

Title: The effect of social exclusion on state paranoia and explicit and implicit self-esteem in a non-clinical sample

Running head: Social exclusion, paranoia, and self-esteem

Authors & affiliations: Stewart, C. a, Rogers, F. a, Pilch, M. a, Stewart, I. a, Barnes-Holmes, Y. b, & Westermann, S. c

- a. School of Psychology, National University of Ireland, Galway
- b. Department of Experimental-Clinical and Health Psychology, Ghent University, Belgium
- c. Department of Clinical Psychology and Psychotherapy, Institute of Psychology, University of Bern, Switzerland

Corresponding author: Electronic mail should be sent to Corinna Stewart at c.stewart2@nuigalway.ie.

Corresponding author Address: School of Psychology, Arts Millennium Building Extension, National University of Ireland, Galway, University Road, Galway City, Ireland.

Abstract

Background and Objectives: The relationship between self-esteem and paranoia may be influenced by social stress. This study aimed to replicate previous research on the impact of social exclusion on paranoia and self-esteem in a non-clinical sample and to extend this work by examining the effect of exclusion on self-esteem at the ‘implicit’ level.

Methods: Non-clinical participants ($N = 85$) were randomly allocated to the Inclusion or Exclusion condition of a virtual ball-toss game (‘Cyberball’). They completed self-report measures of state paranoia and self-esteem, and two implicit measures of self-esteem – the Implicit Association Task (IAT) and Implicit Relational Assessment Procedure (IRAP) – prior to and after exposure to Cyberball.

Results: Social exclusion increased state paranoia. This effect was moderated by distress associated with trait paranoia. Exclusion was also associated with decreased self-reported self-esteem, as well as reduced implicit self-esteem on the IAT. Changes in self-reported self-esteem were associated with state paranoia at post-Cyberball. The IRAP indicated that reductions in implicit self-esteem may be due to increases in ‘Me-Negative’ and ‘Others-Positive’ biases (rather than reductions in ‘Me-Positive’ bias).

Limitations: The current study involved a non-clinical sample and so findings cannot be generalized to clinical paranoia.

Conclusions: These findings are consistent with previous evidence that paranoia is associated with negative self-evaluations, whereas positive self-evaluations can persist in paranoia. They also provide support for the suggestion that investigations of self-esteem in paranoia should extend beyond global self-esteem and might benefit from a distinction between positive and negative components.

The effect of social exclusion on state paranoia and explicit and implicit self-esteem in a non-clinical sample

1. Introduction

Paranoia involves the belief that harm is occurring or will occur and that a perceived persecutor is deliberately trying to cause harm (Freeman & Garety, 2000). It is considered a complex multi-dimensional phenomenon involving a continuum from normal human experience to persecutory delusions at the extreme end (e.g., Freeman et al., 2005; Peters, Joseph, Day, & Garety, 2004). Self-esteem has been implicated in a number of prominent theories of the development and maintenance of paranoia, and there is some evidence in this regard. For example, individuals diagnosed with persecutory delusions, as well as those with paranoia-proneness, are characterized by low and fluctuating self-esteem (Kesting & Lincoln, 2013; Tiernan, Tracey, & Shannon, 2014). Furthermore, there is evidence that reduced self-esteem precedes increases in state paranoia (e.g., Thewissen, Bentall, Lecomte, van Os, & Myin-Germeys, 2008). To account for this relationship, Bentall and colleagues (1994; 2001) proposed that individuals with paranoia attribute negative events externally and personally in order to preserve self-esteem. In contrast, Freeman and colleagues argued that paranoid beliefs directly reflect emotional concerns and negative self-concepts (e.g., self as vulnerable; Freeman et al., 2002; Freeman & Garety, 2014).

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 ('Cyberball'; Williams, Cheung, & Choi, 2000), in which participants are either included or excluded by other (computer-generated) 'players' (e.g., Westermann, Kesting, & Lincoln, 2012; Kesting, Bredenpohl, Klenke, Westermann, & Lincoln, 2013; Lincoln, Stahnke, & Moritz, 2014). For example, Kesting et al. (2013) found that excluded participants reported an increase in sub-clinical paranoia, which was mediated

by a decrease in self-esteem and moderated by paranoia-proneness. 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 self-esteem by allowing us to investigate the potential causal role of adverse social events on paranoia, self-esteem, and the relationship between them.

1.1. Measuring self-esteem

Several factors should be considered when measuring self-esteem in the experimental induction of social stress. First, while much previous research has focused on *global* self-esteem (see Kesting & Lincoln, 2013), Barrowclough et al. (2003) emphasized the distinction between positive and negative components 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” (p. 93). Indeed, the review by Kesting and Lincoln indicated that both persecutory delusions in clinical samples and paranoid ideation 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. To date, only global self-esteem, rather than positive and negative self-evaluations, has been investigated using Cyberball.

Second, self-report methods rely on an individual’s ability to access their feelings and on their willingness to report them (Nisbett & Wilson, 1977; Ross & Nisbett, 1991), and thus 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 Implicit Association Task

(IAT; Greenwald, McGhee, & Schwartz, 1998) that place participants under time pressure and therein capture ‘automatic’ responding.

1.2. Implicit self-esteem

The widely used ‘self-esteem IAT’ (Greenwald & Farnham, 2000) 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. On one block of trials, participants are required to categorize ‘Me’ with ‘Positive’ and ‘Others’ with ‘Negative’ (hypothesized as ‘consistent’ with a typical learning history). On an alternative block of trials, participants must categorize ‘Me’ with ‘Negative’ and ‘Others’ with ‘Positive’ (‘inconsistent’ with typical learning). Faster responding on consistent (‘Me’ with ‘positive’) than inconsistent (‘Me’ with ‘negative’) blocks indicates positive implicit self-esteem.

Despite its widespread use, the IAT has limitations, especially as a measure of implicit self-esteem. First, it juxtaposes self and others as contrast categories, thus rendering it impossible to explore self-based associations independently (Karpinski, 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 these. Second, 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.

The Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes et al., 2006) is an alternative measure that may offer some advantages over the IAT. The IRAP is based on a modern behavioral account of human language and cognition called Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001), which argues that the core functional units of

language are relational. Thus, the IRAP focuses on specific relations (e.g., opposition, or comparison) between stimuli, rather than associations (see Hughes, Barnes-Holmes, & DeHouwer, 2011; Hughes, Barnes-Holmes, & Vahey, 2012). While the IAT and IRAP both require participants to respond quickly and accurately in ways deemed consistent or inconsistent with their previous learning histories, the format of IRAP trials is different from the IAT. Consider the following ‘self-esteem’ IRAP, for example. On each trial, one of two label stimuli (e.g., ‘Me’, ‘Others’) is presented top-screen with, for example, 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 a self-esteem bias (‘Me-Positive’, ‘Me-Negative’, ‘Others-Positive’, ‘Others-Negative’), unlike the IAT’s single metric. This enhanced specificity may be useful in distinguishing between the positive and negative self-evaluations in global self-esteem.

The IRAP has been increasingly used in experimental-clinical research and has provided some important insights (see Vahey, Nicholson, & Barnes-Holmes, 2015, for a meta-analysis). For example, Remue, DeHouwer, Barnes-Holmes, Vanderhasselt, and De Raedt (2013) found that individuals with self-reported dysphoria had lower *actual* self-esteem and higher *ideal* self-esteem than controls. Such research highlights the utility of the IRAP in parsing out how self-concepts may be related to clinically-relevant phenomena. Furthermore, some studies have demonstrated 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), making it well suited for the current study, which examined the effects of a particular environmental manipulation on self-esteem.

1.3. The current study

The aim of this study was to investigate the influence of an adverse interpersonal event (social exclusion) on state paranoia and self-esteem, and to assess the relationship between these two variables in this context by replicating previous research on this topic (e.g., Kesting et al., 2013). It was hypothesized that exclusion would be associated with increased state paranoia and that this would be moderated by trait paranoia. That is, we anticipated that individuals with a history of responding to adverse experiences in a paranoid manner (i.e., paranoia-prone) would be more likely to do so in the current context. It was also hypothesized that changes in state paranoia following exclusion would be associated with changes in self-esteem. Specifically, we expected the adverse experience of exclusion to influence both self-esteem and paranoia, as well as to show a relationship between the two¹. This study also tested if social exclusion would affect 'implicit' self-esteem. Recently, some studies have investigated implicit self-esteem within the context of paranoia using the IAT. This research typically examines discrepancies between self-reported and implicit self-esteem, and compares individuals with persecutory delusions with healthy and/or depressed controls (e.g., MacKinnon, Newman-Taylor, & Stopa, 2011; McKay, Langdon, & Coltheart, 2007; Moritz, Werner, & von Collani, 2006; Kesting, Mehl, Rief, Lindenmeyer, & Lincoln, 2011). Findings have been somewhat mixed however. The experimental approach adopted in this study may help to clarify some of these inconsistencies. Specifically, we hypothesized that exclusion would be associated with reduced self-esteem on the IAT and with increased

¹ While changes in self-esteem may precede or mediate changes in paranoia following social stress, from our perspective, the relationship between these variables is not invariably a straightforward causal one. From this point of view, events in the environment play an important role in influencing both levels of paranoia and self-esteem, as well as the relationship between the two (see Stewart, Stewart, & Hughes, 2016). Thus, while in some contexts self-esteem may indeed show causal influence with respect to paranoia, in other contexts (e.g., during deliteralization exercises or given higher state or trait acceptance) it need not.

‘Me-Negative’ and ‘Others-Negative’ responding at the IRAP trial-type level. In light of evidence that positive self-evaluations may be maintained in paranoia (Kesting & Lincoln, 2013), we did not expect exclusion to influence responding on ‘Me-Positive’ IRAP relations.

2. Method

2.1. Participants

Ninety-seven volunteers (63 female, 31 male, 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.

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. 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 to which player to throw the ball. 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’.

2.3. Measures

2.3.1. Paranoia

‘Trait’ paranoia was assessed using the 18-item multi-dimensional Paranoia Checklist (Freeman et al., 2005). Items range from commonplace paranoid thoughts (e.g., “Bad things

are being said about me behind my back”) to less commonplace 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 scale showed excellent internal reliability overall ($\alpha=.94$), as well as on the Frequency ($\alpha=.87$), Conviction ($\alpha=.85$) and Distress ($\alpha=.90$) subscales in the current sample.

Changes in *state* paranoia were assessed using a state-adapted version of the Paranoia Checklist (Westermann et al., 2012). Participants were asked to rate how strongly six of the original Paranoia Checklist items applied to them “*at the moment*”. The sum score was used as an index of state paranoia.

2.3.2. Explicit self-esteem

The 10-item Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1965) is the most widely used measure of global self-esteem. Participants rate their level of agreement with each item on a four-point Likert scale. To assess *state* self-esteem before and after Cyberball, participants were asked to rate their agreement with the items “*at the moment*”. The RSES showed excellent internal reliability in the current sample ($\alpha=.88$).

2.3.3. Implicit self-esteem

The implicit measures used (IAT and IRAP) are computer-based tests that assess reaction time biases. The IAT stimuli are shown in Table 1. Seven trial blocks were presented with two critical test blocks (one consistent and one inconsistent with a positive self-esteem bias). Instructions for each block were presented onscreen.

The attribute/target stimuli for the implicit measures were the same for both the IAT and IRAP and were based on the self-esteem IRAP developed by Remue et al. (2013; 2014). On the IRAP however, the label stimuli presented at the top of the screen were “I am” and “Others are”. Two response options, “True” and “False”, were presented left and right bottom-screen, with positions counter-balanced over trials. In the IRAP, 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-Negative, Others are-Positive”) on inconsistent blocks. Following orally presented instructions, 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 $< 2500\text{ms}$ to progress to the test phase, which included exactly six test blocks.

On both the IAT and IRAP, 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 test blocks of the IAT and IRAP were included in subsequent analyses.

2.3.4. Cyberball manipulation check

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, 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 had 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?”

Table 1.

Stimuli presented within the IAT

Me vs. Others target stimuli	
Me	Others
I	Others
Me	Not me
My	Them
Self	They
Positive vs. Negative attribute stimuli	
Positive	Negative
valuable	helpless
happy	guilty
tender	desperate
friendly	sad
hopeful	rejected
competent	failed

2.4. Procedure

Previous research has suggested that presentation order of explicit and implicit measures has minimal effect on the output of either (e.g., Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005); hence, for all participants, we presented implicit first. To allow for the possibility that Cyberball effects might be “washed out” after the first implicit measure post-Cyberball and before the second implicit measure post-Cyberball, order of implicit measures were counter-balanced across participants, as was block order (i.e., consistent-first versus inconsistent-first). After participants completed the IAT and IRAP at baseline, they completed the RSES, Paranoia Checklist, and state-adapted Paranoia Checklist. They were then randomly allocated to either the Cyberball Exclusion or Inclusion condition. Immediately after Cyberball, they completed the IAT and 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. Participants were also prompted to mentally reflect on their experiences of the game between the measures.

2.5. Statistical analysis plan

One-way ANOVAs checked for potential between-groups differences (Exclusion vs. Inclusion) on all variables at baseline and on the Cyberball Questionnaire post-Cyberball. Bonferroni correction for multiple comparisons ($\alpha = .05 / 11 \approx .0045$) was used. Effects of social exclusion on state paranoia and explicit and implicit state self-esteem from baseline to post-Cyberball were examined using mixed 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. Effects with $p \leq .05$ were highlighted as statistically significant, while those with $p \leq .1$ were identified as trending toward significance. Effect sizes are discussed with reference to partial eta squared (η^2) standard guidelines for ANOVA for identifying small (.01), medium (.06), and large (.14) effects.

3. Results

3.1. Data preparation

For purposes of data analysis, accuracy criteria were set at $\geq 75\%$ (latency remained at 2500ms). IRAP data for five participants was lost for technical reasons; another five people failed to maintain IRAP accuracy criterion; one person 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 (28 female, 16 male, 1 intersex; *Mean age* = 24.36, *SD* = 10.51) and 40 in the Exclusion condition (27 female, 12 male, 1 non-binary; *Mean age* = 21.23, *SD* = 5.27). No significant age or gender differences were observed between the Inclusion and Exclusion groups. The data were analyzed using SPSS Statistics 21. 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 unsurprising 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 (>1) to warrant log-transformation.

3.2. Sample characteristics at baseline

The sample reported normative levels of RSES-measured self-esteem at baseline ($M = 18.11$, $SD = 5.15$). There was a moderate range of scores on the Paranoia Checklist (54-169; $M = 94.55$, $SD = 27.67$), as well as on the Frequency (15-58; $M = 27.25$, $SD = 9.87$), Conviction (19-63; $M = 32.58$, $SD = 9.19$), and Distress subscales (18-73; $M = 34.73$, $SD = 12.33$). One-way ANOVAs indicated no significant between-groups differences on any of the measures at baseline (all $ps \geq .046$).

Correlational analyses (Bonferroni $\alpha = .0045$) assessed the relationships among the measures at baseline (Table 2). Significant negative correlations were observed between RSES scores and Paranoia Checklist total, Frequency, Conviction, and Distress subscales, and state paranoia. Significant positive correlations were observed between the Paranoia Checklist total and its subscale scores, as well as state paranoia (all $ps \leq .001$). No significant correlations were observed between the implicit and explicit measures. Positive correlations between the 'Others-Negative' IRAP trial-type and the Frequency ($r = .226$, $p = .037$) and Conviction ($r = .226$, $p = .038$) subscales of the Paranoia Checklist trended towards significance.

Table 2.

Pearson's correlations among measures of paranoia and explicit and implicit self-esteem at baseline (N = 85). Note, Me-Positive, Me-Negative, Others-Positive and Others-Negative represent each of the four IRAP trial-types.

	RSES	IAT	Me-Positive	Me-Negative	Others-Positive	Others-Negative
Paranoia Total	-.590**	-.041	.139	-.135	.043	.175
Paranoia Frequency	-.560**	-.047	-.052	-.118	.086	.226*
Paranoia Conviction	-.481**	-.040	.144	-.133	.007	.226*
Paranoia Distress	-.517**	-.085	.163	-.110	.023	.042
State Paranoia	-.410**	.057	.176	-.133	-.071	.018

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

3.3. Manipulation check

One-way ANOVAs assessed whether exclusion (vs. inclusion) 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 the Bonferroni-corrected significance level ($p = .013$). 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 $ps < .001$).

3.4. Effects of exclusion on state paranoia

A mixed 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 approaching 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 ($M = .96$, $SD = .15$; $p = .035$) for the Exclusion group only, and a significant between-groups difference post-Cyberball ($p = .024$).

3.5. Effects of exclusion on explicit self-esteem

A mixed ANOVA revealed a significant time X condition interaction effect on the RSES, $F(1,83) = 4.750$, $p = .032$, $\eta^2 = .054$. 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 ($M = 17.5$, $SD = 6.41$) (Figure 1).

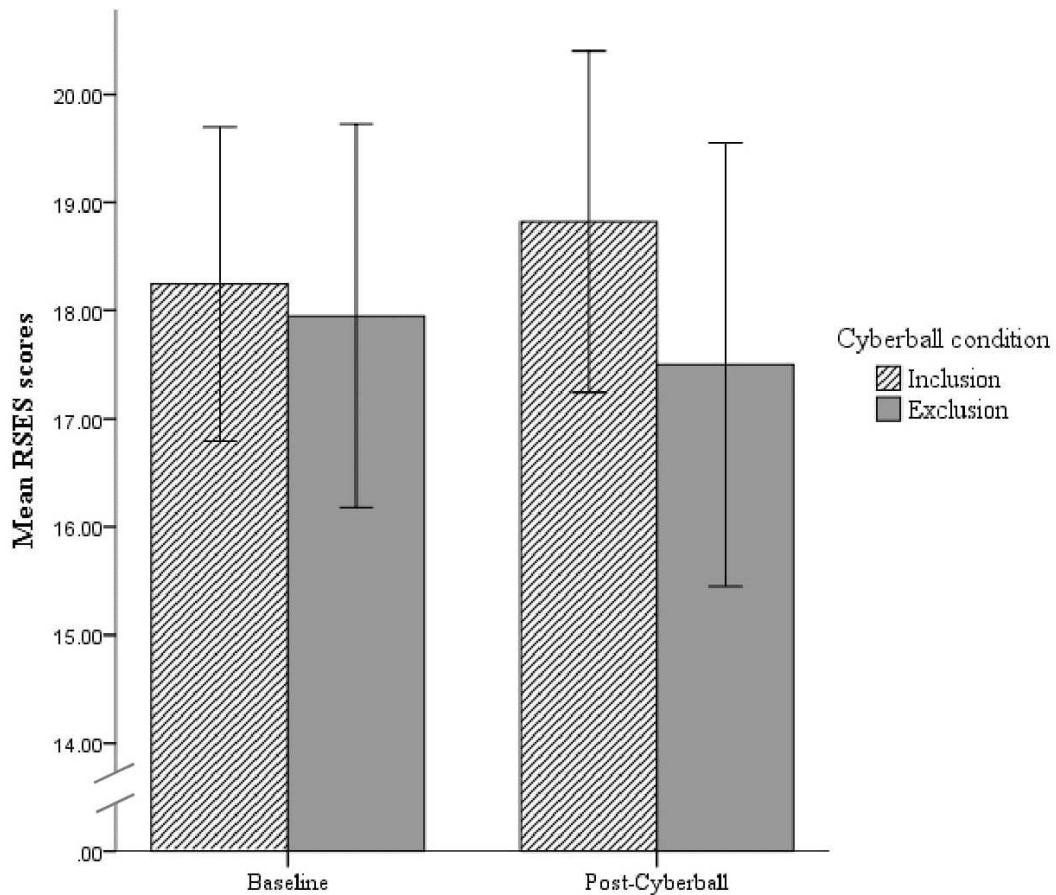


Figure 1.

Changes in mean self-reported state self-esteem from baseline to post-Cyberball for Inclusion and Exclusion groups. Standard error bars represent 95% confidence intervals.

3.6. Effects of exclusion on implicit self-esteem

One-way ANOVAs indicated no effect of order of implicit measures on IAT scores or IRAP trial-type scores at baseline or post-Cyberball.

3.6.1. IAT

IAT scores for the Inclusion and Exclusion groups at baseline and post-Cyberball are presented in Figure 2. One-sample t-tests indicated that both groups showed positive implicit self-esteem at baseline and post-Cyberball ($p < .001$). That is, in both groups, participants responded on average more quickly when asked to categorize 'Me' with 'Positive' and 'Others' with 'Negative' than 'Me' with 'Negative' and 'Others' with 'Positive'. A mixed ANOVA revealed a significant time X condition interaction effect, $F(1,83) = 5.397, p = .023$,

$\eta^2 = .061$, as well as a main effect for time, $F(1,83) = 19.064$, $p < .001$, $\eta^2 = .187$). Post-hoc analysis indicated a significant reduction in IAT score from baseline to post-Cyberball for the Exclusion group only ($p < .001$), indicative of a reduction in implicit self-esteem.

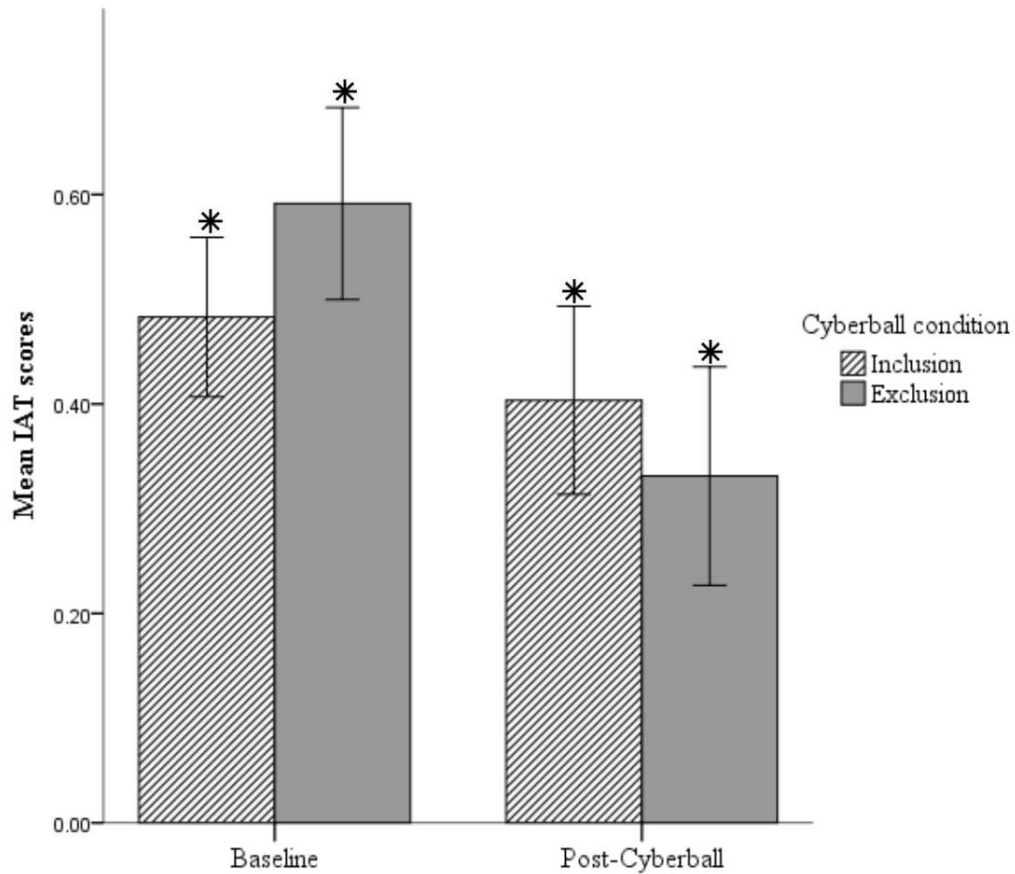


Figure 2.

Changes in implicit state self-esteem as measured by mean IAT scores from baseline to post-Cyberball for the Inclusion and Exclusion groups, with standard error bars (95% confidence intervals). Asterisks denote scores significantly different from zero.

3.6.2. IRAP

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 3. 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 “trial-types”. One-sample t-tests indicated that both the Inclusion and

Exclusion groups demonstrated positive self-esteem at baseline and post-Cyberball on the 'Me-Positive' trial-type (all $ps < .001$). The other three trial-types were not significantly different from zero at baseline. Post-Cyberball, the 'Others-Positive' trial-type was significant for both groups (all $ps < .001$), indicative of an 'Others-Positive' bias. An effect trending towards significance was observed on the 'Me-Negative' trial-type for the Exclusion group post-Cyberball ($p = .081$), indicative of a 'Me-Negative' bias. A mixed ANOVA investigated between-groups differences on the four IRAP trial-types from baseline to post-Cyberball. 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' bias. A between-groups difference on the 'Me-Negative' trial-type post-Cyberball trended towards significance ($p = .066$), suggesting more negative self-bias for the Exclusion group.

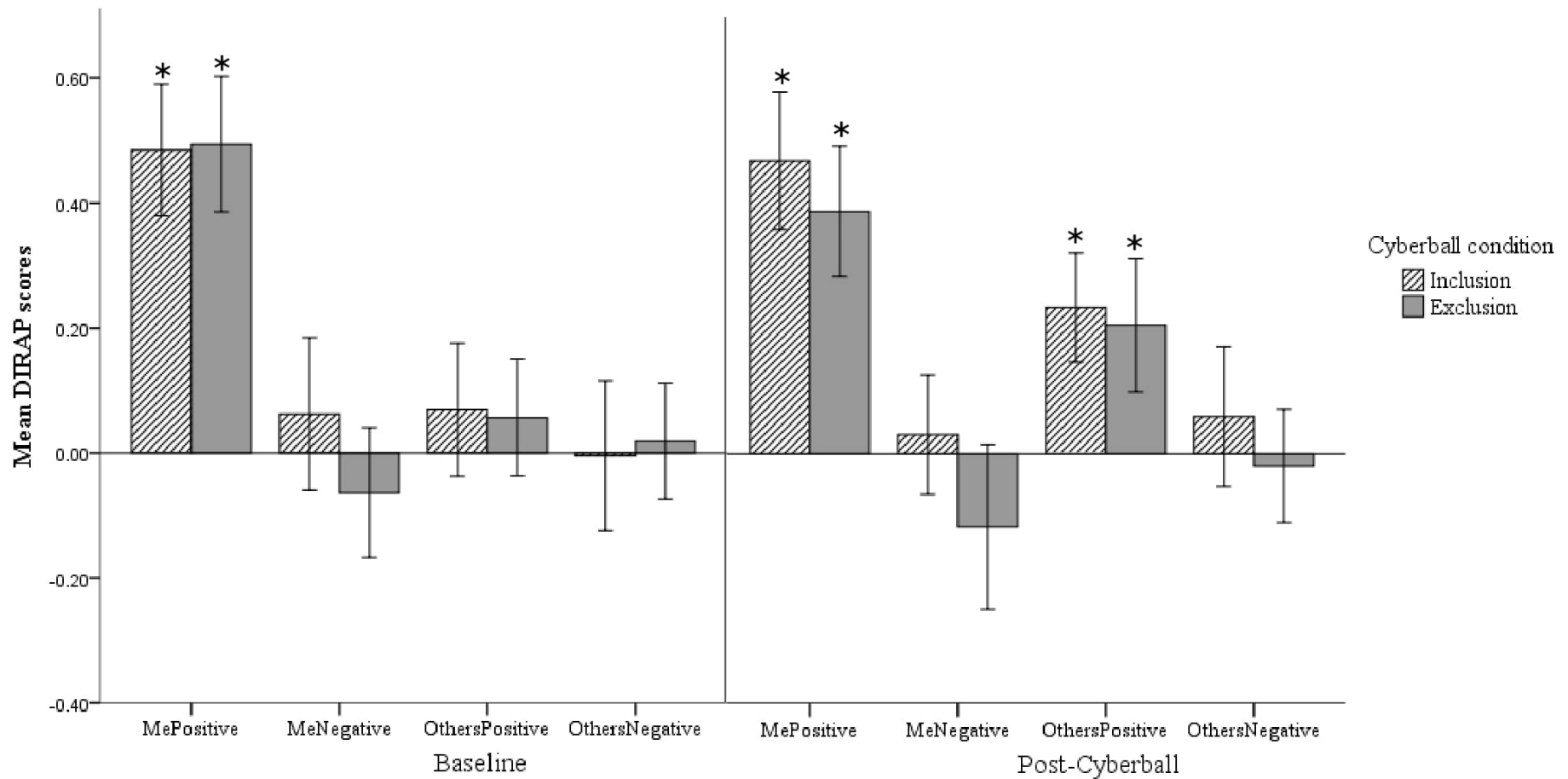


Figure 3.

D-IRAP scores for IRAP trial-types (Me-Positive, Me-Negative, Others-Positive, Others-Negative) at baseline (left-hand side) and post-Cyberball (right-hand side) for the Inclusion and Exclusion groups, with standard error bars (95% confidence intervals). Asterisks represent IRAP trial-types significantly different from zero.

3.7. Associations between state paranoia post-Cyberball, and changes in explicit and implicit self-esteem

Correlation analyses (with Bonferroni correction $\alpha = .0045$) were conducted to examine the relationships between changes in self-reported and implicit self-esteem (as measured by RSES, IAT and IRAP trial-type change scores) and state paranoia post-Cyberball. Negative correlations were observed between state paranoia and RSES change scores ($p = .003$), and trended towards significance between state paranoia and ‘Me-Positive’ trial-type change scores ($p = .021$). A negative correlation between change scores on the ‘Me-Negative’ and ‘Others-Positive’ trial-types trended towards significance ($p = .017$). When examined separately, the negative correlation between ‘Me-Negative’ and ‘Others-Positive’ change scores reached significance for the Exclusion ($r = -.492, n = 40, p = .001$) but not the Inclusion group ($r = -.085, n = 45, p = .577$).

3.8. Moderating effect of trait paranoia dimensions on changes in state paranoia post-Cyberball

To investigate whether the effect of Cyberball condition 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 those in the Exclusion group with a Distress score of ≥ 31 , which was increasingly strong for higher Distress scores. 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. Discussion

This study has shown that social exclusion during Cyberball was associated with increases in state-paranoia and decreases in 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, replicating findings from previous studies (e.g., Kesting et al., 2013). This study also sought to examine whether exclusion might impact implicit self-esteem. On the implicit measures, all participants demonstrated high implicit self-esteem at baseline and post-Cyberball, as indicated by significant IAT scores and effects on the ‘Me-Positive’ trial-type on the IRAP. As hypothesized, the Exclusion group showed decreases in implicit self-esteem from baseline to post-Cyberball on the IAT. However, no changes on the ‘Me-Positive’ trial-type were observed on the IRAP post-Cyberball. Rather, participants who were excluded demonstrated an increased ‘Me-Negative’ bias and differed from included participants on this trial-type at post-Cyberball (though not significantly so). Interestingly, both groups demonstrated a significant ‘Others-Positive’ bias post-Cyberball. 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 they viewed the other ‘players’ negatively. Perhaps they viewed others as stronger or more powerful than previously, a suggestion supported by negative correlation between decreases on ‘Me-Negative’ and increases on ‘Others-Positive’ for excluded participants only.

The results obtained on the IRAP may clarify the IAT findings. That is, the observed decreases in implicit self-esteem on the IAT in the Exclusion group may be attributed to the increased ‘Me-Negative’ and ‘Others-Positive’ responding on the IRAP, rather than a

decrease in positive self-esteem per se. These findings are consistent with previous evidence that paranoia is associated with negative self-evaluations, whereas positive self-evaluations can persist in persecutory delusions and paranoid ideation (Kesting & Lincoln, 2013). These findings may also shed some light on existing theories of persecutory delusions. For instance, the association between ‘Me-Negative’ and ‘Others-Positive’ responding following social exclusion could be considered in terms of the purported ‘poor-me’ paranoia (Trower & Chadwick, 1995), whereby individuals perceive others as ‘bad’ and the self as a victim. However, we did not observe an association between state paranoia and these IRAP trial-types, making it difficult to clarify the relationship between paranoia and this particular pattern of responses to the self and others in this context. More generally, the findings that global self-esteem (as measured by the RSES and IAT) decreased and ‘Me-Negative’ responding increased in response to social exclusion while state paranoia increased could be interpreted in accordance with Bentall et al.’s (1994; 2001) argument that paranoia functions to preserve self-esteem and avoid activation of negative self-beliefs by making externalizing, personalizing attributions for negative events. On the other hand, increased paranoia following social stress may be directly reflective of increased negative self-regard, as proposed by Freeman et al. (2002). Future research examining the consequences of responding to adverse interpersonal events with increased paranoid ideation may help shed light on the function of paranoia in the context of social stress. For example, Lincoln et al. (2014) investigated the impact of a paranoid (versus self-blame or neutral) explanation after being excluded during Cyberball in a non-clinical sample. Results indicated that the paranoid explanation increased self-esteem more than self-blame but that greater self-esteem recovery was observed for the neutral compared to the paranoid explanation.

The finding that the distress associated with trait paranoia at baseline moderated the effect of social exclusion on state paranoia is consistent with previous research by Kesting et

al. (2013). Firstly, this finding supports the necessity for a multi-dimensional approach to paranoia as factors such as the distress associated with paranoid beliefs may be more relevant than the content of the belief alone with respect to the impact paranoia may have on functioning (e.g., Peters et al., 1999b; 2004). Secondly, this finding suggests that the affective component of paranoia may be pertinent to stress-sensitivity. Bell and Freeman (2014) proposed that ‘interpersonal sensitivity’, which refers to “feeling vulnerable in the presence of others due to the expectation of criticism or rejection” (p. 441), may be crucial to paranoia (e.g., Freeman et al., 2008; Masillo et al., 2012). Indeed, 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; Selten, van der Ven, Rutten, & Cantor-Graae, 2013; Valmaggia et al., 2015). Kesting and Lincoln (2013) proposed that negative self-beliefs, perceiving oneself as low in social rank and not accepted by others, and perceiving others negatively is likely shaped by adverse interpersonal experiences, which may be reactivated when the individual is exposed to social stress in future. The key point is that the role of self-esteem in paranoia may be strongly influenced by perceptions of others and this crucial relationship may involve specific negative beliefs about the self (e.g., as vulnerable) and others (e.g., as untrustworthy, see Freeman, 2007; Fowler et al., 2006; Freeman, Garety, Kuipers, Fowler, & Bebbington, 2002).

4.1. Limitations

One limitation is that the “Others” category as presented in the IAT and IRAP was left unspecified, hence conceptualizations of “Others” may have varied across participants at both time points (e.g., friends, strangers, disliked others, etc.). For example, participants may have conceptualized “Others” as the Cyberball ‘players’ post-Cyberball, but not at baseline.

Research by Karpinski (2004) and by Pinter and Greenwald (2005) have already highlighted the importance of clarifying the “Others” contrast category within the IAT.

4.2. Conclusions

These findings support the suggestion that investigations of self-esteem in paranoia should extend beyond global self-esteem to examine positive and negative components (Barrowclough et al., 2003). Measures of global self-esteem and reliance solely on self-report measures may not allow the methodological precision needed to investigate this domain. The non-relative nature of the IRAP may offer advantages over measures like the RSES and IAT in this regard. Also, the subtle changes in negative self-regard captured by the IRAP support the idea that fluctuations in self-esteem (and negative self-evaluations in particular) may be more pertinent to paranoia than global self-esteem as argued by Bentall et al. (2001). Furthermore, as these fluctuations were associated with adverse interpersonal events, future studies should examine how such experiences impact responding to the self and others and how this relates to paranoia in the short-term and across time.

Acknowledgements, declaration of interest, and role of funding organizations:

Preparation of this paper was supported by a NUI Galway Doctoral Scholarship to Corinna Stewart.

References

- 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.
- 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.
- Bell, V., & Freeman, D. (2014). A pilot trial of cognitive behavioural therapy for interpersonal sensitivity in individuals with persecutory delusions. *Journal of Behavior Therapy and Experimental Psychiatry, 45*, 441-446.
DOI: 10.1016/j.jbtep.2014.06.001.
- Bentall, R. P., Corcoran, R., Howard, R., Blackwood, N., & Kinderman, P. (2001). Persecutory delusions: A review and theoretical interpretation. *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., 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.

- Fowler, D., Freeman, D., Smith, B., Kuipers, E., Bebbington, P., Bashforth, H., et al. (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
- 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., & Garety, P. A. (2000). Comments on the content of persecutory delusions: Does the definition need clarification? *British Journal of Clinical Psychology*, *39*, 407-414.
- 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., Fowler, D., & Bebbington, P. E. (2002). A cognitive model of persecutory delusions. *British Journal of Clinical Psychology*, *41*, 331–347.
- Freeman, D., Pugh, K., Antley, A., Slater, M., Bebbington, P., 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
- 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., 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, Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197-216. DOI: 10.1037/0022-3514.85.2.197
- 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.
- 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. <http://dx.doi.org/10.1037/h0100910>
- Hughes, S., Barnes-Holmes, D., & DeHouwer, 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 on to 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
- 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
- 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
- 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
- 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

- 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.
- 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
- 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
- Moritz, S., Werner, R., & von Collani, G. (2006). The inferiority complex in paranoia readdressed: A study with the Implicit Association Test. *Cognitive Neuropsychiatry*, *11*, 402-15. DOI: 10.1080/13546800444000263
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231-259.
- 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
- 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
- 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.
- 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
- Stewart, C., Stewart, I., & Hughes, S. (2016). A contextual behavioral approach to the study of (persecutory) delusions. *Journal of Contextual Behavioral Science*, 5, 235-246. DOI: 10.1016/j.jcbs.2016.09.002
- 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
- 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
- Trower, P., & Chadwick, P. D. (1995). Pathways to defence of the self: A theory of two types of paranoia. *Clinical Psychology*, 2, 263-278.
- 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.

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

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

Williams, K. D. (2009). Ostracism: A temporal need-threat model. *Advances in Experimental Social Psychology*, 41, 275-314. DOI: 10.1016/S0065-2601(08)00406-1

Williams, K. D., Cheung, C., & Choi, W. (2000). Cyberostracism: Effects of being ignored over the Internet. *Journal of Personality and Social Psychology*, 79, 748-762.