this research has sought to identify repertoires of AARR pertinent in a given domain, for example, OCD (Nicholson & Barnes-Holmes, 2012b; Nicholson, McCourt, & Barnes-Holmes, 2013), dysphoria (Remue and colleagues, 2013; 2014), cocaine dependence (Carpenter, Martinez, Vadhan, Barnes-Holmes, & Nunes, 2012), body dissatisfaction and disordered eating (Parling, Cernvall, Stewart, Barnes-Holmes, & Ghaderi, 2012), and non-clinical voice-hearing (e.g., McEnteggart, Barnes-Holmes, Egger, & Barnes-Holmes, 2016). This growing body of research suggests that the IRAP may be able to parse out individual patterns of relational responding pertinent to clinically-relevant domains. Typically, such studies are preceded by preliminary experimental work to identify which relational responses seem most relevant in a given domain. This thesis presents a first step toward investigating paranoia from an RFT perspective using the IRAP. We drew upon the existing cognitive-clinical literature to help determine which relational responses may be pertinent to this phenomenon.

## 1.6. Cognitive approaches to paranoia

Two of the most prominent theories of paranoia are Freeman and colleagues' (2002) 'Threat anticipation model' and Bentall and colleagues' (1994; 2001) 'Defensive model'. Aspects of these models have informed the development of the studies within the current thesis and brief overviews of each are outlined below.

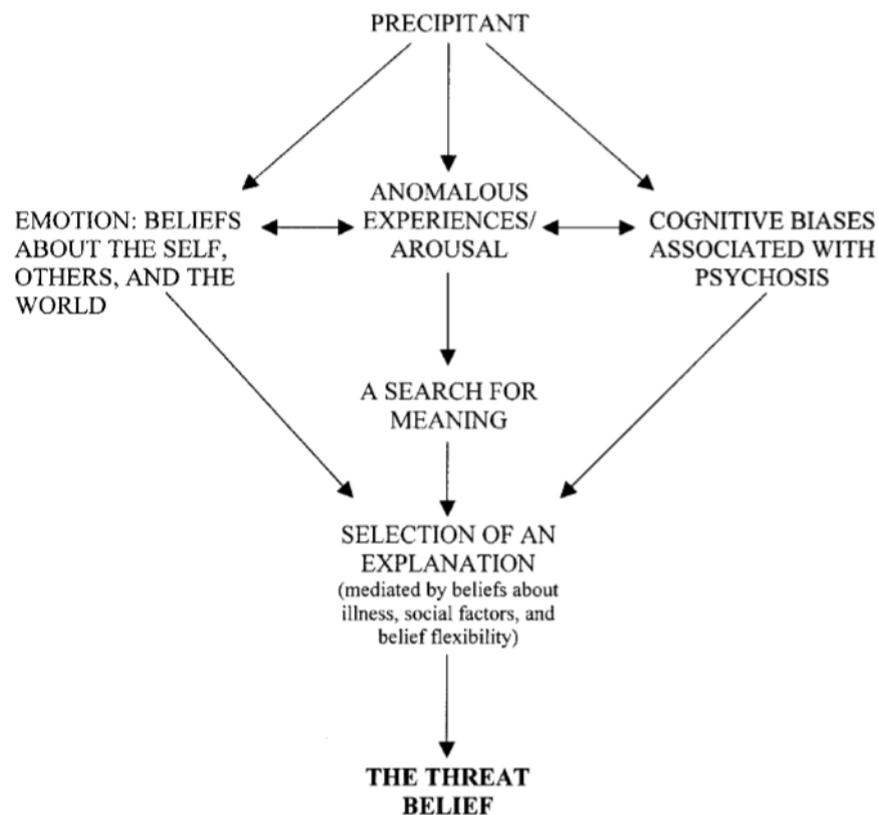
### 1.6.1. Threat anticipation model of paranoia

Freeman and colleagues (2002) conceptualize persecutory beliefs as attributions (i.e., a causal explanation for events) and proposed that paranoia arises from a complex interaction amongst: internal physiological experiences (e.g., heightened arousal, sleep disturbance, illicit drug use); anomalous experiences (e.g., perceptual anomalies, depersonalization); external experiences (e.g., behavior of others); cognitive reasoning biases (e.g., jumping to conclusions); and the (often adverse) environment (see Figure 1.2). In such a context, paranoia is argued to emerge from the search for meaning. According to Freeman et al., anomalous experiences may occur: *directly* (e.g., via difficult interpersonal relationships, isolation); *indirectly* via emotional disturbance (e.g., anxiety); or *cognitive biases* (e.g., reasoning biases, confirmatory bias), all of which may lead to external events that are unusual, ambiguous, negative, or neutral (though often socially significant) also becoming incorporated into the search for meaning. As a result, the individual searches for an explanation of, or meaning in, the triggering (or recent) event, while bringing to bear pre-existing schemas<sup>5</sup> (e.g. regarding self, others, and the world). For instance, a persecutory belief is likely to be formed if the individual already believes they are vulnerable or because they view the world as dangerous, perhaps as a result of previous traumatic experiences (e.g., Beards & Fisher, 2014; Bentall et al., 2014; Murphy, Murphy, & Shevlin, 2015). Freeman et

---

<sup>5</sup> Schemas or schematic beliefs have been defined in the cognitive literature as *psychological* (i.e. mental) structures for processing, categorizing and evaluating stimuli, events and experiences in one's environment (e.g. DiMaggio, 1997). They are considered to be relatively stable patterns of thought that allow one to understand and interpret experiences, and they may also influence or govern behavior (Beck, 1967; Beck et al., 1990).

al. also proposed that a heightened emotional state may lead the individual to interpret the coalescence of feelings (e.g. anxiety), external experiences (e.g. the facial expressions of others) and cognitions (e.g. “others don’t like me”) as evidence of threat. They also suggested that a delusion may reflect an attribution which in turn leads to further attributions. For example, someone may attribute another’s glance at them as “they’re watching me”, and this may facilitate subsequent delusional beliefs (e.g., “some people are out to get me”).



**Figure 1.2.** Summary of the formation of a persecutory delusion according to the threat anticipation model. Reprinted from “A cognitive model of persecutory delusions” by D. Freeman et al. (2002), *British Journal of Clinical Psychology*, 41, 334. Copyright 2002 by the British Psychological Society.

### 1.6.2. Defensive model of paranoia

Bentall and colleagues (1994; 2001) argued that paranoia results from defensive efforts against negative affective processes and cognitions, rather than as a direct reflection of emotional concerns (as per Freeman et al.’s model). Bentall et al. proposed that individuals

who experience persecutory delusions have latent negative beliefs about the self that are vulnerable to activation by negative or threatening life events. They reasoned that studies showing a discrepancy between ‘implicit’ and ‘explicit’ self-esteem might provide support for this assumption. Specifically, they hypothesized that individuals with persecutory delusions have normal or elevated levels of explicit self-esteem but low levels of implicit self-esteem. In simple terms, this model suggests that individuals seek to preserve self-esteem and avoid discrepancies between actual and ideal self, by making externalizing, personalizing attributions for negative events (i.e., attributional biases).

### **1.6.3. Self-esteem, self- and others-evaluations, and paranoia**

These models highlight the central role of ‘schemas’ about the self (e.g., as vulnerable), others (e.g., as devious), and the world (e.g., as dangerous) in the development and maintenance of paranoia. Indeed, this factor is frequently targeted in cognitive approaches to paranoia, with low or unstable ‘self-esteem’ (an evaluative component of the self-concept) considered to be a vulnerability factor (Freeman, 2007) and causal (Bentall et al., 1994; 2001) in the onset and maintenance of paranoia. For instance, Freeman et al. (2002) suggested that existing beliefs are drawn upon to make sense of one’s experiences and that a persecutory belief is likely to be formed if the individual already believes that they are vulnerable or “a soft-target” (p. 335) or because they view other people and the world as hostile. Garety and colleagues (2001) suggested that persecutory thoughts may be held with greater conviction if they are consistent with firmly-held beliefs about the self (e.g., “I am weak”) and others (e.g., “others are untrustworthy”) and that, once formed, such ideas are likely to be considered further confirmation of these negative beliefs, perpetuating paranoia. In their Attribution-self-representation model, Bentall and colleagues suggested that there is a dynamic and mutually influential relationship between attributions for negative or threatening events and self-representations, whereby attributions influence self-representations which in

turn influence future attributions (Bentall et al., 2001; Bentall & Kaney, 2005). For example, attributing negative events to others serves to protect one's self-esteem, thus increasing the likelihood that external and personal attributions (i.e., paranoid beliefs) for similar events will be made in future.

Understanding the relationship between specific self-concepts and paranoia may provide important insights into the latter. Indeed, in his synthesis of the cognitive-clinical literature, Freeman (2016) proposed that self-beliefs (especially negative self-beliefs) are one of the key processes implicated in the development and maintenance of paranoia. Self-esteem has been defined as “a positive or negative attitude toward a particular object, namely, the self” (Rosenberg, 1965, p. 30) and is recurrently targeted in the cognitive-clinical research on paranoia. In relation to the aforementioned cognitive theories, a number of systematic reviews indicate that paranoia is directly associated with negative self-concepts (as per Freeman et al.'s model), “without the need to evoke defensive processes” (as per Bentall et al.'s model; Freeman & Garety, 2014, p. 1181). Furthermore, results from studies investigating explicit self-esteem, implicit self-esteem, and their discrepancy are varied but do not suggest that paranoia enhances self-esteem (see Kesting & Lincoln, 2013). Rather, persecutory delusions and paranoia in the general population are characterized by low and fluctuating self-esteem (Kesting & Lincoln, 2013; Tiernan, Tracey, & Shannon, 2014). Freeman (2007) suggested a circular relationship between self-esteem and paranoia, in which low self-esteem constitutes a vulnerability factor for paranoia, which in turn further reduces self-esteem. Bentall et al. (2001) suggested that the observed inconsistencies in the self-esteem data may be due to failure to take into account the complexity of self-representations, dynamic aspects of the self, and that fluctuations in self-esteem may be more pertinent to paranoia than global self-esteem.

#### **1.6.4. Considerations regarding self-concepts in the context of paranoia**

The focus of the current thesis is the examination of self-concepts (i.e., relational responding to the self) in relation to paranoia. However, several factors should be considered regarding the conceptualization and measurement of self-esteem and other self-concepts in relation to paranoia. First, while much previous research has focused on *global* self-esteem, a distinction can be made between the positive and negative components of self-esteem because “individuals may hold both strong positive and strong negative views about the self at the same time, with the two dimensions not only making independent contributions to global self-esteem but also making separate contributions to behavior and affect” (Barrowclough et al., 2003, p. 93). Indeed, persecutory delusions and paranoia in the general population are associated with specific negative self-evaluations; while positive self-evaluations are maintained (or are at least less impaired) in clinical samples (Kesting & Lincoln, 2013; Tiernan et al., 2014).

Second, more research is needed on specific dysfunctional self-concepts in relation to paranoia beyond self-esteem. Specifically, negative beliefs about the self *and others* may be more relevant to paranoia than self-esteem (Fowler et al., 2006; Freeman, 2007; Freeman, 2016; Smith et al., 2006). For instance, Fowler and colleagues (2006) found that negative evaluations of others were strongly and uniquely associated with paranoia in a non-clinical sample, in combination with negative self-evaluations and anxiety (see also Oliver, O’Connor, Jose, McLachlan, & Peters, 2011). Global self-esteem contributed no unique variance to paranoia however. Fowler et al. suggested that, rather than a lack of self-esteem, appraisals of oneself as weak and others as devious imply a potentially very dangerous social position in which one is both vulnerable and threatened by others. This view is encapsulated in the paranoia hierarchy (Freeman et al., 2005), whereby negative self-evaluations may lead to feelings of being different, isolated, inferior and hence vulnerable and paranoia is likely to

flourish when one perceives the self as vulnerable (Freeman, 2016; Freeman & Garety, 2014).

Third, research on the influence of environmental factors on dysfunctional self-concepts and paranoia is limited (Barrowclough et al., 2003; Klippel et al., 2017). A growing body of evidence suggests that experiences of adversity may play an important role in the development of persecutory delusions (e.g., Beards & Fisher, 2014; Bentall & Fernyhough, 2008; Bentall et al., 2014). Adverse interpersonal experiences likely influence negative beliefs about the self (e.g., as vulnerable) and others (e.g., as hostile), one's relationship to others, and future expectations. Indeed, a core process in paranoid belief systems is the expectation that negative interpersonal interactions will be experienced in the future (Bentall & Fernyhough, 2008). Cognitive researchers have suggested that self-schemas may moderate or mediate the formation and maintenance of paranoia in the context of adverse experiences (e.g., Kesting & Lincoln, 2013). For example, Freeman and Garety (2014) suggested that negative beliefs about the self and others are one pathway via which adverse events lead to paranoia.

Fourth, research to date has mainly involved cross-sectional studies, which limit examination of the dynamic relationships between self- and other-evaluations, paranoia, and the influence of environmental events. Experimental research may allow for a more nuanced and *live* investigation of these factors. Recently, a number of paranoia-induction tasks have been used to examine the influence of adverse experiences on paranoid thoughts, self-esteem, and other affective processes (e.g., Bodner & Mikulincer, 1998; Ellet & Chadwick, 2007; Fenigstein & Venable, 1992; Flower, Newman-Taylor, & Stopa, 2013; Freeman, Evans, Lister, Antley, Dunn, & Slater, 2014; Hesse, Schroeder, Scheeff, Klingberg, & Plewnia, 2017; Kesting, Bredenpohl, Klenke, Westermann, & Lincoln, 2013; Lincoln, Stahnke, & Moritz, 2014; Veling, Pot-Kolder, Counotte, van Os, & van der Gaag, 2016; Westermann,

Kesting, & Lincoln, 2012). For example, Ellet and Chadwick (2007) found that high self-awareness during unsolvable tasks triggered paranoid thoughts in a non-clinical sample regardless of task feedback. Kesting et al. (2013) found that social exclusion was associated with increases in non-clinical paranoia, which was mediated by a decrease in self-esteem and moderated by paranoia-proneness. Further research involving similar paradigms may provide important insight into the influence of adverse experiences on specific self-concepts and paranoia.

Finally, much of this research has relied on self-report methods, which depend on an individual's ability to access their feelings and on their willingness to report them (Nisbett & Wilson, 1977; Ross & Nisbett, 1991). Thus, such measures are susceptible to response biases (e.g., self-presentation), particularly on socially or personally sensitive topics (see Greenwald, Poehlman, Uhlmann, & Banaji, 2009). Measuring 'automatic' or 'implicit' cognitions may help circumvent such biases (see Greenwald et al., 2002). This can be achieved using implicit measures, such as the IAT and IRAP, that place participants under time pressure and therein capture 'automatic' responding.

### **1.7. The current thesis**

In the current thesis, we used the IRAP to investigate patterns of relational responding towards the self (and others) that may be pertinent to paranoia. The five considerations listed above provided a coherent, overarching and ongoing guide to the current experimental work. In relation to the first two considerations, we hoped that the individual trial-type format of the IRAP and the use of relational terms as response options would afford greater precision in differentiating individual patterns of relational responding to the self (e.g., as 'positive' and 'negative'; 'vulnerable' and 'safe') than *global* self-esteem. We hoped to learn something of the coherence of the target relations (using 'True' versus 'False' as response options), their

strength (subtracting consistent responding from inconsistent responding) and their potential relationship with standardized explicit measures of paranoia, including predictive validity.

In relation to the third and fourth considerations, we used both cross-sectional and experimental methods to (i) identify the repertoires of relational responding that may be most pertinent to paranoia, and (ii) investigate the influence of adverse events on these patterns of relational responding (in terms of paranoia, responding to the self and others, and their interaction). The latter was based on previous evidence that the IRAP may capture subtle changes in responding following a brief intervention (e.g., Hooper, Villatte, Neofotistou, & McHugh, 2010; Kishita, Muto, Ohtsuki, & Barnes-Holmes, 2014; McEntegart, Barnes-Holmes, & Adekuoroye, 2016). Hence, we investigated the influence of threat-induction tasks that have been used in previous cognitive-clinical research with non-clinical samples on relational responding to the self on the IRAP (as well as self-report measures). In Study 2, we included unsolvable tasks under conditions of high or low self-awareness, as well as incorporating failure feedback or no feedback (see Ellet & Chadwick, 2007). In Studies 3 and 5, we modeled social exclusion using a virtual ball toss game called ‘Cyberball’ (Williams, Cheung, & Choi, 2000), in which participants were either included or excluded by other (computer-generated) players (see Kesting et al., 2013; Lincoln et al., 2014; Westermann et al., 2012).

In relation to the fifth consideration, we hoped that the latency-response demand of the IRAP would circumvent self-presentation strategies. However, given the extensive and highly valuable nature of the literature using self-report measures to study paranoia, we employed a number of the key measures here (e.g. the Rosenberg Self-Esteem Scale, RSES; Rosenberg, 1965). By utilizing both types of measure, we hoped to gain a broader understanding of the specific repertoires of relational responding that are relevant to paranoia

(i.e., BIRRs and EERRs) and how these patterns of AARR may be influenced by adverse experiences.

### **1.7.1. Thesis summary**

The current thesis comprised five studies are presented ( $N = 354$ ): In each study, paranoia was measured using the Paranoia Checklist (Freeman et al., 2005), the most widely-used multi-dimensional measure of paranoia in the general population. In Study 1 ( $N = 84$ ) we examined the relationship between non-clinical paranoia, relational responding to the self as ‘positive’ and ‘negative’ on the IRAP, and self-reported self-esteem using the RSES. In Study 2, we investigated the influence of a threat-induction task (Ellet & Chadwick, 2007) on ‘state paranoia’ and responding on the IRAP and RSES in the same group of participants as Study 1. While the findings from Study 2 demonstrated that relational responding (as measured by the IRAP and self-report) can be influenced by environmental events, the observed absence of specific effects for task condition on paranoia indicated that an alternative threat-induction task was required to achieve greater precision for influencing paranoia. As paranoia can be considered essentially as a perception of *interpersonal* threat, we decided to employ a paradigm that involved an adverse interpersonal experience (i.e., Cyberball) in Study 3 ( $N = 97$ ). Specifically, we examined the effect of social exclusion on state paranoia, and responding on the IRAP and RSES.

Based on the findings from Study 3, we hypothesized that the paranoia may involve more specific patterns of relational responding to the self (e.g., as vulnerable) and others (e.g., as untrustworthy) than general positive and negative regard. Hence, in Study 4 ( $N = 73$ ), we explored relational responding to the self as ‘safe’ versus ‘vulnerable’, and to others as ‘trustworthy’ versus ‘devious’, in individuals with high non-clinical paranoia compared to controls. We screened 472 people using the Paranoia Checklist and invited those who scored within the top and bottom 16<sup>th</sup> percentiles to complete two IRAPs and additional self-report

measures. This method was deemed preferable for identifying individuals from the general population with high non-clinical paranoia and controls as this approach creates groups that are more clearly distinguished in terms of paranoia level (see Flower, Newman-Taylor & Stopa, 2015, for a similar approach). Findings from Study 4 suggested that relational responding to the self as ‘vulnerable’ may be pertinent to paranoia. Hence, in Study 5 ( $N = 100$ ) we investigated the influence of an adverse interpersonal experience (i.e., social exclusion using Cyberball) on state paranoia and relational responding to the self as ‘vulnerable’ (versus ‘safe’) on the IRAP and self-report measures (and responding to others as ‘trustworthy’ and ‘devious’).

Taken together, the findings from the studies in this thesis support the assertion that the IRAP may be a useful measure for identifying specific patterns of relational responding to the self that are relevant to paranoia. Indeed, the IRAP demonstrated predictive validity in this regard (Study 1) as well as sensitivity to environmental events, including a threat induction task (Study 2) and an adverse interpersonal experience (i.e., social exclusion; Studies 3 and 5).

### **1.7.2. Ethical considerations**

All aspects of the present research were approved by the Research Ethics Committees at the National University of Ireland, Galway (see Appendix G), and University College Dublin (see Appendix H). For each study, participants were given a Participant Information Sheet and provided informed consent. They were made aware that their participation would remain confidential and that they were free to withdraw at any point during the study without providing a reason for doing so. Studies 2, 3 and 5 involved mild deception regarding the true purpose of these experiments (i.e., to examine the impact of adverse experiences on state paranoia). In keeping with the British Psychology Society (BPS) *Code of Ethics and Conduct*, this was considered necessary to preserve the integrity of the research (BPS, 2009).

As outlined in the BPS *Code of Ethics and Conduct* Section 3.4, participants were fully debriefed at the end of each of the studies (or if they chose to withdraw prior to this). Special care was taken for studies involving deception to ensure that participants understood the purpose of the study and why deception was used. Participants were also provided with Debriefing Forms, which included contact information for the researcher as well as contact details for the University counseling services.

## **Chapter 2: The relationship between paranoia, self-reported self-esteem, and relational responding to the self on the IRAP**

### **2.1. Introduction**

The aim of this research is to investigate (non-clinical) paranoia in the general population from a functional-analytic perspective using a Relational Frame Theory (RFT) measure called the Implicit Relational Assessment Procedure (IRAP). As outlined in Chapter 1, previous IRAP research in the clinical domain has involved identifying patterns of arbitrarily applicable relational responding (AARR) pertinent to various clinically-relevant phenomena (see Vahey et al., 2015) and examining how responding on the IRAP may covary with clinically-relevant criterion variables, such as self-reported clinically-relevant phenomena (e.g., phobia; Nicholson & Barnes-Holmes, 2012a), known groups differences (e.g., dysphoria versus controls; Remue et al., 2013; 2014), performance on behavioral approach tasks (e.g., BATs related to disgust and obsessive compulsive (OC) tendencies; Nicholson et al., 2012b; 2013), and susceptibility to interventions that target contextual factors pertinent to these domains (e.g., the effects of a mood induction procedure on depressive symptoms; Hussey & Barnes-Holmes, 2012). Typically, these studies have been preceded by preliminary experimental work to identify which patterns of AARR seem most relevant to a given domain.

Following our review of the cognitive-clinical literature, we decided to focus on self-concepts (i.e., relational responding to the self) as a useful starting point for this research. A wealth of evidence suggests that ‘self-beliefs’ are one of the key processes in the development and persistence of paranoia (see Freeman, 2016). Self-esteem in particular has been implicated in prominent cognitive-clinical models and research in this domain. For example, Freeman (2007) proposed that low or unstable self-esteem is a vulnerability factor in the onset and maintenance of paranoia, while Bentall et al. (1994; 2001) deemed it causal,

with paranoia considered to preserve self-esteem and reduce discrepancies between *actual* and *ideal* self. Moreover, evidence from three recent systematic reviews has indicated that persecutory delusions in clinical samples and paranoia in the general population are associated with low *global* self-esteem (e.g., Garety & Freeman, 2013; Kesting & Lincoln, 2013; Tiernan et al., 2014).

Much of this research has relied on the use of self-report measures (e.g., Rosenberg Self-esteem Scale; RSES, Rosenberg, 1965). However, a few studies have utilized so-called ‘implicit’ measures (e.g., self-esteem Implicit Association Test; IAT, Greenwald & Farnham, 2000; or Go/No-go Association Task; GNAT, Nosek & Banaji, 2001). These studies have typically examined discrepancies between self-reported and implicit self-esteem and compared individuals with persecutory delusions with healthy and/or depressed controls (e.g., MacKinnon, Newman-Taylor, & Stopa, 2011; Moritz, Werner, & von Collani, 2006; Kesting, Mehl, Rief, Lindenmeyer, & Lincoln, 2011). Findings have been inconsistent however, with some studies observing lower implicit self-esteem in individuals with persecutory delusions compared to controls (e.g., Moritz et al., 2006; McKay, Langdon, & Coltheart, 2007) while others found no differences (e.g., Kesting et al., 2011; MacKinnon et al., 2011). Similarly, some have observed discrepancies between self-reported and implicit self-esteem (e.g., Valiente et al., 2011), while others have not (Kesting et al., 2011; MacKinnon et al., 2011; Vázquez, Diez-Alegría, Hernández-Lloreda, & Moreno, 2008). To date, only one published study has examined implicit self-esteem (using the IAT) in relation to paranoia within the general population (Cicero & Kerns, 2011). These findings indicated that paranoia was associated with low self-reported self-esteem but not with implicit self-esteem. There was no association between self-reported and implicit self-esteem, nor was paranoia associated with discrepancies between the two. Given these mixed findings, further research is needed to understand the relationship between self-esteem and paranoia.

There are two important theoretical factors to consider when examining self-esteem in relation to paranoia in the context of this thesis. The first reiterates an important point made previously concerning differences in assumptions regarding measurement between mechanistic- and functionally-oriented researchers. That is, cognitive researchers conceptualize responses on implicit and self-report measures as reflecting structurally distinct mental representations or associations in memory, underpinned by dual processes (e.g., associative or propositional). In contrast, RFT researchers appeal to the single process of AARR and view responses on both measurement types as a function of the individual's learning history with respect to the targeted stimuli as well as features of the current measurement context (see Barnes-Holmes et al., 2010; Hughes et al., 2011; 2012). Specifically, implicit measures are considered to capture brief and immediate relational responses (BIRRs) and self-report measures, extended and elaborated relational responding (EERRs). Hence, the current research is not concerned with examining discrepancies between implicit and explicit self-esteem in the context of paranoia. Rather, our aim is to identify BIRRs and EERRs to the self that may be particularly pertinent to paranoia. The second consideration pertains to recommendations that global self-esteem be subdivided into positive and negative components (e.g., Barrowclough et al., 2003). Recent systematic reviews have indicated that negative self-evaluations seem particularly pertinent to paranoia (Kesting & Lincoln, 2013; Tiernan et al., 2014; see also Freeman, 2016). The format of the IRAP means that it produces four trial-types, which can parse out individual patterns of relational responding to the self (i.e., as 'positive' and 'negative'). This enhanced specificity may be useful in the context of the current research for distinguishing between positive and negative self-evaluations in addition to global self-report measures of self-esteem (e.g., RSES).

### **2.1.1. Hypotheses**

The aim of Study 1 was to examine the relationships between self-reported and implicit self-esteem and paranoia in the general population. Specifically, we wanted to investigate whether specific patterns of AARR to the self (as ‘positive’ and negative’) on the IRAP were related to paranoia (as measured by the Paranoia Checklist). We hypothesized that paranoia would be negatively associated with self-reported self-esteem. As this was the first study to investigate implicit self-esteem in the context of paranoia using the IRAP we refrained from making specific predictions. However, based on previous evidence that paranoia is associated with negative self-regard while positive self-regard is maintained (e.g., Kesting & Lincoln, 2013), we expected paranoia to be more strongly associated with negative than positive self-evaluations on the IRAP. We also explored the predictive validity of self-reported self-esteem and relational responding on the IRAP in this regard. Specifically, we investigated whether responses on the self-report measure of self-esteem (i.e., the RSES) and IRAP trial-types would predict scores on the Paranoia Checklist.

## **2.2. Method**

### **2.2.1. Participants**

Eighty-four participants (22 males, 60 females, 2 unspecified; *Mean age* = 21.54, *SD* = 6.20) took part in this study. These were predominantly NUI Galway Psychology undergraduate students who participated in exchange for course credit. Other NUI Galway students were recruited via advertisements.

### **2.2.2. Measures**

**Paranoia Checklist.** ‘Trait’ paranoia was assessed using the 18-item multi-dimensional Paranoia Checklist (Freeman et al., 2005). The items range from commonplace paranoid thoughts (e.g., “Bad things are being said about me behind my back”) to less common ideas (e.g., “There is a possibility of a conspiracy against me”), with each item rated

on a five-point scale for frequency, conviction, and distress. The Checklist demonstrated excellent internal reliability ( $\alpha = .942$ ) and for the Frequency ( $\alpha = .877$ ), Conviction ( $\alpha = .881$ ), and Distress ( $\alpha = .910$ ) subscales in the current sample.

**Rosenberg Self-Esteem Scale (RSES).** The 10-item RSES (Rosenberg, 1965) is the most widely used measure of global self-esteem. Participants were asked to rate their level of agreement with each statement on a four-point Likert scale (ranging from ‘Strongly Agree’ to ‘Strongly Disagree’) “*at the moment*”<sup>6</sup>. The RSES demonstrated excellent internal reliability in the current sample ( $\alpha = .895$ ).

**The IRAP.** The IRAP stimuli were based on the self-esteem IRAP used by Remue and colleagues (2013; see Table 2.1). In relation to the label stimuli (“I am”, “I am not”), participants were instructed:

*“When you see “I” on the screen, whether it’s “I am” or “I am not”, I want you to think about yourself, [name], as you are now. Not how you want to be or how you think you should be but how you truly see yourself now”.*

This was to ensure that participants responded to “I” as referring to them specifically and that the IRAP captured *actual* rather than *ideal* self-esteem.

Correct responding was determined by whether the participant was completing a consistent or inconsistent block. On consistent blocks, correct responding involved agreeing with (i.e., pressing ‘True’ for) “I am-*Positive*” and “I am not-*Negative*” and disagreeing with (i.e., pressing ‘False’ for) “I am-*Negative*” and “I am not-*Positive*” (in line with a positive self-esteem bias), while the reverse pattern of responding was required on inconsistent blocks (see Figure 2.1). The positions of the two response options, “True” and “False”, were counter-balanced across trials. Prior to each block of trials, participants were informed by a

---

<sup>6</sup> Participants in Study 1 also participated in Study 2; the RSES was state-adapted for the purposes of Study 2 and so the same instructions were utilized in Study 1 for the purposes of consistency.

message on the screen, “During the next block the previous correct and wrong answers are reversed”.

Participants received a minimum of one and a maximum of four pairs of practice blocks, with 24 trials per block. They were required to meet criteria of  $\geq 80\%$  accuracy and mean latency  $\leq 2000$ ms to progress to the test phase, which included exactly six test blocks. If an incorrect response was emitted on either task, a red ‘X’ appeared in the center of the screen and remained on-screen until the correct response was emitted. If a correct response was emitted, the screen was cleared for 400ms before the presentation of the next trial. If participants failed to respond within 2000ms, the message “Too slow!” appeared on-screen until they emitted a response. Accuracy and latency feedback were presented at the end of each block of trials. Only data from the test blocks were included in subsequent analyses.

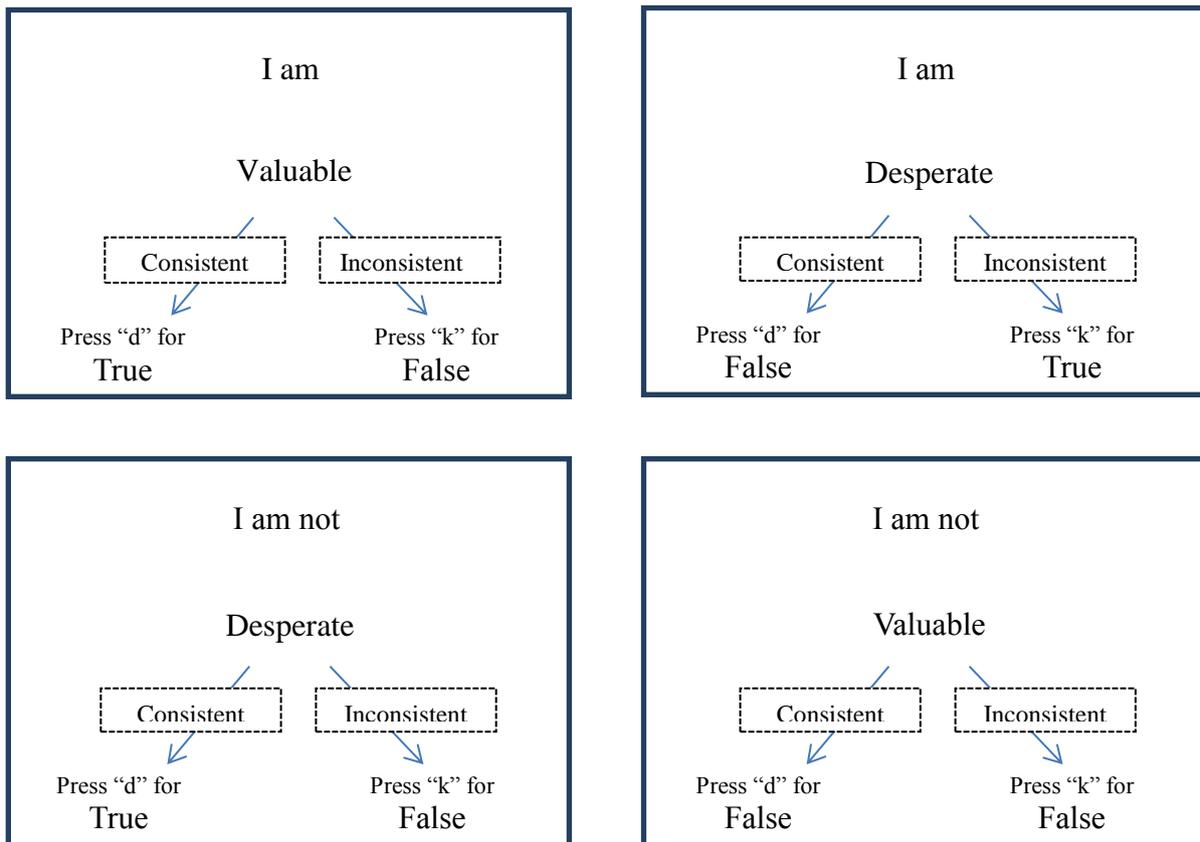
IRAP block order (i.e., consistent- or inconsistent-first) was counter-balanced across participants. Instructions were presented verbally by the researcher, who stayed with the participant during the practice phase and then moved behind a screen once they achieved criteria to proceed to the test phase. If the participant failed to achieve criteria during the practice phase, their participation was ended then and they were thanked and debriefed.

### **2.2.3. General Procedure**

Previous research has suggested that presentation order of self-report and implicit measures has minimal effect on the output of either (e.g., Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005); hence, all participants were presented with the IRAP first followed by the RSES and the Paranoia Checklist (in that order).

**Table 2.1.** IRAP stimuli for Study 1.

Label stimuli	
I am	I am not
Target stimuli	
<i>Positive</i>	<i>Negative</i>
Valuable	Helpless
Happy	Guilty
Kind	Desperate
Friendly	Sad
Hopeful	Rejected
Competent	A failure
Response options	
True	False



**Figure 2.1.** Examples of the four trial-types in the IRAP. Correct answers for responding on consistent and inconsistent blocks of trials are indicated by the arrows (which were not presented during the IRAP procedure).

#### 2.2.4. Statistical analysis plan

A one-way ANOVA checked for potential IRAP block order effects (i.e., consistent- vs. inconsistent-first) on IRAP trial-types. Correlational analyses were conducted between Paranoia Checklist total and subscale scores and RSES and IRAP trial-type scores. Bonferroni correction for multiple comparisons ( $\alpha = .05 / 9 \approx .0055$ ) was used. A multiple regression analysis was conducted to examine the predictive validity of the RSES and IRAP trial-types on Paranoia Checklist total scores.

### 2.3. Results

#### 2.3.1. Data preparation

For the purposes of statistical analysis (and to mitigate effects of attrition), the IRAP accuracy criterion was adjusted to  $\geq 71\%$ <sup>7</sup> (the latency criterion remained at  $\leq 2000$ ms). Two participants withdrew following consent and three failed to meet criteria during the IRAP practice phase. Nine failed to maintain accuracy criteria during the test phase; their data was excluded from subsequent analyses. Thus, 70 participants (19 males, 49 females, 2 unspecified; *Mean age* = 22.01, *SD* = 6.50) were included in the final analyses.

Scoring of the IRAP was conducted using the standardized algorithm for transforming the difference in latencies between consistent and inconsistent blocks of trials into D-IRAP scores as per Nicholson and Barnes-Holmes (2012a). The D-IRAP scores provide an adjusted metric of the size of the difference in reaction times when participants are asked to respond “True” relative to “False” on the four combinations of label and target stimuli, or IRAP ‘trial-types’, (i.e., ‘Me-Positive’, ‘Me-Negative’, ‘Not me-Positive’, ‘Not me-Negative’). IRAP block order (i.e., consistent- vs. inconsistent-first) did not impact responding on the IRAP trial-types (all *ps* > .07).

---

<sup>7</sup> Ten participants scored  $\geq 71\%$  accuracy on one test block during the test phase and were included in the final analysis. Participants who scored  $< 71\%$  during the test phase were removed from subsequent analyses.

The data were analyzed using SPSS Statistics 21. Kolmogorov-Smirnov tests indicated no violations of normality assumptions for the RSES or IRAP trial-types. However, normality assumptions were violated for the Paranoia Checklist ( $p \leq .007$ ). Inspection of histograms indicated positive skew, with most participants demonstrating low paranoia total and subscale scores, which was unsurprising given that the sample was recruited from the general (non-clinical) population. However, skewness and kurtosis were within the acceptable range of  $\pm 2$  for paranoia total, Frequency and Distress subscale scores<sup>8</sup>.

Means, SDs, and the ranges for Paranoia Checklist total and subscale scores, RSES, and IRAP trial-types for the sample are reported in Table 2.2. Higher Paranoia Checklist total and subscale scores indicate higher ‘trait’ paranoia and associated Frequency, Conviction, and Distress. Higher RSES scores indicate higher self-reported self-esteem. Positive D-IRAP scores indicate faster responding on consistent relative to inconsistent blocks of trials (e.g., the observed *Mean* ‘Me-Positive’ scores indicate faster agreement relative to disagreement for ‘Me-Positive’).

---

<sup>8</sup> Kurtosis was 4.83 for Conviction and 2.40 for Distress scores however.

**Table 2.2.** Ranges, Means, and SDs, for Paranoia Checklist total and subscales, RSES, and D-IRAP scores for each IRAP trial-type ( $N = 70$ ).

	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>SD</b>
<b>Paranoia Total</b>	57	160	92.54	23.87
<b>Paranoia Frequency</b>	18	53	27.84	8.31
<b>Paranoia Conviction</b>	19	67	31.47	8.78
<b>Paranoia Distress</b>	18	80	33.23	12.21
<b>RSES</b>	3	30	18.61	5.30
<b>Me-Positive</b>	-.159	1.325	.528	.300
<b>Me-Negative</b>	-.871	.645	-.059	-.367
<b>Not me-Positive</b>	-.810	.892	.063	.417
<b>Not me-Negative</b>	-.861	1.119	.248	.407

### 2.3.2. Correlational analyses

A correlation matrix was calculated using Spearman’s Rho to examine the relationships between RSES, IRAP trial-types, and paranoia scores (see Table 2.3). Significant negative correlations were observed between RSES and Paranoia Checklist total and Frequency, Conviction, and Distress subscale scores (all  $ps < .001$ ). Negative correlations were also observed between the ‘Not me-Negative’ IRAP trial-type and Paranoia Checklist total and Frequency and Distress subscales, indicating that stronger agreement with ‘Not me-Negative’ was associated with lower Paranoia Checklist total ( $p < .001$ ), and Frequency ( $p = .006$ ), and Distress ( $p = .001$ ) scores. A similar trend was observed between

the ‘Not me-Negative’ trial-type and Conviction subscale scores ( $p = .015$ ). There were no correlations between the RSES and IRAP trial-types<sup>9</sup>.

**Table 2.3.** Correlations between RSES, IRAP trial-types, and Paranoia Checklist total and subscale scores ( $N = 70$ ).

	RSES	Me-Positive	Me-Negative	Not me-Positive	Not me-Negative
<b>Paranoia Total</b>	-.538**	.029	-.022	-.042	-.409**
<b>Paranoia Frequency</b>	-.513**	.002	.016	-.060	-.322*
<b>Paranoia Conviction</b>	-.482**	-.062	-.089	.031	-.290*
<b>Paranoia Distress</b>	-.372**	.099	-.027	-.084	-.382**
<b>Not me-Negative</b>	.195	.164	.141	.113	
<b>Not me-Positive</b>	.025	.094	.276*		
<b>Me-Negative</b>	.168	-.209			
<b>Me-Positive</b>	-.033				

\*  $p < .05$       \*\*  $p < .002$

### 2.3.3. Predictive validity of the RSES and IRAP trial-types

A multiple regression analysis was conducted to examine whether RSES scores and IRAP trial-types predicted paranoia (as measured by Paranoia Checklist total score). Demographic variables were also entered into this model as age was negatively associated with paranoia and there was a significant difference between males ( $M = 83.11$ ,  $SD = 21.04$ ) and females ( $M = 96.02$ ,  $SD = 24.02$ ) in Paranoia Checklist total score,  $t(66) = -2.056$ ,  $p = .044$ , with females demonstrating higher Checklist scores compared to males. Results indicated that this model explained 36.1% of the variance in Paranoia Checklist total scores,  $F(7, 69) = 6.677$ ,  $p < .001$ . RSES scores, ‘Not me-Negative’ trial-type scores, and age made

<sup>9</sup> There were also significant negative correlations between age and Paranoia Checklist total ( $r = -.430$ ,  $p < .001$ ), Frequency ( $r = -.425$ ,  $p < .001$ ), Conviction ( $r = -.519$ ,  $p < .001$ ), and Distress ( $r = -.256$ ,  $p = .032$ ) scores.

significant unique contributions to Checklist total scores (see Table 2.4). RSES scores made a larger contribution than the ‘Not me-Negative’ trial-type to the model; RSES uniquely contributed 16.32% to the model whereas ‘Not me-Negative’ contributed 7.12%.

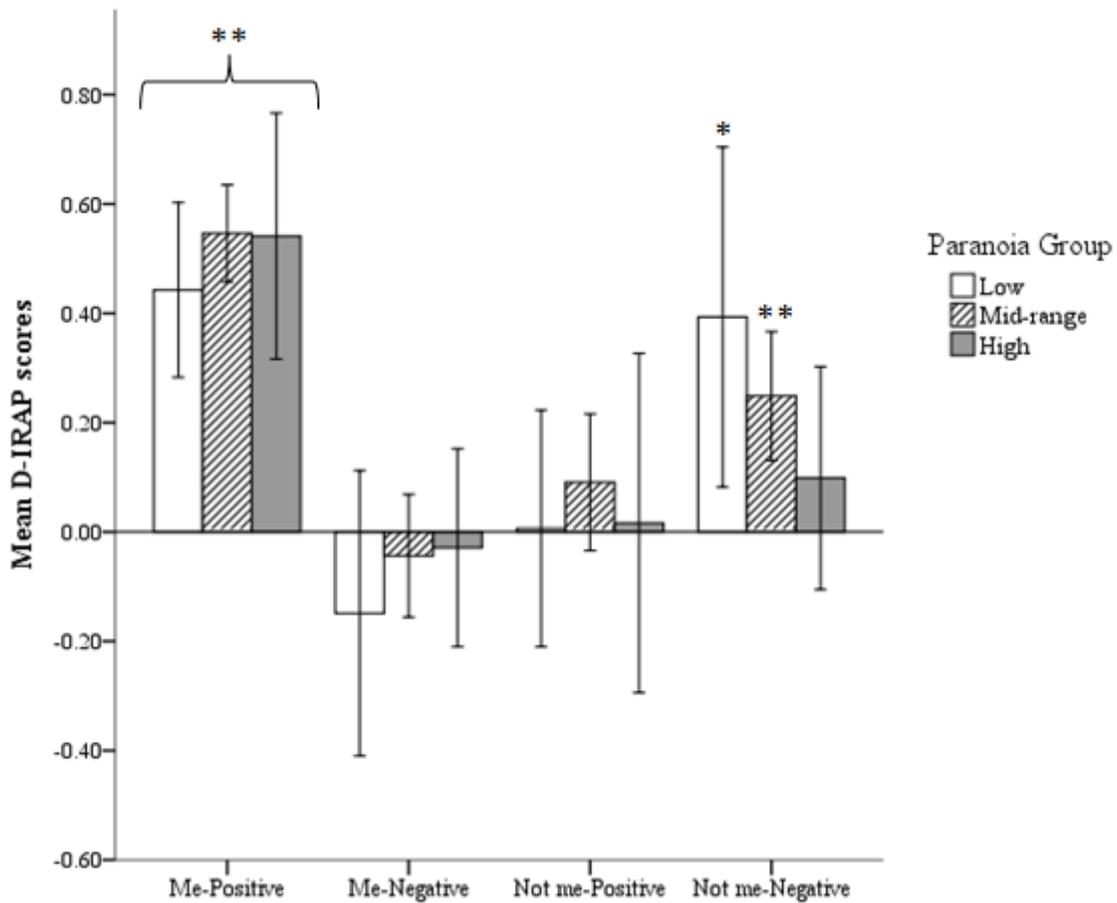
**Table 2.4.** Results from multiple regression analyses using age, gender, RSES scores and IRAP trial-types to predict total score on the Paranoia Checklist.

<b>Dependent variable</b>	<b><i>B</i></b>	<b><i>SE B</i></b>	<b><i>Beta</i></b>	<b><i>p</i></b>
Age	-.855	.365	-.233	.022
Gender	3.990	5.034	.083	.431
Me-Positive	7.729	8.332	.097	.357
Me-Negative	6.779	6.902	.104	.330
Not me-Positive	-2.798	6.022	-.049	.644
Not me-Negative	-16.360	5.905	-.279	.007
RSES	-1.981	.471	-.439	<.001

#### **2.3.4. IRAP trial-type responses as a function of paranoia**

For a more nuanced investigation of the relationship between relational responses to the self on the IRAP and paranoia, we compared responses on the IRAP trial-types across differing levels of paranoia. Freeman (2007) suggested that 10 to 15 per cent of the general population regularly experience paranoid thoughts; thus, we divided participants into three groups on the basis of scoring within one SD of the mean ( $N = 46$ ), one SD above the mean (i.e., top 16<sup>th</sup> percentile,  $N = 12$ ), and one SD below the mean (bottom 16<sup>th</sup> percentile;  $N = 12$ ). This method was deemed preferable to the alternative of using tertiles to select high, mid-range, and low paranoia groups as this approach creates groups that are more clearly distinguished in terms of paranoia level (see Flower et al., 2015, for a similar approach). The mean D-IRAP scores for each trial-type by paranoia group are presented in Figure 2.2. One sample t-tests indicated that the ‘Me-Positive’ trial-type was significantly different from zero for all groups ( $ps < .001$ ), indicative of significantly faster ‘Me-Positive’ agreement (relative

to disagreement). However, only participants in the low paranoia ( $p = .018$ ) and mid-range ( $p < .001$ ) groups demonstrated significant ‘Not me-Negative’ agreement on this trial-type.



\* Sig. at  $\alpha = .05$  \*\* Sig. at  $\alpha \leq .001$

**Figure 2.2.** D-IRAP scores for each trial-type for low, mid-range, and high paranoia groups, with standard error bars (95% confidence intervals). Note, D-IRAP scores above zero indicate faster responding on consistent (i.e., faster ‘Me-Positive’ agreement and ‘Me-Negative’ disagreement) relative to inconsistent blocks of trials, whereas scores below zero indicate the opposite pattern of responding (i.e., faster responding on inconsistent trials). Asterisks indicate trial-types significantly different from zero.

## 2.4. Discussion

The purpose of Study 1 was to explore the relationships between paranoia, self-reported self-esteem (as measured by the RSES) and relational responding to the self as ‘positive’ and ‘negative’ on the IRAP. Findings indicated that paranoia was strongly associated with low self-reported self-esteem, corroborating findings from previous research

(e.g., Martin & Penn, 2001; Thewissen et al., 2008). Paranoia was also associated with the 'Not me-Negative' trial-type on the IRAP, with stronger "I am not-Negative" agreement associated with lower paranoia. Indeed, when IRAP responses were examined across different levels of paranoia, participants in the low and mid-range groups demonstrated significant 'Not me-Negative' agreement, whereas those highest in paranoia did not. Furthermore, both RSES scores and 'Not me-Negative' D-IRAP scores were significant predictors of paranoia, although the RSES was a stronger predictor than the 'Not me-Negative' trial-type.

These findings in relation to the IRAP demonstrate that the IRAP is capable of parsing out individual relational responses to the self and can distinguish between the positive and negative components of self-esteem. Whilst a better predictor of paranoia in this sample, as a global measure of self-esteem, the RSES cannot capture subtle differences in relation to positive and negative self-evaluations. Hence, although both measures showed that self-esteem is indeed relevant to paranoia, the IRAP appeared to provide a more nuanced picture in this regard. Specifically, the IRAP findings highlight the potential importance of negative self-evaluations in predicting paranoia. Furthermore, the observed patterns of responding on the 'Not me-Negative' trial-type suggests that negation of negative self-evaluations may be relevant when situating individuals along the continuum of paranoia. This finding supports those from the wider cognitive-clinical literature that negative self-evaluations are relevant to persecutory delusions and (non-clinical) paranoia (e.g., Collett, Pugh, Waite, & Freeman, 2016; Udachina, Thewissen, Myin-Germeys, Fitzpatrick, O'Kane, & Bentall, 2009; Garety & Freeman, 2013; Kesting & Lincoln, 2013; Tiernan et al., 2014) and covary with and can predict persecutory delusions and paranoia (e.g., Fowler et al., 2006; Freeman, Pugh, et al., 2014; Oliver et al., 2011; Palmier-Claus, Dunn, Drake, & Lewis, 2011; Smith et al., 2006). When used in tandem, the RSES and IRAP could be considered complementary measures of

self-esteem in the context of paranoia. However, the IRAP may offer additional utility in this regard by identifying individual relational responses to the self that are particularly pertinent to this phenomenon.

Regarding the relationship between self-reported and ‘implicit’ self-esteem, we did not observe any correlations between the RSES and IRAP trial-types in this study. Correlations between self-reported and implicit self-esteem have varied between previous studies using the IRAP. For instance, Remue et al. (2014) observed a significant positive correlation between the RSES and *actual* (but not *ideal*) self-esteem on the IRAP (as measured by the overall D- IRAP score) in participants high in self-reported dysphoria only. Timko, England, Herbert and Forman (2010; Experiment 2) and Vahey, Barnes-Holmes and Barnes-Holmes (2009) used a feeling thermometer scale (adapted from Greenwald & Farnham, 2000) to measure self-reported self-esteem, which asked participants to indicate how warmly they felt towards themselves. Vahey et al. observed a significant positive correlation between the feeling thermometer and overall D-IRAP score. Timko et al. did not report correlation analysis between the two self-esteem measures. These few studies suggest that the IRAP scores used (e.g., trial-types vs. overall D-IRAP score<sup>10</sup>) may influence results as both Remue et al. (2014) and Vahey et al. (2009) used the overall D-IRAP score in these analyses. As the purpose of this study was to examine the positive and negative components of self-esteem with the IRAP, it seemed redundant to examine the relationship between the RSES and overall D-IRAP score.

Taken together, the findings from Study 1 provide support for the potential utility of the IRAP as a measure of specific facets of self-esteem (or patterns of relational responding to the self) in the context of (non-clinical) paranoia and that it may show predictive validity in this regard. They also support findings from the wider cognitive-clinical literature that

---

<sup>10</sup> The overall D-IRAP score is defined as the differences in mean response latencies between consistent and inconsistent blocks of trials across all three IRAP test blocks.

negative self-evaluations are particularly relevant to paranoia, whereas positive self-regard may be maintained (e.g., Freeman, 2016; Kesting & Lincoln, 2013; Tiernan et al., 2014). Indeed, all participants demonstrated significant ‘Me-Positive’ agreement on this IRAP trial-type, whereas ‘Not me-Negative’ responses predicted paranoia. The next step for this research will be to identify the contextual factors that may facilitate prediction-and-influence over these patterns of relational responding to the self, paranoia, and the relationship between the two (seen as an important analytic requirement from the functional-analytic perspective). For instance, adverse events may increase paranoia and relational responding to the self as ‘negative’. Such patterns of relational responding could be considered ‘coherent’ in response to such events. That is, typical responding on this trial-type of the IRAP (i.e., in the absence of such events or in *neutral* contexts) may involve faster responding when required to select “I am-Negative-False” relative to “I am-Negative-True” as this likely coheres (or is ‘consistent’) with typical prior learning history in the general (non-clinical) population. However, adverse experiences could influence this relational response such that participants may demonstrate faster responding for “I am-Negative-True” as this could be deemed a coherent or ‘consistent’ response within this context (i.e., an adverse event). Furthermore, responding to the self as ‘negative’ may be associated with AARR related to paranoia (e.g., self-reported paranoia or changes in paranoia following this event) and, taken together, could be considered a coherent ‘relational network’ within this context.

## **Chapter 3: The effects of self-focused attention and task failure on state paranoia, self-reported self-esteem, and relational responding to the self on the IRAP**

### **3.1. Introduction**

Study 1 demonstrated that (non-clinical) paranoia in the general population may be associated with responding to the self as ‘negative’. Specifically, findings on the Implicit Relational Assessment Procedure (IRAP) showed that negation of ‘negative’ self-evaluations (i.e., agreeing ‘Not me-Negative’) was associated with lower paranoia, suggesting that the relation between these stimuli, ‘Not me’ and ‘Negative’ (a relation of distinction between the self and negative evaluations), was relationally coherent for individuals with low to mid-range scores on the Paranoia Checklist in particular. From a functional-analytic perspective however, negative self-evaluations are not considered a complete explanation of paranoia. Rather, the relationship between arbitrarily applicable relational responding (AARR) in terms of paranoia and AARR to the self (e.g., negative evaluations) are conceptualized as patterns of behavior in context (i.e., a behavior-behavior relation; see Hayes & Brownstein, 1986; Hayes et al., 2012). Thus, while negative self-evaluations may predict paranoia (as shown in Study 1), the relationship between these patterns of AARR are themselves based on contextual factors (e.g., contingencies of reinforcement). Hence, functional-analytic researchers aim to determine how these behaviors influence each other and which contextual variables affect them as well as their relationship. By manipulating events in the environment, we may be able to influence these repertoires of AARR, their probability and/or strength, and investigate their interaction in this context.

As a measure of AARR ‘in flight’, the IRAP may be able to capture subtle between-groups differences in AARR due to differing experimental manipulations (e.g., threat-induction tasks), or changes in AARR across time (e.g., from baseline to post-intervention). Indeed, responding on the IRAP has been shown to be influenced by environmental events

(e.g., Bast & Barnes-Holmes, 2015; Cullen et al., 2009; Hooper et al., 2010; Kishita et al., 2014; Hussey & Barnes-Holmes, 2012; McEnteggart et al., 2016) and has demonstrated sensitivity to other ‘moderating’ factors in this regard. For example, Hussey and Barnes-Holmes (2012) found that a mood induction procedure impacted IRAP responses for individuals with self-reported mild to moderate levels of (non-clinical) depressive symptoms, but not their non-depressed counterparts, suggesting that the effects of such events on IRAP responses may be moderated by one’s learning history brought to bear in the current context.

Regarding paranoia, Freeman and colleagues have suggested that negative self-beliefs are often developed in the context of adverse experiences and, by the same token, negative self-beliefs are one route via which adverse events lead to paranoia (Freeman, 2007; 2016; Freeman et al., 2002; Freeman & Garety, 2014). For example, Freeman (2016) reasoned that causing a reduction in self-confidence in vulnerable individuals leads to an increase in paranoid thoughts and there is some evidence in this regard (e.g., Atherton et al., 2016). However, much of the research to date has involved cross-sectional studies, which potentially limits examination of the dynamic relationships amongst self-evaluations, paranoia, and the influence of environmental factors (e.g., adverse experiences). Experimental research may allow for a more nuanced investigation of these factors and help identify environmental variables that may influence paranoia, AARR to the self, and their interaction.

There is evidence that paranoia can be temporarily induced in non-clinical participants using threat-induction paradigms (e.g., Bodner & Mikulincer, 1998; Ellet & Chadwick, 2007; Flower et al., 2015). Previous experimental research has shown that focusing attention on the self may increase the experience of self as the target of others’ thoughts and actions (Fenigstein, 1984) and the experience of paranoid thoughts (Fenigstein & Venable, 1992). More recently, Ellet and Chadwick (2007) examined the effects of task failure under conditions of high and low self-awareness on paranoia. In their study,

participants were exposed to a series of unsolvable tasks under conditions of high self-awareness (whereby a camera was pointed at participants during task completion and they could see themselves on a monitor) or low self-awareness (i.e., no camera present) and were provided with either no feedback or failure feedback on task performance. Findings indicated that high self-awareness alone triggered paranoia regardless of task feedback.

### **3.1.1. Hypotheses**

The aim of Study 2 was to examine the influence of an adverse experience on momentary (or ‘state’) paranoia, self-reported self-esteem, and relational responding to the self on the IRAP in participants from the general population. We used a modified version of Ellet and Chadwick’s (2007) paradigm for our threat-induction procedure. Specifically, we exposed participants to four unsolvable tasks under four conditions (i.e., High self-awareness or Low self-awareness with either Failure feedback or No feedback on task performance; with Low self-awareness + No feedback as the control condition) and measured changes in self-reported momentary or ‘state’ paranoia after each task. We then measured self-reported self-esteem and relational responding to the self as ‘positive’ and ‘negative’ on the IRAP after all four tasks were completed.

Given Ellet and Chadwick’s findings, we expected that state paranoia would increase across exposures to the unsolvable tasks and that higher increases in state paranoia would be observed in High- relative to Low self-awareness conditions. Given evidence that individuals with high non-clinical paranoia experience more paranoid thoughts in contexts of task failure and high self-awareness compared to individuals with low paranoia (Flower et al., 2015), we also hypothesized that changes in state paranoia would be moderated by ‘trait’ paranoia. That is, we anticipated that individuals’ learning histories would influence responding in the current context such that those with a history of responding to adverse events in a paranoid manner would be more likely to do so in the current context compared to individuals without

such histories. We also hypothesized that self-reported self-esteem would be lower in the Failure feedback conditions than the No feedback conditions and lowest in the high self-awareness + Failure feedback condition, as high self-awareness would likely amplify the effects of task failure. It was expected that all participants would demonstrate ‘Me-Positive’ agreement on the IRAP; however, it was hypothesized that participants in the Failure feedback conditions would show faster responding for ‘Me-Negative’ agreement (relative to disagreement). It was expected that this would be strongest in the High self-awareness + Failure feedback condition as this was deemed the most adverse condition (see Flower et al., 2015).

## **3.2. Method**

### **3.2.1. Participants**

The volunteers who completed Study 2 were the same participants from Study 1<sup>11</sup>. Eighty-four volunteers (22 males, 60 females, 2 unspecified; *Mean age* = 21.54, *SD* = 6.20) took part. This sample size was similar to comparable studies conducted by Bodner and Mikulincer (1998) and Ellet and Chadwick (2007).

### **3.2.2. Measures**

**Paranoia Checklist.** ‘Trait’ paranoia was assessed using the 18-item multi-dimensional Paranoia Checklist (Freeman et al., 2005). Each item was rated on a five-point scale for frequency, conviction, and distress (internal consistency for all self-report measures was reported in Study 1).

**State-adapted Paranoia Checklist.** The state-adapted Paranoia Checklist captures current paranoid thinking (Westermann, Kesting, & Lincoln, 2012). The scale contains six

---

<sup>11</sup> That is, for Study 1, participants completed the IRAP, RSES and Paranoia Checklist. For Study 2, the same participants were then randomly allocated to one of four conditions (see “General Procedure”) and completed the unsolvable tasks and state-adapted Paranoia Checklist, followed by the IRAP and RSES again.

items from the original Checklist<sup>12</sup> (e.g., “I am under threat from others”) with amended instructions that read “How strongly do the following thoughts apply to you *at the moment?*” Participants rated their responses on a five-point scale (from 1 = Not at all to 5 = Strongly). The sum score of these items was used as an index of *state* paranoia.

**State-adapted Rosenberg Self-Esteem Scale (RSES).** In the current study, the RSES (Rosenberg, 1965) was adapted so that participants were asked to rate their level of agreement on a four-point Likert scale with each statement ‘*at the moment*’.

**The IRAP.** The IRAP used in Study 2 was the same one used in Study 1 (see Table 2.1). On consistent blocks, correct responding involved agreeing “I am-*Positive*” and “I am not-*Negative*” and disagreeing “I am-*Negative*” and “I am not-*Positive*”, while the reverse pattern of responding was required on inconsistent blocks. The positions of the two response options, “True” and “False”, were counter-balanced over trials. Prior to each block of trials, participants were informed by a message on the screen, “During the next block the previous correct and wrong answers are reversed”.

Instructions for IRAP completion were delivered verbally by the researcher during Study 1. IRAP instructions were not delivered again for the purposes of Study 2. However, participants were still provided with a practice phase prior to the test phase of the IRAP in Study 2. Specifically, participants received a minimum of one and a maximum of four pairs of practice blocks, with 24 trials per block presented in random order. The same accuracy (i.e.,  $\geq 80\%$ ) and mean latency (i.e.,  $\leq 2000\text{ms}$ ) criteria applied for progression to the test phase, which included exactly six test blocks. If an incorrect response was emitted on either (the practice or test) task, participants were presented with a red ‘X’ which remained on-screen until the correct response was emitted. If a correct response was emitted, the screen was cleared for 400ms before the presentation of the next trial. If participants failed to

---

<sup>12</sup> According to Westermann et al. (2012), these items were found to be the most sensitive of the 18 items contained within the Paranoia Checklist to a threat-induction task (i.e., social stress) in previous research (Kesting et al., 2013).

respond within 2000ms, the message “Too slow!” appeared on-screen until they emitted a response. Accuracy and latency feedback were presented at the end of each block of trials. Only data from the test blocks were included in subsequent analyses.

IRAP block order (i.e., consistent- or inconsistent-first) was counter-balanced across participants so that participants who completed the IRAP used in Study 1 with consistent-first then completed the IRAP in Study 2 with consistent-first, and the same for inconsistent-first. If the participant failed to achieve criteria during the practice phase, their participation was ended at this point and they were thanked and debriefed.

**Unsolvable Tasks.** Participants were presented with four unsolvable tasks, which they were told were “concept learning” tasks. Each task consisted of 10 trials. On each trial, two different geometric figures were presented on either side of the computer screen. The geometric figures consisted of four dimensions each of which had two associated values: 1. Color (Red or Black); 2. Shape (Square or Circle); 3. Letter (A or T); 4. Letter size (Large or Small). Participants were instructed that one value of one of the dimensions (e.g., Red) would always be correct and that their task was to identify this consistently correct value out of the eight possibilities within 10 trials. For each of the 10 trials in a task, participants were asked to select the figure that included the correct value. In fact, no such correct value existed and thus in this sense the tasks were ‘unsolvable’. The figures remained on-screen until the participant made a response. The next trial appeared after an inter-trial interval of 300ms. At the end of the 10<sup>th</sup> trial, participants were asked to indicate which one of the eight values they thought was the correct value for that task.

### **3.2.3. General Procedure**

Participants completed all aspects of Study 2 in the following order: The Paranoia Checklist (see Study 1); Ellet and Chadwick’s (2007) unsolvable tasks combined with the

state-adapted Paranoia Checklist (after each task) under one of four conditions; the IRAP; and the state-adapted RSES.

The threat induction procedure was described to participants as four “concept-learning” tasks and they were required to complete the state-adapted Paranoia Checklist after each task. The threat-induction procedure here was a modified version of the one used by Ellet and Chadwick (2007, Experiment 1), in which the presence of the experimenter was not manipulated (their findings indicated that this did not impact state paranoia). Participants were randomly allocated to one of four conditions: 1. Low self-awareness (LSA) with No (task) feedback (control condition); 2. LSA with Failure feedback; 3. High self-awareness (HSA) with No feedback; and 4. HSA with Failure feedback. Within the HSA conditions, a camera pointed directly toward participants during the unsolvable tasks, such that they could view themselves on a monitor placed in front of them next to the computer screen. No comment was made about the presence of the camera and participants did not ask questions about it. The camera was not present within the LSA conditions. For the Failure feedback conditions, participants were presented with a feedback message that said either “Correct” or “Wrong” after each trial within the unsolvable tasks. As previously described, no correct value existed – instead, for each of the 10-trial problems, 5 random correct responses and 5 random incorrect responses were generated. Participants were then asked which of the eight values they thought was correct for that task, and were subsequently presented with the message “That is the wrong answer”, regardless of which value they selected. Participants in the No feedback conditions did not receive any feedback for their responses. Once they finished the four tasks, participants then completed the IRAP and state-adapted RSES. They were then thanked and debriefed.

### 3.2.4. Statistical analysis plan

A one-way ANOVA checked for potential IRAP block order effects (i.e., consistent- vs. inconsistent-first) on IRAP trial-types. A mixed factorial ANOVA was conducted to examine the effects of Self-awareness and Task feedback on state-adapted Paranoia Checklist scores across time. Two 2 X 2 ANOVAs were conducted to examine the effects of Self-awareness and Task feedback on state-adapted Paranoia Checklist scores at Time 4 and RSES scores. A two-way between-groups MANOVA examined the effects of Self-awareness and Task feedback on D-IRAP scores for each trial-type. Correlational analyses were conducted between Paranoia Checklist total and subscales, the state-adapted Paranoia Checklist at Time 4, RSES, and IRAP trial-type scores. Bonferroni correction for multiple comparisons ( $\alpha = .05 / 10 \approx .005$ ) was used.

## 3.3. Results

### 3.3.1. Data preparation

For the purposes of statistical analysis (and to mitigate effects of attrition), the IRAP accuracy criterion was adjusted to  $\geq 71\%$ <sup>13</sup> (the latency criterion remained at  $\leq 2000$ ms). Two participants withdrew following consent, one did not understand the “concept learning” task and was removed, and two were removed due to technical issues with the camera. Six participants failed to meet accuracy criteria on the IRAP practice phase and seven failed to maintain accuracy criteria during the test phase. Hence, 66 participants were included in the data analysis: 16 in the LSA + No feedback condition (3 males, 12 females, 1 unspecified; *Mean age* = 20.03; *SD* = 2.43), 17 in the LSA + Failure feedback condition (4 males, 13 females; *Mean age* = 22.09; *SD* = 5.03), 16 in the HSA + No feedback condition (4 males, 11 females, 1 unspecified; *Mean age* = 20.69; *SD* = 5.23), and 17 in the HSA + Failure feedback condition (5 males, 12 females; *Mean age* = 21.27; *SD* = 6.76).

---

<sup>13</sup> Four participants scored  $\geq 71\%$  accuracy on one block during the test phase and were included in the final analysis. Participants who scored  $< 71\%$  during the test phase were removed from subsequent analyses.

The data were analyzed using SPSS Statistics 21. Kolmogorov-Smirnov tests indicated no violations of normality assumptions for the RSES or IRAP trial-types. However, normality assumptions were violated for the Paranoia Checklist ( $p \leq .007$ ), with histograms indicating positive skew (i.e., most participants demonstrated low paranoia total and subscale scores; see Study 1). However, skewness and kurtosis were within the acceptable range of  $\pm 2$  for paranoia total, Frequency and Distress subscale scores<sup>14</sup>. Means and SDs for Paranoia Checklist total and subscale scores, state paranoia measured after each of the four unsolvable tasks, and RSES scores for the sample are presented in Table 3.1. A one-way ANOVA indicated no between-groups differences on the Paranoia Checklist total or subscale scores (all  $ps > .3$ ). There were no significant age or gender differences between groups either.

### **3.3.2. Effects of the threat-induction task on state paranoia across time**

The effect of the threat-induction task on state paranoia was examined using a 2 X 2 X 4 mixed factorial ANOVA, with Self-awareness (HSA & LSA) and Task feedback (Failure feedback & No feedback) as the between-group factors and Time (4 time points) as the within-group factor. Mauchly's Test of Sphericity indicated that this assumption had been violated so degrees of freedom were corrected for with Greenhouse-Geisser estimates of Sphericity ( $\epsilon = .676$ ). There was a significant main effect for Time,  $F(2.028, 125.753) = 4.530, p = .012, \eta^2 = .068$ . There was a significant linear effect for Time,  $F(1, 62) = 6.065, p = .017, \eta^2 = .089$ , indicating that state paranoia increased across the four time points for all participants. There were no significant main effects for Self-awareness,  $F(1, 62) = .260, p = .612, \eta^2 = .004$ , or Task feedback,  $F(1, 62) = .683, p = .412, \eta^2 = .011$ , nor a Self-awareness X Task feedback interaction effect,  $F(1, 62) = .557, p = .458, \eta^2 = .009$ , for state paranoia across time. Thus, we did not examine whether the effects of Self-awareness and/or Task

---

<sup>14</sup> As per Study 1, Kurtosis was 4.83 for Conviction and 2.40 for Distress scores.

feedback on state paranoia were moderated by ‘trait’ paranoia (i.e., Paranoia Checklist total and subscale scores).

### **3.3.3. Effects of the threat-induction task on state paranoia at Time 4 and RSES**

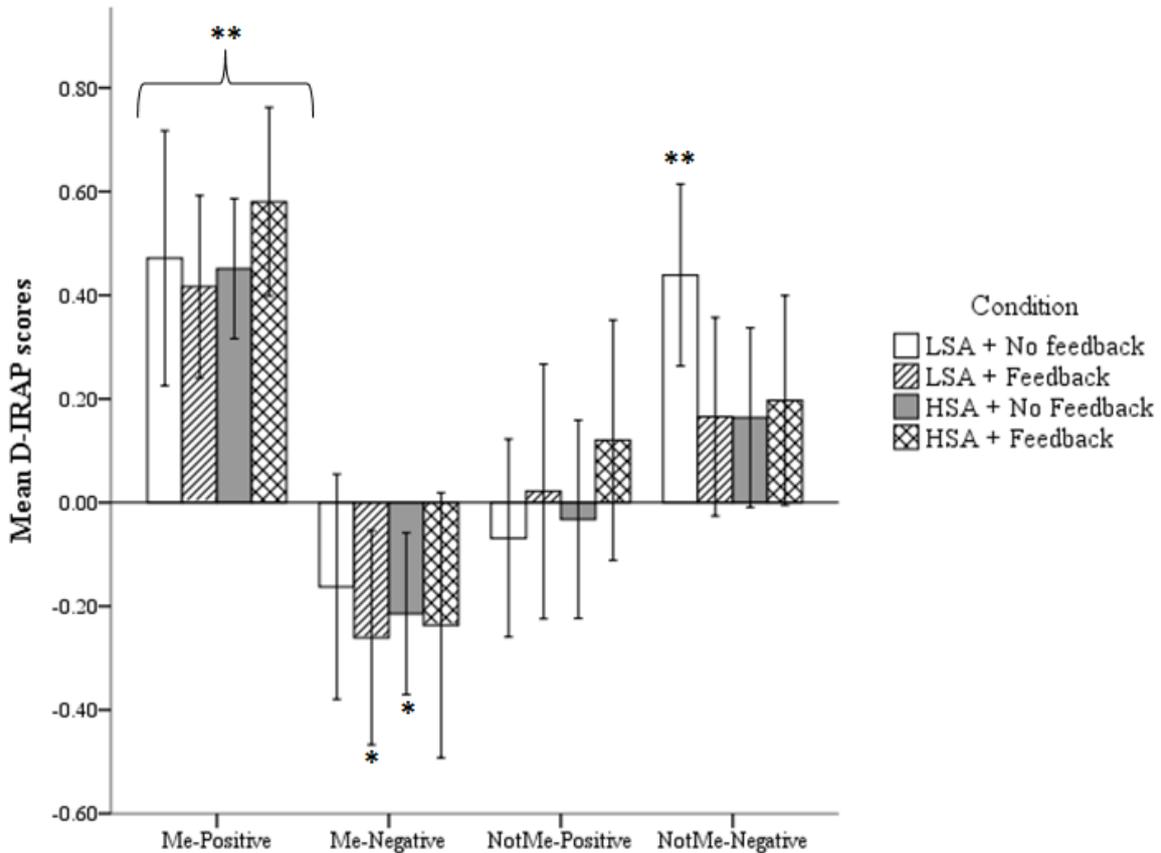
The effects of the threat-induction task on state paranoia at Time 4 (as measured by the state-adapted Paranoia Checklist) and RSES scores were analyzed using 2 (Self-awareness: HSA & LSA) X 2 (Task feedback: Failure feedback & No feedback) ANOVAs. Levene’s tests showed no violation of normality assumptions. The results indicated that there were no main effects for Self-awareness,  $F(1, 65) = .109, p = .742, \eta^2 = .002$ , or Task feedback,  $F(1, 65) = 1.027, p = .315, \eta^2 = .016$ , or interaction effect,  $F(1, 65) = .603, p = .441, \eta^2 = .010$ , for state paranoia at Time 4. Similarly, there were no main effects for Self-awareness,  $F(1, 62) = 1.952, p = .167, \eta^2 = .031$ , or Task feedback,  $F(1, 62) = 1.150, p = .288, \eta^2 = .018$ , or interaction effect,  $F(1, 62) = 1.144, p = .289, \eta^2 = .018$ , for RSES scores.

**Table 3.1.** Means and SDs for Paranoia Checklist total and subscales, state paranoia across the four time points, and RSES scores across the four threat-induction task conditions.

<b>Condition</b>	<b>Paranoia Frequency</b>	<b>Paranoia Conviction</b>	<b>Paranoia Distress</b>	<b>Paranoia Total</b>	<b>T1 Paranoia</b>	<b>T2 Paranoia</b>	<b>T3 Paranoia</b>	<b>T4 Paranoia</b>	<b>RSES</b>
<b>LSA + No feedback</b>	27.88 (10.14)	32.63 (12.25)	33.69 (16.46)	94.19 (31.77)	8.18 (2.99)	8.63 (3.28)	8.75 (2.91)	9.19 (3.92)	16.94 (4.65)
<b>LSA + Feedback</b>	27.71 (8.09)	30.94 (7.49)	31.18 (8.67)	89.82 (22.64)	8.53 (2.07)	8.41 (2.15)	8.59 (2.18)	9.00 (2.21)	16.94 (5.92)
<b>HSA + No feedback</b>	31.13 (10.18)	34.31 (11.31)	36.00 (13.88)	101.44 (30.30)	9.31 (3.91)	9.38 (3.38)	9.44 (3.39)	10.06 (3.82)	17.38 (5.24)
<b>HSA + Feedback</b>	25.82 (5.58)	30.18 (5.63)	34.65 (11.06)	90.65 (15.12)	8.41 (2.15)	8.29 (2.37)	8.53 (2.53)	8.65 (2.64)	20.24 (5.73)

### 3.3.4. Effects of the threat-induction task on relational responding on the IRAP

Responding on IRAP trial-types for each of the four threat-induction task conditions are presented in Figure 3.1. One-sample t-tests indicated that all participants demonstrated significantly faster ‘Me-Positive’ agreement (relative to disagreement) on this trial-type, all  $ps \leq .001$ . Participants in the HSA + No feedback ( $p = .010$ ) and LSA + Failure feedback ( $p = .017$ ) conditions demonstrated significantly faster ‘Me-Negative’ agreement, with participants responding more quickly on inconsistent (i.e. agreeing “I am-Negative”) relative to consistent blocks of trials. This trended towards significance for the HSA + Failure feedback condition ( $p = .067$ ). This was non-significant for the LSA + No feedback condition ( $p = .132$ ). Only participants in the LSA + No feedback condition demonstrated ‘Not me-Negative’ agreement on this trial-type ( $p \leq .001$ ), although this trended towards significance for the HSA + Failure feedback ( $p = .055$ ) and HSA + No feedback conditions ( $p = .062$ ). No effects were observed on the ‘Not me-Positive’ trial-type. A two-way between-groups MANOVA was performed to investigate the effects of Self-awareness and Task feedback on IRAP responding across the four trial-types. No violations of the assumptions of normality were observed. There were no significant effects for Self-awareness (all  $ps > .171$ ), or Task feedback, (all  $ps > .177$ ), or interaction effects on IRAP trial-types (all  $ps > .086$ ).



\* Sig. at  $\alpha = .05$  \*\* Sig. at  $\alpha \leq .001$

**Figure 3.1.** Mean D-IRAP scores for each IRAP trial-type by experimental condition, with standard error bars (95% confidence intervals): LSA + No Feedback; LSA + Failure feedback; HSA + No feedback; HSA + Failure feedback. Note, D-IRAP scores above zero indicate faster responding on consistent (i.e., faster ‘Me-Positive’ agreement and ‘Me-Negative’ disagreement) relative to inconsistent blocks of trials, whereas scores below zero indicate the opposite pattern of responding (i.e., faster responding on inconsistent trials). Asterisks denote trial-types significantly different from zero.

### 3.3.5. Correlational analyses

A correlation matrix was calculated using Spearman’s Rho to examine associations between Paranoia Checklist total and subscale scores, state-adapted Paranoia Checklist scores at Time 4, RSES scores, and IRAP trial-types (with Bonferroni correction,  $\alpha = .005$ ; see Table 3.2). Significant positive correlations were observed between Paranoia Checklist total and subscale scores and state paranoia at Time 4, with higher state paranoia scores at Time 4 associated with higher Paranoia Checklist total, Frequency, Conviction and Distress scores at

baseline. Significant negative correlations were observed between the RSES and Paranoia Checklist total and subscale scores and state-adapted Paranoia Checklist scores at Time 4, with lower RSES scores associated with higher state paranoia scores at Time 4 and higher Paranoia Checklist total, Frequency, Conviction and Distress scores at baseline. No significant correlations were observed between any of the self-report measures and the IRAP trial-types<sup>15</sup>.

**Table 3.2.** Correlations between Paranoia Checklist total and subscales, state paranoia at Time 4, RSES, and IRAP trial-type scores ( $N = 66$ ).

	<b>State Paranoia</b>	<b>RSES</b>	<b>Me-Positive</b>	<b>Me-Negative</b>	<b>Not me-Positive</b>	<b>Not me-Negative</b>
<b>Paranoia Checklist</b>	.591**	-.513**	-.173	-.150	.047	-.168
<b>Paranoia Frequency</b>	.581**	-.495**	-.002	-.038	.070	-.096
<b>Paranoia Conviction</b>	.591**	-.441**	-.083	-.164	-.035	-.128
<b>Paranoia Distress</b>	.421**	-.355**	-.209	-.158	-.006	-.172
<b>State Paranoia</b>		-.422**	-.023	-.116	-.040	-.147
<b>RSES</b>			.200	-.024	-.148	.080

\* Sig. at  $\alpha = .05$  \*\* Sig. at  $\alpha \leq .001$

<sup>15</sup> We also examined correlations between Paranoia Checklist total and subscale scores and change scores for the state-adapted Paranoia Checklist, RSES and IRAP trial-types (calculated for the RSES and IRAP by subtracting ‘baseline’ scores from Study 1 from post-threat-induction tasks scores in Study 2, and state-adapted Paranoia Checklist scores at Time 1 from scores at Time 4 in Study 2). Changes in state paranoia scores were negatively associated with changes in RSES scores,  $r = -.272$ ,  $N = 65$ ,  $p = .028$ , with larger state paranoia change scores associated with smaller RSES change scores. RSES change scores were not associated with Paranoia Checklist total and subscale scores. Change scores for ‘Not me-Negative’ responding on the IRAP were positively associated with Paranoia Checklist total and Frequency scores, with larger ‘Not me-Negative’ change scores associated with higher Paranoia Checklist total scores ( $r = .251$ ,  $N = 65$ ,  $p = .044$ ) and more frequent paranoia ( $r = .255$ ,  $N = 65$ ,  $p = .040$ ) at baseline. Also, changes in ‘Not me-Negative’ responding was positively associated with changes in ‘Me-Negative’ responding on the IRAP,  $r = .319$ ,  $N = 65$ ,  $p = .010$ . Based on the direction of responding (i.e., stronger agreement or disagreement) on these IRAP trial-types in Studies 1 and 2, this suggests that increased ‘Me-Negative’ agreement was associated with reduced ‘Not me-Negative’ agreement. However, these associations were non-significant.

### **3.4. Discussion**

The purpose of Study 2 was to investigate the influence of environmental factors (i.e., a threat-induction task) on (non-clinical) paranoia, self-reported self-esteem, and relational responding to the self as ‘positive’ and ‘negative’ on the IRAP as well as the interaction of these behavior-behavior relations within this context. The results provide preliminary support for the hypothesis that relational responding regarding paranoia can be influenced via an adverse environmental event. Similar to the findings of Ellet and Chadwick (2007), we observed a linear increase in state paranoia across the four unsolvable tasks; however, there were no specific effects for self-awareness or task failure on state paranoia across time or at Time 4. The observed lack of effects for self-awareness in relation to paranoia is somewhat surprising given Ellet and Chadwick’s findings. One reason for this may be procedural differences between the two studies. In Ellet and Chadwick’s study, participants were aware that the study explored the possible induction of paranoid thoughts. In Study 2, participants were not informed of the purpose of the study to avoid potential demand characteristics. Indeed, during debriefing participants reported that they were not aware of the aim of the study or that it was about paranoia.

There were also differences in measurement of paranoia between the studies. Ellet and Chadwick used the Paranoia Scale (PS; Fenigstein & Venable, 1992) to compare paranoia scores across conditions and the Paranoia and Depression Scale (PDS; Bodner & Mikulincer, 1998) to examine changes in state paranoia across time. The effect of HSA on paranoia was observed on the PS in their study. The PS has a broad conception of paranoia and many items are not overtly persecutory (e.g., ‘My parents and family find more fault with me than they should’, ‘No one really cares much what happens to you’, ‘People often disappoint me’). In Study 2, the six-item state-adapted Paranoia Checklist was used, which comprises items that capture commonplace paranoid thoughts and thoughts of a more

clinically-relevant nature. The observed increases in paranoia across the unsolvable tasks in Study 2 demonstrates that the state-adapted Paranoia Checklist is capable of capturing changes in state paranoia. It remains unclear why between-groups differences were not observed however. Although participants demonstrated increased state paranoia following the threat-induction task, it is not clear what form of perceived threat was prompted by the task. The lack of a manipulation check was an omission and without this it is not possible to ascertain whether the HSA conditions increased participants' self-focused attention or whether participants in the feedback conditions attributed failure feedback to themselves or to (potential issues with) the task.

No significant main effects for self-awareness or task failure were observed in relation to RSES scores following the threat-induction task. However, we observed a significant negative correlation between RSES and state paranoia scores (at Time 4) as well as a negative correlation between change scores on both measures (although this was non-significant), suggesting that these patterns of relational responding (in terms of paranoia and self-esteem) were associated with each other in this context. Specifically, lower self-reported self-esteem scores were associated with higher state paranoia scores following the threat-induction task; however, greater changes in state paranoia scores across the unsolvable tasks (from Time 1 to Time 4) were associated with smaller changes in self-reported self-esteem scores (from 'baseline' in Study 1 to post-threat-induction in Study 2).

Similarly, there were no significant effects for self-awareness or task failure on the IRAP. Still, an interesting pattern of responding was observed on the IRAP. As hypothesized, all participants demonstrated significant 'Me-Positive' agreement. However, participants in the control condition (i.e., LSA + No task feedback) demonstrated a different pattern of relational responding to the self in terms of negative self-evaluations (i.e., 'Me-Negative' and 'Not me-Negative') compared to participants in the other task conditions. Specifically,

participants in the HSA conditions and LSA + Failure feedback condition demonstrated stronger agreement with ‘Me-Negative’ (relative to ‘Me-Negative’ disagreement) on this trial-type. This pattern of responding could be considered consistent or ‘coherent’ within this context. That is, relational responding to the self as ‘negative’ ‘makes sense’ following repeated exposure to unsolvable tasks under conditions of high self-awareness and/or failure feedback. Indeed, participants in the control condition did not demonstrate ‘Me-Negative’ agreement and were the only group to demonstrate significant ‘Not me-Negative’ agreement (relative to disagreement) on this trial-type.

There were no significant associations between state paranoia and relational responding on the IRAP. Interestingly, changes in ‘Not me-Negative’ responding (from Study 1 to Study 2) were positively associated with ‘trait’ paranoia and the frequency associated with experiencing paranoid thoughts on the Paranoia Checklist (although not significantly so), with greater changes in ‘Not me-Negative’ responding associated with higher Checklist total and Frequency scores at baseline. Furthermore, changes in ‘Not me-Negative’ responding were associated with changes in ‘Me-Negative’ responding. When relational responding on these two IRAP trial-types were compared between Studies 1 and 2, participants in the HSA conditions and LSA + Failure feedback condition demonstrated increased ‘Me-Negative’ agreement and reduced strength of ‘Not me-Negative’ agreement from Study 1 to Study 2 (i.e., post-threat-induction) and these changes were associated with each other (although not significantly so).

Taken together, the findings from Study 2 support the assertion that the IRAP may be able to capture subtle differences in relational responding to the self (e.g., as ‘negative’) across groups or conditions and may also capture changes in relational responding following intervention (e.g., by comparing responses on the ‘Me-Negative’ and ‘Not me-Negative’ trial-type between Studies 1 and 2). While there were no significant between-groups

differences on the IRAP, the observed patterns of responding suggest that positive self-regard may be maintained following an adverse event whereas negative self-evaluations may be affected, increased or strengthened following such experiences, consistent with previous findings (e.g., Kesting & Lincoln, 2013). Hence, the IRAP appeared to capture seemingly contrasting but coexisting patterns of responses to the self (as both ‘positive’ and ‘negative’).

In summary, while the threat-induction task used in Study 2 did lead to increases in state paranoia, our findings did not replicate those of Ellet and Chadwick (2007). The lack of specific effects for self-awareness (or task feedback) suggests that this may not be the most suitable paradigm for investigating the effects of adverse events on paranoia and related self-evaluations. The absence of a manipulation check is unfortunate and without this we cannot be certain that the features of the task (e.g., HSA, failure feedback) were interpreted by participants as intended. Furthermore, the lack of correlations between state paranoia and relational responding on the IRAP means that the relationship between paranoia and specific relational responses to the self in this context is also unclear. Paranoia is essentially a perception of threat and *interpersonal* threat more specifically. Accordingly, paradigms involving adverse interpersonal experiences may be a more suitable alternative for exploring the impact of aversive events on paranoia, relational responding to the self, and their interaction.

## **Chapter 4: The effects of social exclusion on state paranoia, self-reported self-esteem, and relational responding to the self on the IRAP**

### **4.1. Introduction**

The results from Study 2 support the assertion that, as an operant, arbitrarily applicable relational responding (AARR) is indeed sensitive to environmental contingencies, and that both self-report measures and the Implicit Relational Assessment Procedure (IRAP) can capture differences or changes in AARR following manipulation of contextual features. Specifically, the findings from Study 2 showed that an adverse environmental event (i.e., a threat-induction task) can influence patterns of relational responding regarding momentary or ‘state’ paranoia. In addition, differential patterns of brief and immediate relational responding (BIRR) to the self as ‘negative’ were observed on the IRAP in relation to the different arrangements of environmental contingencies (i.e., task completion under conditions of high or low self-awareness and no task feedback or failure feedback). Responding on the IRAP indicated that ‘Me-Positive’ responding may be unaffected by such adverse events; however, participants who completed unsolvable tasks under conditions of high self-awareness or low self-awareness and failure feedback showed stronger ‘Me-Negative’ agreement and reduced ‘Not me-Negative’ agreement compared to controls, who demonstrated no effects for ‘Me-Negative’ responding and significant ‘Not me-Negative’ agreement. However, as no significant between-groups differences were observed in relation to state paranoia, the threat-induction task used in Study 2 may lack the precision needed to predict-and-influence paranoia for the purposes of this research. Hence, Study 3 sought to examine the effect of an alternative threat-induction task on state paranoia, self-reported self-esteem, and relational responding to the self (and others) as ‘positive’ and ‘negative’ on the IRAP in participants from the general population.

Paranoia is essentially a perception of *interpersonal* threat (see Freeman & Garety, 2000). Furthermore, Bentall and Fernyhough (2008) have argued that a core process in paranoid belief systems is the expectation that negative interpersonal interactions will be experienced in future. These expectations and sensitivities may emerge through a history of adverse experiences (e.g., neglect, bullying, childhood trauma; see Bentall et al., 2014; Murphy et al., 2015; Selten, van der Ven, Rutten, & Cantor-Graae, 2013; Valmaggia et al., 2015). From a Relational Frame Theory (RFT) perspective, AARR may explain how such paranoid ‘belief systems’ (or ‘relational networks’) develop from prior experiences. For example, following an adverse interpersonal event (e.g., being betrayed by someone), AARR can transform the functions of interpersonal stimuli (e.g., certain social stimuli and events may be characterized by aversive functions) and can put the individual into contact with hypothetical future events (e.g., future betrayal via temporal relations), such that, although this event may never happen, the individual may still respond to it. For example, they may experience distress via AARR and may come to fear and avoid certain (social) stimuli. Hence, threat-induction tasks that involve aversive interpersonal events may be more appropriate for examining the influence of particular environmental factors on paranoia, relational responding to the self and others, and their interaction in experimental settings.

Recently, a number of researchers have investigated the relationship between paranoia and self-esteem in non-clinical samples by examining the putative role of social exclusion induced through a virtual ball toss game called ‘Cyberball’ (Williams, Cheung, & Choi, 2000), in which participants are either included or excluded by other (computer-generated) ‘players’ (e.g., Westermann et al., 2012; Kesting et al., 2013; Lincoln et al., 2014). For example, Kesting et al. (2013) found that excluded participants reported an increase in (non-clinical) state paranoia, which was mediated by a decrease in self-esteem and moderated by paranoia-proneness (i.e., ‘trait’ paranoia measured by the Paranoia Checklist). These findings

suggest that the experimental induction of social stress, as in Cyberball, may be useful in studying possible relationships among aversive interpersonal experiences, paranoia, and relational responding to the self and others by allowing us to investigate the potential causal role of an adverse social event on these patterns of relational responding and the relationships between them.

#### **4.1.1. Hypotheses**

The aim of Study 3 was to investigate the influence of an adverse *interpersonal* event (i.e., social exclusion) on state paranoia, self-reported self-esteem, and relational responding to the self and others as ‘positive’ and ‘negative’ on the IRAP. Based on the findings reported by Kesting et al. (2013), we hypothesized that exclusion would lead to increased state paranoia and that the effect of exclusion on state paranoia would be moderated by trait paranoia. Specifically, we assumed that individuals’ learning histories would influence responding in the current context such that those with a history of responding to adverse interpersonal events in a paranoid manner would be more likely to do so following exclusion than individuals without such histories. It was also hypothesized that exclusion would be associated with reduced self-reported self-esteem and that changes in self-reported self-esteem following exclusion would be associated with changes in state paranoia. On the IRAP, it was hypothesized that exclusion would be associated with increased ‘Me-Negative’ and ‘Others-Negative’ agreement, as these patterns of responding to the self and others as ‘negative’ were deemed consistent or coherent in this context (i.e., following exclusion by others). In light of the finding from Study 2 that adverse experiences did not impact ‘Me-Positive’ responding on the IRAP and evidence that positive self-evaluations may be maintained in paranoia (Kesting & Lincoln, 2013), we did not expect exclusion to influence ‘Me-Positive’ responding.

## **4.2. Method**

### **4.2.1. Participants**

Ninety-seven volunteers (31 males, 63 females, 1 intersex, 2 non-binary; *Mean age* = 23.03, *SD* = 8.83) participated. Most were NUI Galway undergraduate Psychology students, who took part for course credit. Other NUI Galway students were recruited via advertisements. Forty-eight participants were randomly allocated to the ‘Exclusion’ (experimental) condition and 49 to the ‘Inclusion’ (control) condition.

### **4.2.2. Social exclusion using Cyberball**

Cyberball (Williams et al., 2000) is a virtual ball toss game that has increasingly been used to examine social exclusion. Participants were told they would be involved in an ‘online game’ with other ‘players’, although these were actually computer-generated avatars (one male, called Shane, and one female, called Katie). In the Exclusion condition, participants received the ball twice from the other players at the start of the game, but did not receive the ball again. In the Inclusion condition, participants received the ball an equal number of times to the other players. Each condition involved 30 ball tosses among three players (including the participant). In both conditions, the participant was free to decide which player to throw the ball to. The version of Cyberball used in this study (4.0) included features intended to increase the credibility of the paradigm, including names and photographs of the other players and a ‘chat box’, which participants could type into to ‘chat’ to the other ‘players’ (one of the other players, Shane, typed “Hi!” after the game began but no other messages were sent by the other ‘players’ for the rest of the game).

### **4.2.3. Measures**

**Paranoia Checklist.** Trait paranoia was assessed using the Paranoia Checklist (Freeman et al., 2005). The scale showed excellent internal reliability overall ( $\alpha = .942$ ), as

well as on the Frequency ( $\alpha = .866$ ), Conviction ( $\alpha = .858$ ) and Distress ( $\alpha = .895$ ) subscales in the current sample.

**State-adapted Paranoia Checklist.** Changes in *state* paranoia were assessed using the state-adapted version of the Paranoia Checklist (Westermann et al., 2012). The sum score was used as an index of state paranoia.

**State-adapted Rosenberg Self-Esteem Scale (RSES).** To assess *state* self-esteem before and after Cyberball, participants were asked to rate their agreement with the items “*at the moment*” on the 10-item RSES. The RSES showed excellent internal reliability in the current sample ( $\alpha = .886$ ).

**Cyberball Questionnaire.** The Cyberball Questionnaire (Williams, et al., 2000; Williams, 2009) was administered post-Cyberball to assess group differences (Inclusion and Exclusion) in need satisfaction (belonging, self-esteem, meaningful existence, and control) and mood (happiness, sadness, shame, anger, and fear) and as a manipulation check. Participants rated agreement with statements related to need satisfaction on a five-point Likert scale. The mood section of the questionnaire was adapted as per Kesting et al. (2013), so that emotions were rated on a 10-point scale (1 = “Does not apply”, 10 = “Applies strongly”) and an additional item (“I am frustrated”) was included. For the manipulation check, participants were asked to rate how “ignored” and “excluded” they felt during the game on the 5-point Likert scale. They were also asked “Assuming that the ball should be thrown to each person equally (33% per player), what percentage of throws did you receive?”

**The IRAP<sup>16</sup>.** The target stimuli were based on those used in Studies 1 and 2. However, the label stimuli presented in the IRAP in Study 3 were “I am” and “Others are” (see Table 4.1 for stimuli). Participants were required to respond according to the rule “I am-Positive, Others are-Negative” on consistent blocks, and according to the reverse rule (“I am-

---

<sup>16</sup> The Implicit Association Task (IAT) was also included in this study as a measure of ‘implicit self-esteem’ for the purposes of publication. The IAT data is not included in this chapter though as the focus of the thesis was on relational responding to the self as measured by the IRAP.

Negative, Others are-Positive”) on inconsistent blocks<sup>17</sup>. The positions of the response options, “True” and “False”, were counter-balanced across trials. Prior to each block of trials, participants were informed by a message on the screen, “During the next block the previous correct and wrong answers are reversed”.

Participants received between one and four pairs of practice blocks, each with 24 trials presented in quasi-random order. They had to meet criteria of  $\geq 80\%$  accuracy and mean latency  $\leq 2500$ ms to progress to the test phase, which included exactly six test blocks. If a correct response was emitted, the screen cleared for 400ms before presentation of the next trial. If an incorrect response was made, a red ‘X’ was presented and remained onscreen until the correct response was made. Accuracy and latency feedback were presented at the end of each block. Only data from the test blocks were included in subsequent analyses.

IRAP block order (i.e., consistent- or inconsistent-first) was counter-balanced across participants. Instructions were presented verbally by the researcher, who stayed with the participant during the practice phase and then moved behind a screen once they achieved criteria to proceed to the test phase. If the participant failed to achieve criteria during the practice phase, their participation was ended then and they were thanked and debriefed.

---

<sup>17</sup> There were some differences between the features of the IRAP used in Study 3 and the IRAPs used in the other studies within the thesis: 1. The IRAP used in Study 3 involved the 2012 version of the IRAP program whereas the other studies involved the 2009 program; 2. Participants were given explicit ‘rules’ for responding on the IRAP during the instructions phase of Study 3 but no rules for responding on the IRAP were given in the other studies; 3. The referents for the label stimuli (“I” and “Others”) were not specified in Study 3; 4. Latency criterion for the IRAP in Study 3 was  $\leq 2500$ ms but was  $\leq 2000$ ms for the other studies (although this was extended to  $\leq 2500$ ms for Study 4); and 5. The “Too Slow!” prompt was not presented if the participant failed to respond within 2500ms in Study 3. These variables were not introduced or tested in a systematic manner. Rather, they emerged “organically” during the research process. However, following advice from leading IRAP experts, it was decided that the 2009 IRAP program was the more suitable version, that no explicit rules for responding should be provided, that referents for the label stimuli (e.g., “I” and “Others”) be clarified explicitly for participants, and that the “Too Slow!” prompt be retained.

**Table 4.1.** IRAP stimuli for Study 3.

<b>Label stimuli</b>	
I am	Others are
<b>Target stimuli</b>	
<i>Positive</i>	<i>Negative</i>
Valuable	Helpless
Happy	Guilty
Tender	Desperate
Friendly	Sad
Hopeful	Rejected
Competent	Failed
<b>Response options</b>	
True	False

#### 4.2.4. General Procedure

At baseline, participants completed the IRAP, RSES, Paranoia Checklist, and state-adapted Paranoia Checklist in this order. They were then randomly allocated to either the Cyberball Exclusion or Inclusion condition. Immediately after Cyberball, they completed the IRAP again followed by the RSES, state-adapted Paranoia Checklist, and Cyberball Questionnaire. To help maintain the effect of Cyberball during completion of all subsequent measures, an A5 paper screenshot of the three avatars from Cyberball (with the ball being thrown between the two other ‘players’) was placed between the participant and keyboard for the remainder of the study. The researcher also prompted participants to mentally reflect on their experiences of the game between the measures.

#### 4.2.5. Statistical analysis plan

One-way ANOVAs checked for potential block order (i.e., consistent- vs. inconsistent-first) effects on IRAP trial-types. One-way ANOVAs checked for potential between-groups differences (Exclusion vs. Inclusion) on all variables at baseline and on the Cyberball Questionnaire post-Cyberball (as a manipulation check). Bonferroni correction for

multiple comparisons ( $\alpha = .05 / 11 \approx .0045$ ) was used. Effects of social exclusion on state paranoia, self-reported self-esteem, and relational responding on the IRAP from baseline to post-Cyberball were examined using mixed factorial ANOVAs with post-hoc tests. Cyberball condition (Inclusion vs. Exclusion) was the between-groups variable and time (baseline vs. post-Cyberball) was the within-group variable. Correlation analyses were conducted to examine associations between trait and state paranoia, self-reported self-esteem, and relational responding on the IRAP at baseline, and to examine associations between trait paranoia and state paranoia at post-Cyberball and changes in self-reported self-esteem and relational responding on the IRAP (from baseline to post-Cyberball) (Bonferroni correction was used for each correlational analysis,  $\alpha = .05 / 10 = .005$ ).

### 4.3. Results

#### 4.3.1. Data preparation

For purposes of data analysis, accuracy criterion was set at  $\geq 75\%$  (latency criterion remained at  $\leq 2500\text{ms}$ )<sup>18</sup>. IRAP data for five participants was lost due to technical problems, five participants failed to maintain IRAP accuracy criterion during the test phase, one withdrew from the study, and another was excluded due to familiarity with Cyberball. Hence, 85 participants were included in subsequent data analysis: 45 in the Inclusion condition (16 males, 28 females, 1 intersex; *Mean age* = 24.36, *SD* = 10.51) and 40 in the Exclusion condition (12 males, 27 females, 1 non-binary; *Mean age* = 21.23, *SD* = 5.27). No significant age or gender differences were observed between the Inclusion and Exclusion groups. Kolmogorov-Smirnov tests indicated no violations of normality assumptions for the RSES. However, normality assumptions were violated for the Paranoia Checklist and state-adapted Paranoia Checklist ( $p \leq .001$ ). Inspection of histograms indicated positive skew on both measures, with most participants showing low trait and state paranoia, which was expected

---

<sup>18</sup> 10 participants scored  $\geq 75\%$  accuracy on at least one block during the test phase and were included in the final analysis (participants scoring  $< 75\%$  accuracy during the test phase were excluded from subsequent analyses).

given that the sample was recruited from the general, non-clinical population. Nonetheless, skew in the case of state paranoia data was considered sufficiently strong ( $>2$ ) to warrant log-transformation.

#### 4.3.2. Sample characteristics at baseline

One-way between-groups ANOVAs (with Bonferroni correction,  $\alpha = .005$ ) indicated no significant between-groups differences on any of the measures at baseline (all  $ps \geq .046$ ). Correlational analyses (Bonferroni correction,  $\alpha = .005$ ) assessed the relationships among the measures at baseline (Table 4.2). Significant positive correlations were observed between the Paranoia Checklist total and subscale scores and state paranoia (all  $ps \leq .001$ ). Significant negative correlations were observed between RSES scores and Paranoia Checklist total, Frequency, Conviction, and Distress subscales, and state paranoia, with higher RSES scores associated with lower paranoia scores. No significant correlations were observed between the IRAP trial-types and self-report measures. However, non-significant positive correlations were observed between the ‘Others-Negative’ trial-type and the Frequency ( $r = .226, p = .037$ ) and Conviction ( $r = .226, p = .038$ ) subscales of the Paranoia Checklist.

**Table 4.2.** Correlations between measures of trait and state paranoia, RSES scores, and IRAP trial-types at baseline ( $N = 85$ ).

	<b>RSES</b>	<b>Me-Positive</b>	<b>Me-Negative</b>	<b>Others-Positive</b>	<b>Others-Negative</b>
<b>Paranoia Total</b>	-.590**	.139	-.135	.043	.175
<b>Paranoia Frequency</b>	-.560**	-.052	-.118	.086	.226*
<b>Paranoia Conviction</b>	-.481**	.144	-.133	.007	.226*
<b>Paranoia Distress</b>	-.517**	.163	-.110	.023	.042
<b>State Paranoia</b>	-.410**	.176	-.133	-.071	.018

\*  $p < .05$  \*\*  $p \leq .001$  (prior to Bonferroni correction)

### **4.3.3. Manipulation check**

One-way between-groups ANOVAs assessed whether exclusion during Cyberball had the intended negative impact on mood and related variables. The Exclusion group showed less belonging, self-esteem, meaningful existence, and control compared to the Inclusion group on the Cyberball Questionnaire (all  $ps < .001$ ). Excluded participants also showed greater shame, sadness, anger, and frustration and less happiness than included participants (all  $ps < .001$ ), and the difference in fear approached significance ( $p = .013$ ). The Exclusion group also reported greater feelings of being ignored and excluded, as well as receiving fewer ball tosses ( $M = 7.48$ ) than the Inclusion group ( $M = 28.09$ ; all  $ps < .001$ ).

### **4.3.4. Effects of exclusion on state paranoia**

A mixed factorial ANOVA explored between-groups differences in (log-transformed) state-adapted Paranoia Checklist scores from baseline to post-Cyberball. Levene's test of equality of error variances indicated violation of normality assumptions, so results should be interpreted with caution. A significant group X time interaction effect was observed,  $F(1,83) = 5.297, p = .024, \eta^2 = .06$ . Effects trending towards significance were observed for group,  $F(1,83) = 3.484, p = .065, \eta^2 = .04$ , and time,  $F(1,83) = 3.131, p = .081, \eta^2 = .036$ . Post-hoc comparisons indicated a significant increase in state paranoia from baseline ( $M = .93, SD = .14$ ) to post-Cyberball ( $M = .96, SD = .15$ ) for the Exclusion group only ( $p = .035$ ), and a significant between-groups difference at post-Cyberball ( $p = .024$ ).

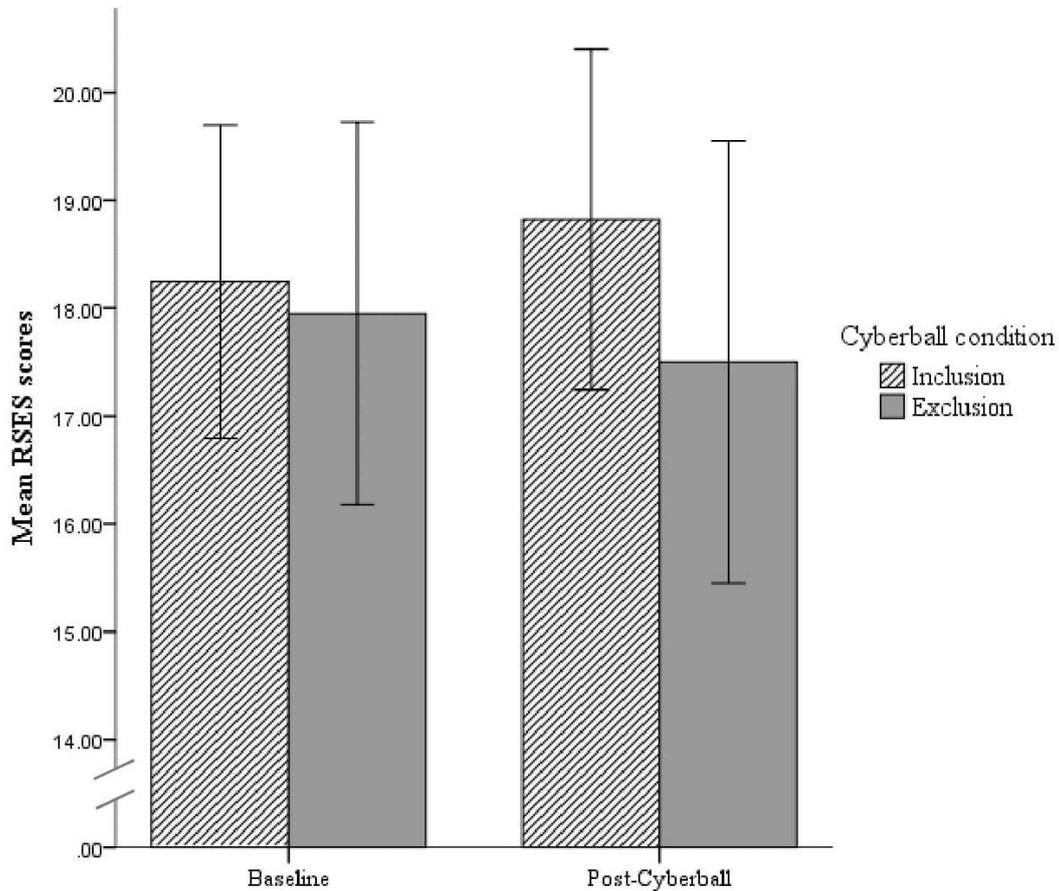
### **4.3.5. Moderating effect of trait paranoia on changes in state paranoia post-Cyberball**

To investigate whether the effect of Cyberball condition (Inclusion or Exclusion) on state paranoia was moderated by proneness to (different dimensions of) paranoia, we performed three moderation analyses using PROCESS for SPSS (Hayes, 2016) with Cyberball condition, mean Frequency/ Distress/ Conviction score on the Paranoia Checklist at baseline and the interaction term of Cyberball condition X Frequency/ Distress/ Conviction

as predictors and state paranoia at post-Cyberball as the dependent variable. State paranoia at baseline was included as a covariate. A significant Cyberball condition X Distress ( $p = .028$ ) moderation effect was found and was probed using the Johnson-Neyman technique to identify regions of significance. Results revealed a significant positive effect for exclusion on state paranoia scores at post-Cyberball for those in the Exclusion group with a Distress score of  $\geq 31$ , which was increasingly strong for higher Distress scores (i.e., participants with increasingly higher Distress scores at  $\geq 31$  showed increasingly higher state paranoia scores following exclusion). This applied to 55% of the sample. Thus, exclusion during Cyberball led to an increase in state paranoia in participants more distressed by paranoid thoughts at baseline.

#### **4.3.6. Effects of exclusion on explicit self-esteem**

A mixed factorial ANOVA explored between-groups differences in RSES scores from baseline to post-Cyberball. Levene's test showed no violation of normality assumptions. A significant group X time interaction effect was observed,  $F(1,83) = 4.750, p = .032, \eta^2 = .054$ . The Inclusion group showed an increase in RSES score from baseline ( $M = 18.24, SD = 4.82$ ) to post-Cyberball ( $M = 18.82, SD = 5.25$ ), while the Exclusion group showed a decrease in RSES score from baseline ( $M = 17.95, SD = 5.54$ ) to post-Cyberball ( $M = 17.5, SD = 6.41$ ) (see Figure 4.1).



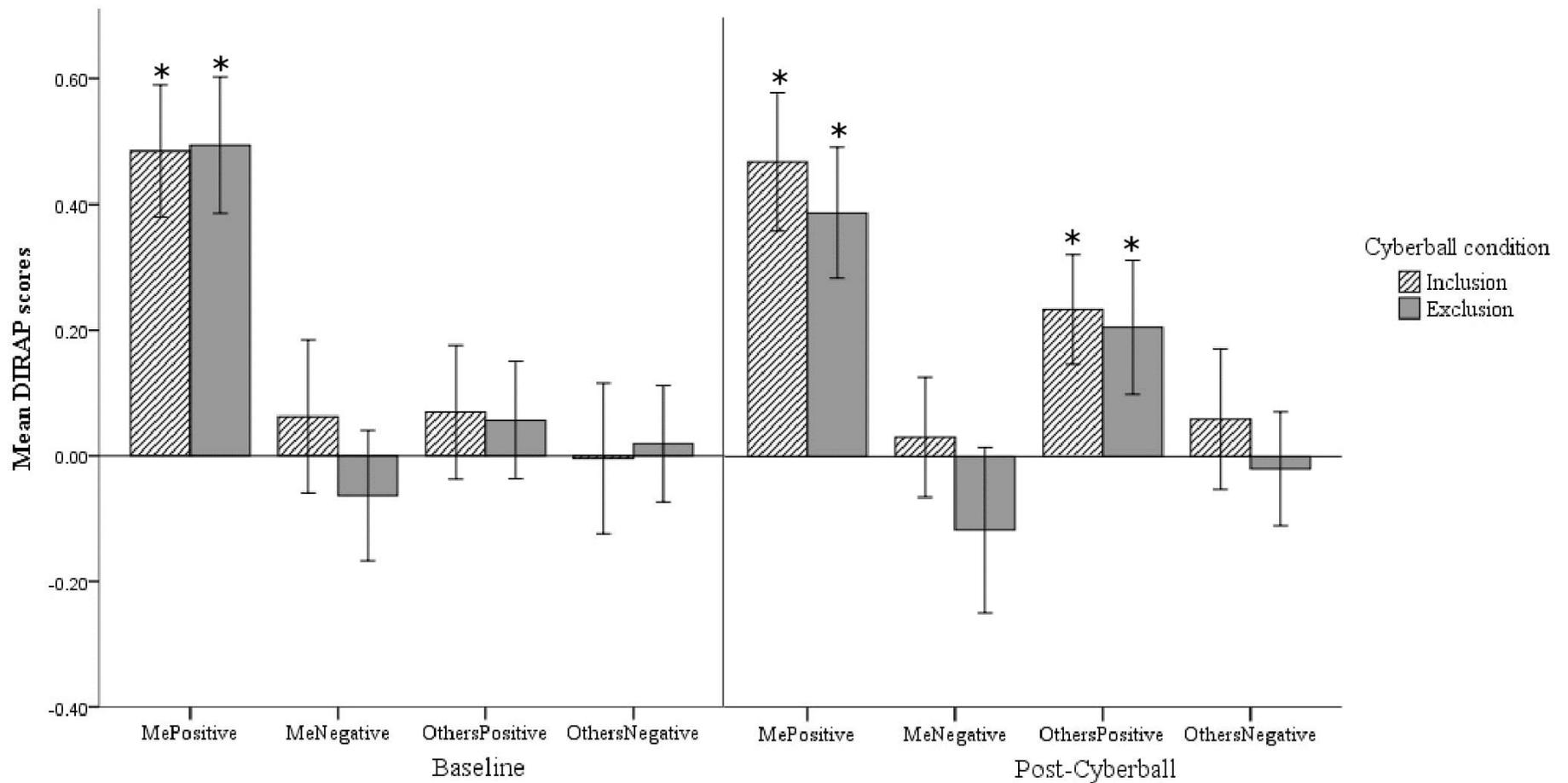
**Figure 4.1.** Changes in mean RSES scores from baseline to post-Cyberball for Inclusion and Exclusion groups. Standard error bars represent 95% confidence intervals.

#### 4.3.7. Effects of exclusion on IRAP responses

We assumed that positive evaluations of the self and others would be consistent or coherent with participants’ pre-experimentally established learning histories; however, due to the format of the IRAP, responding on consistent blocks of trials during the procedure required agreeing ‘Me-Positive’ and ‘Others-Negative’ and disagreeing ‘Me-Negative’ and ‘Others-Positive’ (with the reverse pattern of responding required for inconsistent blocks). Hence, in order to conduct appropriate analysis and interpretations of ‘Me’ and ‘Others’ trial-types, the ‘Others’ trial-types (i.e., ‘Others-Positive’ and ‘Others-Negative’) were inverted (by multiplying them by -1) so that D-IRAP scores greater than zero now shared a common interpretation (see Hussey et al., 2015). That is, such scores indicated a bias towards

confirming or agreeing ‘Me-Positive’ and ‘Others-Positive’ and disconfirming or disagreeing ‘Me-Negative’ and ‘Others-Negative’.

Mean D-IRAP scores for each of the four IRAP trial-types for the Inclusion and Exclusion groups at baseline and post-Cyberball are presented in Figure 4.2. One-sample *t*-tests indicated that both the Inclusion and Exclusion groups demonstrated significant ‘Me-Positive’ agreement (relative to disagreement) at baseline and at post-Cyberball on this trial-type (all *ps* < .001). The other three trial-types were not significantly different from zero at baseline. Post-Cyberball, all participants (regardless of Inclusion or Exclusion) demonstrated a significant ‘Others-Positive’ bias (all *ps* < .001), with faster agreement (relative to disagreement) for ‘Others-Positive’. An effect trending in the direction of significance was observed on the ‘Me-Negative’ trial-type for the Exclusion group post-Cyberball (*p* = .081), with stronger ‘Me-Negative’ agreement (relative to disagreement), indicative of a ‘Me-Negative’ bias. There were no effects on the ‘Others-Negative’ trial-type at either time point. A mixed factorial ANOVA investigated between-groups differences on the four IRAP trial-types from baseline to post-Cyberball. Levene’s test and Mauchly’s test of sphericity indicated no violations of assumptions of normality or sphericity respectively. Results showed main effects for time,  $F(1,83) = 30.455, p < .001, \eta^2 = .268$ , and group,  $F(1,83) = 4.810, p = .031, \eta^2 = .055$ . Post-hoc tests showed significant increases in D-IRAP scores on the ‘Others-Positive’ trial-type from baseline to post-Cyberball for both Inclusion (*p* = .014) and Exclusion (*p* = .022) groups, indicating enhanced ‘Others-Positive’ agreement at post-Cyberball. A between-groups difference on the ‘Me-Negative’ trial-type post-Cyberball trended towards significance (*p* = .066), indicating a ‘Me-negative’ bias for the Exclusion group and no effect for the Inclusion group.



**Figure 4.2.** D-IRAP scores for IRAP trial-types at baseline and post-Cyberball for the Inclusion and Exclusion groups, with standard error bars (95% confidence intervals). Note, D-IRAP scores above zero indicate faster agreement for ‘Me-Positive’ and ‘Others-Positive’ (relative to disagreement) and faster disagreement for ‘Me-Negative’ and ‘Others-Negative’ (relative to agreement), whereas scores below zero indicate the opposite pattern of responding (i.e., faster ‘Me-Positive’ and ‘Others-Positive’ disagreement and ‘Me-Negative’ and ‘Others-Negative’ agreement). Asterisks represent trial-types significantly different from zero.

#### 4.3.8. Correlational analyses

Correlations (with Bonferroni correction,  $\alpha = .005$ ) between Paranoia Checklist total and subscale scores, state paranoia at post-Cyberball and RSES and IRAP trial-type change scores are shown in Table 4.3. Change scores were calculated for RSES scores and the IRAP trial-types by subtracting baseline scores from post-Cyberball scores. State paranoia at post-Cyberball correlated positively with Paranoia Checklist total, Frequency, Conviction, and Distress subscale scores (all  $ps < .001$ ). Negative correlations were observed between state paranoia at post-Cyberball and RSES change scores ( $p = .003$ ), with higher state paranoia scores associated with smaller changes in RSES scores. When analyzed separately, this association was only significant for the Exclusion group,  $r = -.412$ ,  $n = 40$ ,  $p = .008$ . There was a non-significant negative correlation between state paranoia at post-Cyberball and ‘Me-Positive’ trial-type change scores ( $p = .021$ ), with higher state paranoia scores associated with smaller changes in ‘Me-Positive’ scores. Again, when analyzed separately, this association was only observed for excluded participants,  $r = -.392$ ,  $n = 40$ ,  $p = .012$ . ‘Me-Positive’ change scores also correlated negatively with the Distress subscale of the Paranoia Checklist ( $p = .002$ ) and trended towards significance on the Checklist total score ( $p = .009$ ), with higher paranoia total and Distress scores associated with smaller changes in ‘Me-Positive’ scores. A negative correlation between change scores on the ‘Me-Negative’ and ‘Others-Positive’ trial-types trended towards significance ( $p = .017$ ). When examined separately, this reached significance for the Exclusion ( $r = -.492$ ,  $n = 40$ ,  $p = .001$ ) but not the Inclusion group ( $r = -.085$ ,  $n = 45$ ,  $p = .577$ ), with greater changes in ‘Others-Positive’ scores associated with smaller changes in ‘Me-Negative’ scores for excluded participants only.

**Table 4.3.** Correlations between Paranoia Checklist total and subscale scores, state paranoia at post-Cyberball, and change scores (from baseline to post-Cyberball) for RSES and IRAP trial-types ( $N = 85$ ).

	<b>RSES</b>	<b>Me- Positive</b>	<b>Me-Negative</b>	<b>Others-Positive</b>	<b>Others-Negative</b>
<b>Paranoia Total</b>	-.219*	-.280*	.016	-.004	-.016
<b>Paranoia Frequency</b>	-.109	-.193	.032	-.113	-.100
<b>Paranoia Conviction</b>	-.200	-.195	.014	.010	-.113
<b>Paranoia Distress</b>	-.256*	-.328**	-.001	.073	.129
<b>State Paranoia</b>	-.326**	-.249*	-.037	.120	.000
<b>Others-Negative</b>	-.009	-.121	.068	.039	
<b>Others-Positive</b>	-.174	-.135	-.258*		
<b>Me-Negative</b>	-.019	.191			
<b>Me-Positive</b>	.011				

\*  $p < .05$  \*\*  $p \leq .002$

#### 4.4. Discussion

Study 3 showed that social exclusion during Cyberball increased state paranoia and decreased self-reported self-esteem, that the effect of exclusion on state paranoia was moderated by distress associated with experiencing paranoid thoughts at baseline, and that state paranoia at post-Cyberball was associated with changes in self-reported self-esteem for excluded participants, demonstrating that relational responding regarding paranoia and self-esteem were related in this context and also replicating findings from previous research (e.g., Kesting et al., 2013).

Study 3 also sought to examine whether social exclusion might influence relational responding to the self and others as ‘positive’ and ‘negative’ on the IRAP. As hypothesized,

participants who were excluded during Cyberball did not show changes in ‘Me-Positive’ responding, with all participants (regardless of Inclusion or Exclusion) demonstrating significant ‘Me-Positive’ agreement (relative to disagreement) at baseline and at post-Cyberball. Hence, the adverse interpersonal event utilized in this experiment (social exclusion) did not impact (i.e., reduce the probability or strength of) relational responding to the self as ‘positive’ in this context. Rather, excluded participants demonstrated increased ‘Me-Negative’ responding following exclusion and differed from included participants on this trial-type at post-Cyberball (though not significantly so). Specifically, excluded participants demonstrated stronger ‘Me-Negative’ agreement, suggesting that this relation (of coordination) between the self and negative evaluations was coherent in this context. Similar to Study 2, these findings support the assertion that negative self-evaluations are particularly pertinent to paranoia and may be influenced by adverse events, whereas positive self-regard may be maintained (see also Kesting & Lincoln, 2013; Tiernan et al., 2014).

Interestingly, both groups demonstrated a significant ‘Others-Positive’ bias at post-Cyberball, with all participants showing significant agreement (relative to disagreement) with ‘Others-Positive’. While this is understandable for the Inclusion group, it is surprising for the Exclusion group. Responses on the Cyberball Questionnaire and general feedback from excluded participants indicated that they viewed the other ‘players’ negatively. Perhaps they viewed others as stronger or more powerful than previously, a suggestion supported by the significant association between change scores on the ‘Me-Negative’ and ‘Others-Positive’ trial-types for excluded participants only, whereby greater changes in ‘Others-Positive’ scores were associated with smaller changes in ‘Me-Negative’ scores. However, we did not observe any significant correlations between state paranoia and responses on these trial-types, making it difficult to clarify the relationship between paranoia and this particular pattern of responding to the self (as ‘negative’) and others (as ‘positive’) in this context.

The finding that distress associated with experiencing paranoid thoughts in general at baseline moderated the effect of social exclusion on state paranoia is consistent with previous research by Kesting et al. (2013). Importantly, this finding supports the necessity for a multi-dimensional approach to paranoia as factors such as the distress associated with paranoid thoughts or beliefs may be more relevant than the content of the belief alone with respect to the effect that adverse interpersonal experiences may have on relational responding regarding paranoia (and perhaps the impact that paranoia may have on functioning more generally; e.g., Green et al., 2008; Peters et al., 1999b; 2004). Taken together, this finding coupled with the observed increases in responding to oneself as ‘negative’ and others as ‘positive’ following social exclusion provides some insight into the complex patterns of relational responding implicated in this context. That is, one’s learning history may indeed influence or moderate relational responding regarding paranoia in the context of social stress. Moreover, individuals may respond to such events with increased negative self-regard but also with increased positive regard for others.

The observed association between changes in ‘Me-Negative’ and ‘Others-Positive’ responding for excluded participants is an interesting one and suggests that relational responding to the self as ‘negative’ in this context is related to specific patterns of relational responding towards others; namely, responding to others as ‘positive’. However, the ‘positive’ stimuli incorporated in the IRAP in Study 3 related to general positive regard, making it difficult to ascertain the nature of the relationship between responding to the self and others in this context. For example, this association might reflect a wider pattern of responding whereby excluded participants responded to others as ‘more powerful’ than them in some way. Further research is needed to explore more specific functions (e.g., interpersonal threat) regarding relational responding to the self and others and their relationship in the context of paranoia beyond general positive and negative regard. In

addition, the referents for “I” and “Others” were not clarified during the IRAP instructions in this study. Hence, conceptualizations of “Others” may have varied across participants (e.g., friends, strangers, disliked others, etc.) and both time points. For example, participants may have conceptualized “Others” as the Cyberball ‘players’ post-Cyberball, but not at baseline. This highlights the importance of specifying the referents for IRAP (label) stimuli in this context.

Within the wider cognitive-clinical literature, some researchers have proposed that ‘interpersonal sensitivity’, which refers to “feeling vulnerable in the presence of others due to the expectation of criticism or rejection” (Bell & Freeman, 2014, p. 441), may be crucial to paranoia (e.g., Freeman et al., 2008; Freeman & Garety, 2014; Masillo et al., 2012). Indeed, social evaluative concerns (e.g., feelings of vulnerability, fears of rejection) are purported to form the basis for paranoia within the Paranoia Hierarchy (Freeman et al., 2005; see Figure 1.1). This concept of interpersonal sensitivity is likely closely linked to negative beliefs about the self, which are also considered pertinent to paranoia in both clinical and the general populations (e.g., Bebbington et al., 2013; Bentall et al., 1994; 2001; Freeman et al., 2002; Fowler et al., 2006; 2012; Salvatore et al., 2012) and may influence paranoid responding in interpersonal contexts (e.g., Freeman et al., 2003; Freeman, Pugh, Vorontsova, Antley, & Slater, 2010).

From an RFT perspective, perhaps ‘paranoia-prone’ individuals have a history of responding to themselves not only in terms of general negative regard, but specifically as vulnerable. Furthermore, these individuals may also be characterized by patterns of responding to others as being more powerful than them or as untrustworthy or devious. These patterns of AARR may be pertinent to responding in terms of interpersonal threat (i.e., paranoia) in the context of social stress. For instance, Green and colleagues (2006) found that persecutory delusions involving ideas of severe threat were coupled with frequent feelings of

vulnerability in a clinical sample. Furthermore, specific aspects of delusional content were found to be associated with emotional distress, such that individuals who reported feeling less powerful in the face of perceived persecution exhibited higher depression and lower self-esteem. Garety and colleagues (2001) suggested that persecutory beliefs may be held with greater conviction if they are consistent with firmly-held beliefs about the self (e.g., “I am weak”) and others (e.g., “Others are untrustworthy”) and that, once formed, such ideas are likely to be considered further confirmation of these negative beliefs, thereby perpetuating paranoia. From an RFT approach, these patterns of relational responding may be initially established via aversive interpersonal experiences. As these particular patterns of responding to the self (as ‘vulnerable’) and others (as ‘devious’) are increasingly derived, they may also become increasingly coherent and elaborated into relational networks that could be considered typical or characteristic of paranoia. Furthermore, as derivation increases, the responses occasioned by these relational networks (e.g., fear, avoidance), particularly in adverse or ambiguous interpersonal contexts, may come to be elicited in an increasingly ‘automatic’ manner. Hence, the next step within this thesis will be to explore more specific patterns of relational responding to the self (e.g., as vulnerable) and others (e.g., as devious and trustworthy) in the context of paranoia and to examine how adverse interpersonal experiences may influence these repertoires and their interaction.

## **Chapter 5: Relational responding to the self (as ‘vulnerable’ versus ‘safe’) and others (as ‘devious’ versus ‘trustworthy’) on the IRAP in individuals with high non-clinical paranoia and controls**

### **5.1. Introduction**

The findings from the previous studies have highlighted the relevance of relational responding to the self as ‘negative’ to paranoia. The findings from Study 3 suggested that, in the context of social stress, responding to the self as ‘negative’ may be related to responding to others (as ‘positive’). However, the stimuli included in the Implicit Relational Assessment Procedure (IRAP) referred to general positive and negative evaluations, making it difficult to clarify the nature of this relationship. Hence, the aim of Study 4 was to explore more specific relational responses to the self and others and the relationship between the two in relation to paranoia.

Recently, several researchers have suggested that self-esteem may not be the key concept in relation to paranoia; “rather it is specific negative beliefs about the self and others that are important” (Freeman, 2007, p. 435). Negative self-evaluations (e.g., “I am weak”) may create a sense of vulnerability and that one is a potential (“soft”) target of hostility. Indeed, Freeman et al. (2013) suggested that paranoia can be considered an extension of negative ideas about the self as vulnerable (see also Bebbington et al., 2013; Freeman et al., 2005; Salvatore et al., 2012). Such self-evaluations, in combination with negative evaluations of others (e.g., as devious or hostile), likely influence how one adapts and responds to the social world and may set the stage for paranoia to flourish (e.g., “Others would hurt me if given a chance”). For example, Gracie et al. (2007) found that negative beliefs about the self and others mediated the relationship between trauma and paranoia in participants from the general (non-clinical) population.

From a Relational Frame Theory (RFT) perspective, ‘paranoia-prone’ individuals may have a history of responding to themselves as vulnerable and to others as untrustworthy or devious. Such patterns of relational responding may be initially established via aversive interpersonal experiences and subsequently be elaborated into ‘relational networks’ that could be considered typical or characteristic of paranoia. Furthermore, as these relational responses are increasingly derived, they may also be considered more coherent and elicited with greater automaticity, particularly in adverse or ambiguous interpersonal contexts.

Existing measures of self-evaluation tend to capture the presence of positive self-evaluations or their absence (e.g., the Rosenberg Self-Esteem Scale; RSES). Hence, Fowler and colleagues (2006) developed the Brief Core Schema Scales (BCSS) to examine ‘schematic beliefs’ regarding the self and others in a more systematic way. Specifically, the clinically-derived BCSS was designed to assess strongly-held negative evaluations of the self and others, as well as positive self- and others-evaluations in both clinical and the general populations (e.g., Fowler et al., 2006; 2012; Gracie et al., 2007; MacKinnon et al., 2011; Oliver et al., 2011; Smith et al., 2006). Using the BCSS, Fowler et al. (2006) found that participants with psychosis showed greater negative self- and others-evaluations compared to their non-clinical counterparts. At the same time, positive self- and others-evaluations were similar to those reported by the non-clinical sample. They also found that negative others-evaluations predicted paranoia in the non-clinical sample, in addition to positive others-evaluations, anxiety, and negative self-evaluations (in order of contribution to the variance in paranoia scores). Based on these findings, the authors suggested that (non-clinical) paranoia occurs in a specific context of anxiety toned with evaluations of interpersonal threat and personal vulnerability (see also Oliver et al., 2011).

There are two important things to note here. The first reiterates a crucial point made previously – measures that differentiate between distinct dimensions of self- and others-

evaluations may help clarify how these are implicated in paranoia with greater specificity than global measures – an assertion supported by the IRAP findings in the previous studies of this thesis. Second, the IRAP is not only capable of parsing out individual relational responses (e.g., responding to the self as ‘positive’ and ‘negative’) but has demonstrated a level of specificity that may help identify distinct patterns of relational responding pertinent in a given (clinically-relevant) domain. For example, Remue et al. (2013) investigated self-esteem in individuals with self-reported dysphoria and controls. Participants were presented with two IRAPs involving ‘positive’ and ‘negative’ evaluations as target stimuli; however, one IRAP measured *actual* self-esteem (i.e., “I am” and “I am not”) and the other, *ideal* self-esteem (i.e., “I want to be” or “I don’t want to be”). The findings showed that dysphoric individuals had higher *ideal* self-esteem and lower *actual* self-esteem relative to controls.

### **5.1.1. Hypotheses**

The aim of Study 4 was to investigate more specific patterns of relational responding to the self and others in the context of non-clinical paranoia – namely, responding to the self as ‘safe’ and ‘vulnerable’ and to others as ‘trustworthy’ and ‘devious’. We used two separate IRAPs to do so (i.e., ‘Self’ and ‘Others’ IRAP). We compared participants with high levels of non-clinical paranoia and controls to examine whether the two groups would be differentiated by their responses on the IRAPs.

Stimuli from the Others-Negative and Others-Positive scales of the BCSS were directly incorporated into the ‘Others’ IRAP as these stimuli were designed to specifically target evaluations of others pertinent to paranoia (see Fowler et al., 2006; see Table 5.2). However, the Self scales within the BCSS appear to capture general positive (e.g., “talented”, “valuable”) and negative (e.g., “unloved”, “worthless”) self-evaluations, with only some items relating to vulnerability specifically (e.g., “weak”, “vulnerable”; see Appendix E). Hence, for the purposes of Study 4, we identified stimuli that better encapsulated

‘vulnerability’ and ‘safety’ for the ‘Self’ IRAP (see Table 5.1). Self-report measures were used to capture more extended and elaborated relational responding (EERR) to the self and others, including the BCSS, self-reported ratings of the ‘Self’ IRAP stimuli, and RSES.

Based on previous findings that non-clinical paranoia is associated with negative self- and others-evaluations in the general population and that clinical and non-clinical samples demonstrate similar positive self- and others-evaluations (Fowler et al., 2006), it was expected that all participants (regardless of paranoia level) would demonstrate faster agreement (than disagreement) for ‘Me-Safe’ on the ‘Self’ IRAP and ‘Others-Trustworthy’ on the ‘Others’ IRAP. However, it was hypothesized that participants with high non-clinical paranoia would demonstrate faster agreement for ‘Me-Vulnerable’ on the ‘Self’ IRAP and ‘Others-Devious’ on the ‘Others’ IRAP, whereas controls would not. On the self-report measures, it was expected that participants with high non-clinical paranoia would demonstrate higher ‘Me-Vulnerable’ ratings, Negative-Self and Negative-Others scores on the BCSS, and Anxiety, Depression and Stress scores, and lower ‘Me-Safe’ ratings, Positive-Self and Positive-Others scores on the BCSS, and RSES scores relative to controls.

## **5.2. Method**

### **5.2.1. Participants**

Participants were predominantly undergraduate NUI Galway Psychology students who completed the Paranoia Checklist screening measure in exchange for course credit. Other NUI Galway students were recruited via email invitation. Normative percentile scores from the Paranoia Checklist were used to determine the high non-clinical paranoia (top 16<sup>th</sup> percentile; 1 SD above the Mean) and control (bottom 16<sup>th</sup> percentile: 1 SD below the Mean) groups. Selecting participants who are one SD  $\pm$  Mean was deemed preferable to the alternative of using a median split because the former approach creates groups that are more

clearly distinguished in terms of paranoia level (see Flower et al., 2015, for a similar approach).

We aimed to screen at least 300 participants to ensure that there would be sufficient numbers within each group (i.e., 35-40 participants<sup>19</sup>) to complete the IRAPs and additional self-report measures, whilst allowing for attrition. A total of 472 individuals completed the online Paranoia Checklist. The Mean Checklist total score for the total sample was  $M = 99.67$  ( $SD = 28.72$ ). Hence, cut-off scores were approximately  $\leq 71$  for controls and  $\geq 128$  for high non-clinical paranoia. Based on these criteria, 88 respondents met criterion for the control (low paranoia) group and 88 for the high non-clinical paranoia group. Of these, 38 controls (13 males, 25 females;  $Mean\ age = 25.34$ ,  $SD = 8.51$ ), and 35 individuals with high non-clinical paranoia (5 males, 30 females;  $Mean\ age = 20.66$ ,  $SD = 2.72$ ) returned to complete the ‘Self’ and ‘Others’ IRAPs and additional self-report measures. Additional course credit was offered in exchange for participation in the second part of the study.

### 5.2.2. Measures

**Paranoia Checklist.** The Paranoia Checklist (Freeman et al., 2005) was used to assess ‘trait’ paranoia and as a screening measure to identify individuals scoring within the top (high non-clinical paranoia) and bottom (controls) 16<sup>th</sup> percentiles. The Checklist showed excellent internal reliability overall ( $\alpha = .975$ ), as well as on the Frequency ( $\alpha = .930$ ), Conviction ( $\alpha = .936$ ) and Distress ( $\alpha = .955$ ) subscales in the current sample.

**Brief Core Schema Scales.** The BCSS (Fowler et al., 2006) contains 24 items concerning beliefs about the self and others that are assessed on a five-point rating scale (ranging from 0 = Do not believe it, to 4 = Believe it totally). Four scores were obtained for Negative-Self, Positive-Self, Negative-Others, and Positive-Others scales. The BCSS showed

---

<sup>19</sup> Following a recent meta-analysis of criterion effects for the IRAP in the clinical domain, Vahey et al. (2015) proposed that a sample size of  $N = 29-37$  is required in order to achieve statistical power of .80 when testing a continuous first-order correlation between a clinically-focused IRAP effect (i.e., D-IRAP total and/or trial-type scores) and a given criterion variable (e.g., Paranoia Checklist scores).

excellent internal reliability for Negative-Self ( $\alpha = .901$ ), Positive-Self ( $\alpha = .892$ ), Negative-Others ( $\alpha = .907$ ), and Positive-Others ( $\alpha = .913$ ) in the current sample.

**Self-report ratings of ‘Self’ IRAP stimuli.** The target stimuli from the ‘Self’ IRAP were rated using the same state-adapted response format as the BCSS (see Tables 5.1 for stimuli). The scales showed excellent reliability for ‘Me-Safe’ ( $\alpha = .875$ ) and ‘Me-Vulnerable’ ( $\alpha = .859$ ) in the current sample.

**Rosenberg Self-Esteem Scale.** Global self-reported self-esteem was measured using the RSES (Rosenberg, 1965). The RSES showed excellent internal reliability in the current sample ( $\alpha = .915$ ).

**Depression, Anxiety and Stress Scale-21.** The DASS-21 (Lovibond & Lovibond, 1995) is a non-diagnostic tool that measures the severity of a range of symptoms common to both depression and anxiety. Participants were asked to indicate the presence of a symptom over the previous week. Each item was scored from 0 (Did not apply to me at all) to 3 (Applied to me very much or most of the time). The scales showed excellent internal reliability for Depression ( $\alpha = .904$ ), Anxiety ( $\alpha = .853$ ), and Stress ( $\alpha = .908$ ) in the current sample.

**The IRAP.** Two separate IRAPs were used for the purposes of this study. The ‘Self’ IRAP assessed responding to self as ‘vulnerable’ and ‘safe’ (see Table 5.1). The ‘Others’ IRAP assessed responding to others as ‘trustworthy’ and ‘devious’ (see Table 5.2).

**Table 5.1.** Stimuli for ‘Self’ IRAP for Study 4 (and for Study 5).

<b>Label stimuli</b>	
I am	I am not
<b>Target stimuli</b>	
<i>Safe</i>	<i>Vulnerable</i>
Secure	Vulnerable
Safe	Weak
Capable	Exploited
Resilient	Insecure
Sure of myself	Uneasy
Self-assured	Fearful
<b>Response options</b>	
True	False

**Table 5.2.** Stimuli for ‘Others’ IRAP for Study 4.

<b>Label stimuli</b>	
Others are	Others are not
<b>Target stimuli</b>	
<i>Trustworthy</i>	<i>Devious</i>
Fair	Hostile
Good	Harsh
Trustworthy	Judgmental
Accepting	Bad
Supportive	Devious
Truthful	Nasty
<b>Response options</b>	
True	False

Within each IRAP, one of two label stimuli (e.g., “I am” or “I am not” within the ‘Self’ IRAP) was presented at the top of the screen, one of the target stimuli was presented in the center of the screen (e.g., “Vulnerable”, “Safe”), and two response options (True, False) at the bottom right- and left-hand corners of the screen. The positions of “True” and “False” were counter-balanced across trials.

Correct responding on the ‘Self’ IRAP required participants to agree “I am-*Safe*” and “I am not-*Vulnerable*” and disagreeing “I am-*Vulnerable*” and “I am not-*Safe*” on consistent blocks of trials, while on inconsistent blocks participants were required to respond according to the reverse pattern (agreeing “I am-*Vulnerable*” and “I am not-*Safe*” and disagreeing “I am-*Safe*” and “I am not-*Vulnerable*”). Correct responding on the ‘Others’ IRAP required participants to agree “Others are-*Trustworthy*” and “Others are not-*Devious*” and disagreeing “Others are-*Devious*” and “Others are not-*Trustworthy*” on consistent blocks of trials, while on inconsistent blocks participants were required to respond according to the reverse pattern (agreeing “Others are-*Devious*” and “Others are not-*Trustworthy*” and disagreeing “Others are-*Trustworthy*” and “Others are not-*Devious*”). The order of IRAP presentation (i.e., ‘Self’ IRAP-first versus ‘Others’ IRAP-first) was counter-balanced across all participants, as was IRAP block order, such that participants who completed their first IRAP with consistent-first completed the second IRAP with consistent-first, and the same for inconsistent-first. Following verbal instructions from the researcher, participants received a minimum of one and a maximum of four pairs of practice blocks. Each block contained 24 trials presented in quasi-random order. For the ‘Self’ IRAP, participants were instructed:

*“When you see “I” on the screen, whether it’s “I am” or “I am not”, I want you to think about yourself, [name], as you are now. Not how you want to be or how you think you should be but how you truly see yourself now”.*

For the ‘Others’ IRAP, participants were instructed:

*“When you see “Others” on the screen, whether it’s “Others are” or “Others are not”, I want you to think about other people who you do not know, like strangers on the street, rather than friends, family or even people you know and do not like.”*

Participants were required to meet criteria of  $\geq 80\%$  accuracy and mean latency  $\leq 2000$ ms to progress to the test phase, which included exactly six test blocks. If the

participant failed to respond within 2000ms, the message “Too Slow!” appeared in red below the target stimulus and remained on screen until a response was emitted. If a correct response was made, the screen was cleared for 400ms before the presentation of the next trial. If an incorrect response was made, a red ‘X’ was presented and remained on screen until the correct response was made. Accuracy and latency feedback were presented at the end of each block of trials. Only data from the test blocks were included in subsequent analyses.

### **5.2.3. General Procedure**

Participants who met criteria for high non-clinical paranoia and controls on the Paranoia Checklist completed the ‘Self’ and ‘Others’ IRAPs followed by the DASS, self-reported ‘Me-Safe’ and ‘Me-Vulnerable’ ratings, BCSS, and RSES in that order.

### **5.2.4. Statistical analysis plan**

Two independent samples t-tests checked for potential between-groups differences in Paranoia Checklist total scores for ‘returners’ (i.e., eligible participants who returned to complete the IRAPs and other self-report measures) and ‘non-returners’ for the control and high non-clinical paranoia groups. 2 X 2 ANOVAs checked for potential IRAP order (i.e., ‘Self’ or ‘Others’ IRAP-first) and block order effects on IRAP trial-types. One-way ANOVAs checked for between-group differences (controls vs. high non-clinical paranoia) on all variables, with Bonferroni correction for multiple comparisons. A correlation matrix was calculated to explore the relationships between Paranoia Checklist total and subscale scores, ‘Self’ and ‘Others’ IRAP trial-type scores, and self-report measures. Bonferroni correction for multiple comparisons ( $\alpha = .004$ ) was used.

## **5.3. Results**

### **5.3.1. Data preparation**

Independent samples t-tests indicated no significant between-groups differences for ‘returners’ and ‘non-returners’ on Paranoia Checklist total scores for individuals who met

criteria for the control,  $t(86) = -.430, p = .67$ , or high non-clinical paranoia,  $t(86) = 1.663, p = .10$ , groups. For the purposes of data analysis, IRAP accuracy criterion was set at  $\geq 71\%$  and latency criterion was extended to  $< 2500\text{ms}^{20}$ . Six participants (4 controls, 2 high paranoia) failed to meet accuracy and/or latency criteria on the IRAP practice phase. Four (1 control, 3 high paranoia) failed to maintain accuracy criterion during the IRAP test phase for both IRAPs. Two participants (both controls) violated accuracy criterion on one test block of the ‘Self’ IRAP and two (both controls) violated accuracy criterion on at least one test block of the ‘Others’ IRAP. To maximize all available data, these participants were included in subsequent data analysis but their D-IRAP scores for the IRAP on which the accuracy criterion was violated (i.e., ‘Self’ or ‘Others’) were excluded by excluding cases pairwise.

Hence, 63 participants were included in the final data analysis; 33 controls (11 males, 22 females; *Mean age* = 25.75, *SD* = 8.78; *Mean Paranoia Checklist total score* = 64.12, *SD* = 5.18, *Range* = 54–74) and 30 high non-clinical paranoia (3 males; 27 females; *Mean age* = 20.32, *SD* = 2.29; *Mean Paranoia Checklist total score* = 142.63, *SD* = 16.01, *Range* = 122–177). There was a significant between-groups difference for age,  $t(35.77) = 3.367, p = .002$ , with controls shown to be significantly older than participants high in non-clinical paranoia (*Mean difference* = 5.43 years). Pearson Chi-square (with continuity correction) indicated non-significant gender differences between controls and high paranoia,  $\chi^2(1, n = 63) = 3.692, p = .055$ , with more males in the control group. On the DASS, controls scored within the normal range for Depression, Anxiety and Stress whereas participants with high non-clinical paranoia scored within the moderate range for each variable.

---

<sup>20</sup> 5 participants scored  $\geq 71\%$  on at least one test block of one IRAP (‘Self’ or ‘Others’) and were included in the final analysis. Participants who scored  $< 71\%$  during the test phase on both the ‘Self’ and ‘Others’ were removed from subsequent analyses. 12 participants scored  $> 2000\text{ms}$  (but  $< 2500\text{ms}$ ) on at least one test block of the ‘Self’ and/or ‘Others’ IRAP.

### 5.2.2. Sample characteristics

To investigate the items on the Paranoia Checklist endorsed by the control and high non-clinical paranoia groups, we explored the proportion of participants within each group that reported experiencing each item on at least a weekly basis on the Frequency subscale of the Paranoia Checklist (as per Freeman et al., 2005). This was to provide insight into the types of ideas most frequently experienced by participants in both groups and whether participants in the high non-clinical paranoia group endorsed items across all levels of the Paranoia Hierarchy (see Figure 1.1), including ideas of moderate to severe threat (see Table 5.3). Results indicated that the most commonly endorsed items across both groups typically involved social evaluative concerns (e.g., “There might be negative comments being circulated about me”) but the proportion of participants who endorsed these items were substantially higher for the high non-clinical paranoia group relative to controls. Furthermore, participants with high non-clinical paranoia endorsed ideas across all levels of the Paranoia Hierarchy, including ideas of mild (e.g., 43.3% of participants with high non-clinical paranoia endorsed “People deliberately try to irritate me”), moderate (e.g., 50% of participants endorsed “Someone I know has bad intentions towards me”), and more severe threat (e.g., 23.3% endorsed “Someone I don’t know has bad intentions towards me”, and 13.4% endorsed “I am under threat from others”). However, no participants endorsed ideas regarding hidden messages about them in the media, television, or radio (an idea that may be more relevant to or commonly experienced within clinical populations).

**Table 5.3.** Paranoia Checklist items endorsed by participants in the control ( $n = 33$ ) and high non-clinical paranoia ( $n = 30$ ) groups. Endorsement was defined as a Frequency rating (i.e., “How often have you had the thought?”) of  $\geq$  weekly (i.e., a score of  $\geq 3$ ). Percentages represent proportion of each group (control and high non-clinical paranoia) that rated each item with a Frequency score of  $\geq$  weekly.

Item	Controls (Weekly %)	High non-clinical paranoia (Weekly %)
1. I need to be on my guard against others	15.1	76.7
2. There might be negative comments being circulated about me	6.1	83.3
3. People deliberately try to irritate me	0	43.3
4. I might be being observed or followed	3	33.3
5. People are deliberately trying to upset me	0	36.6
6. People communicate about me in subtle ways	3	66.7
7. Strangers and friends looks at me critically	6.1	80
8. People might be hostile towards me	0	60
9. Bad things are being said about me behind my back	0	73.3
10. Someone I know has bad intentions towards me	0	50
11. I have a suspicion that someone has it in for me	0	30
12. People would harm me if given the opportunity	0	26.7
13. Someone I don't know has bad intentions towards me	0	23.3
14. There is a possibility of a conspiracy against me	0	10
15. People are laughing at me	6.1	73.3
16. I am under threat from others	0	13.4
17. I can detect coded messages about me in the press/TV/radio	0	0
18. My actions and thoughts can be controlled by others	6	16.7

Means and SDs for controls and participants with high non-clinical paranoia for each measure are presented in Table 5.4. Kolmogorov-Smirnov tests indicated violations of normality assumptions for the Depression and Anxiety scales of the DASS, self-reported 'Me-Vulnerable' ratings, and the Negative-Self, Negative-Others and Positive-Others scales of the BCSS (all  $ps \leq .018$ )<sup>21</sup>. Inspection of the histograms indicated positive skewness for Depression and Anxiety, with most participants scoring low on these scales. However, skewness and kurtosis were in the acceptable range of  $\pm 2$ . Similarly, 'Me-Vulnerable' ratings and Negative-Self and Negative-Others BCSS scores were positively skewed, which was unsurprising given that participants were recruited from the general population. Skewness and kurtosis were in the acceptable range for 'Me-Vulnerable', Negative-Others and Positive-Others but skew was  $> 2$  for Negative-Self. Hence, non-parametric tests were used for correlation analyses and for between-group comparisons for these variables.

One-way between groups ANOVAs were conducted to examine between-groups differences on each variable (with Bonferroni adjustment for multiple comparisons;  $\alpha \approx .004$ ). Levene's test indicated violation of assumptions of homogeneity of variance for 'Me-Vulnerable' ratings, Negative-Self and Negative-Others scores on the BCSS, and Depression and Anxiety scores (all  $ps \leq .023$ ). Hence, Mann-Whitney U tests were conducted to examine between-groups differences on these variables. Results revealed no significant between-groups differences on any trial-types of the 'Self' or 'Others' IRAPs (all  $ps \geq .157$ ). Controls reported significantly higher 'Me-Safe' ratings ( $p = .001$ ), Positive-Self ( $p < .001$ ), Positive-Others ( $p = .002$ ), and RSES ( $p < .001$ ) scores compared to participants with high non-clinical paranoia. Participants with high non-clinical paranoia scored significantly higher than controls on the Stress scale of the DASS ( $p < .001$ ). Mann-Whitney U tests revealed significant between-groups differences for 'Me-Vulnerable' ratings (controls  $Md = 2.00$ ; high

---

<sup>21</sup> Normality assumptions were also violated for the Paranoia Checklist total and subscale scores ( $p \leq .001$ ). However, a bimodal distribution was to be expected given that participants were selected on the basis that they scored within either the top or bottom 16<sup>th</sup> percentiles (i.e.,  $1 SD \pm M$ ).

non-clinical paranoia  $Md = 7.00$ ), Negative-Self (controls  $Md = 1.00$ ; high non-clinical paranoia  $Md = 4.00$ ), Negative-Others (controls  $Md = 4.00$ ; high non-clinical paranoia  $Md = 8.50$ ), Depression (controls  $Md = 4.00$ ; high non-clinical paranoia  $Md = 14.00$ ), and Anxiety (controls  $Md = 2.00$ ; high non-clinical paranoia  $Md = 12.00$ ), with participants with high non-clinical paranoia scoring higher than controls on each measure (all  $ps < .001$ ).

**Table 5.4.** Means and SDs for ‘Self’ and ‘Others’ IRAP trial-types and self-report measures for participants high in non-clinical paranoia compared with controls ( $N = 63$ ).

	<b>Controls</b>	<b>High paranoia</b>
<b>Me-Safe</b>	.517 (.348)	.428 (.403)
<b>Me-Vulnerable</b>	-.094 (.384)	-.228 (.442)
<b>Not me-Safe</b>	.102 (.412)	-.055 (.441)
<b>Not me-Vulnerable</b>	.232 (.327)	.194 (.442)
<b>Others-Trustworthy</b>	.516 (.312)	.514 (.278)
<b>Others-Devious</b>	-.179 (.343)	-.223 (.467)
<b>Others not-Trustworthy</b>	.074 (.373)	-.023 (.442)
<b>Others not-Devious</b>	.091 (.434)	.218 (.324)
<b>Me-Safe ratings</b>	16.42 (4.19)	12.47 (4.79)
<b>Me-Vulnerable ratings</b>	2.64 (2.09)	8.73 (5.50)
<b>BCSS Negative-Self</b>	1.42 (2.03)	6.20 (5.57)
<b>BCSS Positive-Self</b>	14.18 (4.77)	9.53 (4.45)
<b>BCSS Negative-Others</b>	4.06 (3.55)	8.43 (4.87)
<b>BCSS Positive-Others</b>	15.12 (4.04)	11.70 (4.43)
<b>RSES</b>	20.82 (4.85)	14.43 (6.10)
<b>Depression</b>	4.79 (5.34)	14.20 (9.76)
<b>Anxiety</b>	4.24 (5.93)	13.67 (8.49)
<b>Stress</b>	7.82 (7.62)	18.27 (8.64)

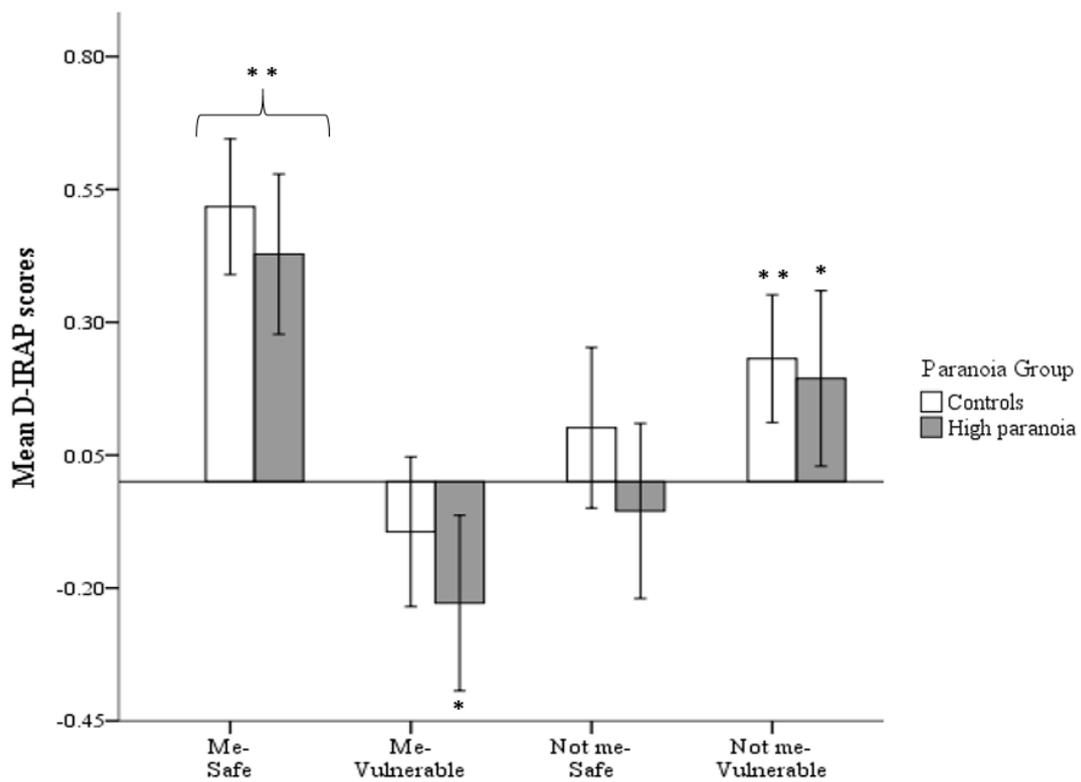
### 5.2.3. IRAP trial-type responses for high non-clinical paranoia and controls

A 2 X 2 ANOVA revealed no effect for IRAP order (i.e., ‘Self’ vs. ‘Others’ IRAP-first) on D-IRAP scores for each trial-type on either IRAP ( $p = .33$ ). An effect trended towards significance for block order ( $p = .053$ ), with significant differences in mean D-IRAP

scores between participants who completed consistent blocks first ( $M = .360$ ) and those who completed inconsistent blocks first ( $M = .585$ ) on the ‘Me-Safe’ trial-type ( $p = .022$ ) and on the ‘Not me-Safe’ trial-type (consistent-first  $M = -.141$ ; inconsistent-first  $M = .204$ ;  $p = .001$ ). However, when this was examined separately for high non-clinical paranoia and controls, this was non-significant. The mean D-IRAP scores for each trial-type for the ‘Self’ and ‘Others’ IRAPs by group (high non-clinical paranoia and controls) are presented in Figures 5.1 and 5.2 respectively.

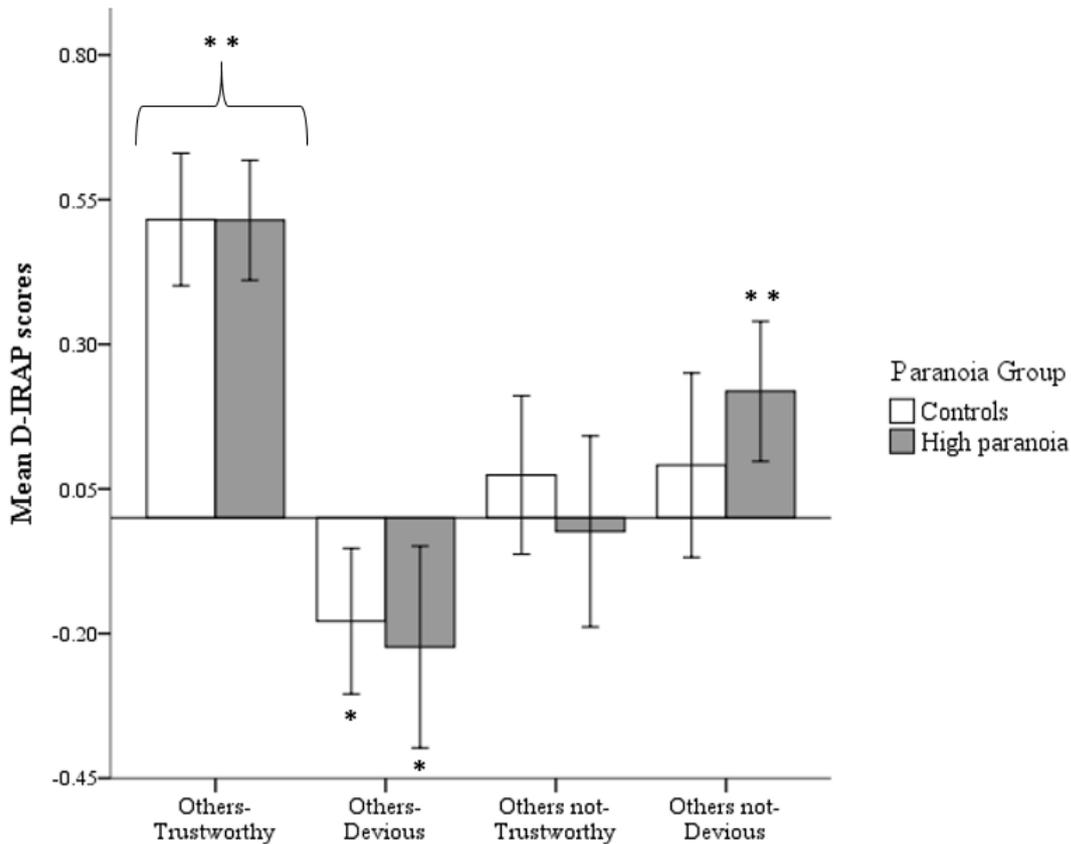
On the ‘Self’ IRAP, one sample t-tests indicated that the ‘Me-Safe’ trial-type was significantly different from zero for both groups ( $ps < .001$ ), with all participants (regardless of paranoia level) demonstrating faster ‘Me-Safe’ agreement. The ‘Not me-Vulnerable’ trial-type also differed significantly from zero for both controls ( $p < .001$ ) and participants with high non-clinical paranoia ( $p = .023$ ), with faster ‘Not me-Vulnerable’ agreement. Only participants with high non-clinical paranoia demonstrated a significant effect on the ‘Me-Vulnerable’ trial-type ( $p = .008$ ), with participants in this group showing faster ‘Me-Vulnerable’ agreement (relative to disagreement), indicative of a ‘Me-Vulnerable’ bias. There were no significant effects on the ‘Not me-Safe’ trial-type.

On the ‘Others’ IRAP, the ‘Others-Trustworthy’ trial-type differed significantly from zero ( $p < .001$ ) for all participants (regardless of paranoia level), with faster ‘Others-Trustworthy’ agreement (relative to disagreement). The ‘Others-Devious’ trial-type was also significantly different from zero for both controls ( $p = .007$ ) and participants with high non-clinical paranoia ( $p = .014$ ), with all participants showing faster ‘Others-Devious’ agreement (relative to disagreement), indicative of an ‘Others-Devious’ bias. Only participants with high non-clinical paranoia demonstrated a significant effect on the ‘Others not-Devious’ trial-type ( $p = .001$ ), with faster ‘Others not-Devious’ agreement (relative to disagreement). There were no effects on the ‘Others not-Trustworthy’ trial-type.



\* Sig. at  $\alpha = .05$  \*\* Sig. at  $\alpha \leq .001$

**Figure 5.1.** D-IRAP scores for ‘Self’ IRAP trial-types for controls and participants with high non-clinical paranoia, with standard error bars (95% confidence intervals). Note, D-IRAP scores above zero indicate faster responding on consistent (i.e., faster ‘Me-Safe’ agreement and ‘Me-Vulnerable’ disagreement) relative to inconsistent blocks of trials, whereas scores below zero indicate the reverse pattern (i.e., faster responding on inconsistent trials). Asterisks indicate trial-types significantly different from zero.



Sig. at  $\alpha = .05$  \*\* Sig. at  $\alpha \leq .001$

\*

**Figure 5.2.** D-IRAP scores for ‘Others’ IRAP trial-types for controls and participants with high non-clinical paranoia, with standard error bars (95% confidence intervals). Note, D-IRAP scores above zero indicate faster responding on consistent (i.e., faster ‘Others-Trustworthy’ agreement and ‘Others-Devious’ disagreement) relative to inconsistent blocks of trials, whereas scores below zero indicate the reverse pattern (i.e., faster responding on inconsistent trials). Asterisks indicate trial-types significantly different from zero.

#### 5.2.4. Correlational analyses

A correlation matrix was calculated (using Spearman’s Rho with Bonferroni correction,  $\alpha = .05 / 22 \approx .0023$ ) to examine the relationships between Paranoia Checklist total and subscale scores, ‘Self’ and ‘Others’ IRAP trial-types, self-reported ‘Me-Safe’ and ‘Me-Vulnerable’ ratings, BCSS, RSES, and Depression, Anxiety, and Stress scores (see Table 5.5). There were no significant correlations between Paranoia Checklist total and subscale scores and ‘Self’ and ‘Others’ IRAP trial-types, although non-significant negative

associations were observed between the 'Not me-Safe' trial-type and Paranoia Checklist total score ( $r = -.287, p = .025$ ) and Distress ( $r = -.278, p = .030$ ), with lower Paranoia total and Distress scores associated with stronger 'Not me-Safe' disagreement. A non-significant negative correlation was also observed between the 'Me-Vulnerable' trial-type and Conviction ( $r = -.255, p = .045$ ), with lower Conviction scores associated with stronger 'Me-Vulnerable' disagreement. On the self-report measures, Paranoia total and Frequency, Conviction and Distress subscale scores were significantly and positively associated with 'Me-Vulnerable' ratings, Negative-Self, Negative-Others, Depression, Anxiety and Stress scores and negatively associated with 'Me-Safe' ratings, Positive-Self, Positive-Others, and RSES scores.

On the 'Self' IRAP, the 'Me-Safe' trial-type was negatively associated with Negative-Self scores ( $r = -.359, p = .004$ ), with lower Negative-Self scores associated with stronger 'Me-Safe' agreement. In addition, non-significant positive associations were observed between this trial-type and self-reported 'Me-Safe' ratings ( $r = .268, p = .037$ ) and Positive-Self scores ( $r = .262, p = .041$ ), with higher self-reported 'Me-Safe' ratings and Positive-Self scores associated with stronger 'Me-Safe' agreement on the IRAP. The 'Me-Safe' trial-type was also non-significantly and negatively associated with self-reported 'Me-Vulnerable' ratings ( $r = -.261, p = .042$ ), with lower 'Me-Vulnerable' ratings associated with stronger 'Me-Safe' agreement. There was also a significant positive correlation between the 'Me-Vulnerable' and 'Not me-Safe' trial-types ( $r = .401, p = .001$ ), with stronger 'Me-Vulnerable' disagreement associated with stronger 'Not me-Safe' disagreement on the IRAP. The 'Me-Vulnerable' trial-type was also non-significantly and negatively associated with Anxiety scores ( $r = -.279, p = .030$ ), with lower Anxiety scores associated with stronger 'Me-Vulnerable' disagreement. The 'Not me-Safe' trial-type was negatively associated with Stress scores ( $r = -.353, p = .005$ ), with lower Stress scores associated with stronger 'Not me-Safe'

disagreement. The 'Not me-Vulnerable' trial-type was non-significantly and positively associated with self-reported 'Me-Safe' ratings ( $r = .342, p = .007$ ), with higher 'Me-Safe' ratings associated with stronger 'Not me-Vulnerable' agreement. There were no significant associations between the 'Others' IRAP trial-types and self-report measures.

Self-reported 'Me-Safe' ratings were positively associated with Positive-Self, Positive-Others, and RSES scores and negatively associated with 'Me-Vulnerable' ratings, Negative-Self, Negative-Others, Depression, Anxiety, and Stress scores. Self-reported 'Me-Vulnerable' ratings were positively associated with Negative-Self, Negative-Others, Depression, Anxiety, and Stress scores and negatively associated with Positive-Self, Positive-Others, and RSES scores.

**Table 5.5.** Correlations between Paranoia Checklist total and subscale scores, ‘Self’ and ‘Others’ IRAP trial-types, self-reported ‘Me-Safe’ and ‘Me-Vulnerable’ ratings, BCSS, RSES, and DASS scores ( $N = 63$ ). Note, 20 = Depression, 21 = Anxiety, and 22 = Stress scores (DASS).

		5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.
1.	Paranoia Total	-.136	-.164	-.287*	-.110	-.036	-.056	-.041	.140	-.506**	.636**	.638**	-.446**	.455**	-.356**	-.511**	.593**	.565**	.565**
2.	Paranoia Frequency	-.227	-.117	-.181	-.067	-.083	-.089	-.053	.140	-.436**	.620**	.665**	-.456**	.518**	-.399**	-.495**	.580**	.621**	.495**
3.	Paranoia Conviction	-.155	-.255*	-.226	-.104	-.042	-.137	-.111	.116	-.421**	.606**	.667**	-.435**	.510**	-.459**	-.509**	.565**	.623**	.521**
4.	Paranoia Distress	-.061	-.139	-.278*	-.097	-.001	-.002	.028	.148	-.483**	.620**	.557**	-.434**	.364**	-.275*	-.460**	.506**	.585**	.564**
5.	Me-Safe		.085	.127	.241	.058	-.081	.129	.017	.268*	-.261*	-.359**	.262*	-.110	.113	.233	-.101	-.119	-.016
6.	Me-Vulnerable			.401**	-.174	-.024	.244	.019	-.179	.070	-.128	-.135	.098	-.171	.057	.139	-.080	-.279*	-.216
7.	Not me-Safe				-.188	-.114	-.102	.036	-.237	.171	-.226	-.171	-.017	-.186	-.175	.125	-.221	-.237	-.353*
8.	Not me-Vulnerable					-.197	.016	-.128	.104	.342*	-.237	-.140	.191	-.077	.112	.172	-.097	-.090	-.025
9.	Others-Trustworthy						.159	.127	.038	.048	-.081	-.103	.032	-.014	.068	.065	.058	.121	.070
10.	Others-Devious							.208	.102	-.037	-.034	-.068	-.008	-.221	.183	.039	-.101	-.151	.025
11.	Others not-Trustworthy								.018	.075	-.123	-.100	.001	-.199	.169	.241	-.143	-.112	-.125
12.	Others not-Devious									-.024	.004	.106	.071	.047	-.148	-.006	.040	.199	.240
13.	Me-Safe ratings										-.648**	-.627**	.687**	-.266*	.355**	.799**	-.688**	-.539**	-.575**

Table 5.5 continued.

	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.
14. Me-Vulnerable ratings											.788**	-.497**	.513**	-.330*	-.651**	.614**	.717**	.689**
15. Negative-Self												-.633**	.519**	-.470**	-.745**	.684**	.731**	.605**
16. Positive-Self													-.173	.416**	.784**	-.593**	-.485**	-.430**
17. Negative-Others														-.509**	-.369**	.465**	.510**	.456**
18. Positive-Others															.441**	-.346*	-.455**	-.393**
19. RSES																-.764**	-.617**	-.579**

## 5.4. Discussion

The purpose of Study 4 was to explore more detailed patterns of relational responding to the self and others beyond general ‘positive’ and ‘negative’ evaluations – specifically, relational responding to the self as ‘safe’ and ‘vulnerable’ and to others (in general) as ‘trustworthy’ and ‘devious’ in individuals with high non-clinical paranoia and controls. Based on previous findings within the cognitive-clinical literature (e.g., Fowler et al., 2006; Gracie et al., 2007), it was expected that individuals with high non-clinical paranoia may relate in accordance with networks involving the self as ‘vulnerable’ and others as ‘devious’, which may be extensively derived and, consequently, more coherent for this group.

On the ‘Self’ IRAP, all participants (regardless of paranoia level) demonstrated significant ‘Me-Safe’ agreement. In line with predictions, participants with high non-clinical paranoia demonstrated significantly faster agreement (relative to disagreement) for ‘Me-Vulnerable’, indicative of a ‘Me-Vulnerable’ bias. Also, all participants (regardless of paranoia level) showed significant ‘Not me-Vulnerable’ agreement. The observed responses on the ‘Me-Vulnerable’ and ‘Not me-Vulnerable’ trial-types for participants with high non-clinical paranoia indicated a seemingly contradictory pattern of responding to the self in terms of vulnerability, with these participants agreeing both ‘I am *Vulnerable*’ and ‘I am not *Vulnerable*’. However, they showed significantly higher self-reported ‘Me-Vulnerable’ ratings compared to controls, suggesting that individuals with high non-clinical paranoia may typically respond to themselves as ‘vulnerable’. Furthermore, these patterns of relational responding to the self showed differential relationships with paranoia (although these did not reach statistical significance). Specifically, stronger ‘Not me-Safe’ disagreement was associated with lower paranoia total and Distress scores, and stronger ‘Me-Vulnerable’ disagreement was associated with lower Conviction scores. Interestingly, Garety and colleagues (2001) proposed that paranoid thoughts and beliefs may be held with greater

conviction if they are consistent with firmly-held beliefs about the self (e.g., as bad, weak, or vulnerable). These findings provide some preliminary support for this suggestion, with lower paranoia and associated distress and conviction associated with distinguishing the self from being ‘vulnerable’ and ‘not safe’.

The findings on the ‘Self’ IRAP indicate that responding to the self as ‘vulnerable’ may be particularly pertinent to paranoia and that individuals with high non-clinical paranoia may be differentiated from controls in this regard. Nonetheless, the observed correlations between responding on the ‘Self’ IRAP, self-reported ‘Me-Safe’ and ‘Me-Vulnerable’ ratings, and BCSS scores suggest a more complex pattern of relational responding to the self, with self-evaluations on these measures shown to be interrelated and perhaps reflective of a more elaborate relational network involving the self as ‘safe’ and ‘vulnerable’. For instance, the IRAP showed that stronger ‘Me-Vulnerable’ disagreement was associated with stronger ‘Not me-Safe’ disagreement. Also, stronger ‘Not me-Vulnerable’ agreement on the IRAP was associated with higher self-reported ‘Me-Safe’ ratings. Furthermore, stronger ‘Me-Safe’ agreement on the IRAP was associated with lower self-reported ‘Me-Vulnerable’ ratings and Negative-Self scores and higher self-reported ‘Me-Safe’ ratings and Positive-Self scores (although not significantly so). Interestingly, responses on the ‘Self’ IRAP were not associated with evaluations of Others, either on the ‘Others’ IRAP or BCSS Others scales. However, self-reported ‘Me-Safe’ ratings were positively associated with Positive-Others and negatively associated with Negative-Others, and ‘Me-Vulnerable’ ratings were positively associated with Negative-Others and negatively associated with Positive-Others scores.

Regarding responding on the ‘Others’ IRAP, all participants (regardless of paranoia level) showed significant ‘Others-Trustworthy’ and ‘Others-Devious’ agreement (relative to disagreement). Only participants with high non-clinical paranoia showed significant ‘Others not-Devious’ agreement, indicative of two co-occurring but seemingly contradictory patterns

to others as ‘devious’ in this group. However, participants with high non-clinical paranoia showed significantly higher Negative-Others and lower Positive-Others scores on the BCSS compared to controls, suggesting that they viewed others (in general) as more ‘devious’ and less ‘trustworthy’.

The absence of between-groups differences on the ‘Others’ IRAP or significant correlations between the ‘Others’ IRAP trial-types and Others scales of the BCSS makes it difficult to draw any firm inferences about the observed responses on the ‘Others’ IRAP. In addition, there were no associations between the ‘Others’ IRAP and Paranoia Checklist or other self-report measures. However, the Others scales of the BCSS correlated in the expected direction with the other self-report measures. Specifically, higher Negative-Others scores were associated with higher Paranoia Checklist total and subscale scores, self-reported ‘Me-Vulnerable’ ratings, and Negative-Self scores, and with lower ‘Me-Safe’ ratings, Positive-Self, Positive-Others, and RSES scores. Higher Positive-Others scores were significantly associated with lower Paranoia Checklist total and subscale scores, ‘Me-Vulnerable’ ratings, and Negative-Self scores, and with higher ‘Me-Safe’ ratings, Positive-Self scores, and RSES scores.

One possible explanation for the lack of between-groups differences on the ‘Others’ IRAP and correlations between the ‘Others’ IRAP and self-report measures is that, while the intended conceptualization of “Others” within this IRAP was specified for participants (i.e., “strangers on the street”), this categorization may have been too broad and thus elicited both positive and negative functions for “Others”. Hence, either the conceptualization of “Others” needed to be more specific and relevant to paranoia or the expected pattern of responding to others as ‘devious’ in high non-clinical paranoia may only be elicited within particular contexts (e.g., following adverse interpersonal events).

In summary, the findings from Study 4 show that non-clinical paranoia is associated with specific patterns of relational responding to the self and others and that responding to the self as ‘vulnerable’ may be particularly relevant to high non-clinical paranoia, corroborating findings from previous research with similar populations (e.g., Fowler et al., 2006; Gracie et al., 2007; Oliver et al., 2011). Cognitive-clinical theories have emphasized the relevance of specific evaluations of the self (as ‘vulnerable’) and others (as ‘hostile’) to paranoia (e.g., Freeman et al., 2002; Freeman & Garety, 2014; Kesting & Lincoln, 2013; Salvatore et al., 2012). For example, Salvatore et al. proposed that paranoia may explain feelings of vulnerability by reasoning that this is because others are hostile. They also suggested that activation of the threat system (e.g., within stressful interpersonal situations) may function to reduce one’s perceived vulnerability and reinforce beliefs that others have hostile intentions. However, in doing so, this further reinforces the notion of oneself as vulnerable in relation to others (e.g., “I am threatened by others”). In a similar vein, Kesting and Lincoln (2013) proposed that negative self-beliefs, perceiving oneself as low in social rank, and evaluating others negatively is likely shaped by adverse interpersonal experiences, which may be reactivated when the individual is exposed to social stress in future.

From the RFT perspective, such responses to the self as ‘vulnerable’ and others as ‘devious’ may be initially established by direct adverse interpersonal experiences. These relations may subsequently be elaborated (via AARR) into more complex relational networks involving paranoid beliefs, such as those suggested by Salvatore and colleagues. However, from the functional-analytic perspective, these are considered behavior-behavior relations and such relational networks are not considered causal in terms of paranoia or related behavior. Rather, these relational networks may be reinforced by their coherence (i.e., ‘sense-making’) and may be strengthened or made more coherent with increased derivation. Furthermore, adverse interpersonal events likely increase the probability and coherence of responding to

the self as 'vulnerable' and to others as 'devious' and reduce the probability of responding to the self as 'safe' and others as 'trustworthy', as the latter would be incoherent within contexts where one is threatened by others. Hence, the next step in this research will be to examine the influence of such events on these particular patterns of relational responding to the self and others, paranoia, and the relationships between them.

## **Chapter 6: The effect of social exclusion on state paranoia, self-reported safety and vulnerability, and relational responding to the self as ‘safe’ and ‘vulnerable’ on the IRAP**

### **6.1. Introduction**

The findings from Study 4 indicated a potentially complex relational network involving interrelated patterns of relational responding to the self (e.g., as vulnerable, safe) and others (e.g., as devious, trustworthy) in relation to non-clinical paranoia. The findings also highlighted the particular importance of responding to the self as ‘vulnerable’ in this regard. These patterns of responding may be initially established by direct aversive experiences and subsequently elaborated into relational networks via arbitrarily applicable relational responding (AARR). Indeed, a wealth of evidence has shown that psychosis generally and paranoia specifically are associated with adverse interpersonal experiences (e.g., interpersonal trauma, neglect, bullying; Bebbington et al., 2004; Bentall et al., 2014; Read, Agar, Argyle, & Aderhold, 2003; Read, van Os, Morrison, & Ross, 2005). Furthermore, some studies have indicated that the effect of these experiences on paranoia may be influenced or mediated by negative beliefs about the self and others (e.g., Bentall & Fernyhough, 2009; Campbell & Morrison, 2007; Fisher, Appiah-Kusi, Grant, 2012; Gracie et al., 2007; Morrison, Frame, & Larkin, 2003; Murphy et al., 2015). For example, Murphy et al. found that trauma-related cognitions characterized by negative self-other evaluations mediated the relationship between adverse early life experiences and current psychotic experiences. These ‘belief systems’ (or relational networks) may in turn influence how individuals interact with their social environment (see Freeman et al., 2002; Garety et al., 2001; Kesting & Lincoln, 2013; Salvatore et al., 2012).

From a functional-analytic perspective, the identification of relevant environmental factors (e.g., adverse social events) is critical for predicting-and-influencing paranoia and

related behavior. We propose that adverse events may influence paranoia and related relational responding to the self and others in three ways. Firstly, and as mentioned previously, such events may influence the development of these relational networks. This may serve an adaptive function as, for example, threat-anticipation may help individuals to detect and avoid future interpersonal harm. Secondly, once established, these networks may interact with environmental factors to influence or increase the probability of paranoia. That is, ‘paranoia-prone’ individuals who typically view themselves as ‘vulnerable’ may be more likely to respond to adverse interpersonal events with paranoia than those without such histories. Thirdly, adverse interpersonal events may also strengthen or reinforce these relational networks, particularly for paranoia-prone individuals, as responding to such events with increased paranoia and responding to the self as ‘vulnerable’ and others as ‘devious’ may be considered coherent within such contexts and with their prior learning histories (e.g., “I was right. I am vulnerable. You can’t trust anyone.”). Importantly, these responses may be considered adaptive within adverse interpersonal contexts. However, they become maladaptive when they generalize to contexts where there is no interpersonal threat.

### **6.1.1. Hypotheses**

The aim of Study 5 was to investigate the influence of an adverse interpersonal event on paranoia and specific patterns of responding to the self (as safe and vulnerable) and others (as trustworthy and devious). As social exclusion using ‘Cyberball’ was shown to be a suitable and effective paradigm for increasing state paranoia in Study 3, we decided to use it for the purposes of Study 5. Specifically, in Study 5 we examined the influence of social exclusion on state paranoia, responding to the self as ‘vulnerable’ and ‘safe’ on the IRAP, and their interaction. In addition, we investigated the effect of exclusion on self-reported ‘Me-Safe’ and ‘Me-Vulnerable’ ratings, Positive-Self and Negative-Self on the Brief Core Schemas Scales (BCSS; Fowler et al., 2006), and self-esteem on the Rosenberg Self-Esteem

Scale (RSES; Rosenberg, 1965). We also investigated the effect of exclusion on self-reported responding to others as ‘trustworthy’ and ‘devious’ using the Positive-Others and Negative-Others scales of the BCSS respectively.

As per Study 3, we hypothesized that exclusion would lead to increased state paranoia and that the effect of exclusion on state paranoia would be moderated by ‘trait’ paranoia. In other words, we expected that individuals’ learning histories would influence responding in the current context such that those with a history of responding to adverse interpersonal events in a paranoid manner would be more likely to do so following exclusion than individuals without such histories. Similar to Study 3, we expected that exclusion would be associated with reduced self-reported self-esteem and that changes in self-reported self-esteem following exclusion would be associated with changes in state paranoia. Based on the finding that exclusion was associated with increased ‘Me-Negative’ responding on the IRAP in Study 3, we expected that excluded participants would show increased Negative-Self responding on the BCSS. However, given previous evidence that positive self-evaluations may be maintained in paranoia (Kesting & Lincoln, 2013) and the findings that ‘Me-Positive’ responding on the IRAP was unaffected by adverse events in Studies 2 and 3, we did not expect exclusion to influence Positive-Self responding on the BCSS. We expected that exclusion would be associated with increased Negative-Others and reduced Positive-Others on the BCSS.

Salvatore and colleagues (2012) argued that “it is as much the absence of safeness as the presence of threat that activates threat-based social mentalities” in relation to paranoia (p. 253). Hence, we expected that exclusion would impact relational responding to the self as ‘safe’ and ‘vulnerable’ on both the IRAP and self-reported ‘Me-Safe’ and ‘Me-Vulnerable’ ratings. Specifically, we expected exclusion to lead to increased vulnerability and reduced safety on both measures.

## 6.2. Method

### 6.2.1. Participants

One hundred volunteers (23 males, 77 females; *Mean age* = 20.96, *SD* = 4.11) participated. Participants were predominantly NUI Galway and University College Dublin (UCD) undergraduate Psychology students who took part in exchange for course credit. Fifty participants were randomly allocated to the ‘Exclusion’ (experimental) condition and 50 to the ‘Inclusion’ (control) condition.

### 6.2.2. Social exclusion using Cyberball

All features of Cyberball (version 4.0; Williams et al., 2000) used in Study 5 were identical to those in Study 3.

### 6.2.3. Measures

**Paranoia Checklist.** ‘Trait’ paranoia was measured using the Paranoia Checklist (Freeman et al., 2005). The scale showed excellent internal reliability overall ( $\alpha = .951$ ), as well as on the Frequency ( $\alpha = .877$ ), Conviction ( $\alpha = .873$ ) and Distress ( $\alpha = .923$ ) subscales in the current sample.

**State-adapted Paranoia Checklist.** Changes in state paranoia post-Cyberball were measured using the six-item state-adapted Paranoia Checklist (Westermann et al., 2012). The sum score of these items were used as an index of ‘state’ paranoia.

**State-adapted Brief Core Schema Scales (BCSS).** The 24-item BCSS (Fowler et al., 2006) was used to measure self-reported beliefs about the self and others as positive and negative. For the purposes of the current study, the BCSS was state-adapted to capture current beliefs about the self and others. Participants were instructed “Try to judge the beliefs on how you currently view yourself and others *at the moment*.” The subscales showed excellent internal reliability for Negative-Self ( $\alpha = .755$ ), Positive-Self ( $\alpha = .847$ ), Negative-Others ( $\alpha = .850$ ), and Positive-Others ( $\alpha = .860$ ).

**State-adapted self-reported ‘Me-Safe’ and ‘Me-Vulnerable’ ratings.** The ‘Self’ IRAP developed for Study 4 was used for the purposes of the current study (see Table 5.1). The items from the IRAP were rated using the same state-adapted response format as the BCSS. The self-report scales showed excellent internal reliability for ‘Me-Safe’ ( $\alpha = .836$ ) and ‘Me-Vulnerable’ ( $\alpha = .785$ ) in this sample.

**State-adapted Rosenberg Self-Esteem Scale (RSES).** For the purposes of the current study, the RSES was state-adapted so that participants were asked to rate their level of agreement with each statement ‘*at the moment*’. The internal reliability was excellent ( $\alpha = .897$ ) in the current sample.

**Cyberball Questionnaire.** The Cyberball Questionnaire (Williams, et al., 2000; Williams, 2009) was used to examine between-groups differences in need satisfaction and mood, and as a manipulation check following exposure to Cyberball.

**The IRAP.** The ‘Self’ IRAP developed for Study 4 was used for the purposes of the current study as this IRAP targeted relational responding to the self as ‘safe’ and ‘vulnerable’ specifically (see Table 5.1). The format, criteria, and instructions for completing the IRAP were identical to that those for completion of the ‘Self’ IRAP in Study 4. IRAP block order was counter-balanced across participants. The researcher stayed with the participant during the practice phase and then moved behind a screen once they achieved criteria to proceed to the test phase. If the participant failed to achieve criteria during the practice phase, their participation was ended then and they were thanked and debriefed.

#### **6.2.4. General Procedure**

At baseline, participants completed the IRAP followed by self-reported ‘Me-Safe’ and ‘Me-Vulnerable’ ratings, BCSS, RSES, the Paranoia Checklist, and the state-adapted Paranoia Checklist (in this order). Participants were then randomly allocated to either the Inclusion or Exclusion condition of Cyberball. Immediately after Cyberball, participants

completed the IRAP again, followed by the self-reported 'Me-Safe' and 'Me-Vulnerable' ratings, BCSS, RSES, state-adapted Paranoia Checklist, and the Cyberball questionnaire. In order to maintain the effect of Cyberball during completion of all subsequent measures, an A5 paper screenshot of the three avatars from Cyberball (with the ball being thrown between the two other 'players') was placed between the participant and the keyboard for the remainder of the study. Participants were also prompted by the researcher to mentally reflect on their experiences of the game between each of the measures.

#### **6.2.5. Statistical analysis plan**

One-way ANOVAs checked for potential block order effects on IRAP trial-types. One-way ANOVAs checked for potential between-groups differences (Exclusion vs. Inclusion) on all variables at baseline. Bonferroni correction for multiple comparisons ( $\alpha = .05 / 16 \approx .003$ ) was used. One-way ANOVAs also checked for between-groups differences on the Cyberball Questionnaire post-Cyberball ( $\alpha = .05 / 13 \approx .004$ ). Effects of social exclusion on state paranoia, self-reported 'Me-Safe' and 'Me-Vulnerable' ratings, BCSS, RSES, and relational responding on the IRAP from baseline to post-Cyberball were examined using mixed factorial ANOVAs with post-hoc tests. Cyberball condition (Inclusion vs. Exclusion) was the between-groups variable and time (baseline vs. post-Cyberball) the within-group variable. Correlation analyses were conducted to examine associations between trait and state paranoia, self-reported 'Me-Safe' and 'Me-Vulnerable' ratings, BCSS, RSES, and relational responding on the IRAP at baseline, and to examine associations between trait paranoia and changes in state paranoia, self-reported 'Me-Safe' and 'Me-Vulnerable' ratings, BCSS, RSES and relational responding on the IRAP (from baseline to post-Cyberball) (Bonferroni correction was used for each correlation analysis).

## 6.3. Results

### 6.3.1. Data preparation

For the purposes of data analysis, IRAP accuracy criterion was set at  $\geq 71\%$  (latency remained at  $\leq 2000\text{ms}$ )<sup>22</sup>. Five participants failed to meet criteria during the IRAP practice phase and three failed to maintain accuracy criterion during the IRAP test phase. Two participants withdrew, one participant's data was lost due to a technical issue<sup>23</sup>, and one was excluded due to familiarity with Cyberball. In their study, Kesting et al. (2013) excluded participants who correctly guessed that the purpose of the study was to investigate the effects of exclusion on paranoia. Seven participants (4 in the Inclusion and 3 in the Exclusion conditions) correctly guessed that the purpose of the study was to investigate paranoia (open-ended feedback questions were included at the end of the study within the automated questionnaires) and were excluded. Hence, 81 participants were included in subsequent data analyses: 39 in the Inclusion condition (9 males, 30 females; *Mean age* = 20.54, *SD* = 3.55) and 42 in the Exclusion condition (9 males, 33 females; *Mean age* = 21.26, *SD* = 4.80). No significant age or gender differences were observed between the Inclusion and Exclusion groups.

Kolmogorov-Smirnov tests indicated that the assumptions of normality had been violated for self-reported 'Me-Vulnerable' ratings, Negative-Self, Negative-Others, Paranoia Checklist subscales (but not total score), and state paranoia (all  $ps \leq .001$ ). Inspection of histograms indicated positive skew on all measures, with most participants showing low scores on each, which was expected given that the sample was recruited from the general,

---

<sup>22</sup> 9 participants scored  $\geq 71\%$  accuracy on at least one block during the test phase and were included in the final analysis. Participants who scored  $< 71\%$  during the test phase were removed from subsequent analyses.

<sup>23</sup> The automated questionnaires 'crashed' during the experiment and all questionnaire data was lost for this participant.

non-clinical population. However, skewness and kurtosis were within the acceptable range of  $\pm 2$ <sup>24</sup>.

### 6.3.2. Sample characteristics at baseline

One-way between-groups ANOVAs (with Bonferroni correction,  $\alpha = .003$ ) indicated no significant between-groups differences on measures of trait and state paranoia, self-reported 'Me-Vulnerable' ratings, BCSS, RSES, or IRAP trial-types at baseline (all  $ps \leq .031$ ). However, between-groups differences on self-reported 'Me-Safe' ratings trended towards significance ( $p = .005$ ), with included participants showing lower 'Me-Safe' ratings ( $M = 14.46$ ,  $SD = 3.85$ ) compared to their excluded counterparts ( $M = 17.12$ ,  $SD = 4.37$ ).

Correlational analyses (with Bonferroni correction,  $\alpha = .003$ ) explored the relationships between trait and state paranoia and self-reported 'Me-Safe' and 'Me-Vulnerable' ratings, BCSS, RSES, and IRAP trial-types at baseline (see Table 6.1). Significant positive correlations were observed between Paranoia Checklist total, Frequency and Conviction subscales, and state paranoia and 'Me-Vulnerable' ratings, Negative-Self, and Negative-Others, with higher trait paranoia and associated Frequency and Conviction and state paranoia associated with higher 'Me-Vulnerable' ratings, Negative-Self, and Negative-Others scores. There was also a significant positive correlation between the Distress subscale and 'Me-Vulnerable' ratings, with higher paranoia Distress associated with higher 'Me-Vulnerable' ratings. There were significant negative correlations between Paranoia Checklist total, Frequency and Conviction subscales and state paranoia and 'Me-Safe' ratings, with lower trait paranoia and associated Frequency and Conviction and state paranoia associated with higher 'Me-Safe' ratings. Paranoia Frequency was negatively associated with Positive-Self, with lower paranoia Frequency associated with higher Positive-Self scores. Paranoia Frequency and Conviction were negatively associated with Positive-Others, with lower

---

<sup>24</sup> Although kurtosis for Negative-Self was 3.445.

paranoia Frequency and Conviction associated with higher Positive-Others scores. Significant negative correlations were observed between Paranoia Checklist total and subscales and state paranoia and RSES, with higher trait and state paranoia associated with lower RSES scores. No significant associations were observed between the IRAP trial-types and the self-report measures.

**Table 6.1.** Correlations between measures of trait and state paranoia and IRAP trial-types, self-reported ‘Me-Safe’ and ‘Me-Vulnerable’ ratings, BCSS, and RSES scores at baseline ( $N = 81$ ).

	<b>Paranoia Total</b>	<b>Paranoia Frequency</b>	<b>Paranoia Conviction</b>	<b>Paranoia Distress</b>	<b>State Paranoia</b>
<b>IRAP Me-Safe</b>	-.040	.000	-.083	-.029	-.017
<b>IRAP Me-Vulnerable</b>	.003	.079	.010	-.053	.092
<b>IRAP Not me-Safe</b>	-.026	-.082	-.017	.012	-.002
<b>IRAP Not me-Vulnerable</b>	-.014	.011	.007	-.041	-.109
<b>Me-Safe ratings</b>	-.362**	-.404**	-.374**	-.237*	-.418**
<b>Me-Vulnerable ratings</b>	.491**	.521**	.436**	.387**	.446**
<b>BCSS Negative-Self</b>	.382**	.446**	.353**	.264*	.412**
<b>BCSS Positive-Self</b>	-.314*	-.387**	-.324*	-.182	-.256*
<b>BCSS Negative-Others</b>	.387**	.495**	.451**	.179	.533**
<b>BCSS Positive-Others</b>	-.262*	-.371**	-.349**	-.068	-.223*
<b>RSES</b>	-.469**	-.526**	-.409**	-.356**	-.413**

\*  $p < .05$  \*\*  $p \leq .001$  (prior to Bonferroni correction)

### 6.3.3. Manipulation check

One-way between-groups ANOVAs (Bonferroni correction,  $\alpha = .004$ ) assessed whether exclusion (versus inclusion) during Cyberball had the intended negative impact on mood and related variables. Excluded participants showed significantly less belonging, self-

esteem, meaningful existence, and control compared to their included counterparts (all  $ps \leq .004$ ). The Exclusion group also demonstrated greater embarrassment, anger, and frustration and less happiness than the Inclusion group (all  $ps < .001$ ). Differences in sadness approached significance ( $p = .008$ ). Excluded participants also reported greater feelings of being ignored and excluded and receiving fewer ball tosses ( $M = 6.55, SD = 3.55$ ) than included participants ( $M = 28.74, SD = 6.48$ ) (all  $ps < .001$ ).

#### **6.3.4. Effects of exclusion on state paranoia**

A mixed ANOVA explored between-groups differences in state-adapted Paranoia Checklist scores from baseline to post-Cyberball. Levene's test of equality of error variances indicated violation of normality assumptions at post-Cyberball ( $p < .002$ ), so results should be interpreted with caution<sup>25</sup>. There was a significant group X time interaction effect,  $F(1,79) = 7.735, p = .007, \eta^2 = .089$ , and a main effect for time,  $F(1,79) = 6.895, p = .01, \eta^2 = .080$ . Post-hoc comparisons indicated a significant increase in state paranoia from baseline ( $M = 8.17, SD = 2.45$ ) to post-Cyberball ( $M = 9.95, SD = 4.89$ ) for the Exclusion group only ( $p = .006$ ), and a significant between-groups difference at post-Cyberball ( $p = .051$ ).

#### **6.3.5. Moderating effect of trait paranoia on changes in state paranoia post-Cyberball**

To investigate whether the effect of Cyberball condition (Inclusion or Exclusion) on changes in state paranoia was moderated by proneness to paranoia, we performed three moderation analyses using PROCESS for SPSS (Hayes, 2016) with Cyberball condition, mean Frequency/ Distress/ Conviction score on the Paranoia Checklist at baseline and the interaction term of Cyberball condition X Frequency/ Distress/ Conviction as predictors and

---

<sup>25</sup> Inspection of state-adapted Paranoia Checklist scores for Inclusion and Exclusion at both time points indicated similar variation or dispersion in scores at baseline ( $SD = 2.40$  for Inclusion and  $SD = 2.45$  for Exclusion). However, there was greater variation in scores for Exclusion at post-Cyberball ( $SD = 4.89$ ) compared to Inclusion ( $SD = 2.46$ ), suggestive that while exclusion during Cyberball led to increased state paranoia, there was relatively large variation in scores in this group. The results from the moderation analysis indicated that changes in state paranoia scores following exclusion were moderated by trait paranoia (i.e., no effect for exclusion on state paranoia for participants with frequency scores  $< 24.56$ , conviction scores  $< 27.53$ , and distress scores  $< 28.58$ , but increasingly higher state paranoia scores for excluded participants who scored above these thresholds), which would partially explain this variation in post-Cyberball state paranoia scores in the Exclusion group.

changes in state paranoia (from baseline to post-Cyberball) as the dependent variable. Results indicated significant moderation effects for each dimension of trait paranoia. That is, there were significant Cyberball condition X Frequency ( $p = .007$ ), Cyberball condition X Conviction ( $p = .006$ ), and Cyberball condition X Distress ( $p = .054$ ) moderation effects.

These effects were probed using the Johnson-Neyman technique to identify regions of significance. Results revealed a significant positive effect for exclusion on state paranoia change scores for participants in the Exclusion condition with a Frequency score of  $\geq 24.56$ , which was increasingly strong for higher Frequency scores (i.e., participants with increasingly higher Frequency scores at  $\geq 24.56$  showed increasingly larger state paranoia change scores following exclusion). This applied to 59.26% of the sample. There was a significant positive effect for exclusion on state paranoia change scores for those in the Exclusion condition with a Conviction score of  $\geq 27.53$ , which was increasingly strong for higher Conviction scores (i.e., participants with increasingly higher Conviction scores at  $\geq 27.53$  showed increasingly larger state paranoia scores following exclusion). This applied to 60.49% of the sample. Finally, there was a significant positive effect for exclusion on state paranoia change scores for those in the Exclusion group with a Distress score of  $\geq 28.58$ , which was increasingly strong for higher Distress scores (i.e., participants with increasingly higher Distress scores at  $\geq 28.58$  showed increasingly larger state paranoia change scores following exclusion). This applied to 53.09% of the sample. Thus, exclusion during Cyberball led to an increase in state paranoia in participants who experienced paranoid thoughts more frequently and were more convinced of and distressed by paranoid thoughts at baseline.

### **6.3.6. Effects of exclusion on self-reported ‘Me-Safe’ and ‘Me-Vulnerable’ ratings**

Mixed ANOVAs investigated between-groups differences from baseline to post-Cyberball on ‘Me-Safe’ and ‘Me-Vulnerable’ ratings (see Figure 6.1). Levene’s test indicated

no violations of assumptions of equality of error variances for either set of ratings. There was a significant group X time interaction effect for ‘Me-Vulnerable’,  $F(1,79) = 4.037, p = .048, \eta^2 = .049$ , with included participants demonstrating reduced ‘Me-Vulnerable’ ratings from baseline ( $M = 5.38, SD = 3.48$ ) to post-Cyberball ( $M = 4.41, SD = 3.76$ ), and excluded participants demonstrating increased ratings from baseline ( $M = 4.81, SD = 3.83$ ) to post-Cyberball ( $M = 5.19, SD = 4.93$ ). There was a significant group X time interaction effect for ‘Me-Safe’,  $F(1,79) = 9.219, p = .003, \eta^2 = .104$ , with included participants demonstrating no changes in ‘Me-Safe’ ratings from baseline ( $M = 14.46, SD = 3.85$ ) to post-Cyberball ( $M = 14.51, SD = 4.28$ ), whereas excluded participants demonstrated reduced ratings from baseline ( $M = 17.12, SD = 4.37$ ) to post-Cyberball ( $M = 14.83, SD = 5.50$ ). There was also a main effect for time,  $F(1,79) = 8.427, p = .005, \eta^2 = .096$ . Post-hoc tests indicated a significant reduction in ‘Me-Safe’ ratings from baseline to post-Cyberball for excluded participants only ( $p = .001$ ).

### **6.3.7. Effects of exclusion on Negative- and Positive-Self and Others BCSS scores**

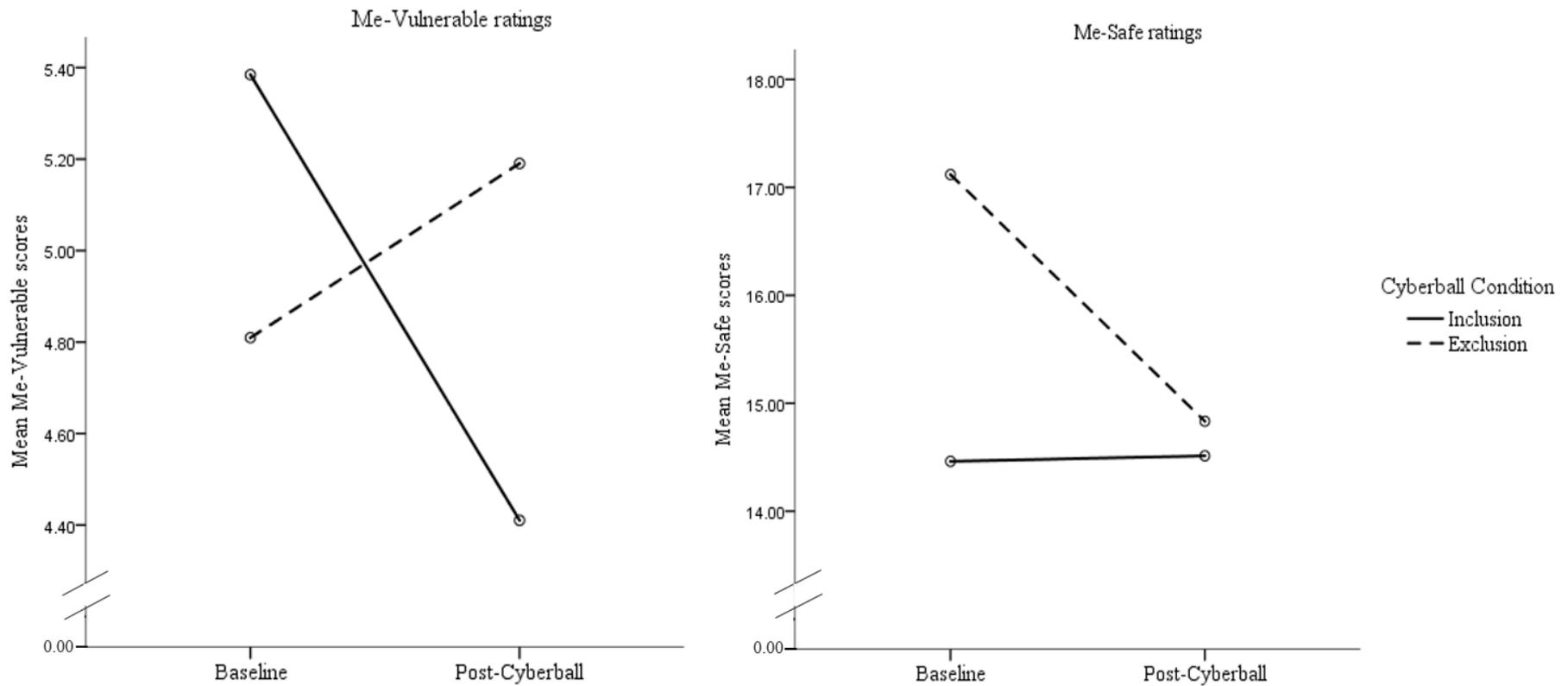
Mixed ANOVAs investigated between-groups differences from baseline to post-Cyberball in Negative-Self and Positive-Self (see Figure 6.2) and Negative-Others and Positive-Others (see Figure 6.3) BCSS scores. Levene’s test indicated no violation of the assumptions of equality of error variances for Negative-Self, Positive-Self, or Positive-Others. However, assumptions of equality of error variances were violated for Negative-Others at post-Cyberball ( $p = .045$ )<sup>26</sup>. Hence, results for Negative-Others should be interpreted with caution. On the Self scales, there was a significant group X time interaction

---

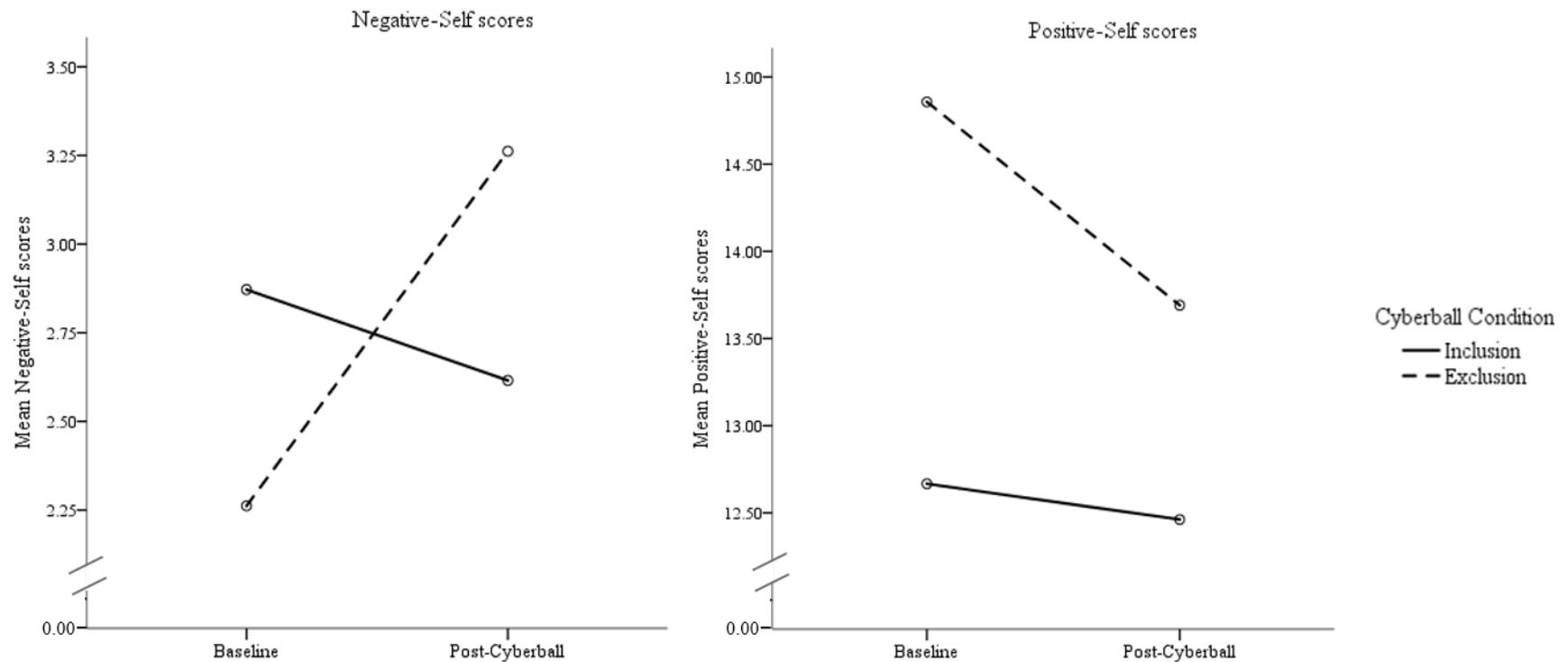
<sup>26</sup> Similar to the observed violations of equality of error variances for state-adapted Paranoia Checklist scores, closer inspection of Negative-Others scores indicated greater variance in scores for Exclusion ( $SD = 5.21$ ) relative to Inclusion ( $SD = 3.68$ ) at post-Cyberball. We investigated whether this might be partially due to the potential moderating effect of paranoia-proneness on the effect of exclusion on Negative-Others change scores. A Cyberball condition X Conviction trended towards significance ( $p = .082$ ). Johnson-Neyman technique revealed a significant positive effect for exclusion on Negative-Others change scores for participants in the Exclusion condition with a Conviction score of  $\geq 30.59$ , which was increasingly strong for higher Conviction scores (i.e., participants with increasingly higher Conviction scores at  $\geq 30.59$  showed increasingly larger Negative-Others change scores following exclusion). This applied to 43.21% of the sample.

effect for Negative-Self,  $F(1,79) = 4.165, p = .045, \eta^2 = .050$ , with included participants showing no changes in Negative-Self scores from baseline ( $M = 2.87, SD = 3.02$ ) to post-Cyberball ( $M = 2.62, SD = 3.28$ ), whereas excluded participants demonstrated an increase in scores from baseline ( $M = 2.26, SD = 2.74$ ) to post-Cyberball ( $M = 3.26, SD = 4.25$ ). There was a main effect for time for Positive-Self,  $F(1,79) = 4.740, p = .032, \eta^2 = .057$ , with significant decreases in scores for excluded participants from baseline ( $M = 14.86, SD = 4.75$ ) to post-Cyberball ( $M = 13.69, SD = 5.53$ ) only ( $p = .030$ ).

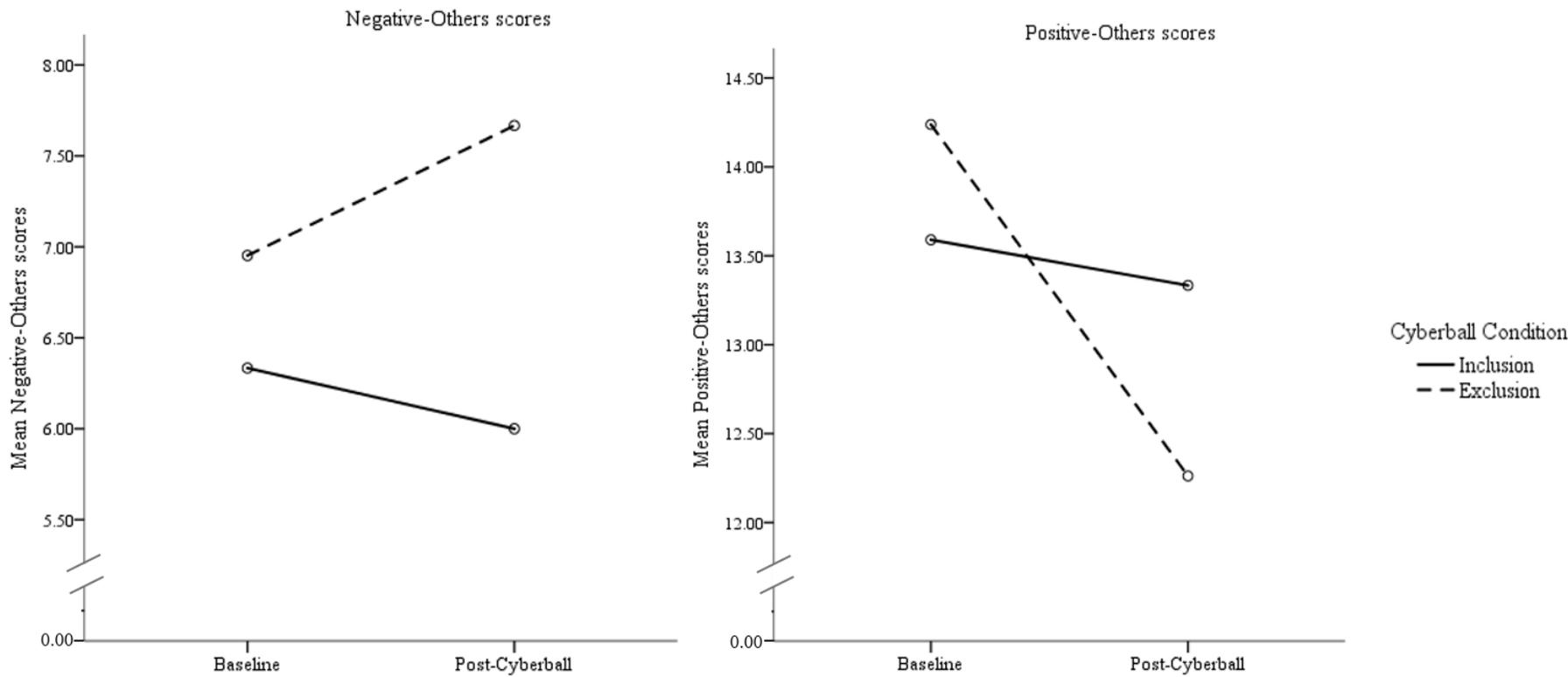
On the Others scales, there was a significant interaction effect for Positive-Others,  $F(1,79) = 8.942, p = .004, \eta^2 = .102$ , with included participants showing no changes in Positive-Others scores from baseline ( $M = 12.67, SD = 4.16$ ) to post-Cyberball ( $M = 12.46, SD = 4.35$ ), whereas excluded participants demonstrated decreased Positive-Others scores from baseline ( $M = 14.24, SD = 3.66$ ) to post-Cyberball ( $M = 12.26, SD = 4.47$ ). There was also a main effect for time,  $F(1,79) = 15.070, p < .001, \eta^2 = .160$ , with post-hoc tests showing significant decreases in Positive-Others scores from baseline to post-Cyberball for excluded participants only ( $p < .001$ ). There was no significant group X time interaction effect for Negative-Others,  $F(1, 79) = 2.813, p = .097, \eta^2 = .034$ , nor main effects for time ( $p = .544$ ) or group ( $p = .197$ ).



**Figure 6.1.** Changes in mean self-reported ‘Me-Vulnerable’ (left-hand side) and ‘Me-Safe’ (right-hand side) ratings from baseline to post-Cyberball for Inclusion and Exclusion groups. Please note differences in scales due to differences in mean rating scores; that is, while the same response format was used for ‘Me-Vulnerable’ and ‘Me-Safe’ ratings, higher scores on ‘Me-Vulnerable’ indicate greater ‘vulnerability’ whereas higher scores on ‘Me-Safe’ indicate greater ‘safety’.



**Figure 6.2.** Changes in mean Negative-Self (left-hand side) and Positive-Self (right-hand side) scores on the BCSS from baseline to post-Cyberball for Inclusion and Exclusion groups. Please note differences in scales due to differences in mean rating scores; that is, while the same response format was used for Negative-Self and Positive-Self scales, higher scores on Negative-Self indicate greater ‘negativity’ towards the self whereas higher scores on Positive-Self indicate greater ‘positivity’ towards the self.



**Figure 6.3.** Changes in mean Negative-Others (left-hand side) and Positive-Others (right-hand side) scores on the BCSS from baseline to post-Cyberball for Inclusion and Exclusion groups. Please note differences in scales due to differences in mean rating scores; that is, while the same response format was used for Negative-Others and Positive-Others scales, higher scores on Negative-Others indicate greater ‘negativity’ towards others whereas higher scores on Positive-Others indicate greater ‘positivity’ towards others.

### 6.3.8. Effects of exclusion on self-esteem

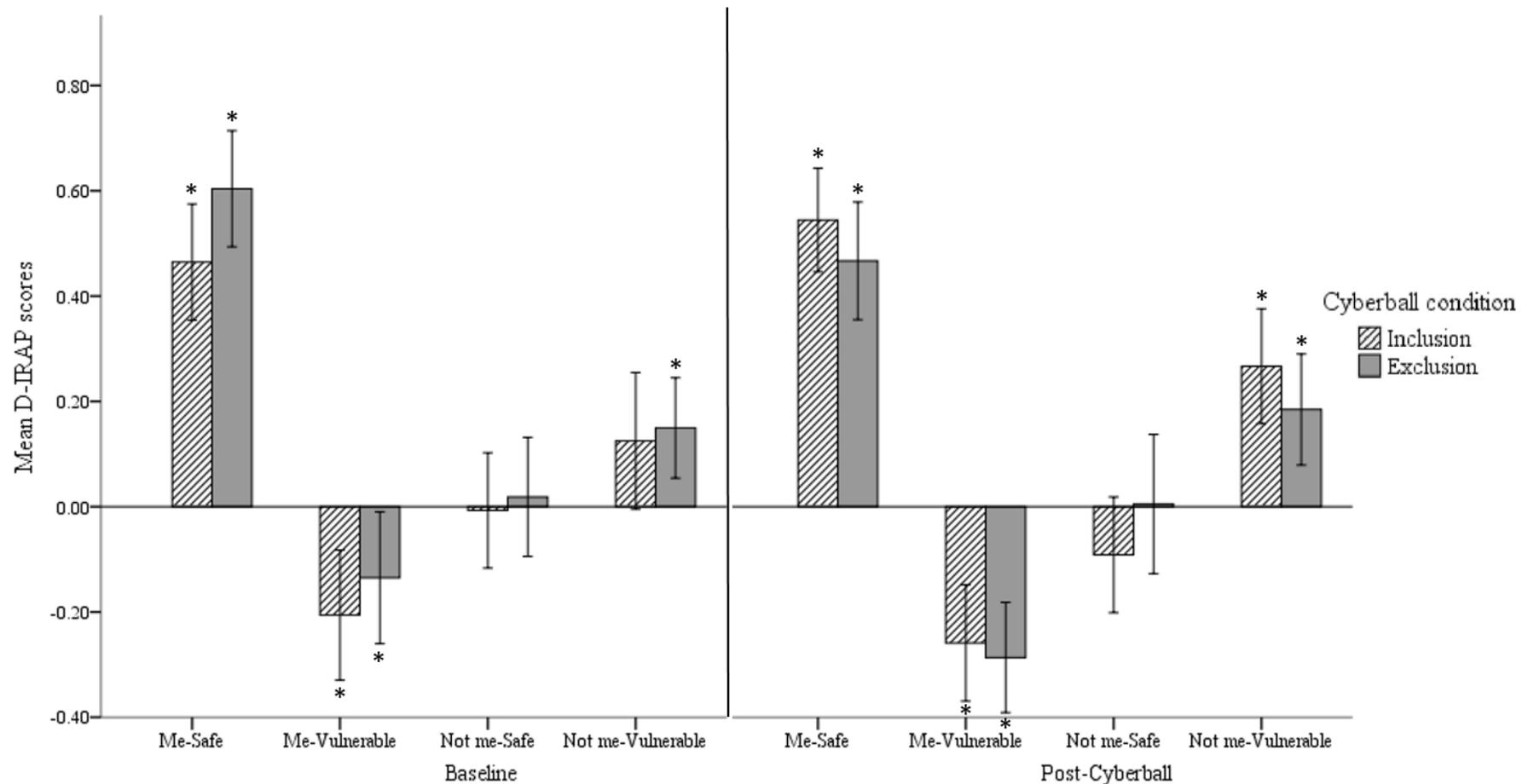
A mixed ANOVA indicated no between-groups differences from baseline to post-Cyberball for RSES scores. An interaction effect trended towards significance,  $F(1,79) = 3.023, p = .086$ , with included participants demonstrating increased self-esteem and excluded participants demonstrating decreased self-esteem from baseline to post-Cyberball but the effect size was small,  $\eta^2 = .037$ .

### 6.3.9. Effects of exclusion on IRAP responses

Mean D-IRAP scores for each of the four IRAP trial-types for the Inclusion and Exclusion groups at baseline and post-Cyberball are presented in Figure 6.4. One-sample *t*-tests indicated that, at baseline, all participants demonstrated significant effects on the ‘Me-Safe’ trial-type ( $ps < .001$ ), with faster agreement (relative to disagreement) for ‘Me-Safe’. All participants also showed significant effects on the ‘Me-Vulnerable’ trial-type ( $p = .002$  and  $p = .035$  for Inclusion and Exclusion groups respectively). Interestingly, participants demonstrated faster agreement (relative to disagreement) for ‘Me-Vulnerable’. Only excluded participants demonstrated a significant effect on the ‘Not me-Vulnerable’ trial-type ( $p = .003$ ), although this trended towards significance for the Inclusion group ( $p = .059$ ), with faster agreement for ‘Not me-Vulnerable’ (relative to disagreement). At post-Cyberball, both included and excluded groups demonstrated significant ‘Me-Safe’ agreement. All participants also showed significant ‘Me-Vulnerable’ agreement and significant ‘Not me-Vulnerable’ agreement on these trial-types at post-Cyberball (all  $ps \leq .001$ ). There were no effects on the ‘Not me-Safe’ trial-type at either time point.

Mixed ANOVAs investigated between-groups differences in relational responding on the IRAP from baseline to post-Cyberball. Levene’s test indicated no violations of assumptions of equality of error variances. There was a significant group X time interaction effect for ‘Me-Safe’,  $F(1,79) = 5.110, p = .027, \eta^2 = .061$ , with included participants

demonstrating increased and excluded participants demonstrating decreased ‘Me-Safe’ responding from baseline to post-Cyberball. On the ‘Me-Vulnerable’ trial-type, there was a significant main effect for time,  $F(1,79) = 4.246$ ,  $p = .043$ ,  $\eta^2 = .051$ , with all participants showing increased ‘Me-Vulnerable’ responding from baseline to post-Cyberball. Post-hoc analyses indicated a significant increase in ‘Me-Vulnerable’ responding for the Exclusion group only ( $p = .047$ ). There were no effects on the ‘Not me-Safe’ or ‘Not me-Vulnerable’ trial-types for either group.



**Figure 6.4.** D-IRAP scores for IRAP trial-types at baseline (left-hand side) and post-Cyberball (right-hand side) for the Inclusion and Exclusion groups, with standard error bars (95% confidence intervals). Note, D-IRAP scores above zero indicate faster responding on consistent (i.e., faster ‘Me-Safe’ agreement and ‘Me-Vulnerable’ disagreement) relative to inconsistent blocks of trials, whereas scores below zero indicate the reverse pattern (i.e., faster responding on inconsistent trials). Asterisks represent trial-types significantly different from zero.

### 6.3.10. Correlational analyses

Correlational analyses (Bonferroni correction,  $\alpha = .004$ ) were conducted to examine the relationships between trait paranoia and changes in state paranoia, self-reported 'Me-Safe' and 'Me-Vulnerable' ratings, Negative-Self, Positive-Self, Negative-Others, and Positive-Others scales of the BCSS, RSES, and relational responding on the 'Me-Safe' and 'Me-Vulnerable' IRAP trial-types (see Table 6.2). Change scores were calculated for state paranoia, 'Me-Safe' and 'Me-Vulnerable' ratings, BCSS, RSES, and the IRAP trial-types by subtracting baseline scores from post-Cyberball scores. In relation to trait and state paranoia, state paranoia change scores were positively associated with Frequency ( $r = .289, p = .009$ ), Conviction ( $r = .248, p = .025$ ) and Distress subscales ( $r = .234, p = .036$ ) and Paranoia Checklist total scores ( $r = .284, p = .010$ ), although not significantly so. Changes in state paranoia were also positively associated with changes in 'Me-Vulnerable' ratings, Negative-Self, and Negative-Others, and negatively associated with changes in 'Me-Safe' ratings, Positive-Self, Positive-Others, and RSES. Conviction was positively associated with changes in 'Me-Vulnerable' ratings, Negative-Self, and Negative-Others. Distress was positively associated with changes in Others-Negative. Paranoia Checklist total was positively associated with 'Me-Vulnerable' ratings and Negative-Self.

In relation to responding to the self as 'safe' and 'vulnerable', changes in 'Me-Safe' ratings were negatively associated with changes in 'Me-Vulnerable' ratings, Negative-Self, and Negative-Others, and positively associated with changes in Positive-Self, Positive-Others, and RSES. Changes in 'Me-Vulnerable' ratings were positively associated with changes in Negative-Self, and Negative-Others, and negatively associated with changes in Positive-Self, Positive-Others, and RSES. There were no significant correlations among changes in relational responding on the IRAP trial-types and change scores on any of the self-report measures.

**Table 6.2.** Correlations between trait paranoia, state paranoia change scores (from baseline to post-Cyberball) and change scores for Me-Safe, Me-Vulnerable, BCSS, RSES and IRAP trial-types ( $N = 81$ ).

	<b>IRAP Me-Safe</b>	<b>IRAP Me-Vulnerable</b>	<b>Me-Safe</b>	<b>Me-Vulnerable</b>	<b>BCSS Neg-Self</b>	<b>BCSS Pos-Self</b>	<b>BCSS Neg-Others</b>	<b>BCSS Pos-Others</b>	<b>RSES</b>
<b>State Paranoia</b>	-.158	-.102	-.490**	.633**	.710**	-.349**	.506**	-.504**	-.518**
<b>Paranoia Frequency</b>	-.170	.013	-.037	.252*	.307*	-.078	.201	-.070	-.230*
<b>Paranoia Conviction</b>	-.126	.096	-.093	.359**	.380**	-.176	.332**	-.038	-.279*
<b>Paranoia Distress</b>	-.155	.136	-.174	.247*	.292*	-.202	.293**	-.119	-.259*
<b>Paranoia Total</b>	-.169	.100	-.125	.313**	.359**	-.178	.310*	-.092	-.286*
<b>RSES</b>	.116	-.134	.506**	-.559**	-.581**	.467**	-.210	.420**	
<b>BCSS Pos-Others</b>	.168	-.047	.618**	-.423**	-.471**	.458**	-.431**		
<b>BCSS Neg-Others</b>	-.104	-.081	-.378**	.461**	.434**	-.370**			
<b>BCSS Pos-Self</b>	-.109	-.173	.649**	-.325**	-.416**				
<b>BCSS Neg-Self</b>	-.121	-.101	-.452**	.685**					
<b>Me-Vulnerable</b>	-.117	.104	-.432**						
<b>Me-Safe</b>	.110	-.140							
<b>IRAP Me-Vulnerable</b>	.185								

\*  $p < .05$  \*\*  $p \leq .002$  (prior to Bonferroni correction)

## 6.4. Discussion

Study 5 investigated the effect of social exclusion during Cyberball on state paranoia, relational responding to the self (as ‘safe’ and ‘vulnerable’) and others (as ‘trustworthy’ and ‘devious’), and the relationships between them. In line with predictions and the findings from Study 3, exclusion led to increased state paranoia and this effect was moderated by the frequency, conviction, and distress associated with experiencing paranoia in general at baseline.

An interesting pattern emerged regarding responding to the self as ‘vulnerable’ on the IRAP and self-report ratings. On the IRAP, all participants showed significant ‘Me-Vulnerable’ agreement at baseline, which increased following Cyberball (though this increase was only significant for Exclusion). This was unexpected at baseline, and at post-Cyberball for included participants. However, it could be considered coherent or consistent within this context. That is, participants were told that they would be playing an online game with strangers, hence vulnerability could be considered a ‘natural’ response in such settings. Nonetheless, it is interesting that included participants continued to demonstrate significant ‘Me-Vulnerable’ agreement at post-Cyberball. Generally, their feedback indicated that they viewed the other ‘players’ positively (although some reported that they felt the other players were not very engaging). Yet, included participants showed reduced self-reported ‘Me-Vulnerable’ ratings from baseline to post-Cyberball. ‘Me-Vulnerable’ ratings and responding on the IRAP mirrored each other for excluded participants, however, with this group showing increased ‘Me-Vulnerable’ ratings from baseline to post-Cyberball.

These differential patterns of responding on the IRAP and self-report ratings for the two groups might reflect differences in terms of brief and immediate relational responses (BIRRs) and extended and elaborated relational responses (EERRs) that could be considered coherent within each context (i.e., Inclusion and Exclusion). That is, while included

participants' BIRRs involved 'Me-Vulnerable' agreement, their subsequent EERRs (i.e., reduced 'Me-Vulnerable' ratings) could reflect additional and contradictory relational responses based on their experiences of Cyberball (e.g., "The other players included me", "They were fair"). In contrast, excluded participants' experiences may have confirmed their BIRRs (e.g., "The other players excluded me", "They were mean"), a suggestion supported by the observed correspondence in 'Me-Vulnerable' responding on both measures in this group.

All participants showed significant 'Me-Safe' agreement on the IRAP at baseline and at post-Cyberball, demonstrating that the IRAP captured individual but co-existing patterns of relational responding to the self as both 'safe' and 'vulnerable'. Included participants demonstrated increased 'Me-Safe' agreement on the IRAP and no changes in self-reported 'Me-Safe' ratings from baseline to post-Cyberball. In contrast, excluded participants demonstrated reduced 'Me-Safe' agreement on the IRAP and lower 'Me-Safe' ratings from baseline to post-Cyberball. In other words, while excluded participants demonstrated significant 'Me-Safe' agreement on the IRAP at both time points, the strength of this relational response reduced following exclusion. Similarly, excluded participants still demonstrated relatively high 'Me-Safe' ratings but these were reduced following exclusion. Also, increased state paranoia was associated with increased 'Me-Vulnerable' ratings and reduced 'Me-Safe' ratings, and changes in 'Me-Vulnerable' and 'Me-Safe' ratings were negatively associated with each other. These findings suggest that vulnerability, reduced safety, and the presence of interpersonal threat are pertinent to paranoia (see Freeman, 2016; Salvatore et al., 2012).

Excluded participants showed reduced RSES scores from baseline to post-Cyberball, whereas included participants showed increased RSES scores (although this did not reach statistical significance). Excluded participants also showed increased Negative-Self scores

and reduced Positive-Self scores from baseline to post-Cyberball, whereas included participants showed no changes in either scores. Changes in RSES, Negative-Self, and Positive-Self scores were associated with changes in state paranoia, with increased state paranoia associated with increased Negative-Self and decreased Positive-Self and RSES scores. These findings highlight the utility of differentiating between specific (e.g., positive and negative) self-evaluations in the context of paranoia as this provides more detailed information regarding the effect of adverse experiences on responding to the self. Furthermore, changes in RSES, Negative-Self, Positive-Self, and self-reported 'Me-Safe' and 'Me-Vulnerable' change scores also correlated with each other in the expected directions, suggesting a complex relational network regarding responding to the self in this context.

Exclusion also led to significant decreases in Positive-Others scores, with excluded participants rating Others as less "trustworthy" and "good" following Cyberball. Changes in Positive-Others scores were negatively associated with changes in state paranoia from baseline to post-Cyberball. There was also a small increase in Negative-Others scores following Exclusion but this was non-significant. There was large variation in Negative-Others scores for excluded participants at post-Cyberball, suggesting that Exclusion had varying effects on Negative-Others responding. The effect of exclusion on Negative-Others scores was moderated by the conviction associated with experiencing paranoid thoughts in general at baseline, which may explain some of the variance in these scores. Changes in Negative-Others scores were positively associated with changes in state paranoia from baseline to post-Cyberball. This could be considered in terms of coherent behavior-behavior relations within this context, with exclusion impacting relational responding regarding paranoia, responding to others as 'negative', and the relationship between the two. Indeed, items on the Negative-Others scale include responding to others as "hostile" and "bad", which could be considered consistent with items from the state-adapted Paranoia Checklist

(e.g., “I am under threat from others”). Indeed, Negative-Others responding could be considered an inherent component of paranoia. Put simply, paranoia entails responding to others as ‘negative’.

The findings that the dimensions of frequency, conviction and distress associated with experiencing paranoia in general moderated the effect of exclusion on state paranoia is consistent with the findings from Study 3 and previous research by Kesting et al. (2013), whereby distress was found to moderate the effect of exclusion on state paranoia. As mentioned above, we also found that conviction moderated the effect of exclusion on Negative-Others responding. Furthermore, changes in ‘Me-Vulnerable’ ratings, Negative-Self, and Negative-Others scores were associated with conviction and with changes in state paranoia. Importantly, these findings support the need to adopt a multi-dimensional approach to paranoia as factors such as the conviction and distress associated with experiencing paranoia may influence the effects that adverse interpersonal experiences have on relational responding regarding paranoia and responding to the self and others.

Interestingly, Garety et al. (2001) proposed that paranoid beliefs may be held with greater conviction if they are consistent with firmly-held beliefs about the self (e.g., as ‘weak’ or ‘vulnerable’) and others (e.g., as ‘untrustworthy’ or ‘hostile’). They reasoned that, once formed, paranoid ideas may be considered further confirmation of these negative self and others beliefs and might contribute to the persistence of paranoia. From an RFT perspective though, these are considered behavior-behavior relations. Such relations are influenced by historical (e.g., one’s prior learning history) and current environmental variables (e.g., adverse interpersonal events) and may be reinforced by their coherence or ‘sense-making’. Indeed, the moderating influence of paranoia-proneness (i.e., frequency, conviction, and distress associated with experiencing paranoia generally) on the effect of exclusion on state paranoia and Negative-Others responding provides some support in this regard. Put simply,

individuals who experienced paranoid thoughts more frequently and were more convinced of and distressed by them were more likely to respond to this adverse event (i.e., social exclusion) in a paranoid manner and to others as ‘negative’ – a pattern of relational responding that could be deemed coherent or consistent with this context (i.e., Exclusion).

There were no significant associations between responding on the IRAP and self-report measures of paranoia or responding to the self and others. The reason for this is unclear. The stimuli incorporated into the IRAP were the exact same as the self-reported ‘Me-Safe’ and ‘Me-Vulnerable’ ratings. Moreover, the patterns of responding on the ‘Me-Safe’ and ‘Me-Vulnerable’ IRAP trial-types and corresponding self-report ratings directly mirrored each other for excluded participants. Also, in Study 4, ‘Me-Vulnerable’ responding on the IRAP was associated with conviction and ‘Not me-Safe’ responding was associated with Paranoia Checklist total and distress (though not significantly so). At the same time, however, there were no significant associations between the ‘Self’ IRAP trial-types and the other self-report measures in Study 4. Although, Study 4 involved participants who were high in non-clinical paranoia and controls; hence the observed (non-significant) associations between responding on the IRAP and measures of paranoia may have been driven by participants with high non-clinical paranoia. Perhaps the observed lack of associations between the ‘Me-Vulnerable’ and ‘Me-Safe’ IRAP trial-types and self-report ratings in Study 5 (despite correspondence between them for excluded participants in particular) was due to the comparatively low levels of trait paranoia ( $M = 91.74$ ,  $SD = 27.25$ ) within this sample.

## Chapter 7: General Discussion

### 7.1. Overview

The current thesis presents a first step towards bringing a functional-analytic perspective to the study of non-clinical paranoia in the general population. Relational Frame Theory (RFT), a functional-analytic theory of human language and cognition, and an RFT-based measure – the Implicit Relational Assessment Procedure (IRAP) – provided the theoretical framework and a methodology with which to do so. Over the past two decades, RFT has expanded rapidly into many areas of important concern, including psychological suffering (see Dymond & Roche, 2013; Stewart, 2016; Törneke, 2010) and paranoia specifically (Stewart et al., 2016). From this perspective, we define paranoia and related phenomena (e.g., self-evaluations) functionally as behaviors, and specifically as arbitrarily applicable relational responding (AARR). The IRAP has been used to target repertoires of AARR that have been established by prior learning histories, including those involved in psychological suffering, and has been increasingly used in experimental-clinical research (see Vahey et al., 2015, for a meta-analysis). Much of this research has sought to identify repertoires of AARR relevant to clinically-relevant phenomena (e.g., dysphoria, OCD, non-clinical voice-hearing).

We began by drawing upon the existing cognitive-clinical literature to help determine which relational responses may be pertinent to non-clinical paranoia. Relational responding to the self was identified as a key factor in this regard, with self-evaluations such as self-esteem and other specific self-evaluations (e.g., as negative, vulnerable) featuring in prominent cognitive-clinical models of paranoia (e.g., Bentall et al., 1994; 2001; Freeman et al., 2002; Garety et al., 2001) and research in this domain (e.g., Fowler et al., 2006; Freeman, 2016; Kesting & Lincoln, 2013; Salvatore et al., 2012; Tiernan et al., 2014). For example, low or unstable self-esteem has been conceptualized as both a vulnerability factor (Freeman,

2007) and causal (Bentall et al., 1994; 2001) in the onset and maintenance of paranoia. However, from the functional-analytic perspective, paranoia and related relational responding to the self are considered patterns of behavior (i.e., AARR) in context. Hence, the overall goal of this research involved examining the relationships between these behaviors and how current contextual variables and prior learning histories affect them. We used the IRAP and a variety of self-report measures to examine individual relational responses to the self and others (e.g., as devious) with a view to parsing out their relationships with each other and with paranoia. We also investigated the impact of adverse environmental events (e.g., social stress) on these patterns of relational responding as well as the influence of individuals' learning histories (e.g., 'paranoia-proneness') on the effect of these events on paranoia.

## **7.2. Summary of study findings from the current thesis**

In Study 1, we explored the relationship between non-clinical paranoia and relational responding to the self in terms of self-esteem (using the Rosenberg Self-esteem Scale; RSES) and responding to the self as 'positive' and 'negative' on the IRAP. Higher paranoia was associated with low self-reported self-esteem on the RSES and 'Not me-Negative' responding on the IRAP, with both measures predicting paranoia (on the Paranoia Checklist). On the IRAP, negation of negative self-evaluations (i.e., 'Not me-Negative' agreement) was highlighted as being particularly pertinent when situating individuals along the paranoia continuum. These results also supported findings from the wider cognitive-clinical literature that paranoia is associated with low self-esteem and negative self-evaluations in particular in both clinical and the general populations (e.g., Garety & Freeman, 2013; Kesting & Lincoln, 2013; Tiernan et al., 2014).

In Study 2, we examined the influence of a threat-induction task (Ellet & Chadwick, 2007) on 'state' paranoia, self-esteem (on the RSES), and relational responding to the self as 'positive' and 'negative' on the IRAP in the same sample of participants from Study 1.

Specifically, participants completed four unsolvable tasks under conditions of high or low self-awareness and failure feedback or no feedback. The findings showed that all participants (regardless of task condition) showed a linear increase in state paranoia across time, with no specific effects for task condition. ‘Me-Positive’ agreement on the IRAP was maintained from Study 1 and across all conditions. However, participants in the high self-awareness conditions and low self-awareness + failure feedback condition showed ‘Me-Negative’ agreement, whereas participants in the control condition (i.e., low self-awareness + no feedback) did not. Furthermore, controls were the only group to demonstrate significant ‘Not me-Negative’ agreement on this IRAP trial-type. These findings showed that the IRAP can capture subtle differences in AARR in response to environmental events. They also provided further support for the assertion that negative self-evaluations are particularly relevant to paranoia, while positive self-evaluations may be maintained or are at least less impaired (see Kesting & Lincoln, 2013; Tiernan et al., 2014).

Due to the lack of specific effects for task condition on state paranoia in Study 2, in Study 3 we explored an alternative threat-induction task that might afford greater precision for predicting-and-influencing paranoia. As paranoia is essentially a perception of *interpersonal* threat (see Freeman & Garety, 2000), we examined the effect of social exclusion on state paranoia and responding on the RSES and IRAP using a virtual ball-toss game called ‘Cyberball’ (see Kesting et al., 2013; Lincoln et al., 2014; Westermann et al., 2012, for similar approaches). The findings showed that exclusion led to increased state paranoia, reduced self-esteem on the RSES, and increased ‘Me-Negative’ and ‘Others-Positive’ agreement on the IRAP. Moreover, the effect of exclusion on state paranoia was moderated by the distress associated with experiencing paranoid thoughts in general at baseline. This finding replicated previous research (e.g., Kesting et al., 2013) and

underscored the importance of adopting a multi-dimensional approach to paranoia<sup>27</sup>. Crucially, it also demonstrated that one's learning history may influence responding to current environmental events. Changes in responding to the self as 'negative' and others as 'positive' were associated with each other in the Exclusion group only, suggesting that negative self-evaluations in the context of social stress may be related to specific relational responses to others. However, the stimuli included in the IRAP in Study 3 referred to general positive and negative regard, making it difficult to clarify the nature of the relationship between responding to the self and others in this context.

Hence, in Study 4 we explored more specific relational responses to the self (as 'safe' and 'vulnerable') and others (as 'trustworthy' and 'devious') and their relationship with paranoia in individuals with high non-clinical paranoia and controls. We hypothesized that participants with high non-clinical paranoia may have a history of responding to the self as 'vulnerable' and others as 'devious' whereas controls would not. The findings showed that while all participants (regardless of paranoia level) showed significant 'Me-Safe' agreement on the IRAP, participants with high non-clinical paranoia demonstrated a significant 'Me-Vulnerable' bias, supporting our hypothesis. Furthermore, the observed associations between responding to the self on the IRAP and self-report measures indicated a complex relational network regarding the self and paranoia. There were no between-groups differences regarding responding to Others on the IRAP, although individuals with high non-clinical paranoia showed higher Negative-Others and lower Positive-Others scores relative to controls on the self-report measures. Also, self-reported Self- and Others-evaluations correlated with each other and paranoia in the expected directions.

---

<sup>27</sup> In other words, the impact of experiencing paranoid thoughts does not depend simply on their presence but also on the interpretation of such experiences, with the dimensions of preoccupation, conviction, and distress associated with paranoid thoughts deemed particularly pertinent when determining the impact of such experiences on an individual's functioning (see Green et al., 2008; Peters et al., 1999b; 2004).

In Study 5 we examined the influence of social exclusion (using ‘Cyberball’ again) on state paranoia, self-reported self- and others-evaluations, and relational responding to the self as ‘safe’ and ‘vulnerable’ on the IRAP. The results supported those of Study 3, with exclusion leading to increased state paranoia, which was moderated by the frequency, conviction and distress associated with experiencing paranoia in general. The effect of exclusion on self-reported Negative-Others responding was also moderated by conviction. The findings on the IRAP revealed that all participants (regardless of Inclusion or Exclusion) showed ‘Me-Vulnerable’ agreement at baseline and at post-Cyberball, which increased significantly for excluded participants. However, self-reported ‘Me-Vulnerable’ ratings differed between the groups, with included participants showing reduced ‘Me-Vulnerable’ ratings at post-Cyberball (diverging from their responses on the IRAP) and excluded participants showing increased ‘Me-Vulnerable’ ratings (converging with their IRAP responses). All participants showed significant ‘Me-Safe’ agreement on the IRAP at both time points, but the strength of this response decreased for excluded participants, as did their self-reported ‘Me-Safe’ ratings. Furthermore, excluded participants also showed reduced Positive-Others scores at post-Cyberball, indicating that they viewed others as less trustworthy. Changes in self-reported ‘Me-Safe’ and ‘Me-Vulnerable’ ratings, Positive-Self and -Others, and Negative-Self and -Others scores were significantly associated with each other and with changes in state paranoia, suggesting a complex relational network involving interrelated responses to the self, others, and paranoia that could be considered consistent with participants’ experiences of Cyberball (i.e., Inclusion or Exclusion).

### **7.3. The functional-analytic approach to paranoia**

This section considers these findings in terms of the functional-analytic approach to the study of paranoia more generally. We begin by considering the goals of this doctorate research and how the information garnered from these studies might inform our wider goals

of developing our understanding of the behavioral processes involved in paranoia and related phenomena (e.g., relational responding to the self and others) and the contextual factors that may predict-and-influence them. We then turn our attention to the contribution of this research to the RFT and IRAP literature. We interpret these findings using the RFT concepts of derivation, coherence and complexity and consider the utility and precision of the IRAP as a measure of individual relational responses pertinent to paranoia. Specifically, we consider the IRAP as a measure of the strength of repertoires of AARR that have been established by prior learning histories and its sensitivity regarding capturing changes in AARR in response to environmental events.

### **7.3.1. Implications of these findings for the functional-analytic approach to paranoia**

Contextual behavioral science (CBS) provided the conceptual foundation for this research, the goal of which is to achieve prediction-and-influence over behavior with precision, scope and depth. As mentioned above, we define paranoia and related phenomena (e.g., responding to the self and others) as AARR. Accordingly, the two main goals of this research were, firstly, to identify and elucidate the behavior(s) of interest (i.e., AARR regarding paranoia and related responding to the self and others) to provide greater specificity and precision in this regard; and secondly, to investigate whether we could predict-and-influence these behaviors and their interaction by manipulating features of the current context (i.e., environmental variables). We also examined the influence of prior learning histories on the effect of these manipulations on paranoia.

Regarding the first goal, we sought to identify specific repertoires of relational responding to the self and others pertinent to paranoia using the IRAP and self-report measures. One way of doing this is to investigate their covariance with measures of paranoia (e.g., the Paranoia Checklist) and their predictive utility in this regard (see Vahey et al., 2015). The findings from all five studies showed that self-reported self-esteem on the RSES

was significantly and negatively associated with paranoia, with higher paranoia associated with lower self-esteem. The findings from Study 1 showed that both RSES scores and ‘Not me-Negative’ responding on the IRAP predicted Paranoia Checklist scores. Specifically, low self-esteem and distinguishing the self from ‘negative’ evaluations on the IRAP (i.e., agreeing ‘Not me-Negative’) predicted paranoia levels.

Another way of identifying specific relational responses to the self and others relevant to paranoia is via known groups differences; that is, examining whether a measure can discriminate between two groups known to differ on the variable of interest (see Cronbach & Meehl, 1955). Several researchers have suggested that paranoia can be considered an extension of negative ideas about the self as vulnerable and that individuals who experience paranoia may hold beliefs that they are weak, vulnerable, or a ‘soft target’ (Bebbington et al., 2013; Freeman et al., 2005; Freeman et al., 2013; Garety et al., 2001; Salvatore et al., 2012). The findings on the IRAP in Study 4 highlighted the relevance of responding to the self as ‘vulnerable’ to paranoia, with participants with high non-clinical paranoia demonstrating a strong relation of coordination between the self and ‘vulnerable’ (i.e., agreeing ‘Me-Vulnerable’), whereas controls did not.

However, the observed correlations between responding on the IRAP, self-report measures, and paranoia in some studies indicated a more complex pattern regarding relational responding to the self in relation to paranoia. For example, in Study 4, stronger ‘Me-Vulnerable’ disagreement on the IRAP was associated with lower conviction, and stronger ‘Not me-Safe’ disagreement was associated with lower distress in relation to experiencing paranoid thoughts in general (although not significantly so). Thus, specific relational responses to the self may be pertinent to different aspects or dimensions of paranoia. Also, while responding to the self as ‘safe’ on the IRAP did not differentiate high non-clinical paranoia from controls in this study, the latter finding suggested that responding to the self as

‘safe’ may also be relevant to paranoia. Indeed, participants with high non-clinical paranoia showed significantly lower ‘Me-Safe’ ratings relative to controls on the self-report measure.

The findings from Studies 4 and 5 also indicated a complex relational network involving interrelated relational responses to the self and others. For instance, in Study 4, stronger ‘Me-Vulnerable’ disagreement was associated with stronger ‘Not me-Safe’ disagreement on the IRAP; stronger ‘Not me-Vulnerable’ agreement on the IRAP was associated with higher ‘Me-Safe’ ratings; and stronger ‘Me-Safe’ agreement on the IRAP was associated with lower ‘Me-Vulnerable’ ratings and Negative-Self scores and higher ‘Me-Safe’ ratings and Positive-Self scores (though not all were significant). In Study 5, changes in ‘Me-Safe’ and ‘Me-Vulnerable’ ratings were associated with changes in other Positive- and Negative-Self, RSES, and Positive- and Negative-Others scores, all of which were associated with changes in state paranoia. Taken together, these findings show that relational responding to the self as ‘negative’ and ‘vulnerable’ are particularly relevant to paranoia. However, they also highlight the complexity of relational responding to the self and others in this regard and the need to consider relational responding at multiple levels, including individual relational responses and more elaborate relational networks.

Importantly, when considering the role of relational responding to the self and others in relation to paranoia, these are not considered ‘causal’. That is, while specific responses to the self may covary with or predict paranoia, these are considered behavior-behavior relations, which are themselves contextually controlled by environment-behavior relations (e.g., contingencies of reinforcement; see Hayes & Brownstein, 1986). Hence, the second goal of this research was to examine the influence of current and historical environmental factors on these (behavior-behavior) relations. Three studies (2, 3, and 5) examined the impact of adverse environmental events on relational responding in terms of paranoia, relational responding to the self and others, and their interaction. We also examined whether

individuals' learning histories regarding paranoia moderated the effect of these events on these behaviors in two of these studies.

The findings demonstrated that paranoia may be influenced (momentarily) by threat-induction tasks, with increases in paranoia observed following exposure to unsolvable tasks (Study 2) and social exclusion (Studies 3 and 5). Moreover, these events also affected relational responding to the self and others on the IRAP and self-report measures. For example, in Study 3, participants showed increased 'Me-Negative' and 'Others-Positive' agreement on the IRAP and reduced self-esteem on the RSES following exclusion. In Study 5, excluded participants demonstrated increased 'Me-Vulnerable' agreement and reduced 'Me-Safe' agreement on the IRAP and self-report ratings. They also showed reduced Positive-Others responding, indicating that they viewed others as less trustworthy following exclusion. These findings demonstrate that, as an operant, AARR is indeed sensitive to contextual factors (see Barnes-Holmes, Hayes, et al., 2001), as the strength of these specific relational responses were influenced by these adverse events. Moreover, in Study 5, changes in state paranoia were associated with changes in relational responding to the self and others, indicating that exclusion impacted both individual relational responses and behavior-behavior relations.

AARR may also be affected by one's prior learning history. In Studies 3 and 5, the effects of social exclusion on state paranoia were moderated by 'paranoia-proneness', with the distress associated with experiencing paranoia in general moderating the effect of exclusion on state paranoia. Additionally, in Study 5, the frequency and conviction associated with experiencing paranoia also moderated the effect of exclusion on state paranoia. Conviction also moderated the effect of exclusion on Negative-Others responding. Taken together, these findings show that paranoia and related relational responding to the self and others, as well as their interaction, can be influenced by manipulating environmental

variables, and these can be proximal (e.g., current social exclusion), distal (e.g., prior adverse interpersonal experiences), or an interaction of the two. For instance, in Studies 3 and 5, only participants who were more distressed by paranoid thoughts demonstrated increased state paranoia following exclusion. These findings may help explain why some individuals respond to adverse interpersonal experiences with increased paranoia while others do not. Indeed, by examining potential moderators of behavioral effects, more can be learned about the conditions under which certain elements in the environment influence behavior (see De Houwer, 2011).

In summary, precision is one of the goals of CBS and the findings from these studies have helped elucidate some of the specific relational responses to the self that may be particularly pertinent to paranoia; namely, responding to the self as ‘negative’ and ‘vulnerable’. However, they also highlight the importance of examining broader repertoires of relational responding to the self and others in relation to paranoia (i.e., relational networks). These findings also contribute to the broader goals of predicting-and-influencing paranoia and related behavior by demonstrating that these responses and their interaction may be shaped or strengthened by environmental factors, and by adverse interpersonal events in particular. Moreover, by identifying potential moderators of behavior, we may be better able to specify both the types of events that may impact paranoia as well as individuals who may be particularly vulnerable in this regard (e.g., based on their prior learning history).

### **7.3.2. Considering these findings in terms of RFT and the IRAP**

The IRAP was selected as the primary measure for investigating relational responding pertinent to paranoia for two reasons: firstly, there is a growing evidence-base to support the IRAP’s utility as a measure of clinically-relevant repertoires of AARR across a variety of domains (see Vahey et al., 2015); and secondly, IRAP research has made an important contribution to the development of RFT research more generally, including improving our

knowledge and understanding of the dimensions (e.g., derivation, coherence, complexity) and dynamics of AARR (see, for example, the Multidimensional Multilevel, or ‘MDML’, model of Barnes-Holmes et al., 2016). The IRAP captures AARR ‘in flight’ and provides a metric of the strength or persistence of AARR (Hussey et al., 2015). Hence, we suggest that it may be useful for identifying established repertoires of AARR pertinent to paranoia as well as investigating their malleability following modifications to contextual factors.

### **7.3.2.1. Derivation and coherence**

The Relational, Elaboration and Coherence (REC; Barnes-Holmes et al., 2010) model proposes that derivation opportunities (i.e., how ‘rehearsed’ a relational response is) may influence the relative ‘automaticity’ with which certain responses will be emitted in future (Hughes et al., 2012). Moreover, the dimensions of derivation and coherence may covary, with particular patterns of AARR considered increasingly coherent (and perhaps strengthened) with increased derivation opportunities (see Barnes-Holmes et al., 2016). Within the IRAP procedure, two response contingencies are put into competition with each other. If the data show that there is a faster mean level of responding on one contingency relative to the other, then it is presumed that the former is a more highly derived relational response. In other words, we might conclude that the former contingency involves responding in accordance with pre-experimentally established low derivation stimulus relations that may have been derived many times in the past, whereas the latter contingency involves responding with lab-induced high derivation<sup>28</sup> stimulus relations considered inconsistent with one’s assumed prior history of responding. The direction of the observed D-IRAP score provides an

---

<sup>28</sup> The extent to which a relational response has been derived previously can vary along a continuum from low to high. For example, if an individual is told that B is the same as A and that C is the same as B they will likely derive that A is the same as C. The first instance in which they derive this A-C relation could be considered a ‘highly derived’ response given that it is entirely derived from a limited set of prior learning experiences. However, given an ever-increasing number of opportunities to derive (or ‘rehearse’) this relation, the resulting response may come to be increasingly considered a ‘low derivation’ response (see Hughes et al., 2012; Barnes-Holmes et al., 2016).

indication of whether the relation between the stimuli (e.g., coordination between ‘Me’ and ‘Positive’) is relationally coherent or incoherent for participants (i.e., whether faster responding occurs for one response contingency over the other).

Thus, responding on the IRAP may provide insight into individuals’ prior learning history with regards these stimulus relations. Namely, the observed difference in response latency between these two required responses (e.g., faster responding for ‘Me-Positive-True’ relative to ‘Me-Positive-False’) may indicate that this response is relationally coherent and has been derived many times in the past (e.g., participants typically respond to the self as ‘positive’). The findings in Study 4 showed that participants with high non-clinical paranoia demonstrated a significant ‘Me-Vulnerable’ bias on this IRAP trial-type, whereas controls showed no effect. Specifically, participants with high non-clinical paranoia indicated faster agreement (relative to disagreement) for ‘Me-Vulnerable’, suggesting that this may be a coherent and ‘low derivation’ response for this group. In other words, individuals who experience greater paranoia may have a history of responding to the self as vulnerable.

#### **7.3.2.2. Coherence and flexibility**

As an operant, AARR can be shaped by contingencies of reinforcement or punishment. RFT researchers propose that coherence is a conditioned reinforcer for AARR because responding in coherent ways results in both reinforcement from the socio-verbal community (e.g., positive attention for AARR in an internally consistent or coherent manner) and effective action (e.g., being able to accurately, coherently, and contingently describe relations among events; see Bordieri, Kellum, Wilson & Whiteman, 2015; Wray et al., 2012). Furthermore, discriminating that one is deriving coherently (i.e., that one is “correct”) or that certain patterns of relational responding increase the likelihood of certain consequences (e.g., prevention of harm) may increase the probability that similar responses are emitted in future.

Relational responding in terms of paranoia and related responding to the self and others is likely influenced by environmental contingencies (e.g., adverse interpersonal events), which may reinforce particular relational responses and their perceived coherence (e.g., “I was right. I am vulnerable. You can’t trust anyone.”). Hence, while (all things being equal) the IRAP can be used to provide an indication of *typical* relational responding in a given domain (e.g., ‘Me-Vulnerable’ responding in individuals with high paranoia), it can also be used to examine the malleability or *flexibility* of AARR in response to changes in the environment. For instance, the findings from Study 3 showed that all participants showed no ‘Me-Negative’ bias at baseline. However, at post-Cyberball excluded participants demonstrated (non-significant) ‘Me-Negative’ agreement (relative to disagreement), whereas included participants continued to demonstrate no effect on this trial-type. In contrast, all participants demonstrated a ‘Me-Positive’ bias, which remained robust even after exclusion, with both included and excluded participants continuing to demonstrate significant ‘Me-Positive’ agreement at post-Cyberball. These findings demonstrate that certain relational responses can be strengthened following environmental manipulations, whilst others may remain unaffected by such events. The findings from Study 5 indicated that certain relational responses can also be weakened by environmental events. For instance, while all participants demonstrated ‘Me-Safe’ agreement at both time points, excluded participants showed decreased ‘Me-Safe’ agreement from baseline to post-Cyberball.

These findings may help elucidate how adverse interpersonal events may shape or influence such patterns of relational responding. For instance, repeated exposure to adverse events (e.g., bullying) or more severe or traumatic aversive interpersonal experiences (e.g., physical abuse) may set the stage for the development of repertoires of AARR involving paranoia and related responding to the self as vulnerable and others as hostile. Indeed, a wealth of evidence indicates that psychosis generally and paranoia specifically are associated

with such experiences (Bebbington et al., 2004; Bentall et al., 2014; Read et al., 2003; 2005; Varese et al., 2012).

However, from a functional-analytic perspective, when determining whether this behavior (i.e., paranoia) is indicative of ‘psychopathology’ (i.e., impacting one’s functioning and/or psychological suffering), it is important to consider it in terms of the ‘act-in-context’. In Studies 2, 3, and 5, participants were exposed to *actual* adverse events. Hence, the observed changes in relational responding regarding increased paranoia and responding to the self as ‘negative’ and ‘vulnerable’ on the IRAP could be considered consistent or coherent with participants’ experiences. Put simply, it ‘makes sense’ that one would respond to the self as more vulnerable or less safe when treated unfairly by others. Moreover, these responses are probably adaptive within such contexts. Indeed, participants would be justified in not pursuing a relationship with the other Cyberball players and/or avoiding them in future based on this interaction. Paranoia and related responding to the self and others may be considered maladaptive, however, when they become rigid, overgeneralized, and insensitive to direct contingencies (e.g., occurring in neutral or positive interpersonal contexts). From the RFT perspective, the reinforcing functions of coherence may explain why these responses continue in situations where it is inappropriate (see Wray et al., 2012; Hayes et al., 1999; 2012). That is, such responses may be *internally* coherent for the individual (e.g., based on their prior learning history) but would be considered incoherent with the responding of the wider verbal community within such (neutral or positive) contexts (see Stewart et al., 2016). However, RFT research on the topic of coherence is still developing.

### **7.3.2.3. Complexity**

Throughout the current thesis, we utilized self-report measures in addition to the IRAP to investigate repertoires of AARR to the self and others in relation to paranoia. According to the REC model, the IRAP captures brief and immediate relational responses

(BIRRs) while self-report measures capture extended and elaborated relational responding (EERRs) as a function of the assessment context (e.g., time pressure). When used in tandem, they may elucidate more complex networks of interrelated relational responding.

The findings from Study 5 revealed an elaborate and interesting pattern of BIRRs and EERRs regarding ‘Me-Safe’ and ‘Me-Vulnerable’ responding in relation to participants’ experiences of Cyberball. All participants demonstrated significant ‘Me-Vulnerable’ agreement on the IRAP at baseline. However, included participants continued to show a ‘Me-Vulnerable’ bias on the IRAP but reduced ‘Me-Vulnerable’ self-report ratings following Cyberball. In contrast, excluded participants showed increased ‘Me-Vulnerable’ responses on both measures at post-Cyberball. The observed responses on the IRAP and self-report ratings might reflect patterns of BIRRs and EERRs that were consistent within each context. That is, while included participants’ BIRRs involved ‘Me-Vulnerable’ agreement, their subsequent EERRs could reflect additional and contradictory relational responses based on their experiences of Cyberball (e.g., “The other players included me”), whereas, excluded participants’ experiences may have confirmed their BIRRs. In addition, changes in EERRs regarding self- and others-evaluations and paranoia were consistent with each other and with participants’ experiences of Inclusion or Exclusion. For instance, excluded participants demonstrated increased paranoia, ‘Me-Vulnerable’ and Negative-Self and reduced ‘Me-Safe’, Positive-Self and Positive-Others responding, indicative of an intricate relational network of interrelated responses that was internally coherent and consistent within this context.

In summary, using the IRAP (an RFT-based measure that is high in precision), we identified individual relational responses pertinent to paranoia, their strength, and their probability (based on features of the context in which they were assessed). In addition, Studies 2, 3 and 5 yielded interesting information regarding environmental factors or antecedents (e.g., social stress) that may influence these responses and their interaction (i.e.,

environment-behavior interactions). The findings from these studies also pointed to a possible reinforcer for these patterns of AARR – coherence. Furthermore, the observed moderating influence of paranoia-proneness on the effect of exclusion on state paranoia (Studies 3 and 5) and Negative-Others responding (Study 5) highlighted the influence of one’s prior learning history in this regard.

### **7.3.3. Implications for IRAP research: Methodological issues**

While the IRAP demonstrated both precision and utility within the current thesis, a few methodological issues should be considered regarding the use of this measure. Firstly, double-negation (e.g., “I am not-*Negative-False*”) was involved on some trial-types. As this required complex responding under conditions of time and accuracy pressure, this may have placed added cognitive burden on participants, which could affect the reliability of responses on these trial-types and attrition.

Regarding reliability, in everyday language, complex stimulus relations (e.g., “I am not-*Negative-False*”) may be contacted less frequently compared to simpler stimulus relations (e.g., “I am-*Positive-True*”). Hence, such repertoires may be less well established, potentially affecting variability in responding on these trial-types. However, in Study 1, ‘Not me-*Negative*’ responses predicted paranoia. Also, in Study 2, only controls continued to demonstrate significant ‘Not me-*Negative*’ agreement following the threat-induction task, whereas participants in the other conditions did not. Hence, these responses appeared to correspond with participants’ experiences of the task. However, in Study 5, excluded participants demonstrated significant ‘Not me-*Vulnerable*’ agreement, which seemed incoherent within this context (i.e., Exclusion) and with their co-occurring ‘Me-*Vulnerable*’ agreement. Nonetheless, “I am not” was considered the most appropriate contrast category for “I am” in relation to self-evaluations within these studies. Indeed, while the IRAP is non-relative (i.e., interpretation of responding on each trial-type is specific to the class of label

and target stimuli presented within that trial-type), it is not a-contextual (see Hussey, Thompson, McEnteggart, Barnes-Holmes, & Barnes-Holmes, 2015). In other words, individuals' learning histories with regards to the label and target stimuli may influence their behavior within the IRAP "because the label-label and target-target relations contribute to the broader context of that [which] is set within the measure" (Hussey et al., 2015, p. 292). Thus, the choice of contrast category may have an important bearing on responding within the IRAP.

Regarding attrition, Barnes-Holmes et al. (2010) recommended that IRAP procedural considerations such as accuracy and latency criteria be adjusted to the population that is being sampled and the stimuli that are being used in the study. In each study of the current thesis, accuracy criteria were reduced to either 71% or 75% and the latency criterion in Study 4 was also extended to 2500ms to minimize attrition and maximize available data. Following these adjustments, attrition due to failure to meet accuracy and/or latency criteria during the IRAP practice phase or to maintain accuracy during the test phase was relatively low across all five studies (ranging between 5 and 15%). In addition, during the instructions for IRAPs involving double-negation, we advised participants to read the label and target stimuli as one sentence and to select "True" to agree with the statement or "False" to disagree with it. Participants reported finding this strategy helpful. Also, recently developed measures like the 'Natural Language' IRAP (Kavanagh, Hussey, McEnteggart, Barnes-Holmes, & Barnes-Holmes, 2016) and Relational Responding Task (De Houwer, Heider, Spruyt, Roets, & Hughes, 2015) involve presenting whole statements center-screen (rather than label stimuli top-screen and target-stimuli center-screen). Such procedures may circumvent issues regarding double-negation.

A second methodological issue concerns the provision of specific 'rules' for responding on the IRAP (e.g., "I am-Positive, I am not-Negative" and "I am-Negative, I am

not-Positive” on consistent and inconsistent blocks respectively within a self-esteem IRAP). On the one hand, explicit rules may help reduce attrition; on the other hand, however, it has been suggested that the type of rule presented may affect responding on the IRAP trial-types (see Finn, Barnes-Holmes, Hussey, & Graddy, 2016). We only provided such rules in Study 3. Anecdotally, this appeared to facilitate participants’ rapid achievement of fast and accurate responding during the IRAP practice blocks compared to the other studies. While some preliminary research has shown that rules may affect responding on IRAP trial-types (e.g., Finn et al., 2015), this was investigated using non-valenced stimuli (e.g., shapes, colors) and has not been examined systematically with valenced stimuli relevant to clinically-relevant phenomena. Moreover, it is likely that responding on the IRAP naturally entails ‘rule-following’ (e.g., participants may generate rules as a strategy for completing the task under time pressure). Hence, it may be more appropriate to consider the *source* of such rules regarding their potential impact on responding, such as whether these rules are *socially* delivered by the experimenter or are derived by the participants. In the feedback questions presented at the end of the studies that did not involve explicit rules, many participants reported adopting a strategy for completing the IRAP (e.g., “being positive” on consistent blocks and “being negative” on inconsistent blocks).

The IRAP procedure can be challenging for some and, in such cases, it may be prudent to calibrate the instructions and/or criteria according to stimulus-difficulty and participant ability. This may help reduce exclusion of participants who may be of most interest to the research being carried out (e.g., clinical populations). Recently, McEntegart and colleagues (*in press*) outlined recommendations for using the IRAP with individuals with a diagnosis of psychosis. This population may present with experiences (e.g., paranoia, auditory hallucination) and medication side-effects (e.g., poor concentration, fatigue) that might interfere with completion of experimental tasks. In their study, the experimenter

conducted a functional assessment of participants' requirements on a participant-by-participant basis, with procedural modifications made where necessary to support the individual to complete the IRAP.

#### **7.4. The cognitive-clinical approach to paranoia**

This section explores the findings from this research in relation to the cognitive-clinical approach to paranoia. We begin by considering the contribution of this work to the existing literature on self-concepts in relation to paranoia. We then discuss the advantages of adopting an experimental approach in this domain and suggest that the IRAP may provide a novel and useful measure for investigating specific self-evaluations (or repertoires of relational responding to the self) pertinent to paranoia as well as changes in these evaluations (or responses) following experimental manipulations. We then discuss these findings in relation to prominent cognitive-clinical theories of paranoia (e.g., Bentall et al., 1994; 2001; Freeman et al., 2002). Finally, we suggest possible directions for future research. We propose that this work may benefit from the functional-cognitive framework – an approach that marries behavioral principles and cognitive theories to the mutual advantage of researchers operating from the functional-analytic and cognitive-clinical traditions within this domain (see De Houwer, 2011; Hughes, De Houwer, & Barnes-Holmes, 2016; Hughes, De Houwer, & Perugini, 2016; De Houwer, Hughes, & Barnes-Holmes, 2017).

##### **7.4.1. Relevance of these findings to the cognitive-clinical literature**

Three recent systematic reviews have shown that paranoia is associated with low *global* self-esteem and negative self-evaluations in particular, while positive self-regard may be maintained or less impaired (Freeman & Garety, 2014; Kesting & Lincoln, 2013; Tiernan et al., 2014). In each study of the current thesis, paranoia was associated with low self-esteem. Moreover, the findings from Study 1 showed that both self-esteem and responding to the self as 'negative' (i.e., 'Not me-Negative') predicted paranoia. In Studies 1, 2, and 3,

paranoia was not associated with ‘Me-Positive’ responding on the IRAP, nor was ‘Me-Positive’ responding affected by the threat-induction tasks used in Studies 2 and 3; rather, participants demonstrated increased ‘Me-Negative’ agreement following these events.

The current research also extends this literature by examining more specific self-evaluations in relation to paranoia. Several researchers have proposed that individuals who experience paranoia may hold beliefs that they are vulnerable (e.g., Freeman, 2016; Freeman et al., 2002; Garety et al., 2001; Salvatore et al., 2012). The findings from Study 4 provide support for this idea as participants with high non-clinical paranoia demonstrated significant ‘Me-Vulnerable’ agreement on the IRAP and lower self-reported vulnerability relative to controls. Moreover, the finding that stronger ‘Me-Vulnerable’ disagreement was associated with lower paranoia conviction provides some support for the assertion made by Garety et al. (2001) that paranoid beliefs may be held with greater conviction if they are consistent with firmly-held beliefs about the self as vulnerable.

The current research also provides further support for the multi-dimensional approach to paranoia advocated by many cognitive-clinical researchers (e.g., Freeman et al., 2005; Green et al., 2008; Peters et al., 1999b; 2004). They argue that factors such as the frequency of experiencing paranoid thoughts or beliefs as well as the preoccupation, conviction and distress associated with them may be more relevant than the content of the belief alone with respect to the impact that paranoia may have on functioning. The findings from Studies 3 and 5 showed that the distress associated with trait paranoia, as well as frequency and conviction (Study 5), moderated the effect of social exclusion on state paranoia. This finding regarding the moderating influence of distress supports findings from previous research (e.g., Kesting et al., 2013) and suggests that the affective component of paranoia may be pertinent to stress-sensitivity in interpersonal contexts (i.e., feeling vulnerable in the presence of others due to the expectation of criticism or rejection; see Bell & Freeman, 2014).

#### **7.4.2. Contribution to the cognitive-clinical literature: Methodological considerations**

Several cognitive-clinical researchers have proposed that negative self-beliefs are developed in the context of adverse interpersonal experiences, which may lead the individual to feel inferior to others, different and apart, and hence vulnerable, and paranoia is likely to flourish when one perceives the self as vulnerable (Bentall et al., 2014; Freeman, 2016; Freeman & Garety, 2014; Kesting & Lincoln, 2013; Salvatore et al., 2012). However, research on the influence of environmental factors on dysfunctional self-evaluations and paranoia has been limited, with much of the research conducted to date comprised of cross-sectional studies (see Klippel et al., 2017; Valmaggia, Day, & Rus-Calafell, 2016). The current research addresses this gap within the literature by examining the dynamic relationships between adverse experiences, paranoia, and self- and other-evaluations using an experimental approach. Moreover, as the IRAP captures individual relational responses ‘in flight’, this measure may facilitate a *live* investigation of the effects of environmental events on specific self-evaluations pertinent to paranoia.

Studies 3 and 5 highlighted the important role that adverse interpersonal events may have in influencing paranoia and related self- and others-evaluations. Specifically, social exclusion led to increased state paranoia, ‘Me-Negative’ (Study 3) and ‘Me-Vulnerable’ agreement and reduced ‘Me-Safe’ agreement and Positive-Others ratings (Study 5), with changes in paranoia and self-and others-evaluations interrelated in this context (Study 5). More generally, adverse experiences may influence the development of paranoia and related repertoires of responding to the self (as vulnerable) and others (as devious), as well as the probability of paranoid responding in future social interactions. Indeed, recent research involving Virtual Reality has shown that a history of social defeat (e.g., Valmaggia et al., 2015), bullying victimization (e.g., Valmaggia, Day, et al., 2015), and experiences of physical assault (Freeman et al., 2013; Freeman, Antley, et al., 2014) are associated with an

increased likelihood of making paranoid appraisals of ambiguous or neutral social interactions. However, more research is needed to identify the factors that might maintain or reinforce paranoid responses across time and influence their generalization to such contexts (e.g., neutral interpersonal scenarios where no threat is present).

#### **7.4.3. Considering these findings in relation to cognitive-clinical theories of paranoia**

Freeman and colleagues' (2002) 'Threat anticipation model' and Bentall and colleagues' (1994; 2001) 'Defensive model' – two prominent cognitive-clinical models of paranoia – were used to inform the current research. These studies may shed some light on these theories. To briefly restate the core tenets of these models regarding self-concepts and paranoia, Bentall et al. (1994; 2001) proposed that individuals with paranoia attribute negative events externally and personally in order to preserve self-esteem. In contrast, Freeman et al. argued that paranoid beliefs directly reflect emotional concerns and negative self-concepts.

The findings from Studies 1 and 4 provide support for the assertion that paranoia is associated with low self-esteem and negative self-regard. Furthermore, in Study 4, participants with high non-clinical paranoia demonstrated significantly lower self-esteem, Positive-Self, and 'Me-Safe' ratings and higher Negative-Self and 'Me-Vulnerable' ratings relative to controls. This suggests that paranoia is directly associated with and reflects low self-esteem and negative self-concepts (as per Freeman et al.'s model), rather than defensive efforts against 'underlying' negative self-beliefs (as per Bentall et al.'s model).

However, in Bentall et al.'s (2001) most recent revision of their model, fluctuations of self-esteem are instead seen as critical, with paranoia considered partly as a consequence of dysfunctional efforts to regulate self-esteem. Of the few studies that have investigated fluctuations in self-esteem in relation to paranoia, findings (within non-clinical samples) indicate that paranoia is associated with greater instability of self-esteem (e.g., Raes & Van

Gucht, 2009; Thewissen et al., 2008; 2007). Interestingly, changes in self-esteem following the threat-induction task in Study 2 and exclusion in Study 5 were negatively associated with changes in state paranoia, with larger changes in paranoia associated with smaller changes in self-esteem. On the one hand, these findings could be interpreted in accordance with Bentall et al.'s (2001) argument that paranoia functions to preserve self-esteem. On the other hand, increased paranoia following social stress may be directly reflective of increased negative self-regard following these adverse experiences, as proposed by Freeman et al. (2002). Indeed, in Study 5, changes in paranoia were positively associated with changes in Negative-Self and 'Me-Vulnerable' ratings.

From the functional-analytic perspective, however, it is important to identify the consequences of paranoid responses to adverse events in order to ascertain the function of paranoia in these contexts. For example, Lincoln et al. (2014) investigated the impact of a paranoid (versus self-blame or neutral) explanation following social exclusion during Cyberball. Their findings indicated that the paranoid explanation increased self-esteem more than self-blame immediately following the explanation but that greater self-esteem recovery was observed for the neutral compared to the paranoid explanation (at 15 minutes follow-up). Bentall and colleagues suggested that there is a dynamic and mutually influential relationship between paranoia and self-concepts, whereby attributing negative events to others (i.e., paranoia) may serve to protect one's self-esteem, which may reinforce or increase the likelihood that paranoid attributions for similar events will be made in future. Thus, paranoia may be negatively reinforced by reducing distress and negative self-evaluations in the short-term (see Freeman, 2007) but might have detrimental long-term effects. Similarly, Lincoln et al. (2014) suggested that paranoia may be maintained in terms of "a learned reaction pattern" (p. 404), with subtle changes in interpersonal schemas arising from paranoid interpretations

of these interpersonal events further corroborating this response pattern (see also Kesting & Lincoln, 2013).

### **7.5. Implications for future research and clinical intervention: A functional-cognitive approach**

The current research represents an important first step towards investigating (non-clinical) paranoia from a functional-analytic perspective, informed by cognitive-clinical models and research. These studies have helped identify and elucidate behaviors relevant to paranoia (e.g., specific relational responses to the self and others) as well as contextual factors that may influence their probability and strength (e.g., social stress). The next step for this line of research will be to investigate the processes that underpin the generalization of threat anticipation to (social) stimuli that have not been directly paired with unpleasant experiences (behavior that is central to paranoia) as well as the factors that influence or exacerbate this behavior.

We argue that this process is AARR (Stewart et al., 2016). From the RFT perspective, once an adverse event is experienced, AARR allows that event to be related to hitherto entirely unrelated events, so that they acquire some of the same aversive functions as the directly experienced event. Indeed, previous research has shown that AARR may explain how people come to fear and avoid certain stimuli in the absence of prior experience or instruction (e.g., Dymond, Schlund, et al., 2014). AARR may also explain how people come to anticipate future adverse interpersonal experiences. For instance, Munnelly, Martin, Dack, Zedginidze and McHugh (2014) demonstrated transformation of social exclusion functions via AARR using Cyberball. Participants were trained to form two three-member equivalence classes (A1-B1-C1; A2-B2-C2), with stimuli comprised of fictitious games (e.g., C1 = “Boceem” game; C2 = “Casors” game). They were then exposed to exclusion (C2) and inclusion (C1) during Cyberball (Experiment 1). Results showed that participants expected

that they would be excluded from A2 and B2 and included in A1 and B1. In a second experiment, participants were exposed to exclusion (C2) only. The findings indicated that participants expected to be “more” included in, or “less” excluded from, games (A1, B1, and C1) that were unrelated to C2, despite having never directly experienced inclusion during Cyberball (C1).

Previous research has shown that responses on the IRAP can predict future behavior, including performance on behavior approach tasks (e.g., Nicholson & Barnes-Holmes, 2012b) and treatment engagement (Carpenter et al., 2012). The procedures used in Studies 3 and 5 could be combined with the paradigm developed by Munnely et al. to investigate whether relational responding on the IRAP might predict exclusion- or threat-anticipation to derived stimuli (i.e., A2 and B2 games) and avoidance of these stimuli. For example, participants could be invited to choose a second online game to play, with options including games paired with exclusion and novel games. Given the centrality of ‘Me-Vulnerable’ responding to paranoia and that perceiving the self as vulnerable may influence paranoia (see Freeman, 2007; Salvatore et al., 2012), responses on this trial-type might predict threat-anticipation and/or avoidance of A2 and B2.

Individuals who experience paranoia may develop *rules* (involving the self as vulnerable and others as devious) that allow them to navigate their social world (e.g., “Others are untrustworthy so I should avoid them”). Such rules may undermine contact with direct contingencies and reinforce avoidance (e.g., by contacting the consequences stipulated within the rule, such as the absence of harm, via AARR). Indeed, Freeman et al. (2002) proposed that individuals with persecutory delusions engage in safety behaviors to avoid a feared outcome, and when that outcome does not occur, they may (incorrectly) conclude the safety behaviors prevented harm (see Freeman, Garety, & Kuipers, 2001; Freeman et al., 2007). This may provide some insight as to why paranoid beliefs are often resistant to extinction. It

also suggests another avenue for future research – examining factors that undermine avoidance and promote contact with direct contingencies. Such research may have important clinical applications by identifying variables that support individuals who experience paranoia to discriminate whether threat is *actually* present and to respond accordingly (e.g., avoidance or escape when faced with interpersonal threat would be an adaptive response in this context). Indeed, Freeman (2016) argued that “how best to help patients relearn safety must become a research focus” (p. 688).

The results from Studies 3 and 5 have shown how negative patterns of relational responding and behaviour-behaviour relations (i.e., paranoia and responding to the self as vulnerable) can be promoted via adverse interpersonal events. Furthermore, the frequency, distress and conviction associated with experiencing paranoia in general moderated the effect of such experiences on momentary paranoia within these contexts. Future research might examine interventions aimed at counteracting these effects and exploring more adaptive ways of responding to paranoid thoughts as a means of supporting people with paranoia. For instance, responding to paranoid thoughts with strong conviction may lead to increased inflexibility, overgeneralization, and insensitivity to direct contingencies. Therapeutic approaches such as mindfulness and Acceptance and Commitment Therapy (ACT; Hayes et al., 1999; 2012) attempt to disrupt or modify the behavior-regulatory functions of clinically-relevant events by changing the contexts in which they are usually related to behavior (e.g., using cognitive defusion techniques to create ‘psychological distance’ from one’s thoughts). They also encourage contact with other environmental contingencies so that individuals may alter and adapt their behavior accordingly. ACT clarifies that it is not paranoid thoughts themselves that are problematic but how people respond to them (see Morris, Johns, & Oliver, 2013). Similarly, Cognitive Behavioral Therapy (CBT) approaches to treating paranoia encourage individuals to become “detached observers of their fears” by monitoring

and learning about their paranoid thoughts so that they can better understand them and their causes or triggers (Freeman & Garety, 2006, p. 413). CBT may also attempt to change individuals' interpretation of their (e.g., anomalous) experiences, or simply to enhance coping strategies (e.g., using worry-reduction strategies). In some cases, it is possible to review with the individual the evidence for and against different explanations for their experiences and to conduct behavioral experiments to test out paranoid thoughts (see Freeman & Garety, 2006; Freeman & Garety, 2014; Freeman, Bradley, et al., 2016; Freeman, Waite, et al., 2016; Waller et al., 2015).

## **7.6. Conclusions**

Broadly speaking, the current research demonstrates that by conceptualizing paranoia and related responding to the self and others as patterns of behavior in context, we can identify the contextual variables that affect them. Using the IRAP, we identified specific patterns of relational responding to the self that may be pertinent to paranoia and captured subtle changes in these responses following threat-induction tasks. Indeed, the experimental approach adopted throughout this thesis showed that paranoia and related behavior can be influenced by aversive events and suggested that adverse interpersonal experiences in particular may be an important factor in the development of paranoia. By defining paranoia as AARR and attempting to identify the learning histories and environmental factors that support paranoia and related behavior, it may be possible to not only predict these behaviors, but to influence them.

## References

- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5<sup>th</sup> Ed.). Arlington, VA: American Psychiatric Publishing.
- Atherton, S., Antley, A., Evans, N., Černis, E., Lister, R., Dunn, G., Slater, M., & Freeman, D. (2016). Self-confidence and paranoia: An experimental study using an immersive virtual reality social situation. *Behavioural and Cognitive Psychotherapy*, *44*, 56-64. DOI: 10.1017/S1352465814000496
- Augustson, E. M., Dougher, M. J., & Markham, M. R. (2000). Emergence of conditional stimulus relations and transfer of respondent eliciting functions among compound stimuli. *Psychological Record*, *50*, 745-770.
- Barnes-Holmes, D. (2000). Behavioral pragmatism: No place for reality and truth. *The Behavior Analyst*, *23*, 191-202.
- Barnes-Holmes, D., Barnes-Holmes, Y., Hussey, I., & Luciano, C. (2016). Relational Frame Theory: Finding its historical and philosophical roots and reflecting upon its future development: An introduction to Part II. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *The Wiley Handbook of Contextual Behavioral Science* (pp. 117-128). West Sussex, UK: Wiley-Blackwell.
- Barnes-Holmes, D., Barnes-Holmes, Y., Luciano, C., & McEnteggart, C. (*in press*). From the IRAP and REC Model to a Multi-Dimensional Multi-Level framework for analyzing the dynamics of arbitrarily applicable relational responding. *Journal of Contextual Behavioral Science*. DOI: 10.1016/j.jcbs.2017.08.001
- Barnes-Holmes, D., Barnes-Holmes, Y., Power, P., Hayden, E., Milne, R., & Stewart, I. (2006). Do you really know what you believe? Developing the Implicit Relational Assessment Procedure (IRAP) as a direct measure of implicit beliefs. *The Irish Psychologist*, *32*, 169-177.

- Barnes-Holmes, Y., Barnes-Holmes, D., & Smeets, P. M. (2004). Establishing relational responding in accordance with opposite as generalized operant behavior in young children. *International Journal of Psychology and Psychological Therapy*, *4*, 559-586.
- Barnes-Holmes, D., Hayes, S. C., Dymond, S., & O'Hora, D. (2001). Multiple stimulus relations and the transformation of stimulus functions. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.) *Relational Frame Theory: A Post-Skinnerian Account of Human Language and Cognition* (pp. 51-71). New York: Kluwer Academic/Plenum Publishers.
- Bast, D., & Barnes-Holmes, D. (2015). Priming thoughts of failing versus succeeding and performance on the Implicit Relational Assessment Procedure (IRAP) as a measure of self-forgiveness. *The Psychological Record*, *65*, 665-678. DOI: 10.1007/s40732-015-0137-0
- Barrowclough, C., Tarrier, N., Humphreys, L., Ward, J., Gregg, L., & Andrews, B. (2003). Self-esteem in schizophrenia: Relationships between self-evaluation, family attitudes, and symptomatology. *Journal of Abnormal Psychology*, *112*, 92-99.
- Beards, S., & Fisher, H. L. (2014). The journey to psychosis: An exploration of specific psychological pathways. *Social Psychiatry & Psychiatric Epidemiology*, *49*, 1541-1544. DOI: 10.1007/s00127-014-0953-6
- Bebbington, P. E., Bhugra, D., Brugha, T., Singleton, N., Farrell, M., Jenkins, R., Lewis, G., & Meltzer, H. (2004). Psychosis, victimisation and childhood disadvantage. Evidence from the second British National Survey of Psychiatric Morbidity. *British Journal of Psychiatry*, *185*, 220-226.
- Bebbington, P. E., McBride, O., Steel, C., Kuipers, E., Radovanovic, M., Brugha, T., Jenkins, R., Meltzer, H., & Freeman, D. (2013). The structure of paranoia in the general

- population. *British Journal of Psychiatry*, 202, 419-427. DOI: 10.1192/bjp.bp.112.119032
- Bebbington, P. E., & Nayani, T. (1995). The Psychosis Screening Questionnaire. *International Journal of Methods in Psychiatric Research*, 5, 11-19.
- Beck, A. T. (1967). *Depression: Clinical, Experimental, and Theoretical Aspects*. New York: Harper & Row. (Republished as *Depression: Causes and Treatment*. Philadelphia: University of Pennsylvania Press, 1972).
- Beck, A. T., Freeman, A., Pretzer, J., Davis, D. D., Fleming, B., Ottavani, R., Beck, J., Simon, K. M., Padesky, C., Meyer, J. & Trexler, L. (1990). *Cognitive Therapy of Personality Disorders*. New York: Guilford Press
- Bell, V., & Freeman, D. (2014). A pilot trial of Cognitive Behavioural Therapy for interpersonal sensitivity in individuals with persecutory delusions. *Journal of Behaviour Therapy and Experimental Psychiatry*, 45, 441-446. DOI: 10.1016/j.jbtep.2014.06.001.
- Bentall, R. P. (2006). Madness explained: Why we must reject the Kraepelinian paradigm and replace it with a 'complaint-orientated' approach to understanding mental illness. *Medical Hypotheses*, 66, 220-233. DOI: 10.1016/j.mehy.2005.09.026
- Bentall, R. P., Corcoran, R., Howard, R., Blackwood, N., & Kinderman, P. (2001). Persecutory delusions: A review and theoretical integration. *Clinical Psychology Review*, 21, 1143-1192. DOI: 10.1016/S0272-7358(01)00106-4
- Bentall, R. P., de Sousa, P., Varese, F., Wickham, S., Sitko, K., Haarmans, M., & Read, J. (2014). From adversity to psychosis: Pathways and mechanisms from specific adversities to specific symptoms. *Social Psychiatry and Psychiatric Epidemiology*, 49, 1011-1022. DOI: 10.1007/s00127-014-0914-0.

- Bentall, R. P., & Fernyhough, C. (2008). Social predictors of psychotic experiences: Specificity and psychological mechanisms. *Schizophrenia Bulletin*, *34*, 1012-1020.  
DOI: 10.1093/schbul/sbn103
- Bentall, R. P., & Kaney, S. (2005). Attributional lability in depression and paranoia. *British Journal of Clinical Psychology*, *44*, 475-488. DOI: 10.1348/014466505X29602
- 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*, 331-341.
- Bodner, E., & Mikulincer, M. (1998). Learned helplessness and the occurrence of depressive-like and paranoid-like responses: The role of attentional focus. *Journal of Personality and Social Psychology*, *74*, 1010-1023.
- Bordieri, M., Kellum, K. K., Wilson, K. G., & Whiteman, K. C. (2015). Basic properties of coherence: Testing a core assumption of Relational Frame Theory. *The Psychological Record*, *66*, 83-98. DOI: 10.1007/s40732-015-0154-z
- British Psychological Society (2009). *Code of Ethics and Conduct: Guidance published by the Ethics Committee of the British Psychological Society*. Leicester, UK: The British Psychological Society.
- Campbell, M. L. C., & Morrison, A. P. (2007). The relationship between bullying, psychotic-like experiences and appraisals in 14-16-year olds. *Behavior Research and Therapy*, *45*, 1579-1591. DOI: 10.1016/j.brat.2006.11.009
- Carpenter, K. M., Martinez, D., Vadhan, N. P., Barnes-Holmes, D., & Nunes, E. V. (2012). Measures of attentional bias and relational responding are associated with behavioral treatment outcome for cocaine dependence. *American Journal of Drug and Alcohol Abuse*, *38*, 146-154. DOI: 10.3109/00952990.2011.643986

- Cicero, D. C., & Kerns, J. G. (2011). Is paranoia a defense against or an expression of low self-esteem? *European Journal of Personality*, *25*, 326-335. DOI: 10.1002/per.794
- Collett, N., Pugh, K., Waite, F., & Freeman, D. (2016). Negative cognitions about the self in patients with persecutory delusions: An empirical study of self-compassion, self-stigma, schematic beliefs, self-esteem, fear of madness, and suicidal ideation. *Psychiatry Research*, *239*, 79-84. DOI: 10.1016/j.psychres.2016.02.043
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.
- Cullen, C., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). The Implicit Relational Assessment Procedure (IRAP) and the malleability of ageist attitudes. *The Psychological Record*, *59*, 591-620.
- De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice-versa. *Perspectives on Psychological Science*, *6*, 202-209. DOI: 10.1177/1745691611400238
- De Houwer, J., Heider, N., Spruyt, A., Roets, A., & Hughes, S. (2015). The Relational Responding Task: Toward a new implicit measure of beliefs. *Frontiers in Psychology*, *6*, 319. DOI: 10.3389/fpsyg.2015.00319
- De Houwer, J., Hughes, S., & Barnes-Holmes, D. (2017). Psychological engineering: A functional-cognitive perspective on applied psychology. *Journal of Applied Research in Memory and Cognition*, *6*, 1-13. DOI: 10.1016/j.jarmac.2016.09.001
- DiMaggio, P. (1997). Culture and cognition. *Annual Review of Sociology*, *23*, 263-287. DOI: 10.1146/annurev.soc.23.1.263.
- Dymond, S., Schlund, M. W., Roche, B., De Houwer, J., & Freegard G. P. (2012). Safe from harm: Learned, instructed, and symbolic generalization pathways of human threat-avoidance. *PLoS One*, *7*, DOI:10.1080/17470218.2013.800124

- Dymond, S., Schlund, M. W., Roche, B., & Whelan, R. (2014). The spread of fear: Symbolic generalization mediates graded-threat avoidance in specific phobia. *The Quarterly Journal of Experimental Psychology*, *67*, 247-259. DOI: 10.1080/17470218.2013.800124
- Ellet, L., & Chadwick, P. (2007). Paranoid cognitions, failure, and focus of attention in college students. *Cognition and Emotion*, *21*, 558-576. DOI: 10.1080/02699930600758155
- Fenigstein, A., & Venable, P. (1992). Persecutory ideation and self-consciousness. *Journal of Personality and Social Psychology*, *62*, 129-138.
- Fisher, H. L., Appiah-Kusi, E., & Grant, C. (2012). Anxiety and negative self-schemas mediate the association between childhood maltreatment and paranoia. *Psychiatry Research*, *196*, 323-324. DOI: 10.1016/j.psychres.2011.12.004
- Flower, L., Newman-Taylor, K., & Stopa, L. (2015). Cognitive control processes in paranoia: The impact of threat induction on strategic cognition and self-focused attention. *Behavioural and Cognitive Psychotherapy*, *43*, 108-118. DOI:10.1017/S1352465813000891
- Fowler, D., Freeman, D., Smith, B., Kuipers, E., Bebbington, P. E., Bashforth, H., Coker, S., Hodgekins, J., Gracie, A., Dunn, G., & Garety, P. A. (2006). The Brief Core Schema Scales (BCSS): Psychometric properties and associations with paranoia and grandiosity in non-clinical and psychosis samples. *Psychological Medicine*, *36*, 749-759. DOI: 10.1017/S0033291706007355
- Fowler, D., Hodgekins, J., Garety, P.A., Freeman, D., Kuipers, E., Dunn, G., Smith, B., & Bebbington, P. E. (2012). Negative cognition, depressed mood, and paranoia: A longitudinal pathway analysis using structural equation modeling. *Schizophrenia Bulletin*, *38*, 1063-1073. DOI: 10.1093/schbul/sbr019

- Freeman, D. (2006). Delusions in the nonclinical population. *Current Psychiatry Reports*, 8, 191-204. DOI: 10.1007/s11920-006-0023-1
- Freeman, D. (2007). Suspicious minds: The psychology of persecutory delusions. *Clinical Psychology Review*, 27, 425-457. DOI: 10.1016/j.cpr.2006.10.004
- Freeman, D. (2016). Persecutory delusions: A cognitive perspective on understanding and treatment. *The Lancet*, 3, 685-692. DOI: 10.1016/S2215-0366(16)00066-3
- Freeman, D., Antley, A., Ehlers, A., Dunn, G., Thompson, C., Vorontsova, N., Garety, P. A., Kuipers, E., Glucksman, E., & Slater, M. (2014). The use of immersive Virtual Reality (VR) to predict the occurrence 6 months later of paranoid thinking and post-traumatic stress symptoms assessed by self-report and interviewer methods: A study of individuals who have been physically assaulted. *Psychological Assessment*, 26, 841-847. DOI: 10.1037/a0036240
- Freeman, D., Bradley, J., Waite, F., Sheaves, B., DeWeever, N., Bourke, E., McInerney, J., Evans, N., Černis, E., Lister, R., Garety, P. A., Dunn, G. (2016). Targeting recovery in persistent persecutory delusions: A proof of principle study of a new translational psychological treatment (the Feeling Safe Programme). *Behavioural and Cognitive Psychotherapy*, 44, 539–552. DOI: 10.1017/S1352465816000060
- Freeman, D., Evans, N., Lister, R., Antley, A., Dunn, G., & Slater, M. (2014). Height, social comparison, and paranoia: An immersive virtual reality experimental study. *Psychiatry Research*, 218, 348-352. DOI: 10.1016/j.psychres.2013.12.014
- Freeman, D., & Fowler, D. (2009). Routes to psychotic symptoms: Trauma, anxiety and psychosis-like experiences. *Psychiatry Research*, 169, 107-112. DOI: 10.1016/j.psychres.2008.07.009

- Freeman, D., & Garety, P. A. (2000). Comments on the content of persecutory delusions: Does the definition need clarification? *British Journal of Clinical Psychology, 39*, 407-414. DOI: 10.1348/014466500163400
- Freeman, D., & Garety, P. A. (2006). Helping patients with paranoid and suspicious thoughts: A cognitive-behavioural approach. *Advances in Psychiatric Treatment, 12*, 404-415. DOI: 10.1192/apt.12.6.404
- Freeman, D., & Garety, P. A. (2014). Advances in understanding and treating persecutory delusions: A review. *Social Psychiatry and Psychiatric Epidemiology, 49*, 1179-1189. DOI: 10.1007/s00127-014-0928-7
- 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 in a non-clinical population. *British Journal of Psychiatry, 186*, 427-435. DOI: 10.1192/bjp.186.5.427
- Freeman, D., Garety, P. A., & Kuipers, E. (2001). Persecutory delusions: Developing the understanding of belief maintenance and emotional distress. *Psychological Medicine, 31*, 1293-1306.
- 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*, 331-347. DOI: 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*, 89-99. DOI: 10.1016/j.brat.2006.01.014
- Freeman, D., McManus, S., Brugha, T., Meltzer, H., Jenkins, R., & Bebbington, P. E. (2011). Concomitants of paranoia in the general population. *Psychological Medicine, 41*, 923-936. DOI:10.1017/S00332917110001546

- Freeman, D., Pugh, K., Antley, A., Slater, M., Bebbington, P. E., Gittins, M., Dunn, G., Kuipers, E., Fowler, D., Garety, P. A. (2008). A virtual reality study of paranoid thinking in the general population. *British Journal of Psychiatry*, *192*, 258-263. DOI: 10.1192/bjp.bp.107.044677
- Freeman, D., Pugh, K., Dunn, G., Evans, N., Sheaves, B., Waite, F., Černis, E., Lister, R., & Fowler, D. (2014). An early Phase II randomised controlled trial testing the effect on persecutory delusions of using CBT to reduce negative cognitions about the self: The potential benefits of enhancing self-confidence. *Schizophrenia Research*, *160*, 186-192. DOI: 10.1016/j.schres.2014.10.038
- Freeman, D., Pugh, K., Vorontsova, N., Antley, A., & Slater, M. (2010). Testing the continuum of delusional beliefs: An experimental study using virtual reality. *Journal of Abnormal Psychology*, *119*, 83-92. DOI: 10.1037/a0017514
- Freeman, D., Slater, M., Bebbington, P. E., Garety, P. A., Kuipers, E., Fowler, D., Met, A., Read, C., Jordan, J., & Vinayagamorthy, V. (2003). Can virtual reality be used to investigate persecutory ideation? *Journal of Nervous and Mental Diseases*, *191*, 509-514. DOI: 10.1097/01.nmd.0000082212.83842.fe
- Freeman, D., Thompson, C., Vorontsova, N., Dunn, G., Carter, L. A., Garety P. A., Kuipers, E., Slater, M., Antley, A., Glucksman, E., & Ehlers, A. (2013). Paranoia and post-traumatic stress disorder in the months after a physical assault: A longitudinal study examining shared and differential predictors. *Psychological Medicine*, *43*, 2673-2684. DOI: 10.1017/S003329171300038X.
- Freeman, D., Waite, F., Emsley, R., Kingdon, D., Davies, L., Fitzpatrick, R., & Dunn, G. (2016). The efficacy of a new translational treatment for persecutory delusions: Study protocol for a randomised controlled trial (The Feeling Safe Study). *Trials*, *17*, 134. DOI: 10.1186/s13063-016-1245-0.

- Garety, P. A., Kuipers, E., Fowler, D., Freeman, D., & Bebbington, P. E. (2001). A cognitive model of the positive symptoms of psychosis. *Psychological Medicine, 31*, 113-154.  
DOI: 10.1017/S0033291701003312
- Gawronski, B., & Bodenhausen, G. V. (2007a). What do we know about implicit attitude measures and what do we have to learn? In B. Wittenbrink & N. Schwarz (Eds.), *Implicit Measures of Attitudes: Procedures and Controversies* (pp. 59-127). New York: Guilford Press.
- Gawronski, B., & Bodenhausen, G. V. (2007b). Unravelling the processes underlying evaluation: Attitudes from the perspective of the APE model. *Social Cognition, 25*, 687-717. DOI: 10.1521/soco.2007.25.5.687
- Gil, E., Luciano, C., Ruiz, F. J., & Valdivia-Salas, V. (2012). A preliminary demonstration of transformation of functions through hierarchical relations. *International Journal of Psychology and Psychological Therapy, 12*, 1-19.
- Gracie, A., Freeman, D., Green, S., Garety, P. A., Kuipers, E., Hardy, A., Ray, K., Dunn, G., Bebbington, P., & Fowler, D. (2007). The association between traumatic experience, paranoia and hallucinations: Test of the predictions of psychological models. *Acta Psychiatrica Scandinavica, 116*, 280-289. DOI: 10.1111/j.1600-0447.2007.01011.x
- Green, C. E., Freeman, D., Kuipers, E., Bebbington, P. E., Fowler, D., Dunn, G., & Garety, P. A. (2008). Measuring ideas of persecution and social reference: The Green et al. Paranoid Thought Scales (GPTS). *Psychological Medicine, 38*, 101-111.  
DOI: 10.1017/S0033291707001638
- Green, C. E., Garety, P. A., Freeman, D., Fowler, D., Bebbington, P. E., Dunn, G., & Kuipers, E. (2006). Content and affect in persecutory delusions. *British Journal of Clinical Psychology, 45*, 561-577. DOI: 10.1348/014466506X98768

- Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*, 3-25. DOI: 10.1037//0033-295X.109.1.3
- Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*, 1022-1038.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464-1480.
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Task: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17-41. DOI: 10.1037/a0015575
- Hayes, A. (2016). The PROCESS macro for SPSS and SAS.  
<http://processmacro.org/index.html>
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational Frame Theory: A Post-Skinnerian Account of Human Language and Cognition*. New York: Kluwer Academic/Plenum Publishers.
- Hayes, S. C., Barnes-Holmes, D., & Wilson, K. G. (2012). Contextual Behavioral Science: Creating a science more adequate to the challenge of the human condition. *Journal of Contextual Behavioral Science*, *1*, 1-16. DOI: 10.1016/j.jcbs.2012.09.004
- Hayes, S. C., Blackledge, J. T., & Barnes-Holmes, D. (2001). Language and cognition: Constructing an alternative approach within the behavioral tradition. In S. C. Hayes, D. Barnes-Holmes & B. Roche (Eds.) *Relational Frame Theory: A Post-Skinnerian Account of Human Language and Cognition* (pp. 3-20). New York: Kluwer Academic/Plenum Publishers.

- Hayes, S. C., & Brownstein, A. J. (1986). Mentalism, behavior-behavior relations, and a behavior-analytic view of the purposes of science. *The Behavior Analyst, 9*, 175-190.
- Hayes, S. C., Fox, E., Gifford, E. V., Wilson, K. G., Barnes-Holmes, D., & Healy, O. (2001). Derived relational responding as learned behavior. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.) *Relational Frame Theory: A Post-Skinnerian Account of Human Language and Cognition* (pp. 21-50). New York: Kluwer Academic/Plenum Publishers.
- Hayes, S. C., Strosahl, K., & Wilson, K. G. (1999). *Acceptance and Commitment Therapy: An experiential approach to behavior change*. New York: Guilford Press.
- Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2012). *Acceptance and Commitment Therapy: The process and practice of mindful change* (2nd edition). New York: Guilford Press.
- Hesse, K., Schroeder, P. A., Scheeff, J., Klingberg, S., & Plewnia, C. (2017). Experimental variation of social stress in virtual reality: Feasibility and first results in patients with psychotic disorders. *Journal of Behavior Therapy & Experimental Psychiatry, 56*, 129-136. DOI: 10.1016/j.jbtep.2016.11.006
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*, 1369-1385. DOI: 10.1177/0146167205275613
- Hooper, N., Villatte, M., Neofotistou, E., & McHugh, L. (2010). The effects of mindfulness versus thought suppression on implicit and explicit measures of experiential avoidance. *International Journal of Behavioral Consultation and Therapy, 6*, 233-244. DOI: 10.1037/h0100910

- Hughes, S., & Barnes-Holmes, D. (2016a). Relational Frame Theory: The basic account. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *The Wiley Handbook of Contextual Behavioral Science* (pp. 129-178). West Sussex, UK: Wiley-Blackwell.
- Hughes, S., & Barnes-Holmes, D. (2016b). Relational Frame Theory: Implications for the study of human language and cognition. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *The Wiley Handbook of Contextual Behavioral Science* (pp. 179-226). West Sussex, UK: Wiley-Blackwell.
- Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorizing in implicit attitude research: Propositional and behavioral alternatives. *The Psychological Record, 61*, 465-498.
- Hughes, S., Barnes-Holmes, D., & Vahey, N. (2012). Holding onto our functional roots when exploring new intellectual islands: A voyage through implicit cognition research. *Journal of Contextual Behavioral Science, 1*, 17-38. DOI: 10.1016/j.jcbs.2012.09.003
- Hughes, S., De Houwer, J., & Barnes-Holmes, D. (2016). The moderating impact of distal regularities on the effect of stimulus pairings: A novel perspective on evaluative conditioning. *Experimental Psychology, 63*, 20-44. DOI: 10.1027/1618-3169/a000310
- Hughes, S., De Houwer, J., & Perugini, M. (2016). The functional-cognitive framework for psychological research: Controversies and resolutions. *International Journal of Psychology, 51*, 4-14. DOI: 10.1002/ijop.12239
- Hussey, I., & Barnes-Holmes, D. (2012). The Implicit Relational Assessment Procedure as a measure of implicit depression and the role of psychological flexibility. *Cognitive and Behavioral Practice, 19*, 573-582. DOI: 10.1016/j.cbpra.2012.03.002

- Hussey, I., Barnes-Holmes, D., & Barnes-Holmes, Y. (2015). From Relational Frame Theory to implicit attitudes and back again: Clarifying the link between RFT and IRAP research. *Current Opinion in Psychology*, 2, 11-15. DOI: 10.1016/j.copsy.2014.12.009
- Hussey, I., Ní Mhaoileoin, D., Barnes-Holmes, D., Ohtsuki, T., Kishita, N., Hughes, S., & Murphy, C. (2015). The IRAP is non-relative but not a-contextual: Changes to the contrast category influence men's dehumanization of women. *The Psychological Record*, 66, 291-299.
- Hussey, I., Thompson, M., McEnteggart, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2015). Interpreting and inverting with less cursing: A guide to interpreting IRAP data. *Journal of Contextual Behavioral Science*, 4, 157-162. DOI: 10.1016/j.jcbs.2015.05.001
- Johns, L. C., & van Os, J. (2001). The continuity of psychotic experiences in the general population. *Clinical Psychology Review*, 21, 1125-1141. DOI: 10.1016/S0272-7358(01)00103-9
- Johns, L. C., Cannon, M., Singleton, N., Murray, R. M., Farrell, M., Brugha, T., Bebbington, P. E., Jenkins, R., & Meltzer, H., (2004). Prevalence and correlates of self-reported psychotic symptoms in the British population. *The British Journal of Psychiatry*, 185, 298-305. DOI: 10.1192/bjp.185.4.298
- Karpinski, A. (2004). Measuring self-esteem using the Implicit Association Test: The role of the other. *Personality and Social Psychology Bulletin*, 30, 22-34. DOI: 10.1177/0146167203258835
- Kavanagh, D., Hussey, I., McEnteggart, C., Barnes-Holmes, Y., & Barnes-Holmes, D. (2016). Using the IRAP to explore natural language statements. *Journal of Contextual Behavioral Science*, 5, 247-251. DOI: 10.1016/j.jcbs.2016.10.001

- Kesting, M. L., Bredenpohl, M., Klenke, J., Westermann, S., & Lincoln, T. M. (2013). The impact of social stress on self-esteem and paranoid ideation. *Journal of Behavior Therapy and Experimental Psychiatry, 44*, 122-128. DOI:10.1016/j.jbtep.2012.07.010
- Kesting M. L., & Lincoln, T. M. (2013). The relevance of self-esteem and self-schemata in individuals with persecutory delusions. *Comprehensive Psychiatry, 54*, 766-789. DOI: 10.1016/j.comppsy.2013.03.002
- Kesting, M. L., Mehl, S., Rief, W., Lindenmeyer, J., & Lincoln, T. M. (2011). When paranoia fails to enhance self-esteem: Explicit and implicit self-esteem and its discrepancy in patients with persecutory delusions compared to depressed and healthy controls. *Psychiatry Research, 186*, 197-202. DOI: 10.1016/j.psychres.2010.08.036
- Kinderman, P., & Bentall, R. P. (1997). Causal attributions in paranoia and depression: Internal, personal, and situational attributions for negative events. *Journal of Abnormal Psychology, 106*, 341-345.
- Kishita, N., Muto, T., Ohtsuki, T., Barnes-Holmes, D. (2014). Measuring the effect of cognitive defusion using the Implicit Relational Assessment Procedure: An experimental analysis with a highly socially anxious sample. *Journal of Contextual Behavioral Science, 3*, 8-15. DOI: 10.1016/j.jcbs.2013.12.001
- Klippel, A., Myin-Germeys, I., Chavez-Baldini, U. Y., Preacher, K. J., Kempton, M., ...Reininghaus, U. (2017). Modeling the interplay between psychological processes and adverse, stressful contexts and experiences in pathways to psychosis: An experience sampling study. *Schizophrenia Bulletin, 43*, 302-315. DOI: 10.1093/schbul/sbw185
- Levin, M. E., Twohig, M. P., & Smith, B. M. (2016). Contextual Behavioral Science: An overview. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *The*

*Wiley Handbook of Contextual Behavioral Science* (pp. 17-36). West Sussex, UK: Wiley-Blackwell.

- Lincoln, T. M., Stahnke, J., & Moritz, S. (2014). The short-term impact of a paranoid explanation on self-esteem: An experimental study. *Cognitive Therapy and Research, 38*, 397-406. DOI: 10.1007/s10608-014-9600-5
- MacKinnon, K., Newman-Taylor, K., & Stopa, L. (2011). Persecutory delusions and the self: An investigation of implicit and explicit self-esteem. *Journal of Behavior Therapy and Experimental Psychiatry, 42*, 54-64. DOI: 10.1016/j.jbtep.2010.05.003
- Martin, J. A., & Penn, D. L. (2001). Social cognition and subclinical paranoid ideation. *British Journal of Clinical Psychology, 40*, 261-265.
- Masillo, A., Day, F., Laing, J., Howes, O., Fusar-Poli, P., Byrne, M., Bhattacharyya, S., Fiori Nastro, P., Girardi, P., McGuire, P., & Valmaggia, L. (2012). Interpersonal sensitivity in the at-risk mental state for psychosis. *Psychological Medicine, 42*, 1835-1845. DOI: 10.1017/S0033291711002996.
- McEntegart, C., Barnes-Holmes, Y., & Adekuoroye, F. (2016). The effects of a voice hearing simulation on implicit fear of voices. *Journal of Contextual Behavioral Science, 5*, 154-159. DOI: 10.1016/j.jcbs.2016.06.003i
- McEntegart, C., Barnes-Holmes, Y., Egger, J., & Barnes-Holmes, D. (2016). A pilot study of the relations within which hearing voices participates: Towards a functional distinction between voice hearers and controls. *The Psychological Record, 66*, 611-626. DOI: 10.1007/s40732-016-0196-x
- McEntegart, C., Barnes-Holmes, Y., Mellon, M., van Dongen, L., Hendriks, A., Barnes-Holmes, D., Egger, J., & Harte, C. (*in press*). Recommendations for using the IRAP with a medicated in-patient population with a diagnosis of psychosis. *International Journal of Psychology and Psychological Therapy*.

- McHugh, L., Barnes-Holmes, Y., & Barnes-Holmes, D. (2004a). Perspective-taking as relational responding: A developmental profile. *The Psychological Record, 54*, 115-144.
- McKay, R., Langdon, R., & Coltheart, M. (2007). The defensive function of persecutory delusions: An investigation using the Implicit Association Test. *Cognitive Neuropsychiatry, 12*, 1-24. DOI: 10.1080/13546800500363996
- Monestes, J. L., Villatte, M., Stewart, I., & Loas, G. (2014). Rule-based insensitivity and delusion maintenance in schizophrenia. *Psychological Record, 64*, 329-338. DOI: 10.1007/s40732-014-0029-8
- Moritz, S., Werner, R., & von Collani, G. (2006). The inferiority complex in paranoia readdressed: A study with the Implicit Association Test. *Cognitive Neuropsychiatry, 11*, 402-415. DOI: 10.1080/13546800444000263
- Morris, E. M. J., Johns, L. C., & Oliver, J. E. (2013). *Acceptance and Commitment Therapy and Mindfulness for Psychosis*. Wiley-Blackwell: John Wiley & Sons, Ltd.
- Morrison, A. P., Frame, L., & Larkin, W. (2003). Relationships between trauma and psychosis: A review and integration. *British Journal of Clinical Psychology, 42*, 331-353.
- Munnely, A., Martin, G., Dack, C., Zedginidze, A., & McHugh, L. (2014). The transfer of social exclusion and inclusion functions through derived stimulus relations. *Learning & Behavior, 42*, 270-280. DOI: 10.3758/s13420-014-0144-z.
- Murphy, S., Murphy, J., & Shevlin, M. (2015). Negative evaluations of self and others, and peer victimization as mediators of the relationship between childhood adversity and psychotic experiences in adolescence: The moderating role of loneliness. *British Journal of Clinical Psychology, 54*, 326-344. DOI:10.1111/bjc.12077

- Nicholson, E., & Barnes-Holmes, D. (2012a). The Implicit Relational Assessment Procedure (IRAP) as a measure of spider fear. *The Psychological Record*, *62*, 263-277.
- Nicholson, E., & Barnes-Holmes, D. (2012b). Developing an implicit measure of disgust propensity and disgust sensitivity: Examining the role of implicit disgust propensity and sensitivity in obsessive-compulsive tendencies. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*, 922-930. DOI:10.1016/j.jbtep.2012.02.001.
- Nicholson, E., McCourt, A., & Barnes-Holmes, D. (2013). The Implicit Relational Assessment Procedure as a measure of obsessive beliefs in relation to disgust. *Journal of Contextual Behavioral Science*, *2*, 23-30. DOI:10.1016/j.jcbs.2013.02.002.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231-259.
- Nosek, B. A., & Banaji, M. R. (2001). The Go/No-go Association Task. *Social Cognition*, *19*, 625-66.
- O'Hora, D., Pelaez, M., Barnes-Holmes, D., Rae, G., Robinson, K., & Chaudhary, T. (2008). Temporal relations and intelligence: Correlating relational performance with performance on the WAIS-III. *The Psychological Record*, *58*, 569-584.
- Oliver, J., E., O'Connor, J. A., Jose, P. E., McLachlan, K., & Peters, E. (2011). The impact of negative schemas, mood and psychological flexibility on delusional ideation: Mediating and moderating effects. *Psychosis*, *4*, 1-13. DOI: 10.1080/17522439.2011.637117
- Palmier-Claus, J., Dunn, G., Drake, R., & Lewis, S. (2011). The negative and positive self: A longitudinal study examining self-esteem, paranoia and negative symptoms in individuals with first-episode psychosis. *Early Intervention in Psychiatry*, *5*, 150-155. DOI: 10.1111/j.1751-7893.2010.00250.x

- Parling, T., Cernvall, M., Stewart, I., Barnes-Holmes, D., & Ghaderi, A. (2012). Using the Implicit Relational Assessment Procedure to compare implicit pro-thin/anti-fat attitudes of patients with anorexia nervosa and non-clinical controls. *Eating Disorders: The Journal of Treatment and Prevention, 20*, 127-143. DOI: 10.1080/10640266.2012.654056
- Peters, E. R., Joseph, S. A., Day, S., & Garety, P. A. (2004). Measuring Delusional Ideation: The 21-item Peters et al. Delusions Inventory (PDI). *Schizophrenia Bulletin, 30*, 1005-1022.
- Peters, E. R., Joseph, S. A., & Garety, P. A. (1999b). Measurement of delusional ideation in the normal population: Introducing the PDI (Peters et al. Delusions Inventory). *Schizophrenia Bulletin, 25*, 553-576.
- Pinter, B., & Greenwald, A. G. (2005). Clarifying the role of the “other” category in the self-esteem IAT. *Experimental Psychology, 52*, 74-79. DOI: 10.1027/1618-3169.52.1.74
- Raes, F., & Van Gucht, D. (2009). Paranoia and instability of self-esteem in adolescents. *Personality and Individual Differences, 47*, 928-932. DOI: 10.1016/j.paid.2009.07.017
- Read, J., Agar, K., Argyle, N., & Aderhold, V. (2003). Sexual and physical assault during childhood and adulthood as predictors of hallucinations, delusions and thought disorder. *Psychology and Psychotherapy: Theory, Research, and Practice, 76*, 1-22.
- Read, J., van Os, J., Morrison, A. P., & Ross, C. A. (2005). Childhood trauma, psychosis and schizophrenia: A literature review with theoretical and clinical implications. *Acta Psychiatrica Scandinavica, 112*, 330-350. DOI:10.1111/j.1600-0447.2005.00634.x
- Remue, J., DeHouwer, J., Barnes-Holmes, D., Vanderhasselt, M., De Raedt, R. (2013). Self-esteem revisited: Performance on the Implicit Relational Assessment Procedure as a

- measure of self- versus ideal self-related cognitions in dysphoria. *Cognition and Emotion*, 27, 1441-1449. DOI: 10.1080/02699931.2013.786681.
- Remue, J., Hughes, S., De Houwer, J., & De Raedt, R. (2014). To be or want to be: Disentangling the role of actual versus ideal self in implicit self-esteem. *PLoS ONE* 9(9): e108837. DOI:10.1371/journal.pone.0108837
- Roger, D., Jarvis, G., & Najarian, B. (1993). Detachment and coping: The construction and validation of a new scale for measuring coping strategies. *Personality and Individual Differences*, 15, 619-626.
- Rosenberg, M. (1965). *Society and the Adolescent Self-image*. Princeton, NJ: Princeton University Press.
- Ross, L., & Nisbett, R. E. (1991). *The Person and the Situation: Perspectives on Social Psychology*. New York: McGraw-Hill.
- Salvatore, G., Lysaker, P. H., Popolo, R., Procacci, M., Carcione, A., & Dimaggio, G. (2012). Vulnerable self, poor understanding of others' minds, threat anticipation and cognitive biases as triggers for delusional experience in schizophrenia: a theoretical model. *Clinical Psychology and Psychotherapy*, 19, 247-259. DOI: 10.1002/cpp.746
- Selten, J-P., van der Ven, E., Rutten, B. P. F., & Cantor-Graae, E. (2013). The social defeat hypothesis of schizophrenia: An update. *Schizophrenia Bulletin*, 39, 1180-1186. DOI: 10.1093/schbul/sbt134
- Smith, B., Fowler, D., Freeman, D., Bebbington, P. E., Bashforth, H., Garety, P. A., Dunn, G., & Kuipers, E. (2006). Emotion and psychosis: Links between depression, self-esteem, negative schematic beliefs and delusions and hallucinations. *Schizophrenia Research*, 86, 181-188. DOI: 10.1016/j.schres.2006.06.018

- Stewart, C., Stewart, I., & Hughes, S. (2016). A contextual behavioral approach to the study of (persecutory) delusions. *Journal of Contextual Behavioral Science*. DOI: 10.1016/j.jcbs.2016.09.002
- Stewart, I. (2016). The fruits of a functional approach for psychological science. *International Journal of Psychology*, 51, 15-27. DOI: 10.1002/ijop.12184
- Stewart, I., Barnes-Holmes, D., & Roche, B. (2004). A functional-analytic model of analogy using the Relational Evaluation Procedure. *The Psychological Record*, 54, 531-552.
- Stewart, I. & Roche, B. (2013). Relational Frame Theory: An overview. In S. Dymond & B. Roche (Eds.) *Advances in Relational Frame Theory: Research and Application*. CA, USA: Context Press/New Harbinger Publications Inc.
- 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, 143-153. DOI: 10.1037/0021-843X.117.1.143
- Thewissen, V., Myin-Germeys, I., Bentall, R. P., de Graaf, R., Vollebergh, W., & van Os, J. (2007). Instability in self-esteem and paranoia in a general population sample. *Social Psychiatry & Psychiatric Epidemiology*, 42, 1-5. DOI: 10.1007/s00127-006-0136-1
- Tiernan, B., Tracey, C., & Shannon, C. (2014). Paranoia and self-concepts in psychosis: A systematic review of the literature. *Psychiatry Research*, 216, 303-313. DOI: 10.1016/j.psychres.2014.02.003
- Timko, C. A., England, E. L., Herbert, J. D., & Forman, E. M. (2010). The Implicit Relational Assessment Procedure as a measure of self-esteem. *The Psychological Record*, 60, 679-698.
- Törneke, N. (2010). *Learning RFT: An Introduction to Relational Frame Theory and Its Clinical Application*. CA, USA: New Harbinger Publications, Inc.

- Törneke, N., Luciano, C., & Valdivia Salas, S. (2008). Rule-governed behavior and psychological problems. *International Journal of Psychology and Psychological Therapy*, 8, 141-156.
- Udachina, A., Thewissen, V., Myin-Germeys, I., Fitzpatrick, S., O'Kane, A., & Bentall, R. P. (2009). Understanding the relationships between self-esteem, experiential avoidance, and paranoia: Structural Equation Modelling and experience sampling studies. *Journal of Nervous & Mental Disease*, 9, 661-668. DOI: 10.1097/NMD.0b013e3181b3b2ef
- Vahey, N. A., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). A first test of the Implicit Relational Assessment Procedure (IRAP) as a measure of self-esteem: Irish prisoner groups and university students. *The Psychological Record*, 59, 371-388.
- Vahey, N. A., Nicholson, E., & Barnes-Holmes, D. (2015). A meta-analysis of criterion effects for the Implicit Relational Assessment Procedure (IRAP) in the clinical domain. *Journal of Behavior Therapy and Experimental Psychiatry*, 48, 59-65. DOI: 10.1016/j.jbtep.2015.01.004.
- Valiente, C., Cantero, D., Vázquez, C., Sanchez, Á., Provencio, M., & Espinosa, R. (2011). Implicit and explicit self-esteem discrepancies in paranoia and depression. *Journal of Abnormal Psychology*, 120, 691-699. DOI: 10.1037/a0022856
- Valmaggia, L. R., Day, F., Garety, P. A., Freeman, D., Antley, A., Slater, M., Swapp, D., Myin-Germeys, I., & McGuire, P. (2015). Social defeat predicts paranoid appraisals in people at high risk for psychosis. *Schizophrenia Research*, 168, 16-22. DOI: 10.1016/j.schres.2015.07.050
- Valmaggia, L. R., Day, F. L., Kroll, J., Laing, J., Byrne, M., Fusar-Poli, P., & McGuire, P. (2015). Bullying victimisation and paranoid ideation in people at ultra-high risk for psychosis. *Schizophrenia Research*, 168, 68-73. DOI: 10.1016/j.schres.2015.08.029

- Valmaggia, L. R., Day, F., & Rus-Calafell, M. (2016). Using Virtual Reality to investigate psychological processes and mechanisms associated with the onset and maintenance of psychosis: A systematic review. *Social Psychiatry & Psychiatric Epidemiology*, *51*, 921-936. DOI: 10.1007/s00127-016-1245-0
- van Os, J., & Verdoux, H. (2003). Diagnosis and classification of schizophrenia: Categories versus dimensions, distributions versus disease. In R. M. Murray, P. B. Jones, E. Susser, J. van Os, & M. Cannon (Eds.), *The Epidemiology of Schizophrenia* (pp. 364-410). Cambridge, UK: Cambridge University Press.
- 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*, 661-671. DOI: 10.1093/schbul/sbs050
- Vázquez, C., Diez-Alegría, C., Hernández-Lloreda, M. J., Moreno, M. N. (2008). Implicit and explicit self-schema in active deluded, remitted deluded, and depressed patients. *Behavior Therapy and Experimental Psychiatry*, *39*, 587-599. DOI: 10.1016/j.jbtep.2008.01.006
- Veling, W., Pot-Kolder, R., Counotte, J., van Os, J., & van der Gaag, M. (2016). Environmental social stress, paranoia and psychosis liability: A virtual reality study. *Schizophrenia Bulletin*, *42*, 1363-1371. DOI: 10.1093/schbul/sbw031
- Vitale, A., Barnes-Holmes, Y., Barnes-Holmes, D., & Campbell, C. (2008). Facilitating responding in accordance with the relational frame of comparison: Systematic empirical analyses. *The Psychological Record*, *58*, 365-390.
- Waller, H., Emsley, R., Freeman, D., Bebbington, P. E., Dunn, G., Fowler, D., Hardy, A., Kuipers, E., & Garety, P. A. (2015). Thinking Well: A randomised controlled feasibility study of a new CBT therapy targeting reasoning biases in people with

distressing persecutory delusional beliefs. *Journal of Behavior Therapy and Experimental Psychiatry*, 48, 82-89. DOI: 10.1016/j.jbtep.2015.02.007

Westermann, S., Kesting, M. L., & Lincoln, T. M. (2012). Being deluded after being excluded? How emotion regulation deficits in paranoia-prone individuals affect state paranoia during experimentally induced social stress. *Behavior Therapy*, 43, 329-340. DOI: 10.1016/j.beth.2011.07.005

Wilson, K. G., Whiteman, K., & Bordieri, M. (2013). The pragmatic truth criterion and values in contextual behavioral science (pp. 27-47). In S. Dymond & B. Roche (Eds.) *Advances in Relational Frame Theory: Research and Applications*. Context Press.

**Appendix A: The Paranoia Checklist (Freeman et al., 2005)**

<b>Frequency</b> “How often have you had the thought?”	Rarely	Once a month	Once a week	Several times a week	At least once a day
<b>Conviction</b> “How strongly do you believe it?”	Do not believe it	Believe it a little	Believe it somewhat	Believe it a lot	Absolutely believe it
<b>Distress</b> “How upsetting is it for you?”	Not distressing	A little distressing	Somewhat distressing	Moderately distressing	Very distressing

1. I need to be on my guard against others
2. There might be negative comments being circulated about me
3. People deliberately try to irritate me
4. I might be being observed or followed
5. People are trying to make me upset
6. People communicate about me in subtle ways
7. Strangers and friends look at me critically
8. People might be hostile towards me
9. Bad things are being said about me behind my back
10. Someone I know has bad intentions towards me
11. I have a suspicion that someone has it in for me
12. People would harm me if given an opportunity
13. Someone I don't know has bad intentions towards me
14. There is a possibility of a conspiracy against me
15. People are laughing at me
16. I am under threat from others
17. I can detect coded messages about me in the press/TV/radio
18. My actions and thoughts might be controlled by others

**Appendix B:** The state-adapted Paranoia Checklist (Westermann et al., 2012)

**Instructions:** How strongly do the following thoughts apply to you *at the moment*?

---

Not at all

A little

Somewhat

Moderately

Strongly

---

1. People deliberately try to irritate me
2. People are trying to make me upset
3. Someone I don't know has bad intentions towards me
4. There is a possibility of a conspiracy against me
5. People are laughing at me
6. I am under threat from others

### Appendix C: Rosenberg Self-esteem Scale (Rosenberg, 1965)

**Instructions:** Below is a list of statements dealing with your general feelings about yourself.

Select one response option to indicate how much you agree with each statement.

---

Strongly agree	Agree	Disagree	Strongly disagree
<hr/>			
1. On the whole, I am satisfied with myself			
2. At times I think I am no good at all*			
3. I feel that I have a number of good qualities			
4. I am able to do things as well as most other people			
5. I feel I do not have much to be proud of*			
6. I certainly feel useless at times*			
7. I feel I am a person of worth, at least on an equal plane with others			
8. I wish I could have more respect for myself*			
9. All in all, I am inclined to feel that I am a failure*			
10. I take a positive attitude towards myself			

\* denote reverse-scored items

**Appendix D:** Cyberball Questionnaire (Williams et al., 2000; Williams, 2009)

**Instructions:** For each question, please select the number that best represents the feelings you were experiencing during the game.

---

1				5
Not at all	2	3	4	Extremely

---

***Belonging***

1. I felt disconnected\*
2. I felt rejected\*
3. I felt like an outsider\*
4. I felt I belonged to the group
5. I felt the other players interacted with me a lot

***Self-esteem***

6. I felt good about myself
7. My self-esteem was high
8. I felt liked
9. I felt insecure\*
10. I felt satisfied

***Meaningful existence***

11. I felt invisible\*
12. I felt meaningless\*
13. I felt nonexistent\*
14. I felt important
15. I felt useful

**Control**

- 16. I felt powerful
- 17. I felt I had control over the game
- 18. I felt I had the ability to significantly alter events
- 19. I felt I was unable to influence the actions of others\*
- 20. I felt the other players decided everything\*

\* denote reverse-scored items

**Instructions:** For each question, please select the number that best represents the feelings you were experiencing during the game.

---

1	2	3	4	5	6	7	8	9	10
Does not apply									Applies strongly

---

Embarrassed / Ridiculed / Ashamed / Foolish

Frightened / Timid / Afraid / Scared

Sad / Depressed / Miserable / Dejected

Happy / Gay / Cheerful / Delighted

Angry / Annoyed / Mad / Sore

I am frustrated

**Instructions:** For each question, please select the number that best represents the feelings you were experiencing during the game.

---

1	2	3	4	5
Not at all				Extremely

---

I was ignored

I was excluded

Assuming the ball should be thrown to each player equally (33% for each person), what percentage of throws did you receive? \_\_\_\_\_%

**Appendix E: Brief Core Schemas Scales (Fowler et al., 2006)**

**Instructions:** This questionnaire lists beliefs that people can hold about themselves and other people. Please indicate how strongly you hold each belief by selecting one of the 5 options. Try to judge the beliefs on how you currently view yourself and others. Do not spend too long on each belief. There are no right or wrong answers and the first response to each belief is often the most accurate.

0	1	2	3	4
Do not believe it	Believe it slightly	Believe it moderately	Believe it very much	Believe it totally
<b>Negative Self</b>	<b>Positive Self</b>	<b>Negative Others</b>	<b>Positive Others</b>	
I am unloved	I am a respected	Others are hostile	Others are fair	
I am worthless	I am valuable	Others are harsh	Others are good	
I am weak	I am talented	Others are unforgiving	Others are trustworthy	
I am vulnerable	I am successful	Others are bad	Others are accepting	
I am bad	I am good	Others are devious	Others are supportive	
I am a failure	I am interesting	Others are nasty	Others are truthful	

**Appendix F: Depression, Anxiety and Stress Scale-21 (Lovibond & Lovibond, 1995)**

**Instructions:** Please read each statement and indicate how much the statement applied to you over the past week. There are no right or wrong answers. Do not spend too much time on any statement.

---

Did not apply to me at all	Applied to me to some degree, or some of the time	Applied to me to a considerable degree, or a good part of the time	Applied to me very much, or most of the time
----------------------------	---	--	--

---

1. I found it hard to wind down
2. I was aware of dryness of my mouth
3. I couldn't seem to experience any positive feeling at all
4. I experienced breathing difficulty (e.g., excessively rapid breathing, breathlessness in the absence of physical exertion)
5. I found it difficult to work up the initiative to do things
6. I tended to over-react to situations
7. I experienced trembling (e.g., in the hands)
8. I felt that I was using a lot of nervous energy
9. I was worried about situations in which I might panic and make a fool of myself
10. I felt that I had nothing to look forward to
11. I found myself getting agitated
12. I found it difficult to relax
13. I felt down-hearted and blue
14. I was intolerant of anything that kept me from getting on with what I was doing
15. I felt I was close to panic
16. I was unable to become enthusiastic about anything
17. I felt I wasn't worth much as a person
18. I felt that I was rather touchy
19. I was aware of the action of my heart in the absence of physical exertion (e.g., sense of heart rate increase, heart missing a beat)
20. I felt scared without any good reason
21. I felt that life was meaningless

**Appendix G: Research Ethics Committee approval letter from NUI Galway with accompanying Participant Information Sheet, Consent Form, and Participant Debriefing Form**



Leas-Uachtarán  
um Thaighde

Vice President  
for Research

1 November 2016

Ref: 16-Oct-17

Corinna Stewart  
Room G042  
School of Psychology  
NUI Galway

Dear Ms Stewart

**Re: 16-Oct-17 'Implicit and explicit responding to the self as vulnerable and others as untrustworthy in the context of paranoia within the general (non-clinical) population: The effect of social exclusion on responding'.**

I write to you regarding the above proposal which was submitted for ethical review. I am pleased to inform you that at the meeting of the Research Ethics Committee held on 25 October 2016 in the Research and Innovation Building, NUI Galway, it was the decision of the Committee to grant your proposal **APPROVAL**.

When the decision was taken I was chairing the meeting and the following members were also present:

Dr Manus Biggs	Ms Laura Dempsey	Dr Eilionoir Flynn
Dr Brian Hallahan	Dr Martina Kelly	Dr Veronica McCauley
Mr Patrick Towers	Jane Walsh	

All NUI Galway Research Ethic Committee approval is given subject to the Principal Investigator submitting annual and final statements of compliance. The first statement is due on or before 1 November 2017.

See annual and final statement of compliance forms below. Section 7 of the REC's Standard Operating Procedures gives further details, and also outlines other instances where you are required to report to the REC.

Yours sincerely

Allyn Fives  
Chair, Research Ethics Committee

OÉ Gaillimh,  
Bóthar na hOllscoile,  
Gaillimh, Éire

NUI Galway,  
University Road,  
Galway, Ireland

T +353 91 495 312  
F +353 91 494 591

[www.nuigalway.ie/research/vp\\_research](http://www.nuigalway.ie/research/vp_research)

## PARTICIPANT INFORMATION SHEET

### 1 Title of Project: Online social tasks and attitudes to the self and others

### 2 Invitation

You are invited to take part in a research study. Before you decide, it is important that you understand why the research is being done and what it will involve. This *Participant Information Sheet* tells you about the purpose, risks and benefits of this research study. If you agree to take part, we will ask you to sign a *Consent Form*. If there is anything that you are not clear about, we will be happy to explain it to you. Please take as much time as you need to read this information. You should only consent to participate in this research study when you feel you understand what is being asked of you, and you have had enough time to think about your decision. Thank you for reading this.

### 3 Purpose of the Study

This study is concerned with how online social tasks might affect views of ourselves and others. You have been asked to participate as we have invited all NUI Galway students to take part.

### 4 Taking Part – What it Involves

#### Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this *Information Sheet* to keep and be asked to sign a *Consent Form*. If you decide to take part you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect your rights in any way.

#### What will happen to me if I take part?

If you decide to participate you be asked to complete a computer task and four questionnaires. You will then be invited to complete a simple online social task followed by the computer task and questionnaires again. The study will take place in a room located in the School of Psychology. You can complete the study at a time and day that suits you.

#### How long will my part in the study last?

The study will take approximately 1 hour and breaks can be taken throughout.

#### What are the possible benefits in taking part?

Course credits are offered to first and second year psychology students who participate. We hope that participating in this research will be an interesting experience and give you an idea of lab based research in psychology.

#### What are the possible disadvantages and risks of taking part?

While there are no discernible risks, there is a possibility that some of the content of the computer task or questionnaires may be uncomfortable or distressful to some participants. If you become uncomfortable or distressed during the study we can stop at any time and offer you support to alleviate your distress.

**What are my rights?**

Please note that you have the right to considerable, respectful treatment in a pleasant and safe psychological environment. You have the right to full information about the study and the right to ask questions at any time. In addition, all information relating to participation in this study is confidential. Any results or information gathered will be securely stored in a way that protects your identity. Personal information will be stored separately to data and participants will only be identified by a number. Original data files will be stored securely.

**What happens at the end of the study?**

When all participants have been tested (this should be within 6-8 months of your participation), you will receive a summary of one or 2 pages of the main findings. While it could be up to 2 years before final results are published, we would be pleased to include you on an address list to receive publications arising from the study. Only general findings will be reported, without reference to identifiable individual results.

**What happens if I change my mind during the study?**

You are entitled to change your mind about participating in this at any time without disadvantage or penalty.

**Who do I contact for more information or if I have further concerns?**

Please do not hesitate to contact the researcher if you have any further queries – [c.stewart2@nuigalway.ie](mailto:c.stewart2@nuigalway.ie)

If you have any concerns about this study and wish to contact someone in confidence, you may contact: **The Head, School of Psychology, National University of Ireland, Galway.**





## DEBRIEFING PROCEDURE

### **Title of Project: Online social tasks and attitudes to the self and others**

*This procedure is read to the participant. We do not include this information about Cyberball in the Participant Debriefing Form to avoid the possibility of other potential participants finding out about the true nature of this game and the aims of the study before taking part as this may affect responding on the measures after participants play Cyberball.*

Thank you for taking part in this study. Your participation is invaluable to our research. The purpose of this study is to investigate the effects of interpersonal experiences on attitudes to the self and others, self-esteem, suspicious thoughts, and mood.

The game you played during the study, called “Cyberball”, did not involve real players. The two players were actually computer-generated avatars. There are two conditions within this study. In one condition, participants are treated fairly and are included by the other players. In the second condition, participants are treated fairly and are excluded by the other players. We are interested to see if there are differences between the two conditions and how being excluded affects people’s attitudes to themselves and others, as well as suspicious thoughts, mood and self-esteem. We apologize for any discomfort this may have caused you and will try to address any questions or concerns you may have about the study.

You are welcome to contact me (the researcher) with any questions or queries, even after your participation has ended today. If you feel affected by any of the topics related to this research, I will try to offer you support to alleviate any distress that you may feel and to address any concerns that you may have. If you would like to speak to a professional counsellor or psychotherapist about any issues related to the topic of this study, the contact information for the NUI Galway Student Counselling Services is provided the Participant Debriefing Form.

Finally, we ask that you do not tell anyone else who may participate in this study about the nature of Cyberball or the aim of this study as this may affect our results.

Thank you again for your participation.



## **PARTICIPANT DEBRIEFING SHEET**

### **Title of Project: Online social tasks and attitudes to the self and others**

Thank you for taking part in this study. Your participation is invaluable to our research. The purpose of this study is to investigate the effects of interpersonal experiences on attitudes to the self and others, self-esteem, suspicious thoughts, and mood.

We would like to assure you again that your data is confidential and that your responses will not be linked to your name or any personal information. Data from the computer tasks and questionnaires will be de-identified and stored anonymously. Your data will be stored securely and destroyed after 5 years. All information collected as part of this study will be used solely for experimental purposes. The data collected from all participants who take part in this study will be pooled together. Once the study is completed we will produce a report that will describe the findings of the study. You will not be identified in any report or publication. The report will not include any personal details of the people who took part; it will only describe the overall results of the study.

If you have any further questions or concerns, you can contact the researcher, even after your participation has ended. If you feel affected by the topic of this study, the researcher will try to offer you support to alleviate any distress that you may feel and to address any concerns that you may have. If you would like to speak to a professional counsellor about seeking further support, please see the information provided below about the NUI Galway Student Counselling Services. If at a later date, you have questions or concerns about this study, do not hesitate to contact the researcher.

**Researcher: Corinna Stewart**  
Email: [c.stewart2@nuigalway.ie](mailto:c.stewart2@nuigalway.ie)  
Phone: 083 130 5644  
School of Psychology

**NUI Galway Student Counselling Services**  
Email: [counselling@nuigalway.ie](mailto:counselling@nuigalway.ie)  
Phone: 091 492 484  
5 Distillery Road, NUI Galway NUI Galway  
Website: [http://www.nuigalway.ie/student\\_services/counsellors/](http://www.nuigalway.ie/student_services/counsellors/)

## Appendix H: Research Ethics Committee approval from University College Dublin (UCD)

with accompanying Participant Information Sheet, Consent Form, and Participant Debriefing  
Form



### UCD Office of Research Ethics

Roebuck Castle  
University College Dublin  
Belfield, Dublin 4, Ireland

T +353 1 716 8767

### An Oifig Eitic Thaighde UCD

Caisleán Roebuck  
An Colaiste Ollscoile, Baile Átha Cliath  
Belfield, Baile Átha Cliath 4, Éire

hrec@ucd.ie  
www.ucd.ie/researchethics

November 16<sup>th</sup>, 2016

Dr Louise McHugh  
UCD School of Psychology  
Newman Building  
Belfield  
Dublin 4

**Re: HS-16-63-McHugh:** *Implicit and explicit responding to the self as vulnerable and others as untrustworthy in the context of paranoia within the general (non-clinical) population: The effect of social exclusion on responding*

Dear Dr McHugh

Thank you for your response to the Human Research Ethics Committee – Humanities (8 & 14/11/16). The Decision of the Committee is that **approval is granted** for this application which is subject to the conditions set out below.

Please note that **public liability insurance for this study has been confirmed** in accordance with our guidelines.<sup>[1]</sup>

**Your request to access UCD students was also reviewed and granted.** Please ensure that any additional permissions to access participants, whether internal (heads of Schools) or external are obtained before the recruitment of the participants is commenced.

Please note that approval is for the work and the time period specified in the above protocol and is subject to the following:

- Any amendments or requests to extend the original approved study will need to be approved by the Committee. Therefore you will need to submit by email the *Request to Amend/Extend Form* (HREC Doc 10);
- Any unexpected adverse events that occur during the conduct of your research should be notified to the Committee. Therefore you will need to Submit, by email, an *Unexpected Adverse Events Report* (HREC Doc 11);
- You or your supervisor (if applicable) are required to submit a signed *End of Study Report Form* (HREC Doc 12) to the Committee upon the completion of your study;
- This approval is granted on condition that you ensure that, in compliance with the Data Protection Acts 1988 and 2003. If applicable, all data will be managed in accordance with your application and that you will confirm this in your *End of Study Report* (HREC Doc 12).

.../.

- Please note that further **new** submissions from you may not be reviewed until any **End of Study Reports due** have been submitted to the Office of Research Ethics. That is, any earlier study that you received ethical approval for from the UCD HRECs;
- You may require copies of submitted documentation relating to this approved application and therefore we advise that you retain copies for your own records;
- Please note that the granting of this ethical approval is premised on the assumption that the research will be carried out within the limits of the law;
- Please also note that approved applications and any subsequent amendments are subject to a Research Ethics Compliance Review.

The Committee wishes you well with your research and look forward to receiving your End of Study Report. All forms are available on the website [www.ucd.ie/researchethics](http://www.ucd.ie/researchethics) please ensure that you submit the latest version of the relevant form. If you have any queries regarding the above please contact the Office of Research Ethics and please quote your reference in all correspondence.

Yours sincerely,



---

Dr Joan Tiernan  
Chair Human Research Ethics Committee - Humanities

---

<sup>[i]</sup> [http://www.ucd.ie/researchethics/information\\_for\\_researchers/insurance/](http://www.ucd.ie/researchethics/information_for_researchers/insurance/)

## PARTICIPANT INFORMATION SHEET



### **Title of Project: Online social tasks and attitudes to the self and others**

My name is Corinna Stewart and I am a PhD student at NUI Galway under the supervision of Dr. Ian Stewart. I am working with Dr. Louise McHugh in School of Psychology at UCD. I am inviting you to take part in a study about online social tasks and attitudes to the self and others.

### **What is this research about?**

This study investigates online social tasks and attitudes to the self and others. You will be asked to complete a computer task and some questionnaires about your views of yourself and others and to take part in a brief online social game. You do not have to have any experience in online computer games to participate.

### **Why have I been asked to participate?**

This study requires voluntary participants who meet the following criteria:

- Are aged 18 years or over
- Are fluent in English
- Have normal or corrected to normal vision
- Do not have a diagnosis of psychosis or related psychotic illness

You should not participate in this study if any of these criteria apply to you.

### **What will happen if I decide to take part in this research study?**

Participation involves completion of a computer task and questionnaires that measure attitudes to the self and others. You will then be asked to take part in a brief simple online social game with two other players. Afterwards, you will be asked to complete the computer task and questionnaires again. This study takes approximately 1 hour to complete and you can take breaks throughout.

### **How will you protect my privacy?**

The only identifiable information about you that will be collected in this study is your name and signature on the consent form, which is needed to document your agreement to participate. These consent forms will be locked in a secure filing cabinet, which only the researcher will have access to. In order to protect your identity, you will be assigned a participant ID code. Data from the computer task and questionnaires will be de-identified and stored anonymously. All information collected as part of this study will be used solely for experimental purposes and your personal identity will not be revealed in any publication of this research. After 5 years, all participants' data will be destroyed.

### **What are the benefits of taking part in this research study?**

There are no direct benefits to participation. Participation is voluntary and you will not be awarded monetary or other rewards for taking part. However, this research can help improve our knowledge about online social tasks and attitudes to the self and others.

### **What are the potential risks of taking part in this research study?**

There is minimal risk associated with participation. There is a possibility that some components of the study (e.g., questionnaires) may be uncomfortable to some participants. If you feel uncomfortable at any stage during the study, you can stop at any time. You are also

welcome to ask questions about the study at any time, even after your participation has ended.

**Can I change my mind at any stage and withdraw from the study?**

Participants have the right to withdraw at any stage before, during, or after data collection. You do not have to provide a reason for doing so and this will not affect your rights in any way. If you would like to withdraw your data from the study, please contact the researcher. You can withdraw your data until the point where the study is written up for my PhD thesis and/or submitted for publication. Please note, your personal identity will not be revealed in any publication of this research.

**Who do I contact for more information or if I have further concerns?**

If you have any further questions about this study, please do not hesitate to contact me:  
Corinna Stewart      Email: [c.stewart2@nuigalway.ie](mailto:c.stewart2@nuigalway.ie)      Phone: 083 130 5644

This research is supervised by Dr. Louise McHugh in the School of Psychology at UCD:  
[louise.mchugh@ucd.ie](mailto:louise.mchugh@ucd.ie) Phone: 01 7168408

and by Dr. Ian Stewart at the School of Psychology, NUI Galway:  
Email: [ian.stewart@nuigalway.ie](mailto:ian.stewart@nuigalway.ie) Phone: 091 493569

**CONSENT FORM**



**Title of Project: Online social tasks and attitudes to the self and others**

**Name of researcher:** Ms. Corinna Stewart

**DECLARATION**

**Please initial box**

I have read the Participant Information Sheet and have had time to consider whether to take part in this study.

I confirm that I have met all of the study inclusion criteria and

- am aged 18 years or over
- am fluent in English
- have normal or corrected to normal vision
- do not have a diagnosis of psychosis or related psychotic illness

I understand that my participation is voluntary (it is my choice) and that I am free to withdraw from the research at any time without disadvantage.

I understand that, as part of this research project I am voluntarily agreeing to complete a computer task and questionnaires and to take part in an online social task.

I understand that my name will not be linked to my data and that I will not be identified in any publication of this data.

I agree to take part in this research.

**Name of Participant (in block letters):** \_\_\_\_\_

**Signature:** \_\_\_\_\_ **Date:** / /

**Name of researcher (in block letters):** \_\_\_\_\_

**Signature:** \_\_\_\_\_ **Date:** / /

## DEBRIEFING PROCEDURE



### **Title of Project: Online social tasks and attitudes to the self and others**

*This procedure is read to the participant. We do not include this information about Cyberball in the Participant Debriefing Form to avoid the possibility of other potential participants finding out about the true nature of this game and the aims of the study before taking part as this may affect responding on the measures after participants play Cyberball.*

Thank you for taking part in this study. Your participation is invaluable to our research. The purpose of this study is to investigate the effects of interpersonal experiences on attitudes to the self and others, self-esteem, suspicious thoughts, and mood.

The game you played during the study, called “Cyberball”, did not involve real players. The two players were actually computer-generated avatars. There are two conditions within this study. In one condition, participants are treated fairly and are included by the other players. In the second condition, participants are treated fairly and are excluded by the other players. We are interested to see if there are differences between the two conditions and how being excluded affects people’s attitudes to themselves and others, as well as suspicious thoughts, mood and self-esteem. We apologize for any discomfort this may have caused you and will try to address any questions or concerns you may have about the study.

You are welcome to contact me (the researcher) with any questions or queries, even after your participation has ended today. If you feel affected by any of the topics related to this research, I will try to offer you support to alleviate any distress that you may feel and to address any concerns that you may have. If you would like to speak to a professional counsellor or psychotherapist about any issues related to the topic of this study, the contact information for the UCD Student Counselling Services is provided the Participant Debriefing Form.

Finally, we ask that you do not tell anyone else who may participate in this study about the nature of Cyberball or the aim of this study as this may affect our results.

Thank you again for your participation.

## PARTICIPANT DEBRIEFING SHEET



### **Title of Project: Online social tasks and attitudes to the self and others**

Thank you for taking part in this study. Your participation is invaluable to our research. The purpose of this study is to investigate the effects of interpersonal experiences on attitudes to the self and others, self-esteem, suspicious thoughts, and mood.

We would like to assure you again that your data is confidential and that your responses will not be linked to your name or any personal information. Data from the computer tasks and questionnaires will be de-identified and stored anonymously. Your data will be stored securely and destroyed after 5 years. All information collected as part of this study will be used solely for experimental purposes. The data collected from all participants who take part in this study will be pooled together. Once the study is completed we will produce a report that will describe the findings of the study. You will not be identified in any report or publication. The report will not include any personal details of the people who took part; it will only describe the overall results of the study.

If you have any further questions or concerns, you can contact the researcher, even after your participation has ended. If you feel affected by the topic of this study, the researcher will try to offer you support to alleviate any distress that you may feel and to address any concerns that you may have. If you would like to speak to a professional counsellor about seeking further support, please see the information provided below about the UCD Student Counselling Services. If at a later date, you have questions or concerns about this study, do not hesitate to contact the researcher.

**Researcher: Corinna Stewart**  
Email: [c.stewart2@nuigalway.ie](mailto:c.stewart2@nuigalway.ie)  
Phone: 083 130 5644  
School of Psychology  
NUI Galway

**UCD Student Counselling Services**  
Website: <http://www.ucd.ie/studentcounselling>  
Phone: 01 716 3134/ 3143  
University College Dublin Student Centre  
Belfield, Dublin 4