

An Empirical Study of Defensive Avoidance in Paranoia

Michael Moutoussis

University of Manchester, and Institute of Neurology, University College London, UK

Wael El-Deredy

University of Manchester, UK

Richard P. Bentall

University of Liverpool, UK

Background: There is controversy as to whether psychological defensive avoidance is associated with paranoia. **Aims:** To elucidate whether “Poor-me” paranoid patients, who believe that the persecution they perceive is undeserved, show more prominent avoidance of negative thoughts about themselves than healthy and clinical controls. **Method:** The act of avoidance of aversive mental contents was assessed in 27 healthy controls and 48 patients with poor-me, bad-me (perceived to be deserved) or no paranoia. Defensive avoidance was assessed via established questionnaires, a novel task based on self-discrepancy theory and research-clinician ratings. **Results:** Participants in all groups showed substantial levels of verbal defensive avoidance. Paranoia was associated with reduced self-reported tolerance of negative mental contents (high Experiential Avoidance, EA). Contrary to our hypotheses, poor-me and bad-me patients showed similar EA. All participant groups showed similar levels of verbal defensive avoidance. **Conclusion:** The findings do not support an association of psychological avoidance with paranoia.

Keywords: Psychological defence, avoidance, bad-me paranoia, poor-me paranoia, self-discrepancies

Introduction

Whether paranoia stems from defensive avoidance of undesirable aspects of the self has been a matter of great controversy. It is a hypothesis of great clinical and scientific consequence, but despite its sophisticated clinical-theoretical history (e.g. Segal, 1994; Chadwick, Birchwood and Trower, 1996) the empirical tests that it has received have been inconclusive. The present study aimed to test this hypothesis using both established and novel methods.

Reprint requests to Michael Moutoussis, Wellcome Trust Centre for Neuroimaging, Institute of Neurology, 12 Queen Square, London WC1N 3BG, UK. E-mail: fzsemmo@g.n.apc.org

© British Association for Behavioural and Cognitive Psychotherapies 2013

Theories of paranoia as a defence against negative self-representations

Paranoia-as-defence is important in both the psychodynamic (Winters and Neale, 1983) and the cognitive-attributional (Bentall, Corcoran, Howard, Blackwood and Kinderman, 2001) theories of paranoia. The most prevalent psychodynamic account of paranoia derives from object-relations theory (ORT; e.g. Ogden, 1983). According to ORT negative aspects of the self, especially those related to hostility towards important figures, are avoided in paranoia. Patients would not necessarily be aware of such negative mental states; they would avoid them unconsciously, usually by denying that they apply to themselves – “it is important not to be hostile with people, and I’m not a hostile person”. The psychodynamic consensus (e.g. Segal, 1994; Blum, 1994) indicates that paranoid defences would “project” perceived malevolence to others instead of accepting one’s own negative feelings, perceiving others as hostile or unfair.

Attributional theory holds that paranoid patients are prone to negative self-representations; however, they try to block these out and defend their self-esteem by blaming others instead of themselves when negative events happen (a self-serving bias). The theory needs to account for the observation that paranoid patients as a whole tend to have low self-esteem (Freeman et al., 1998). Recognizing this difficulty, Bentall et al. (2001) made use of a distinction between poor-me paranoia (in which persecution is believed to be undeserved) and bad-me paranoia (in which persecution is believed to be deserved) previously made by Trower and Chadwick (1995). They argued that defensive processes are only important in poor-me paranoia. Consistent with this, poor-me paranoid patients, but not bad-me patients, tend to have relatively preserved self-esteem (Chadwick, Trower, Juusti-Butler and Maguire, 2005).

The aversive self-attributes relevant to paranoia can be specified more accurately with the tools of self-discrepancy theory. According to Higgins (1987) specific discrepancies between the standards that one holds and the evaluation of the self are clinically important. Self-discrepancy research has furnished tools for the idiographic specification of areas of self-representation that are likely to be charged in this way. Kinderman and Bentall (1996) demonstrated the importance of self-discrepancies in paranoia. They showed that paranoid patients saw themselves more positively than they thought their own parents saw them. This provides *prima facie* support for a projective-type avoidance: “it is not me who is bad, it is my ill-disposed parents who see me that way”.

Some complexities of the experimental evidence on paranoia-as-defence

Several studies have provided evidence for a defensive function of paranoia. Kinderman, Prince, Waller and Peters (2003) observed an improvement in self-discrepancies of paranoid patients with low depression scores after exposure to threat words. Studies of overt vs. covert self-esteem (McKay, Langdon and Coltheart, 2007) and attributional style (Mehl et al., 2010) have found healthy-like overt characteristics but depressed-like covert ones associated with paranoia. Thus non-depressive paranoia may improve the overt self-image in the presence of a negative underlying self-representation and this may be an effect of deservedness.

Other studies, however, have found evidence against the defence theory of paranoia. Cicero and Kerns (2010), for example, found no evidence for low covert self-esteem in nonclinical paranoia using the implicit association test. Vazquez, Diez-Alegria, Hernandez-Lloreda and Moreno (2008) obtained similar results in paranoid and remitted patients using the

self-referent incidental recall task. Fornells-Ambrojo and Garety (2009) found preserved self-esteem but no increased self-serving bias associated with poor-me paranoia.

The importance of avoidance as a central feature of the defensive mechanism

In seeking to resolve these contradictory findings we note that self-esteem and related overall self-concepts are high-level measures affected by many factors, such as perception of social failure or developmental confounders such as low IQ. They may thus be outcomes too far removed from the detailed mechanisms that sustain paranoid thinking. On the other hand, a theme running through all defence hypotheses is that paranoid thinking is sustained partly by increased avoidance of negative aspects of the self. The term “Experiential avoidance” describes the deliberate avoidance of internal representations as if they were noxious stimuli (Bach and Hayes, 2002). We thus hypothesized that those who believe that they don’t deserve to be persecuted (poor-me) would strongly avoid negative attributes of the self that threaten to become prominent in consciousness. In the case of deserved (bad-me) paranoia, in contrast, specific negative thoughts about the self, consciously recognized as important, would be less avoided than in healthy individuals (attribution and psychodynamic theories would explain this same prediction differently). We note that “antipsychotic” drugs used to treat paranoid syndromes severely blunt avoidance responses in animals; we have argued that they may reduce paranoia through suppressing experiential avoidance (Moutoussis, Williams, Dayan and Bentall, 2007). Experimental evidence for the operation of such defensive avoidance in paranoia would be both scientifically and clinically important. As well as supporting the “defence” hypothesis, it would link for the first time the psychopharmacology of avoidance to the psychological processing of aversive thoughts. Clinically, it would mean that more attention has to be paid to the activation and avoidance of specific negative self-attributes in the achievement of remission and the prevention of relapse of paranoid syndromes. We make no specific claim about what type of negative self-attributes might be implicated in paranoia; to the extent that a patient senses that a negative characteristic may be attributed to her or him, s/he might experience anxiety, which will initially drive avoidance. In the longer term in PMP patients however, successful avoidance would be expected to leave little anxiety to be experienced.

Aims

Predictions of the “defensive avoidance” hypotheses tested in the present study

We set out to test specific predictions stemming from the general hypothesis that poor-me paranoia is characterized by powerful defensive avoidance.

1. If they avoid negative thoughts about the self, poor-me participants (PMP) may express relatively more positive overt views of the self, especially if not depressed, than bad-me participants (BMP). Thus we first predicted that: (a) they should show lesser Actual-Ideal self-discrepancies (AID) than BMP; (b) PMP should also see themselves more positively than they perceive others seeing them (Actual-Other discrepancy; AOD), compared to clinical and healthy controls, consistent with the findings of Kinderman and Bentall (1996); (c) PMP should show higher social desirability scores than BMP, as

social desirability assesses the propensity to express overly positive self-representations (Crowne and Marlow, 1960).

2. Expanding on the findings of Udachina et al. (2009), we predicted that PMP patients would give more avoidant scores than BMP on a validated self-report measure of experiential avoidance.
3. If PMP participants were experimentally exposed to salient self-characteristics and asked to talk about them, we predicted that: (a) Given the choice, they would spend less time talking about the more self-discrepant attributes, which threaten to induce negative emotions. This would be consistent with how antipsychotic agents increase avoidance response latency (Moutoussis et al., 2007); (b) When exploring self-discrepant, negative attributes the content of their answers would show higher levels of avoidance than BMP and healthy people.

Method

Exposing participants to negative, avoided thoughts and probing the defences employed to ward them off might be distressing. Both anxiety about looming negative self-characteristics and specific negative feelings associated with the characteristics themselves might be experienced. Special care was taken to detect and ameliorate any such distress. The study was approved by the relevant NHS Trust and the UK National Research Ethics Service.

Participants

The key independent variables we sought to examine were paranoia, deservedness and mood. We expected these not to be tightly specified by diagnosis or clinical severity. We therefore used diagnostic and severity criteria to recruit participants, but we then allocated them into groups on the basis of additional measures. We invited a group of mental-health service users for whom persecutory ideation was an important clinical problem (“threatened group”, $N = 25$) and a group who were predominantly “low in mood” ($N = 23$). We expected the first group to include mostly poor-me paranoid participants and the second mostly depressed ones, with lesser, mostly bad-me paranoia (cf. Fornells-Ambrojo and Garety, 2005). We matched the clinical participant groups with respect to their psychiatric status and severity of problems: all clinical participants received at least “Enhanced Care Program Approach” treatment, either within a community treatment team or as inpatients. Twenty-seven healthy control participants were recruited from the subject panel of the University of Bangor School of Psychological Sciences so as to approximately match the clinical groups for age and sex. They were free of psychopathology and serious physical disorders.

Exclusion criteria for all study groups were: a primary diagnosis of substance abuse or personality disorder, presence of major medical illness, significant learning disability or cognitive impairment, and lack of fluent English. As defensive mechanisms in mania might also account for some of our hypotheses (Jolley et al., 2006), we also excluded all patients with current mania, hypomania or with a history of bipolar disorder. Overall, roughly half the patients who were invited to participate declined and did not consent for their demographic data to be recorded.

We also recorded participants’ age, gender, prescription (or not) of a therapeutic dose of antidepressants, any anxiolytics or hypnotics, total chlorpromazine-equivalent daily

antipsychotic dosage, Clozapine use, years in full-time education, and peak socioeconomic achievement. Selected items of the Brief Psychiatric Rating Scale, viz. conceptual disorganization, grandiosity, suspiciousness and hallucinatory behaviour were rated by the first author.

As we intended to examine the relation between deservedness and defensive avoidance in paranoia, we split the whole sample according to their scores on the Persecution and Deservedness Scale (PADS). The PADS is an instrument suitable for assessing perceived persecution and the extent to which persecution is felt to be deserved in both clinical and non-clinical populations (Melo, Taylor and Bentall, 2006). Informed by previous research (Melo, Corcoran, Shryane and Bentall, 2009), we selected the “paranoid” group to have persecution scores above 1.2 (out of 4) on the persecution subscale of the PADS. Paranoid participants were categorized as “bad me” if they had PADS-Deservedness scores more than 1.2 (out of 4). These thresholds separated a large low-deservedness subgroup from a sparser high-deservedness subgroup in each clinical recruitment group. The final grouping was: Healthy non-threatened ($N = 23$; 18 male), Clinical non-threatened ($N = 9$; 7 male), bad-me paranoid ($N = 14$; 8 male) and poor-me paranoid ($N = 29$; 22 male).

The confirmed ICD-10 diagnoses (World Health Organization, 1992) of participants had as follows. The poor-me group had 11 participants diagnosed with schizophrenia, 3 with other psychotic disorders, 12 with major depressive episodes (only two with psychotic symptoms other than paranoia), one with adjustment disorder and two that did not meet diagnostic criteria for any disorder (healthy). The most common secondary diagnoses were harmful use of cannabis and alcohol. The bad-me group had 7 participants with depressive episodes, 4 with schizophrenia, 1 with schizoaffective disorder, and 2 not reaching any diagnostic criteria (healthy). Only two patients in this group had significant comorbidity (alcohol, emotionally unstable personality traits). The non-threatened clinical group included 7 patients with schizophrenia and 2 with depressive disorder. The healthy-non-threatened group only contained participants meeting no criteria for any mental disorder.

In order to examine the relation between mood and defensive avoidance in paranoia, we recategorized the sample according to paranoia and depression. The paranoia threshold was as above. The “depressed” group had depression scores of at least 8 (out of 21) on the depression subscale of the Hospital Anxiety and Depression Scale (HADS; Zigmond and Snaith, 1983). This threshold confers satisfactory sensitivity and specificity in separating depressive episode cases from non-cases (Olsson, Mykletun and Dahl, 2005). We thus obtained the following groups: Non-threatened, non-depressed ($N = 25$; 22 male), depressed-paranoid ($N = 23$; 16 male) and non-depressed paranoid ($N = 17$; 14 male). We found only 5 non-threatened depressed participants, possibly due to the severity criteria that we adopted. This last group was too small and so we excluded it from further analyses.

The overall grouping pattern was consistent with previous research (Melo et al., 2009). As a whole, participants were 73% male and their mean age was 40.8 years ($SD = 14.2$). None of the subgroups described above differed significantly from these values either in age or gender.

Self-report questionnaires

The following instruments were used: The PADS (Melo et al., 2006) is a 10-item, psychometrically robust measure of perceived persecution (subscale PADS-P) and beliefs about whether persecution is deserved (subscale PADS-D). We used this scale to separate

PMP from BMP groups as above. The Acceptance and Action Questionnaire, second version, is a 10-item measure of Experiential Avoidance. It measures the propensity to avoid or control aversive internal states (Hayes et al., 2004). The Marlowe-Crowne Social Desirability Scale (Crowne and Marlowe, 1960) measures the propensity to give an overly positive self-representation. Finally, the HADS (Zigmond and Snaith, 1983) is a validated self-report measure of both anxiety (HADS-A) and depression (HADS-D).

Interactive assessment of self-representation and avoidance

A computerized adaptation of the Self-Lines Measure (Francis, Boldero and Sambell, 2006) was central to our hypothesis-testing. Each participant described in their own words 15 idiographically important domains of self-representation. They were asked to describe five characteristics they would “ideally like to” possess (e.g. good-looking), five that they felt they “ought to” possess (e.g. honest) and five characteristics, unrelated if possible to the others, that they “feared they might show and would like to avoid” (e.g. impatient). They were then asked to describe the polar opposites of each of these 15 characteristics, thus obtaining 15 domains each with a more desirable and a more undesirable pole. We expected that contemplating having the undesirable characteristic would be aversive and we later used it to elicit aversive thoughts. For each domain the participant rated “how they really were” and “how others saw them” via visual-analogue scales (VAS) stretching from the desirable (top) to the undesirable (bottom) characteristic displayed on a computer screen. Diverging slightly from the approach of Kinderman and Bentall (1996), participants were first asked to guess as well as they could how people who knew them “reasonably well” might rate them. The views their parents might have about them were explored at a later stage (see below and Appendix). Actual-Ideal and Other-Ideal discrepancies, as well as their difference, the Actual-Other discrepancy, were thus calculated on line (AID, OID and AOD). In this paper discrepancy scores are reported as percentages of the VAS range, a “large” AID or OID referring to a negative number of large magnitude. These scores were utilized in the next part of the study.

We then administered an Engagement with Negative Attributes task based on the idiographically important attributes. Participants were asked to imagine and explain how several self-characteristics might apply to themselves. Each participant was exposed to 30 questions derived from attributes of varying self-discrepancy that they themselves had described as important. Five questions out of 30 centred on a desirable characteristic and the rest on undesirable ones. Each trial was preceded by a simple shape that hinted at the Other-Actual discrepancy rated for the pair to which the upcoming attribute belonged. That is, a warning stimulus would predict the aversiveness of a second stimulus (Figure 1). Participants were told that they could try to learn which stimuli hinted at less pleasant questions so as to skip to the next question if they wished. They were also told they could skip (escape) any question they disliked once it appeared on the screen. In the event, “skipped” questions were extremely rare in all groups (0–2 out of c. 700 per group).

For the purposes of analysis we designed a measure of how much participants exposed themselves to aversive self-characteristics. First, we examined which discrepancies correlated with negative mood and found that AID was more predictive of HADS and PADS scores than AOD (see Results section). The aversive item questions were thus divided into the 10 highest-AID ones (hiAID), the middle 10, and the lowest-discrepancy 5 items. The lowest-discrepancy questions were grouped with the positive-attribute questions (loAID). For each participant the

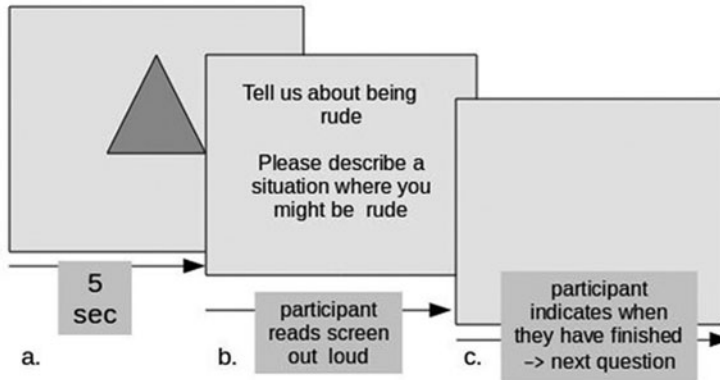


Figure 1. Engagement with negative attributes task. a) Participants first saw one of three simple geometric shapes. They were told that this hinted at how they rated themselves on the characteristic they were going to be asked about; and that if they could guess “What the question might be like” and they’d prefer not to answer it, they could say “skip”. b) Once the participants had read the question containing the (usually negative) attribute and indicated they were ready to answer, the experimenter moved on to a blank screen, timed and sound-recorded the answer. c) The participant indicated when they wanted to move on to the next question, being free to avoid answering the present one completely if they chose to do so.

median time spent talking about loAID items, t_{+med} and the analogous quantity for hiAID items, t_{-med} were computed. The fractional change $\Delta F(t_{med}) = (t_{+med} - t_{-med}) / t_{+med}$ was then derived. Data were log-transformed to approximate normal distributions.

Observer-rated engagement scale (ORES)

Participants’ answers were recorded and later transcribed and rated on a 4-point scale with respect to their content. This scale measured engagement in, vs. avoidance of, thinking about “difficult” aspects of the self. Zero coded for complete avoidance of exploring a potentially important attribute, while 3 coded for a well-engaged response, such that the rater could recognize confidently what the attribute meant for the participant, especially in terms of self-esteem. In terms of inter-rater reliability, 103 randomly selected responses were blindly rated by both MM and RPB. Spearman’s correlation between the two raters’ scores was 0.58, $p < .00001$. For the analysis below, responses were subdivided into low-engagement (ORES 0 or 1) vs. high-engagement (ORES 2 or 3). The inter-rater agreement for this was 82%.

Analysis strategy

We used two-way ANOVA (group \times sex) to examine group differences. We included years-of-education (a proxy for intellectual ability; Matarazzo and Herman, 1984) and age as covariates. All reported ANOVAs conform to this structure. As the study was based on specific hypotheses, especially about the poor-me group, simple contrasts were used as post-hoc tests. We also used linear regression analysis to examine how the variables most closely

Table 1. Means (and standard deviations) for all the measures pertaining to the four groups of participants. The fractional change in response time pertains to time spend answering for low AI discrepancy minus high AI discrepancy stimuli

Paranoia	Healthy <i>M (SD)</i>	Clinical control <i>M (SD)</i>	Bad-me <i>M (SD)</i>	Poor-me <i>M (SD)</i>
Paranoia, PADS-P	0.48 (0.40)	0.88 (0.20)	2.52 (0.84)	2.52 (0.64)
Deservedness, PADS-D	0.44 (0.44)	0.24 (0.68)	2.28 (0.80)	0.60 (0.38)
HADS-Depression	2.52 (2.02)	7.98 (4.41)	9.87 (5.46)	8.40 (5.04)
HADS-Anxiety	4.41 (2.94)	7.98 (2.94)	11.55 (4.20)	11.55 (5.04)
HADS total	7.14 (3.57)	15.96 (6.72)	21.0 (9.24)	20.16 (9.66)
Self-Discrepancies (%)				
Actual-Ideal Discr.	− 24.9 (13.8)	− 27.4 (14.6)	− 42.6 (23.0)	− 42.0 (24.6)
Other-Ideal Discr.	− 26.0 (14.2)	− 25.0 (14.5)	− 46.4 (21.0)	− 38.4 (23.0)
Other-Actual Discr.	1.08 (8.64)	− 2.30 (15.1)	3.79 (28.5)	− 0.62 (16.3)
“Avoidance & Action” and social desirability				
AAQ-2	48.6 (8.4)	34.8 (12.6)	24.0 (11.4)	23.4 (12.6)
Marlowe-Crowne	18.8 (5.0)	22.1 (3.0)	18.1 (5.0)	16.17 (5.6)
Fractional change in response time:				
$\Delta F (t_{med})$	− 0.049 (0.37)	− 0.12 (0.32)	− 0.23 (0.38)	− 0.26 (0.47)
Experimenter-rated fraction of high engagement responses:				
High Eng. fraction	0.76 (0.17)	0.70 (0.22)	0.73 (0.23)	0.71 (0.19)

related to the hypothesized psychological mechanisms related to the symptoms themselves. We performed a multi-stage regression analysis. In the first stage variables widely believed to contribute to symptoms, but weakly related to our central hypotheses, were entered. In a second stage, variables related to the key hypotheses were entered. Alternative explanatory variables were entered last. We employed standard diagnostic statistics to ensure the quality of the regression analysis.

Results

1. Do participants with poor-me paranoia express relatively more positive overt views about the self?

Descriptive statistics about paranoia, deservedness and the key putative explanatory variables are shown in Table 1. The first prediction tested was that PMP would be associated with a relatively preserved overt self-representation.

- Poor-me and bad-me participants had similar mean AIDs, both worse than the healthy-control group. In ANOVA with AID as the dependent variable the main effect of group was significant ($F(3,61) = 3.73, p < .05$). Simple contrast analysis with the healthy-control group as reference showed significant differences with both the BMP ($p < .01$) and the PMP ($p < .005$) groups.
- Poor-me participants did not report significantly greater other-actual discrepancies than controls (ANOVA: $F(3, 61) = 1.68, p = .18$)

- c) There was no evidence according to the social desirability measure for a more positive view of the self in the PMP group, compared to the BMP group (or indeed the healthy one: Tables 1 and 2).

Persecution and depression scores were predicted by self-discrepancies

With OID as the dependent variable, ANOVA did show significant effects ($F(3,61) = 3.26$, $p < .05$). Simple contrast analysis with the healthy-control group as reference revealed a significant difference with the bad-me group ($p < .005$) but only a trend-level difference with the poor-me group ($p = .066$). Interaction with gender was non-significant in all three ANOVAs above (AID: $F(3,61) = 0.65$, $p = .58$; OID: $F(3,61) = 0.65$, $p = .58$; AID: $F(3,61) = 0.87$, $p = .46$).

Persecution scores were significantly correlated with affective variables; e.g. for PADS-P vs. HADS-D, $r = 0.48$, $p < .001$ (corrected for age, gender, socioeconomic achievement and years-in-education). Self-discrepancy measures were highly correlated with depression and anxiety. The partial correlation between HADS-D and AID, for example, was 0.70 ($p < .001$; corrected as above).

Controlling for depression rendered the relationship between PADS-P and AID nonsignificant (uncorrected for HADS-D: $r = -0.42$, $p < .001$; corrected: $r = -0.14$, $p = .23$). On the other hand, controlling for HADS-D reduced the correlation between OID and PADS-P from -0.46 to -0.28 but statistical significance was maintained ($p < .05$).

2. PMP and BMP were both associated with high experiential avoidance

The prediction tested here was that PMP participants would score lower than healthy and clinical controls on the Acceptance and Action Questionnaire (AAQ-2). ANOVA with “acceptance and action” (AAQ-2) as the dependent variable was highly significant for the effect of group ($F(3,61) = 14.76$, $p < .001$). The healthy group differed significantly from all the other groups, but the two paranoid groups both had mean AAQ-2 score = 24 (Table 1). Paranoia was associated with lower AAQ scores, but there was no evidence that AAQ-2 scores differentiated between BMP and PMP.

A further analysis was carried out where the total sample was grouped according to paranoia \times depression, to directly compare paranoia and depression with respect to experiential avoidance and to attempt to replicate the findings of Kinderman et al. (2003). The two-way ANOVA, group \times gender with age and years-in-education as covariates, was highly significant ($F(2,58) = 40.9$, $p < .001$). Depressed paranoid participants scored significantly worse (more avoidantly) on the AAQ than non-depressed paranoid ones ($p < .001$), who in turn scored significantly worse than the non depressed-non paranoid group ($p < .005$). AID and OID discriminated between healthy and depressed paranoid, but not between healthy and non-depressed paranoid groups (Table 2).

3. Behavioural avoidance of negative self-characteristics did not differ in PMP vs. BMP

Our third prediction was that PMP would show greater behavioural avoidance in response to high-discrepancy probes in the “engagement with negative attributes task”, both with

Table 2. Here the sample has been divided with respect to paranoia & depression. Key descriptive statistics are shown. Simple contrast analysis with the non-depressed paranoid group as reference resulted in the significant results shown in bold

	Non-paranoid non-depressed N = 26 M (SD)	Paranoid depressed N = 23 M (SD)	Paranoid non-depressed N = 17 M (SD)
AAQ	47.4 (9.6) p < .005	16.8 (10.8) p < .001	33.0 (9.6) (reference)
AI Discr.	-23.4 (13.4) n.s.	-54.22 (19.8) p < .005	-25.9 (20.0) (reference)
OI Discr.	-25.3 (13.6) n.s.	-50.1(22.8) p < .01	-28.7 (15.1) (reference)
AO Discr.	1.84 (8.62) n.s.	-4.12 (23.2) n.s.	2.80 (13.1) (reference)
MC (Social Desirability)	19.5 (4.6) n.s.	17.2 (5.3) n.s.	16.8 (4.0) (reference)

respect to shortening their self-exposure to aversive attributes and with respect to avoidant, less engaged verbal content compared to BMP and healthy controls.

- ANOVA showed no significant differences between the groups in response shortening, $\Delta F(t_{med})$, for the more aversive questions ($F(3,61) = 0.69, p = 0.56$).
- Using the observer-rated engagement scale (ORES) all 2251 voice-recorded responses were rated by the first author with respect to avoidance or engagement with the self-attribute stimuli. ANOVA with the fraction of high-engagement responses in aversive trials per subject as the dependent variable demonstrated no significant differences between the groups ($F(3,60) = 0.52, p = .67$) although substantial fractions of avoidant responses were seen (group means 24–30%; Table 1).

Regression analyses and the effect of possible confounders

Regression analyses were carried out to investigate whether our results might be explained by other processes of well-recognized relevance to clinical paranoia. As neurocognitive factors related to intellectual ability correlate with levels of paranoia (Bentall et al., 2009), we always entered “years in full-time education”, a proxy for intellectual ability, as a first step in all regression analyses.

We then examined the role of four measures that might reflect defensive avoidance most directly: the Marlowe-Crowne score, AAQ-2 scores, the “fraction of high-engagement responses during aversive trials” and the timing avoidance measure $\Delta F(t_{med})$. As a final step, we entered variables that might confound or alternatively explain the results. We measured many putative explanatory variables for each participant but our total sample of $N = 75$ would not support a simultaneous analysis of all of them. We reasoned that variables that did not contribute to predicting PADS-P scores would be unlikely to explain away any effects of the variables of key interest. We therefore first performed a diagnostic linear regression, entering

Table 3. Regression analysis for auxiliary variables

Linear regression with dependent variable: PADS-Paranoia					
Model variable	B	Std. Err. B	β	Sig. level	VIF
(Constant)	0.29	0.20		0.16	
Concept. Dis.	-0.014	0.04	-0.035	0.75	1.84
Age	0.00	0.002	-0.035	0.69	1.23
Hallucinations	0.015	0.02	0.08	0.46	1.91
CPZ eq. dose	0.00	0.00	0.12	0.24	1.62
HADS-Depression	0.135	0.17	0.11	0.43	3.31
HADS-Anxiety	0.54	0.16	0.48	0.002	3.35
Grandiosity	0.08	0.03	0.255	0.008	1.40
Educ. years	-0.02	0.01	-0.22	0.049	1.88

$R^2 = 0.6$; Adjusted $R^2 = 0.61$

Table 4. Regression analysis: PADS-Persecution. The last variable-entry-stage where a significant change in F was achieved is shown. Note AAQ2-HADS collinearity

Linear Regression with dep. var.: PADS-Paranoia					
Model variable	B	Std. Err. B	β	Sig. level	VIF
(Constant)	0.87	0.198		0.000	
Educ. years	-0.018	0.007	-0.216	< 0.01	1.23
MC score	-0.21	0.126	-0.122	0.097	1.07
AAQ-2	-0.59	0.158	-0.551	< 0.001	4.45
transf. $\Delta F (t_{med})$	0.047	0.055	0.06	0.402	1.05
Hi Eng. Frac.	0.14	0.109	0.092	0.22	1.11
HADS score	0.12	0.173	0.096	0.50	4.07
Grandiosity	0.073	0.025	0.22	< 0.01	1.23

$R^2 = 0.70$; Adjusted $R^2 = 0.67$

only variables that were *not* related to the study hypotheses but might confound results, including age, BPRS measures and antipsychotic dose (Table 3). High Variance Inflation Factors (VIF) for HADS-A and HADS-D suggested that these were substantially collinear and hence their simultaneous analysis might be unreliable. HADS-total and BPRS-grandiosity finally emerged as the important variables to enter in the last step.

It was found that AAQ-2 scores, “years-in-education” and “grandiosity” were the significant predictors of PADS-P scores (Table 4). The adjusted proportion of PADS-P variance explained by the best model was 0.67 (Table 4). The substantial collinearity between AAQ-2 and HADS scores impaired resolving the separate contributions of EA and depression/anxiety to paranoia.

Discussion

We hypothesized that patients who feel subject to undeserved persecution would show prominent defensive avoidance of negative aspects of the self. We note in passing that

the ability to realistically consider one's weaknesses, or to appraise negatively of one's unsuccessful actions, is important for learning, as long as it does not lead to over-generalized negative beliefs. In this study we used multiple measures of the tendency to avoid negative mental experiences. In addition to well-established questionnaires, these included a novel task containing questions of greater clinical validity, similar to clinical exploratory statements. Proximal measures of defensive avoidance (response times, content of answers) and of self-representation (self-discrepancies) were used as more distal outcomes, especially self-esteem, may be confounded by other influences.

We predicted that: 1) Poor-me participants would show relatively preserved social desirability and increased actual-other self-discrepancies; 2) Poor-me participants would show greater subjectively reported experiential avoidance; and 3) When invited to talk about negative self-attributes, PM patients would give shorter, more avoidant answers than BM or healthy participants.

Persecution, self-representation and affect

Our prediction that the poor-me participants would show relatively preserved Actual-Ideal discrepancies was not supported. There was also no evidence that PM patients paid the "price" of wider Other-Actual discrepancies to reduce Actual-Ideal ones. If PM participants had relatively preserved explicit views of themselves they might report being in line with social standards. This was not the case: there was, if anything, a trend in the direction opposite to that predicted (Table 4).

Levels of affective symptoms, both anxiety/depression and grandiosity, were associated with paranoia (cf. Bentall et al., 2009). However, affect did not account for the correlation between paranoia and Other-Ideal discrepancies. OID may therefore be more closely related to paranoia than depression, as affect did account for the association of paranoia with AID (cf. Kinderman and Bentall, 1996). Compared with healthy controls, differences in OID were large and significant for BMP but only trend-level for PMP. This pattern of results supports a "sociometer" (Leary, Tambor, Terdal and Downs, 1995) rather than a "projective" function for self-representation. The "sociometer" explanation is also consistent with the small magnitude of OADs: participants thought they actually were close to the evaluation that others, "who knew them reasonably well", might make.

When paranoid participants were split according to depression scores, the non-depressed group had similar self-discrepancies as the healthy participants (Table 2). This replicates the corresponding findings of Kinderman and Bentall (1996). However, Kinderman and Bentall (1996) found high self-actual-parent-actual (or parent-ideal) discrepancies in paranoid participants, whereas OAD did not discriminate between any groups of interest. It may be that the self-parent discrepancies of paranoid patients do not generalize widely to their representations of self and others. The aetiological importance of the self-parent discrepancies thus awaits further evaluation.

The role of experiential avoidance

Persecuted participants showed a low tolerance of their negative mental contents: low "acceptance and action" (AAQ-2) scores were the best predictor of paranoia, consistent with the findings of Udachina et al. (2009). AAQ-2 scores were not modulated by deservedness,

against our expectation that PMP participants would avoid unpleasant private experiences more than BMP ones. On the other hand, EA increased with depression over and above the effect of paranoia (Table 2). Closer inspection of the AAQ-2 suggests why it may not distinguish between PMP and BMP. Several of its items use the admission that private experiences are feared or disabling as a measure of avoidance. BM participants indeed feel disabled by negative thoughts but may engage in non-accepting preoccupation, not active avoidance, as a form of EA.

We found no evidence that high EA is accompanied by a more favourable self-image in PMP. The high-EA paranoid groups displayed less favourable social desirability and self-discrepancy scores overall. Thus, having a low tolerance of negative mental contents may not be an effective strategy for avoiding them.

Clinical experience suggested that participants might avoid negative thoughts about the self with strategies such as talking vaguely or denying that specific negative self-attributes might possibly apply to them (see Appendix on ORES). We predicted that PM participants would use such avoidant statements more often than the control groups. In the event, such avoidant statements were commonly seen but their occurrence did not significantly differ between groups on the basis of paranoia and deservedness.

Integration with other findings on self-representation in paranoia

How can the present results be reconciled with the findings of more favourable overt self-representations and attributions in undeserved/non-depressive paranoia (McKay et al., 2007; Mehl et al., 2010)? A reconciling explanation may be that non-depressed and/or PM patients may respond to threat with more “fight-like” thoughts associated with images of empowerment (Fornells-Ambrojo and Garety, 2009; Campbell and Morrison, 2007). Depressed and/or bad-me patients may respond with more “learned helplessness-like” thoughts associated with uselessness. These may have opposite effects on explicit self-esteem, irrespective of a psychological-defensive motivation. The functional motivation may be social-psychological: Kramer (1994) provided evidence that in healthy people persecutory ideas are a “sinister attribution bias” that reduces false-negatives in the detection of social adversity.

Study limitations

Engagement with Negative Attitudes task has not been independently validated, and neither has the Observer-rated Engagement Scale. Defensive avoidance is, however, a clinically very well established concept and ratings by trained observer-clinicians were as close to a reference measure as was available to us. Similarly, it would be desirable to utilise better measures of intellectual function. We consider it unlikely, however, that cognitive decline associated with poor-me paranoia rendered participants less avoidant. Thus even if some of our measures were flawed, the overall pattern of results is quite consistent and thus difficult to dismiss.

In this sample AAQ-2 scores were collinear with HADS scores, thus impairing the assessment of how EA and depression/anxiety simultaneously relate to paranoia (Table 4). In the light of the discussion of the role of experiential avoidance above, the AAQ may conflate the measurement of (i) the extent of aversive personal experiences and (ii) the strategies used to deal with them.

The number of clinical control participants with normal levels of perceived threat was small, possibly because in London few such patients are treated beyond the care utilization threshold that we used. This limits drawing conclusions about less severe disorders but does not affect our findings about poor-me paranoia.

Defensive avoidance may not relate to the multiple important attributes of the self that we elicited. It may instead be specific to attributes related to the central paranoid beliefs. Ideally studies of such self-attributes should also draw on a detailed psychotherapeutic formulation/conceptualization.

Too little relevant anxiety may have been present in our experimental context of exposure to self-discrepant attributes, compared to the actual life context of the persecutory beliefs. A procedure, however, that would raise relevant anxieties would face serious ethical problems and might also increase psychological reactance (Kingdon and Turkington, 2005), thus obscuring the picture.

Clinical implications

Helping patients understand their difficulties is a crucial part of clinical practice. A clinician might suggest to a patient that defensive avoidance could be an important cause of their persecutory beliefs: our findings would not support this. Our results encourage attention to patients' threat-related thoughts that may lead to counter-productive efforts to control private experiences. Clinicians may helpfully be aware that feeling threatened is most often associated with an overtly negative self-representation. Sometimes, however, feeling threatened may have positive effects. This conclusion may encourage therapists to explore with patients both "positive" (e.g. angry empowerment) and "negative" (e.g. defensive withdrawal) reactions to threat, as either reaction might perpetuate distress (cf. Pretzer, 2007).

Research implications

First of all research should test the replicability of our results while addressing the limitations identified above. A measure of experiential avoidance enquiring separately about the extent of aversive personal experiences vs. the strategies used to deal with them would be most illuminating.

The procedures we introduced for assessing engagement vs. avoidance of aversive self-thoughts did detect much defensive avoidance and therefore merit further research and development. "Active control groups" that show high levels of defensive avoidance would be useful. Cicero and Kerns (2010), for example, utilised narcissism to delineate a high-defensiveness control group. Better gold standard probes for assessing such defensiveness are also needed.

It may be more fruitful to study paranoia within the native context of idiographically important, threat-related thoughts and feelings, rather than the once-removed context of most studies, including ours. A key control group would be people who *are* concerned about interpersonal threat but *are not* mentally ill, e.g. those undergoing threatening legal proceedings.

Why should so many people be prone to persecutory ideas if not because of psychological-defensiveness? Expanding on the findings of Kramer (1994), research should examine how self-representations interact with other-representations to determine social interaction. Is there

an optimal way of gauging social threat, and what is the normative way of evaluating oneself in a potentially threatening context? A fascinating tool that may bring these strands together is that of game-theoretical tasks, recently introduced in psychopathology research (King-Casas et al., 2008). Such tasks provide in vivo measures constructs such as mistrust, likely to be central to paranoia.

Conclusion

It is difficult to think about aspects of the self where one falls substantially short of one's ideals. When such issues are brought to mind people often employ avoidant thinking. However, the overall pattern of evidence from this study stands against the hypothesis that greater avoidance of negative thoughts about the self motivates poor-me paranoia.

Our findings are likely to help improve the scientific assessment of experiential avoidance and of thinking about difficult aspects of the self. Elucidating the optimal way to think about oneself when feeling under threat may help advance the debate on the role of self-esteem in paranoia.

Acknowledgements

The authors would like to thank Diana Menzies for her help in carrying out this project, Yu Li for coding the computerized "Engagement with Negative Attributes Task", Peter Dayan for helpful discussions, and Charlotte Mackintosh for help with transcribing participant data. The authors have no financial conflict of interest to declare.

References

- Bach, P. and Hayes, S. C.** (2002). The use of acceptance and commitment therapy to prevent the rehospitalization of psychotic patients: a randomized controlled trial. *Journal of Consulting and Clinical Psychology*, 70, 1129–1139.
- Bentall, R. P., Corcoran, R., Howard, R., Blackwood, N. and Kinderman, P.** (2001). Persecutory delusions: a review and theoretical integration. *Clinical Psychology Review*, 21, 1143–1192.
- Bentall, R. P., Rowse, G., Shryane, N., Kinderman, P., Howard, R., Blackwood, N., et al.** (2009). The cognitive and affective structure of paranoid delusions: a trans-diagnostic investigation of patients with schizophrenia spectrum disorders and depression. *Archives of General Psychiatry*, 66, 236–247.
- Blum, H.** (1994). Paranoia: new psychoanalytic perspectives. In J. Oldham and S. Bone (Eds.), *Paranoid Betrayal and Jealousy: the loss and restitution of object constancy*. Maddison, CT: International Universities Press.
- Campbell, M. and Morrison, A.** (2007). The subjective experience of paranoia: comparing the experiences of patients with psychosis and individuals with no psychiatric history. *Clinical Psychology and Psychotherapy*, 14, 63–77.
- Chadwick, P., Trower, P., Juusti-Butler, T.-M. and Maguire, N.** (2005). Phenomenological evidence for two types of paranoia. *Psychopathology*, 38, 327–333.
- Chadwick, P., Birchwood, M. and Trower, P.** (1996). *Cognitive Therapy for Delusions, Voices and Paranoia*. Chichester: John Wiley and Sons.
- Cicero, D. and Kerns, J.** (2010). Is paranoia a defence against or an expression of low self-esteem? *European Journal of Personality*, 25, 326–335. doi: 10.1002/per.794

- Crowne, D. P. and Marlowe, D.** (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349–354.
- Fornells-Ambrojo, M. and Garety, P.** (2005). Bad me paranoia in early psychosis: a relatively rare phenomenon. *British Journal of Clinical Psychology*, 44, 521–528.
- Fornells-Ambrojo, M. and Garety, P. A.** (2009). Understanding attributional biases, emotions and self-esteem in “poor me” paranoia: findings from an early psychosis sample. *British Journal of Clinical Psychology*, 48, 141–162.
- Francis, J., Boldero, J. and Sambell, N.** (2006). Self-lines: a new, psychometrically sound “user-friendly” idiographic technique for assessing self-discrepancies. *Cognitive Therapy Research*, 30, 69–84.
- Freeman, D., Garety, P., Fowler, D., Kuipers, E., Dunn, G., Bebbington, P., et al.** (1998). The London-East Anglia randomized controlled trial of cognitive-behaviour therapy for psychosis. IV: self-esteem and persecutory delusions. *British Journal of Clinical Psychology*, 37, 415–430.
- Hayes, S., Strosahl, K., Wilson, K., Bissett, R., Pistorello, J., Toarmino, D., et al.** (2004). Measuring experiential avoidance: a preliminary test of a working model. *Psychological Record*, 54, 553–578.
- Higgins, E. T.** (1987). Self-discrepancy: a theory relating self and affect. *Psychological Review*, 94, 319–340.
- Jolley, S., Garety, P., Bebbington, P., Dunn, G., Freeman, D., Kuipers, E., et al.** (2006). Attributional style in psychosis: the role of affect and belief type. *Behaviour Research and Therapy*, 44, 1597–1607.
- Kinderman, P. and Bentall, R. P.** (1996). Self-discrepancies and persecutory delusions: evidence for a model of paranoid ideation. *Journal of Abnormal Psychology*, 105, 106–113.
- Kinderman, P., Prince, S., Waller, G. and Peters, E.** (2003). Self-discrepancies, attentional bias and persecutory delusions. *British Journal Clinical Psychology*, 42, 1–12.
- King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P. and Montague, P. R.** (2008). The rupture and repair of cooperation in borderline personality disorder. *Science*, 321, 806–810.
- Kingdon, D. and Turkington, D.** (2005). Cognitive therapy of schizophrenia. In D. Kingdon and D. Turkington, *Case Formulation and Intervening with Delusions* (pp. 96–118). New York: Guilford Press.
- Kramer, R.** (1994). The Sinister Attribution Error: paranoid cognition and collective distrust in organizations. *Motivation and Emotion*, 18, 199–230.
- Leary, M. R., Tambor, E. S., Terdal, S. K. and Downs, D. L.** (1995). Self-esteem as an interpersonal monitor: the sociometer hypothesis. *Journal of Personality and Society Psychology*, 68, 518–530.
- Matarazzo, J. and Herman, D.** (1984). Relationship of education and IQ in the WAIS-R standardizations sample. *Journal of Consulting and Clinical Psychology*, 52, 631–634.
- McKay, R., Langdon, R. and Coltheart, M.** (2007). The defensive function of persecutory delusions: an investigation using the Implicit Association Test. *Cognitive Neuropsychiatry*, 12, 1–24.
- Mehl, S., Rief, W., Lullmann, E., Ziegler, M., Muller, M. J. and Lincoln, T. M.** (2010). Implicit attributional style revisited: evidence for a state-specific “self-decreasing” implicit attributional style in patients with persecutory delusions. *Cognitive Neuropsychiatry*, 15, 451–476.
- Melo, S., Corcoran, R., Shryane, N. and Bentall, R. P.** (2009). The persecution and deservedness scale. *Psychology and Psychotherapy*, 82, 247–260.
- Melo, S., Taylor, J., and Bentall, R. P.** (2006). “Poor me” versus “bad me” paranoia and the instability of persecutory ideation. *Psychology and Psychotherapy: Theory, Research, Practice*, 79, 271–287.
- Moutoussis, M., Williams, J., Dayan, P. and Bentall, R. P.** (2007). Persecutory delusions and the conditioned avoidance paradigm: towards an integration of the psychology and biology of paranoia. *Cognitive Neuropsychiatry*, 12, 495–510.
- Ogden, T.** (1983). The concept of internal object relations. *International Journal of Psychoanalysis*, 64, 227–241.

- Olsson, I., Mykletun, A. and Dahl, A. A.** (2005). The Hospital Anxiety and Depression Rating Scale: a cross-sectional study of psychometrics and case finding abilities in general practice. *BMC Psychiatry*, 5, 46.
- Pretzer, J.** (2007). Cognitive therapy of personality disorders. In A. Beck, A. Freeman and D. Davis (Eds.), *Paranoid Personality Disorder (2nd ed.)* (pp.115–137). New York: Guilford Press.
- Segal, H.** (1994). Paranoia: new psychoanalytic perspectives. In J. Oldham and S. Bone (Eds.), *Paranoid Anxiety and Paranoia*. Madison, CT: International Universities Press Inc.
- Trower, P. and Chadwick, P.** (1995). Pathways to defense of the self: a theory of two types of paranoia. *Clinical Psychology: Science and Practice*, 2, 263–278.
- Udachina, A., Thewissen, V., Myin-Germeys, I., Fitzpatrick, S., O’Kane, A. and Bentall, R. P.** (2009). Understanding the relationships between self-esteem, experiential avoidance, and paranoia: structural equation modelling and experience sampling studies. *Journal of Nervous and Mental Disorders*, 197, 661–668.
- Vazquez, C., Diez-Alegria, C., Hernandez-Lloreda, M. J. and Moreno, M. N.** (2008). Implicit and explicit self-schema in active deluded, remitted deluded, and depressed patients. *Journal of Behavior Therapy and Experimental Psychiatry*, 39, 587–599.
- Winters, K. C. and Neale, J. M.** (1983). Delusions and delusional thinking in psychotics: a review of the literature. *Clinical Psychology Review*, 3, 227–253. doi: 10.1016/0272-7358(83)9001-4
- World Health Organization** (1992). *The ICD-10 International Classification of Mental and Behavioural Disorders: clinical descriptions and diagnostic guidelines (10th ed.)*. Geneva: WHO.
- Zigmond, A. S. and Snaith, R. P.** (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, 67, 361–370.

Appendix

Questions inviting exploration of attributes

Following the “warning stimulus” the computer presented the statement “tell us about being X”, where “X” was one of the idiographically important attributes (Figure 1). Four types of question followed this statement. These were:

1. Are there people you know that might care if you were like that, for example your parents? Please describe what they might think if you appeared X.
2. Please tell us what other people might think about you if you appeared X.
3. Please describe a situation where you might be X.
4. Please describe what would it mean for you if you were X.

Observer-rated engagement scale

Ratings were made as follows:

0. No engagement: e.g. Prompt: “Describe a situation where you might be selfish” Response: “I wouldn’t be selfish deliberately ... I’m trying to think of one where I might be by mistake ... no, I can’t think of one, next question”. 0 also included reversal of the valence with which the term was used, as participants were instructed that these questions were about elaborating on the previous part of the experiment; e.g. if anger previously described as undesirable now became a positive attribute.
1. Low engagement: here the description only fleetingly made contact with the potentially negative (or positive) experiences. It was easy for the rater to see how in the situation presented the participant avoided, “distanced”, ameliorated or justified the attribute – e.g. “I suppose I would lie to prevent somebody else’s suffering”. A rating of 1 included most one-word answers. It also included sophisticated “balanced” accounts, especially of the yes-but variety that left the speaker

and listener with an anodyne experience. It also included answers that circumvented the key point of the question e.g. “what others think . . .” or “a situation . . .”

2. Moderate engagement: This described situations or consequences in a way that lacked affective tone: it might be just factual e.g. “if I was unreliable I wouldn’t get the letters out in time”. It might also describe aversive (or desirable) experiences /situations in vague, non-specific, nondescript, conventional terms.
3. High engagement: Here a palpable sense of threat (or support) to self-esteem, or esteem from others (or of self-evaluation within a situation) was described e.g. “if I was unreliable, other people would think badly of me/avoid me”. The rater could easily imagine that what was described corresponded to an aversive (or desirable) situation.