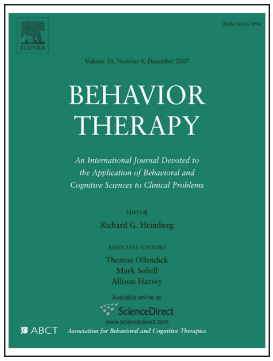
Accepted Manuscript



Negative interpretation biases precede the onset of psychosis

Jenny Yiend, Paul Allen, Natalie Lopez, Irina Falkenberg, Huai-Hsuan Tseng, Philip McGuire

PII: S0005-7894(18)30149-7
DOI: https://doi.org/10.1016/j.beth.2018.11.002
Reference: BETH 856

To appear in: *Behavior Therapy*

Received date: 11 July 2018
Accepted date: 17 November 2018

Please cite this article as: Jenny Yiend, Paul Allen, Natalie Lopez, Irina Falkenberg, Huai-Hsuan Tseng, Philip McGuire , Negative interpretation biases precede the onset of psychosis. Beth (2018), https://doi.org/10.1016/j.beth.2018.11.002

Negative interpretation biases precede the onset of psychosis

Jenny Yiend*, Paul Allen, Natalie Lopez, Irina Falkenberg, Huai-Hsuan Tseng, and Philip McGuire

Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London

*Corresponding author: Dr. Jenny Yiend, PO 63, Institute of Psychiatry, King's College London, De Crespigny Park, London, SE5 8AF. Tel 0207848 5119. *Email*: jenny.yiend@kcl.ac.uk. *Phone*: +44 (020)78485119

Acknowledgements

We thank Yuanyuan Huo, Damla Irez and Katie Davis for assistance with data collection; George Savulich for assistance with preparation of the tasks and Josie Parkhouse for assistance in formatting the manuscript. Philip McGuire acknowledges support from the NIHR Biomedical Research Centre for Mental Health at KCL and SLaM, London UK. The views expressed are those of the authors and not necessarily those of the NHS.

Abstract

This study investigated whether a negative interpretation bias was present in people at high risk for psychosis. People with an At Risk Mental State (ARMS) (n = 21), patients with First Episode Psychosis (FEP) (n = 20), and healthy controls (n = 20) performed three tasks, each of which was designed to measure interpretation bias. Both ARMS and FEP participants showed an attenuated positive bias compared to controls. These findings extend previous results investigating interpretation bias in psychosis by showing that interpretative biases are present before the onset of psychosis, and could therefore contribute to its development. Biased interpretation mechanisms could be a new target for clinical intervention in the early phase of psychosis.

Keywords: interpretation bias, psychosis, at risk mental state, cognitive vulnerability

Negative interpretation biases precede the onset of psychosis

Cognitive biases are defined as 'the selective processing of pathology congruent information that might confirm a pathological belief' (Savulich, Shergill & Yiend, 2012, p. 516). More specifically, interpretation bias is defined as 'a consistent tendency to interpret emotionally ambiguous stimuli, situations, or events in a negative (or positive) manner' (Lee, Mathews, Shergill & Yiend, 2016, p. 26). Such biases may contribute to causing and maintaining psychopathologies (Yiend, 2010) through an explicit, plausible pathway. The suggested mechanism, when applied to psychosis, is that an enhanced tendency to select paranoid material for further processing (be it via attention or interpretation bias), is likely to lead to an artificially increased perception of risk of personal harm in the environment, which will enhance and maintain the matching symptoms, and in turn will promote further biased processing. A cycle of reciprocal causation has been suggested, and the closer the match between the disorder and the focus of the bias, the more potent the effects are likely to be.

Most recently, this putative causal role of pathology-congruent biased processing is supported quite directly by manipulation studies (e.g. studies using cognitive bias modification (CBM) techniques) showing that altering these biases in interpretation results in changes to symptoms or proxy symptoms (e.g. in anxiety: Yiend, Mackintosh, & Mathews, 2005; Mackintosh, Mathews, Yiend, Ridgeway, & Cook, 2006; Mathews, Ridgeway, Cook, & Yiend, 2007). A recent review of meta-analyses of CBM studies is informative about effects on anxiety and depression (Jones & Sharpe, 2017), but additional psychopathologies have also been investigated such as perfectionism and eating disorders (e.g. Yiend, Savulich, Coughtrey, & Shafran, 2011; Yiend, Parnes, Shepherd, Roche, & Cooper, 2014). This research is being translated into therapies which seek to alleviate clinical symptoms, therefore a better understanding of these biases in psychosis may ultimately result in improvements to treatment and prognosis.

Psychosis research has focused on cognitive deficits at the global level, such as impairments in attention, motor skills, working memory and executive function (e.g. Fioravanti, Bianchi, & Cinti, 2012; O'Carroll, 2000). These deficits are a prominent feature of psychosis and reflect generic impairments in cognitive abilities. In light of the complexity and variability of the psychotic symptom profile, there is a call for a single-symptom approach to psychosis research. This involves researching one putative causal factor at a time and examining the specific symptom it is believed to trigger (Freeman, 2011). This, it is argued, allows the development of more effective forms of treatment that target specific psychotic traits. Cognitive biases allow us this level of specificity as they focus on specific cognitive domains, such as interpretation, attention and reasoning (Blanchette & Richards, 2010). Savulich, Shergill, and Yiend (2012) reviewed the literature on paranoia-relevant interpretation in psychosis. They concluded that despite evidence suggesting that paranoia and paranoid psychosis is associated with selective avoidance of threat, generally reduced 'data gathering' and negative interpretations of hallucinations that elicit distress, there has been relatively little research examining the selective information processing biases of the sort that might support or exacerbate the paranoid beliefs themselves. Given the potential aetiological importance of these pathology congruent biases, they called for further research to investigate pathology congruent information processing in psychosis.

The present study compared At Risk Mental State (ARMS) participants and First Episode Psychosis (FEP) patients with healthy controls. FEP is defined as those who have, or are currently, experiencing a first episode of non-organic psychosis, which include "episodes of schizophreniform, manic and depressive psychoses, puerperal psychoses and acute and transient psychoses amongst others" (Macmillan, 2007, p. 1). FEP is therefore a clinical psychosis patient group in the early stages of the illness. The inclusion of the FEP group served, firstly, to replicate and extend the findings of the Savulich, Shergill, and Yiend (2017)

who examined a chronic clinical psychosis sample and reported negatively biased interpretation, specifically related to paranoid material in patients compared to matched healthy controls. Secondly, the FEP group provided a clinical patient sample against which to benchmark the extent of biased interpretation in our pre-clinical ARMS group.

ARMS are those who have not yet experienced their first episode of psychosis but are at risk of developing a psychotic illness. In this study, this was defined as displaying one of the following characteristics outlined in the Comprehensive Assessment of At-Risk Mental State (CAARMS; Yung et al., 2005) measure: attenuated psychotic symptoms, brief limited intermittent psychosis, or having a vulnerability to psychosis such as a first degree relative with psychosis or a schizotypal personality disorder combined with a decline in functioning for over a month at some point in the past year. The inclusion of this group was to test the causal relationship between interpretation biases and clinical psychosis. Within the literature ARMS groups are typically used to investigate the development of highly predictive vulnerability markers for FEP, as these features commonly present prior to illness onset (Knowles & Sharma, 2004). In the present context we expected that if interpretation bias is a causal factor in the development of psychosis, then the ARMS group should display this bias in comparison to matched healthy controls.

More broadly, focusing on the prodromal phase of illness in psychopathology research is important for at least three reasons. First, there is a clear precedent, based on previous prodromal research, that mechanisms (such as interpretation biases) previously reported in clinical samples, should be observable in the prodromal stages of an illness. (e.g. Iacoviello, Alloy, Abramson, & Choi, 2010). Therefore, we predicted that the characteristics of psychosis, including interpretative bias, would be present in the prodromal group. Second, as outlined above, if a mechanism (here interpretation bias) does indeed contribute to the onset and maintenance of a clinical disorder, then it *must* also be present in the prodromal

state, since any causal factor must precede its effect. Although demonstrating this would not be conclusive evidence of a causal role, it would be consistent with this proposition. More importantly, failure to observe the putative causal mechanism in the prodromal stage would lead to a definitive rejection of its hypothesized role in the onset of the disorder. Third, if pathology congruent interpretative biases are found in the prodromal phase, then this presents an exciting new treatment possibility. Experimental interventions targeting these biases (commonly called 'cognitive bias modification' procedures) could be given to those in the ARMS stage, to determine whether there is a beneficial effect on paranoid symptoms and transition rates to the disorder. Indeed one such intervention specifically targeting biased interpretations in clinical paranoia is already undergoing feasibility testing (Savulich et al., under review; Yiend et al., 2017). The present study is a critical precursor because using such interventions without first demonstrating the presence of the target dysfunctional mechanism might not only be ineffective, but could even be harmful (Yiend et al., 2015).

To measure interpretation biases the present study used well-established experimental tests with materials specifically relevant to the paranoid symptoms of psychosis, and which have been validated in previous clinical (Savulich et al., 2017) and subclinical (Savulich et al., 2015) research. Paranoia is estimated to be present in almost 50% of psychosis cases, and persecutory delusions are the second most common symptom (Sartorius et al., 1986). We selected three different tasks to measure interpretation biases in an attempt to provide congruent validity for our findings. One task was the Similarity Rating Task (SRT), which used short emotionally ambiguous scenarios and assessed participants' interpretation of their meaning. In a clinical study Savulich et al. (2017) found that both paranoid and non-paranoid psychosis patient groups were more biased, both in their responses to and interpretations of, emotionally ambiguous information, compared to controls. Their SRT results also revealed a specific bias in the interpretation of ambiguity related to potentially paranoid content in

paranoid patients (Savulich et al., 2017). Savulich et al. (2015) similarly found in a subclinical study that paranoid and low-paranoid groups endorsed non-paranoid interpretations more than paranoid interpretations, with both groups displaying an adaptive interpretation bias. This tendency was significantly weaker in the high than the low-paranoid group, suggesting a pattern of interpretation bias consistent with the maintenance of associated pathology.

A second task, the Scrambled Sentences Task (SST), measured the same mechanism (interpretation of emotional ambiguity) but used single sentence stimuli. Participants viewed a string of mixed up words and had to unscramble them to create a meaningful sentence with two different possibilities, one having a benign meaning and the other a pathological meaning (e.g. threatening, paranoid or negative). In a clinical study, Savulich et al. (2017) reported that both the paranoid and non-paranoid patient groups unscrambled significantly more negative sentences than the control group, and that paranoid patients made significantly more paranoid interpretations in the SST. Savulich et al. (2015) similarly, in a subclinical study reported that the high-paranoid group unscrambled a significantly higher percentage of negative sentences and a significantly lower percentage of positive sentences than the low paranoid group.

The final task, the Emotion Identification Task (EIT), used video clips of faces slowly morphing from a neutral to an emotional expression. The stimuli were therefore emotionally ambiguous at the point of participants' responses. The task measured what level of emotional intensity was required for correct identification of the emotion. Signal detection analyses can be used to separate participants' ability to accurately identify the emotion present, their 'perceptual sensitivity', from a more general tendency or preference for selecting a particular emotion, termed 'response bias'. Interpretation biases have been evidenced using this task in a number of clinical studies (e.g. Joormann & Gotlib, 2006).

In summary, the aim of the present study was to investigate the degree to which pathology-congruent interpretation biases differed between three groups with varying levels of psychosis: patients who have experienced their first episode of psychosis (FEP), patients at risk of developing psychosis (ARMS), and matched healthy controls. Our objective was to understand how interpretation biases may be involved in the progression and maintenance of psychotic symptoms. We tested the following hypotheses: a) that, replicating and extending previous work in chronic psychosis, the FEP group would be more negatively biased in their interpretations of ambiguous information than controls, b) that, consistent with the hypothesis that bias is a causal factor in the development of psychosis, the ARMS group would be more negatively biased in their interpretations of ambiguous information than controls. A comparison of secondary interest was that between the ARMS and FEP groups. Were this to follow the pattern reported in the broader literature on cognition in ARMS, the negative bias in the ARMS group would be less severe than in the FEP group (e.g. Hauser, Zhang, Sheridan et al., 2017). In contrast, were this to follow the pattern seen in the broader cognitive bias literature, the negative bias in the ARMS group would be at least as strong as that seen in the FEP group (e.g. Everaert, Podina & Koster, 2017).

Method

Participants

At Risk Mental State (ARMS) participants (n = 21) were recruited from OASIS (Outreach and Support in South London). The inclusion criteria were fluency in English; age 18 to 35; not meeting criteria for an Axis I disorder according to DSM-IV-TR (American Psychiatric Association, 2000); displaying one of the following characteristics outlined in the CAARMS: attenuated psychotic symptoms, brief limited intermittent psychosis, or have a vulnerability to psychosis such as having a first degree relative with psychosis or a schizotypal personality

disorder combined with a decline in functioning for over a month at some point in the past year. Exclusion criteria were pregnancy, history of neurological illness or serious head injury and excessive alcohol consumption (>21 units/ week for males, >14 units/ week for females).

First Episode Psychosis (FEP) patients (n = 20) were out-patients recruited from the South London and Maudsley NHS Foundation Trust Early Intervention Service. The inclusion criteria were fluency in English; age 18 to 35; not meeting criteria for a non-psychotic Axis I disorder according to DSM-IV-TR (American Psychiatric Association, 2000); and who had experienced their first episode of psychosis in the previous 3 years. Exclusion criteria were as for the ARMS group.

Healthy control participants (n = 20) were recruited from within the staff and student population of King's College London and the local community of South East London via advertising. Inclusion criteria were fluency in English; age 18 to 65. Exclusion criteria were as for patients.

Clinical assessments

ARMS and FEP participants were categorised using the Positive and Negative Symptoms Scale (PANSS; Kay et al. 1987), a 30-item clinical tool measuring severity in schizophrenia. ARMS and HC groups were administered the CAARMS measure to confirm categorisation of participants. Both clinical instruments were administered by fully trained members of the research team.

Demographic and Questionnaire Measures

Participants completed a battery of questionnaires measuring premorbid intelligence, paranoia, anxiety, depression, and stress. An abbreviated version of the Wechsler Adult Intelligence Test Third Edition (WAIS-III; Blyler et al. 2000) was used to measure IQ, comprising the following sub-tests: Digit-Symbol Substitution; Arithmetic; Block Design and

Information. Subtests were chosen based on past research showing their efficacy in accurately, but quickly, estimating IQ in people with schizophrenia (Blyler et al. 2000). The Green et al. Paranoid Thoughts Scale (GPTS; Green et al. 2008) and the Paranoia Scale (Fenigstein & Vanable, 1992) were used to measure paranoia levels. Both scales were used and aggregated for analysis to improve the validity of measurement of the underlying construct of trait paranoia. The Liebowitz Social Anxiety Scale (LSAS; Liebowitz, 1987) was used to measure fear and avoidance of social situations and the Depression Anxiety Stress Scale (DASS; Lovibond & Lovibond, 1995) was used to measure depression, anxiety, and stress.

Experimental Tasks

Tasks and materials are available on the Open Science Framework at https://osf.io/9hrf4/.

Similarity Ratings (SRT). The SRT task (Eysenck *et al.* 1991; Mathews & Mackintosh, 2000) has good reliability (Smith et al., 2017) and here comprised 15 passages designed to measure the degree of negative interpretation bias (taken from Eysenck *et al.* 1991), and 15 designed to measure the degree of paranoid interpretation bias (taken from Savulich *et al.* 2013). First, participants encoded ambiguous scenarios by reading two sentences with 10 secs to complete the final missing word fragment, (e.g. paranoid content example: '*You are collaborating with a new colleague on a work project. Despite your reminders, your colleague misses a critical… m—ting'*: meeting; negative example: '*Sandy gives a speech at her best friend's wedding party. During the speech the crowd begins to …l—gh'*: laugh). To reinforce encoding, participants were asked a neutral comprehension question (i.e. not influencing spontaneous emotional interpretations of the ambiguous passage; e.g. respectively, for examples above: '*Are you collaborating at work?*' Correct response: yes; *'Is Sandy's sister getting married?'* Correct response: no). In the second part of the task (the

Recognition Test) participants rated how similar to the original test passages were two separate, disambiguated sentences (denoted target sentences) using a Likert scale (1 = very different in meaning, 4 = very similar in meaning to the original passage). Each sentence reflected a different possible emotional meaning of the preceding passage (e.g. respectively, for examples above: *´Your colleague is sabotaging you'*: target, paranoid interpretation versus *'Your colleague is forgetful'*: target, non-paranoid interpretation; *´Everyone ridiculed Sandy's speech'*: target, negative interpretation versus *'Everyone enjoyed Sandy's speech'*: target, positive interpretation). 'Foil' sentences acted as control items for emotional response bias, by retaining the same level of emotional meaning, without reflecting interpretations of the passage presented (e.g. respectively, for examples above: *'Your cocktail is spiked´*: paranoid foil versus *'Your cocktail has the wrong ingredients'*: non-paranoid foil; '*The wedding party was below average´*: negative foil versus *'The wedding party was above average'*: positive foil). A bias to respond in a more or less paranoid or negative direction was captured by foils, whereas a bias to make specific interpretations of the previously encoded ambiguity was captured by targets.

Scrambled Sentences Task (SST). The SST (Wenzlaff et al. 1993; Smith et al., 2017) involved reordering 20 (10 related to paranoia content and 10 related to generally negative emotion) strings of words (e.g. paranoid: 'follow policemen try protect to me'; negative: 'happy miserable be I to expect') to construct grammatically correct statements. Each word string can be reordered in one of two possible ways, reflecting different meanings (e.g. 'policemen try to follow me': paranoid interpretation or 'policemen try to protect me': non-paranoid interpretation; 'I expect to be miserable': negative interpretation 'I expect to be happy': positive interpretation). Participants were instructed to reorder each 6-word string into a 5-word, grammatically correct statement. The proportion of paranoid statements constructed, out of the total number of items completed in the 4 minutes allowed, gives an

index of the degree of bias. As is usual participants were asked to perform the task under cognitive load (remembering a six-digit number) which is known to prevent any tendency for participants to suppress or control their bias.

In line with previous reports in the literature using this task (e.g. Yiend et al., 2014; Lee et al., 2016) interpretation bias scores were calculated for each type of content (paranoid, negative) as the proportion of sentences unscrambled to create a paranoid or negative meaning, out of the total attempted. We used the inclusive scoring method in which both exact matches, and matches with similar meaning, were included in this count (e.g. 'someone was aggressive toward me' or 'someone was aggressive' would both be counted as paranoid interpretations). The number of paranoid/negative interpretations was then divided by the total number of sentences attempted (including errors, for example grammatically incorrect sentences) to give a proportionate bias score. Thus a higher value indicated evidence of a more maladaptive (i.e. paranoid or negative) interpretation bias. The same procedure was repeated to calculate adaptive bias scores (i.e. proportion of non-paranoid or positive interpretations). Note that the positive bias score is not simply the inverse of the negative bias score, since some responses are counted as errors and are included only in the denominator.

Emotion Identification Task (EIT). The EIT (Joorman & Gotlib, 2006) measures the ability to recognize emotional expressions using a computerized four-alternative forced-choice task. Participants viewed morphed (Abrosoft FantaMorph software, version 3.5.5) neutral- emotional face stimuli (Ekman and Friesen, 1976) presented in E prime version 2.0 and selected, as quickly as possible, which of four possible emotions (anger, fear, happiness, or sadness) they saw using an E prime v 2.0 serial response box. The task comprised 13 randomised trials of each emotion. To ensure ceiling and floor effects were avoided (Savulich

et al. 2015), neutral - happy clips were shown for 5 secs (20% of the full morph) and neutral – negative emotion clips for 10 secs (40% of the full morph).

D-prime (*d'*), a measure of sensitivity, was calculated using the formula:

$$d'(E) = z(\text{Hits\_E}) - z(\text{FA\_E})$$

where "E" is one of four emotions: anger, fear, happiness, or sadness, and z represents standardised scores for the proportion of 'Hits'(pHits) and the proportion of false alarms (pFA). pHits for each emotion was calculated by dividing the number of hits (i.e. correct identifications of the emotion) by the maximum number possible (=13). The number of false alarms (i.e. false attributions of emotion *x* to a video clip of emotion *y)* was calculated by summing the number of times an emotion was incorrectly identified as one of the three other possible emotions. For example, the number of false alarms for *happiness* would be the number of times *happiness* was selected in response to viewing *anger, fear,* and *sadness*. pFA was calculated by, first, calculating the maximum possible number of false alarms (here 13 trials × 3 emotions = 39), then dividing the total number of false alarms for that emotion by this maximum figure . Z-scores were then calculated for each value of pFA and pHits and used in the above formula to calculate a sensitivity score (*d'*) for each emotion (Macmillan & Creelman, 1991, p. 9).

Procedure

The study received full ethical approval from the National Health Service UK Research Ethics Committee, reference, 11/LO/0623. After completing informed consent procedures, participants received the clinical interview measures and completed self-report questionnaires, before commencing the experimental tasks. Each task was explained verbally prior to administration and a brief practice session was given. For the computerised tasks,

participants were left to work through the items once they had satisfactorily completed the practice. In the case of the SST, the participant worked through three practice items, with help and correction where necessary, and proceeded to the main task once the researcher was confident that they had understood the task requirements.

The study was nested within a wider project on cognitive processing in prodromal psychosis, to be reported separately. Care was taken to avoid cross contamination between tasks and fatigue by ensuring that all tasks reported here were given at the same session, in counterbalanced order, were not preceded by other tasks likely to introduce contamination effects, and that participants were suitably rested prior to beginning the present battery. Average session length for the present battery was 4 hours. Participants were compensated proportionally for their time, and travel expenses at the rate of £10 per hour.

Results

Participant Characteristics

Demographic and Clinical Characteristics. Table 1 shows the demographic and clinical characteristics of the sample. One-way ANOVAs and follow up t-tests for continuous data, and Chi squared or Fisher's exact tests (where cell count was <5) revealed the profile of group differences as shown in Table 1. The groups differed significantly on a number of demographic (e.g. age, IQ and employment) and self-report (e.g. anxiety and depression) measures which are known to be systematically associated with psychosis. We also collected information on receipt of medication (not shown in the Table 1), which revealed that the number of participants using prescription medication for mental health problems (including but not limited to antipsychotics) was 0, 5, and 14 in the HC, ARMS and FEP groups respectively. In line with the recommendations of Miller and Chapman (2001) we did not

attempt to systematically control for these variables so as not to corrupt the grouping variable.

On the clinical measures the profile of the groups was as expected: tests did not reveal significant differences between the two patient groups on their PANSS scores, and the ARMS group showed significantly higher pathology on the CAARMS compared to healthy controls. As expected the groups differed significantly on the DASS, LSAS, PS, and GPT (all *p*s < .03). These differences were driven by the healthy control group; tests did not show significant differences between ARMS and FEP on any of these measures (all *p*s > .3). To give an indication of the level of paranoia in ARMS and FEP we calculated the proportion of each sample who scored more than one standard deviation above the mean of a normative non-clinical sample on the relevant subscale of the GPTS (persecution subscale, *n* = 353, non-clinical mean = 22.1, sd = 9.2; Green et al., 2008). Eleven out 20 (55%) of FEP patients scored above this cut-off and for the ARMS sample the figure was 6/21 (29%).

Similarity Ratings (SRT)

Errors and outliers are not possible by virtue of the design of the task. There were therefore no missing data or exclusions. In the case of one participant the equipment failed during the task and another declined the task; both are therefore excluded from this analysis, leaving the sample size for each group as follows: HC (*n* = 20), ARMS (*n* = 20), and FEP (*n* = 19). With this sample size we achieved 93% power to detect a medium effect size on a within-between interaction at alpha = .05 assuming a repeated measurements correlation of .5. At 80% power, assuming the same other parameters, our analysis was sensitive to detect a medium (f = .25) to small (f = .1) effect (detectable f = .21). Mean response ratings per participant were calculated separately for each condition [Target Type (target, foil), Direction (adaptive (i.e. positive or non-paranoid), maladaptive (i.e. negative or paranoid)) and Content (paranoid, negative)] automatically using E-Prime v2.0 Data Aid software. Means and

standard error of the data are displayed in Table 3a. This dependent measure met the assumptions required for parametric tests.

Analysis of these data followed the precedent set in the previous literature (e.g. Savulich et al., 2017; Lee, Mathews, Shergill, & Yiend, 2016). A mixed model repeated-measures ANOVA was conducted on mean response ratings using the between-subjects variable Group (HC, FEP, ARMS) and within-subjects variables Target Type (target, foil), Content (paranoid, negative) and Direction (adaptive, maladaptive). A number of significant main effects and interactions were qualified by a significant interaction involving all factors, Target Type × Content × Direction × Group, F (2, 56) = 5.492, p = .007, η_p^2 = .164).

In order to interpret the interaction, we followed the precedent of previous literature using this task (e.g. Savulich et al., 2017; Lee et al., 2016), which was also consistent with testing our current hypotheses regarding interpretation biases. We first considered targets and foils separately, since only targets measure interpretation bias (foils capture response bias), using two repeated-measures ANOVA's of Group × Content × Direction. The interaction was significant for Targets, F (2, 56) = 5.758, p = .005, η_p^2 = .171, but not Foils, F (2, 56) = .184, p = .832, η_p^2 = .007, indicating significant group differences in interpretation bias, despite no differences in response bias.

We then examined the factor Content, in line with our predictions that bias effects should be stronger for material matching the paranoid concerns of the patient groups. Separate repeated measures ANOVAs (Group x Direction) for paranoid and negative material, revealed a significant group difference in interpretation bias for paranoid items, F (2, 56) = 9.307, p > .001, η_p^2 = .249, but not negative items, F (2, 56) = 1.858, p = .166, η_p^2 = .062, in line with content specificity. To finally isolate the nature of this group difference in paranoid interpretation bias relevant means are shown in Figure 1. As shown, tests showed no significant group differences in the extent to which they made paranoid interpretations, F (2,

56) = 1.44, p = .246) but did reveal differences in the extent to which they made adaptive, non-paranoid interpretations, F (2, 56) = 5.996, p = .004. Bonferroni corrected individual comparisons revealed no significant differences between ARMS and FEP groups (p = .999), but both groups made significantly less adaptive interpretations (p = .005; p = .042, respectively) compared to HC. Taken together, the pattern of results suggested a difference between healthy controls and both at-risk and first-episode participants in the rating of non-paranoid items, indicating that both clinical/subclinical groups lacked an adaptive bias in their interpretation of ambiguous situations which was characteristic of healthy controls.

Scrambled Sentences

The design of the task is such that there are no outliers and "errors" are incorporated in the calculation of the bias score (see below). The accuracy for recall of the cognitive load was 73%. Three participants' SST data were missing, because they opted out of completing this task, meaning the sample size for analysis was as follows: HC (n = 19), ARMS (n = 21), and FEP (n = 18). With this sample size we achieved 92% power to detect a medium effect size on a within-between interaction at alpha = .05 assuming a repeated measurements correlation of .5. At 80% power, assuming the same other parameters, our analysis was sensitive to detect a medium to small effect (f = .21). Mean bias scores are shown in Table 2 (b). This dependent measure met the assumptions required for parametric tests.

Maladaptive bias. A mixed model, repeated-measures ANOVA was conducted on paranoid/ negative bias score with factors Group (HC, ARMS, FEP) and Content (paranoid, negative). There was no significant Group or Content main effect, F (2, 55) = 1.37, p = .263, η_p^2 = .047 and F (1, 55) = 1.70, p = .197, η_p^2 = .030, respectively. Nor was there a significant Group x Content interaction, F (2, 55) = .762, p = .472, η_p^2 = .027.

Adaptive bias. A mixed model, repeated-measures ANOVA was conducted on non-paranoid/ positive bias scores with factors Group (HC, ARMS, FEP) and Content (paranoid, negative). There was a significant Group main effect, F (2, 55) = 3.73, p = .030, η_p^2 = .079), but no significant Group x Content interaction, F (2, 55) = 1.21, p = .305, η_p^2 = .042. The group main effect reflected a significantly more adaptive bias (irrespective of whether the ambiguous material depicted paranoid or negative content) in the HC group (M = .71, SD = .22) compared to the two patient groups (FEP: M = .54, SD = .20; ARMS: M = .53, SD = .24), ps < .03. Tests comparing the two patient groups did not reveal significant results, p = .94. Similar to the results on the SRT task, these data suggested both clinical/subclinical groups lacked an adaptive bias in their interpretation of ambiguous situations which was characteristic of healthy controls, although in this case the effect applied irrespective of the type of ambiguous material presented.

Emotion Identification Task (EIT)

The design of the task did not permit missing values since morphing clips were shown for a fixed duration after which participants had to select one of the four possible emotions in order to proceed to the next trial (see method). As there were only four response options possible there were no outliers. Five participants' data were missing from the analysis; two datasets were lost due to equipment failure and three chose not to take part in the task. The number of participants in each group was therefore as follows: HC (n = 20), ARMS (n = 17), and FEP (n = 19). With this sample size we achieved 91% power to detect a medium effect size on a within-between interaction at alpha = .05 assuming a repeated measurements correlation of .5. At 80% power, assuming the same other parameters, our analysis was sensitive to detect a medium to small effect (f = .18). The dependent measure (*d'*, sensitivity) met the assumptions required for parametric tests.

A mixed model repeated measures ANOVA was conducted on *d'* (sensitivity) scores with factors Group (HC, ARMS, FEP) and Emotion (Happiness, Anger, Fear, Sadness). Mauchly's test indicated that sphericity was violated (W = .72, p = .004), hence the Greenhouse-Geisser correction was used to adjust degrees of freedom. There was a significant Group main effect, F (2, 53) = 12.28, p < .001, η_p^2 = .317), no significant Emotion main effect, F (3, 159) < .01, p > .999, η_p^2 <.001, and no significant Group x Emotion interaction, F (5.1, 135.4) = .84, p = .540, η_p^2 = .031. The group difference reflected that while statistical comparison between ARMS and FEP groups did not show significant results (*d'* = -.34, -.53 respectively, p = .536), both groups were significantly less sensitive than the HC (*d'* = .79, *ps* < .001). The negative *d'* estimates in both patient groups indicated higher false alarm rates than correct identifications (hit rates) reflecting the high degree of misattributions of emotion typically found in these patients (Johns & McGuire, 1999).

We therefore also analysed false alarm rates across emotion type to identify whether the pattern of misattributions varied across groups or with specific emotions. A mixed model repeated measures ANOVA was conducted on p(FA) values with factors Group (HC, ARMS, FEP) and Emotion (Happiness, Anger, Fear, Sadness). Mauchly's test indicated that sphericity was violated (W = .60, p < .001), hence the Greenhouse-Geisser correction was used to adjust degrees of freedom. There was no significant Group x Emotion interaction, F (4.70, 124.6) = .519, p = .793, η_p^2 = .019, only a significant Group main effect, F (2, 53) = 12.26, p < .001, η_p^2 = .316) and a main effect of Emotion, F (2.40, 124.6) = 16.36, p < .001, η_p^2 = .236. FEP and ARMS groups did not differ significantly on tests of the extent of their misattributions, p(FA) = .20 in both cases, p = .59, but both made significantly more misattributions than HCs, p(FA) = .15, ps < .001. Bonferroni-corrected post-hoc tests comparing misattributions of different emotions (irrespective of group) showed that the most common misattribution was sadness, p(FA) = .26, ps < .001, followed by anger and happy,

p(FA) = .16, .18, p = .57, with fear being least commonly incorrectly selected p(FA) = .11, ps <.005. Results for all three tasks are shown in Figure 1.

Relationship between symptom severity and bias

In order to examine the relationship between bias and paranoia symptom severity we adopted the approach reported in a previous investigation (Lee et al., 2016) within the combined patient sample only (n=41). We first calculated a composite score by z transforming, then averaging, the GPTS and PS scores to give a single combined score for paranoia symptom severity. This was entered as the dependent variable in a multiple linear regression analysis. Independent variables were the measures of paranoid interpretation bias for each task; for the SRT target paranoid, target non paranoid and for the SST paranoid (maladaptive) bias, non paranoid (adaptive) bias. For the EIT we collapsed d' scores across emotion categories and entered a single predictor of sensitivity to detect emotion, in line with the previous results showing no effect of specific emotion category. Results of the regression are shown in Table 3 and Figure 2. The model accounted for 47% of the variance in the paranoia symptom score, F (5, 32) = 4.71, p =.003. Target paranoid score on the SRT (β = .64, t(38) = 3.28, p = .003) was the only significant independent predictor. The remaining variables made no significant contributions to the model (all ts < 1.6, all ps > .13; see Table 3).

Discussion

We found a very similar pattern of results across three independent measures of interpretation bias. On the similarity ratings task we found significant group differences in interpretation bias, despite no differences in response bias. Both ARMS and FEP groups displayed significantly less adaptive interpretations than healthy controls, although these differences were not reflected in overtly paranoid interpretations. These results suggest that both patient

groups lacked the adaptive bias in their interpretation of ambiguous situations that was seen in healthy controls. Furthermore, effects on this task showed some degree of content specificity, in that the group differences were specific to paranoid, but not negative items. On the scrambled sentences task, the pattern was similar. Despite no significant effects on maladaptive bias, the groups differed on adaptive bias, with a more positive bias in the healthy controls than both clinical groups. However, on this task the effects were not content specific, but rather occurred irrespective of whether the ambiguous material depicted paranoid or negative information. In the identification task, both the clinical groups were less sensitive to detect emotion than controls, an effect that again was not moderated by type of emotion. In terms of our key mechanism of interest, interpretation of ambiguity, this result reflected a relative inability in clinical participants to accurately detect partial signals of emotion in the environment, which was confirmed by their correspondingly high emotion misattribution rate. Finally, regression analyses showed that paranoia-relevant interpretations accounted for 47% of the variance in self-reported paranoia in the patient sample.

This pattern of results supported our first hypothesis, that the FEP group would be more negatively biased in their interpretations of ambiguous information than controls. Whilst not displaying an overtly negative bias in their interpretation, the FEP group were significantly less positively biased than controls. These results mirrored those found by Savulich et al (2013) in a sample of subclinical high trait paranoid participants, and replicated and extended those of Savulich et al. (2017) who examined the same measures in a group of patients with chronic schizophrenia. The present results also supported our second hypothesis, that if bias is a causal factor in the development of psychosis then the ARMS group should be more negatively biased in their interpretations of ambiguous information than controls. There was no statistical evidence of differences between the ARMS and FEP groups on any bias measure, consistent with a relatively stable pattern of biased interpretation

across different stages of psychosis.

A number of features of our data merit further discussion. First the results revealed a distinct pattern of reduced positive bias rather than increased negative bias in the clinical groups compared to controls. This is a common finding within the cognitive bias literature, and is usually taken to reflect the existence of a continuum between adaptive (unrealistically positive) and maladaptive (overtly negative) processing (Hirsch and Mathews, 2000; Eysenck et al., 1991). Secondly our results varied in the extent to which biases were specifically displayed on material matching the concerns of the pathology (i.e. 'content specificity'). In one task (similarity rating) effects were specific to emotionally ambiguous material directly related to paranoia, and were not found on more generally negative material. On the other tasks however there was no evidence of content specificity, in that clinical groups showed interpretation biases on all emotionally relevant ambiguous information compared to healthy controls. These findings suggest that while specificity is sometimes apparent when examining interpretation biases in psychosis, it may not be an entirely reliable effect. One possible reason for this is the inevitable idiosyncrasy of the match between the standard stimuli sets an individual participant's personally relevant paranoid concerns. Thus while for some participants the paranoid materials may be especially relevant and elicit content specific effects, this is unlikely to be true for all and could lead to the observed unreliability. Future work could seek to address this by devising ways to select unique, personally relevant sets of material for use in experiments on a participant by participant basis.

A third feature of our results was the lack of emotion specific effects on the EIT task. Despite using anger, fear, sadness and happiness, we found only a generally reduced sensitivity to detect all emotions in the clinical/subclinical groups. This might be attributable to that fact that these stimuli were the least relevant of all the tasks to paranoia specific concerns or to lack of power to detect finer grained differences between emotions. In addition

the stimuli used in this task were visual rather than verbal which could have led to differences in the manifestation of processing biases. One might also argue that the EIT task aligns more closely with general measures of ability to identify emotion, rather than with more specific tests of the biased interpretation of emotional ambiguity, such as the SRT and SST. However one chooses to interpret the mechanisms involved in this task, the pattern of data we found was largely consistent with previous reports on emotion recognition tasks in psychosis in the wider literature. For example, in a recent systematic review and meta-analysis Barkl et al. (2014) reported a generalized effect of significantly poorer accuracy for identifying facial expressions of emotion in early-onset and first-episode psychosis compared to healthy controls, with some evidence that certain emotions (disgust, fear and surprise) were harder to identify than others (sadness and happiness) while some revealed no group differences (anger, neutral).

Our study had a number of limitations. First, the sample size was small. ARMS participants are relatively hard to recruit to research studies, even when there is access to a specialised clinical service. Replication of our findings is essential to confirm our conclusions. Although statistical tests returned no significant differences between ARMS and FEP groups, our small sample size meant that we only had 80% power to detect large effect sizes, dropping to less than 50% for small or medium effects. Furthermore, within the null hypothesis testing approach, it is invalid to conclude in favour of the null hypothesis (H_0), irrespective of power and sample size. In future work researchers could measure the Bayes Factor, which can then be used to evaluate the evidence for and against H_0 (Aczel et al., 2018). This would provide a more definitive conclusion regarding the presence or absence of differences in interpretation bias between ARMS and FEP.

Secondly, our findings were *consistent with* biased interpretation playing a causal role in psychosis, because bias was already present in the at-risk group. These results

incrementally extend the evidence for causality, which to date has relied upon observing associations within the clinical population. However, it is not possible to rule out other explanations, such as an unidentified third factor giving rise to both bias and disorder.

The next incremental step in ascertaining the causal status of interpretation bias might be a longitudinal follow-up of ARMS to test whether initial interpretation biases predicted later transition to clinical psychosis. An easier and more compelling way to definitively demonstrate causality is to manipulate the putative cause and examine whether the predicted effect can then be observed. Fortunately in the field of interpretation bias there is the very real possibility to do this. So called 'cognitive bias modification' techniques are experimental procedures which have been designed to induce biases in an individual participant (for example create a positive or negative bias in those participating in the procedure); thereafter changes in symptoms or proxy symptoms can be measured. Over a decade of work has established the feasibility and safety of these procedures in both healthy (e.g. Mathews, Ridgeway, Cook & Yiend, 2007; Lee, Mathews et al 2015) and clinical samples (Yiend, Parnes et al 2014; Yiend, Lee et al. 2014; Cardi et al., 2015). Future work on biased interpretation in psychosis could seek to test the causal hypothesis by applying an interpretation bias modification procedure specifically designed to target biases relevant to paranoia in healthy, vulnerable or even clinical groups.

A further limitation was that we could not investigate disorder specificity within the current design, i.e. the extent to which our results were characteristic of clinical paranoia distinct from comorbid depression and anxiety. These symptoms are highly correlated and therefore a different experimental design would be required to adequately investigate this question. For example, one might recruit a sample of similarly paranoid people who vary in their level of depression (and/or vice versa) and measure both negativity bias and paranoid bias. If bias is disorder specific then negativity bias alone should vary with depression

severity, when paranoia levels are held constant, whereas paranoid bias should vary with paranoia severity in a sample where depression level was constant.

The present study has some potential clinical implications. The presence of negative biases in interpretation in the high risk state presents exciting new treatment possibilities that could be explored involving modifying interpretation biases, as described above. At present the main psychological intervention used in the ARMS is Cognitive Behavioural Therapy (CBT). However, it has been associated with only moderate effect sizes for delusions (van der Gaag et al 2014). A recent meta-analysis suggests that although CBT may improve the symptoms and reduce the risk of psychosis in the short term, there is only limited evidence that it can prevent the disorder in the longer term. It is currently not possible to exclude the hypothesis that CBT might only delay the onset of psychosis without altering its course (Fusar-Poli et al., 2017). New directions in treatments emphasise briefer, targeted interventions, with a focus on putative causal factors, such as cognitive biases (Freeman, 2011). A trial testing the feasibility of modifying the negative biases specifically relevant to paranoia and paranoid beliefs, as examined in this paper, is currently underway (Yiend et al., 2017). If successful, an intervention targeting these biases offers the potential to reduce distressing paranoia.

Conclusions

In summary, this study examined the biased interpretation of emotionally ambiguous information in a sample of participants in the 'at risk mental state' (ARMS) a sample of first episode psychosis (FEP) patients and healthy controls. Results showed significantly attenuated positive biases in both ARMS and FEP groups compared to controls and that the ARMS group performed similarly to the FEP groups on all three bias tasks. These findings extend previous results investigating interpretation bias in clinical and subclinical psychosis by raising the possibility that interpretative biases could make a causal contribution to the

development of psychosis. We conclude that biased interpretation mechanisms could be a useful target for intervention both prior to the onset of psychosis and in the early stages of the illness.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text review). Washington, DC: Author.

Aczel B., Palfi, B., Szollosi , A., Kovacs, M., Szaszi, B., Szecsi, P., Zrubka, M., Gronau, Q. F., van den Bergh, D. & Wagenmakers, E-J. (2018). Quantifying Support for the Null Hypothesis in Psychology: An Empirical Investigation. *Advances in Methods and Practices in Psychological Science, 1(3)*, 357–366. DOI: 10.1177/2515245918773742.

Baker, S. L., Heinrichs, N., Kim, H., & Hofmann, S. G. (2002). The Liebowitz social anxiety scale as a self-report instrument: a preliminary psychometric analysis. *Behaviour Research and Therapy, 40*(6), 701-715. doi:10.1016/s0005-7967(01)00060-2

Barkl, S. J., Lah, S., Harris, A. W., & Williams, L. M. (2014). Facial emotion identification in early-onset and first-episode psychosis: A systematic review with meta-analysis. *Schizophrenia Research, 159*(1), 62-69. doi:10.1016/j.schres.2014.07.049

Blanchette, I., & Richards, A. (2010). The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Cognition & Emotion, 24*(4), 561-595. doi:10.1080/02699930903132496

Blyler, C. R., Gold, J. M., Iannone, V. N., & Buchanan, R. W. (2000). Short form of the WAIS-III for use with patients with schizophrenia. *Schizophrenia Research, 46*(2-3), 209-215. doi:10.1016/s0920-9964(00)00017-7

Brown, T. A., Chorpita, B. F., Korotitsch, W., & Barlow, D. H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical

samples. *Behaviour Research and Therapy, 35*(1), 79-89. doi:10.1016/s0005-7967(96)00068-x

Cardi, V., Corfield, F., Leppanen, J., Rhind, C., Deriziotis, S., Hadjimichalis, A., . . . Treasure, J. (2015). Emotional Processing, Recognition, Empathy and Evoked Facial Expression in Eating Disorders: An Experimental Study to Map Deficits in Social Cognition. *Plos One, 10*(8). doi:10.1371/journal.pone.0133827

Crawford, J. R., & Henry, J. D. (2003). The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample. *British Journal of Clinical Psychology, 42*(2), 111-131. doi:10.1348/014466503321903544

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.

Eysenck, M. W., Mogg, K., May, J., Richards, A., & Al, E. (1991). Bias in interpretation of ambiguous sentences related to threat in anxiety. *Journal of Abnormal Psychology, 100*(2), 144-150. doi:10.1037//0021-843x.100.2.144

Everaert, J., et al. (2017). A comprehensive meta-analysis of interpretation biases in depression. *Clinical Psychology Review* 58: 33-48. Doi:10.1016/j.cpr.2017.09.005

Faul, F., Erdfelder, E., Lang, A-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39 (2),* 175-191.

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

Fioravanti, M., Bianchi, V., & Cinti, M. E. (2012). Cognitive deficits in schizophrenia: an updated meta-analysis of the scientific evidence. *BMC Psychiatry, 12*(1). doi:10.1186/1471-244X-12-64

Freeman, D. (2011). Improving cognitive treatments for delusions. *Schizophrenia Research, 132*(2-3), 135-139. doi:10.1016/j.schres.2011.08.012

Fusar-Poli, P., Ramella-Cravaro, V., Oliver, D., Kingdon, J., Kotlicka-Antczak , M., Valmaggia, L., ... McGuire, P. (2017). Deconstructing vulnerability for psychosis: Meta-analysis of environmental risk factors for psychosis in subjects at ultra high-risk. *European Psychiatry, 40*, 65-75. DOI: 10.1016/j.eurpsy.2016.09.003

Gaag, M. V., Valmaggia, L. R., & Smit, F. (2014). The effects of individually tailored formulation-based cognitive behavioural therapy in auditory hallucinations and delusions: A meta-analysis. *Schizophrenia Research, 156*(1), 30-37. doi:10.1016/j.schres.2014.03.016

Green, C. E. L., Freeman, D., Kuipers, E., Bebbington, P., Fowler, D., . . . Garety, P. A. (2008). Measuring ideas of persecution and social reference: the Green et al. Paranoid Thought Scales. *Psychological Medicine, 38*(1), 101-111.

Hauser, M., et al. (2017). Neuropsychological Test Performance to Enhance Identification of Subjects at Clinical High Risk for Psychosis and to Be Most Promising for Predictive Algorithms for Conversion to Psychosis: A Meta-Analysis. *Journal of Clinical Psychiatry* 78(1): E28-E40. doi: 10.4088/JCP.15r10197

Hirsch, C. R., & Mathews, A. (2000). Impaired positive inferential bias in social phobia. *Journal of Abnormal Psychology, 109*(4), 705-712. doi:10.1037//0021-843x.109.4.705

Iacoviello, B. M., Alloy, L. B., Abramson, L. Y., & Choi, J. Y. (2010). The early course of depression: A longitudinal investigation of prodromal symptoms and their relation to the symptomatic course of depressive episodes. *Journal of Abnormal Psychology, 119*(3), 459-467. doi:10.1037/a0020114

Jones, E. B., & Sharpe, L. (2017). Cognitive bias modification: A review of meta-analyses. *Journal of Affective Disorders, 223*, 175-183. doi:10.1016/j.jad.2017.07.034

Johns, L. C., & Mcguire, P. K. (1999). Verbal self-monitoring and auditory hallucinations in schizophrenia. *The Lancet, 353*(9151), 469-470. doi:10.1016/s0140-6736(98)05288-x

Joormann, J., & Gotlib, I. H. (2006). Is this happiness I see? Biases in the identification of emotional facial expressions in depression and social phobia. *Journal of Abnormal Psychology, 115*(4), 705-714. doi:10.1037/0021-843x.115.4.705

Kay, S. R. (1990a). Positive-negative symptom assessment in schizophrenia: Psychometric issues and scale comparison. *Psychiatric Quarterly, 61*(3), 163-178. doi:10.1007/bf01064966

Kay, S. R. (1990b). Significance of the Positive-negative Distinction in Schizophrenia. *Schizophrenia Bulletin, 16*(4), 635-652. doi:10.1093/schbul/16.4.635

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophrenia Bulletin, 13*(2), 261-276. doi:10.1093/schbul/13.2.261

Knowles, L., & Sharma, T. (2004). Identifying Vulnerability Markers in Prodromal Patients: A Step in the Right Direction for Schizophrenia Prevention. *CNS*

Spectrums, *9*(08), 595-602. doi:10.1017/s1092852900002765

Lee, J., Mathews, A., Shergill, S., & Yiend, J. (2016). Magnitude of negative interpretation bias depends on severity of depression. *Behaviour Research and Therapy*, *83*, 26-34. doi:10.1016/j.brat.2016.05.007

Lee, J., Mathews, A., Shergill, S., Chan, D. K., Majeed, N., & Yiend, J. (2015). How can we enhance cognitive bias modification techniques? The effects of prospective cognition. *Journal of Behavior Therapy and Experimental Psychiatry*, *49*, 120-127. doi:10.1016/j.jbtep.2015.03.007

Liebowitz, M. R. (1987). Social Phobia. *Anxiety Modern Trends in pharmacopsychiatry*, 141-173. doi:10.1159/000414022

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the depression anxiety stress scales* (2nd ed.). Sydney, Australia: School of Psychology, University of New South Wales.

Mackintosh, B., Mathews, A., Yiend, J., Ridgeway, V., & Cook, E. (2006). Induced Biases in Emotional Interpretation Influence Stress Vulnerability and Endure Despite Changes in Context. *Behavior Therapy*, *37*(3), 209-222. doi:10.1016/j.beth.2006.03.001

Macmillan, I. (2007). Early Intervention in Psychosis: Guidance clarifying EIP and ARMS referral criteria [PDF]. NHS England North.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: a user's guide*. Cambridge: Cambridge University Press.

Mathews, A., & Mackintosh, B. (2000). Induced emotional interpretation bias and anxiety. *Journal of Abnormal Psychology*, *109*(4), 602-615. doi:10.1037//0021-843x.109.4.602

Mathews, A., Ridgeway, V., Cook, E., & Yiend, J. (2007). Inducing a benign

interpretational bias reduces trait anxiety. *Journal of Behavior Therapy and Experimental Psychiatry, 38*(2), 225-236. doi:10.1016/j.jbtep.2006.10.011

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. Journal of Abnormal Psychology, 110(1), 40-48. doi.org/10.1037/0021-843X.110.1.40.

O'carroll, R. (2000). Cognitive impairment in schizophrenia. Advances in Psychiatric Treatment, 6(3), 161-168. doi:10.1192/apt.6.3.161

Page, A. C., Hooke, G. R., & Morrison, D. L. (2007). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in depressed clinical samples. *British Journal of Clinical Psychology, 46*(3), 283-297. doi:10.1348/014466506x158996

Phillips, L. K., & Seidman, L. J. (2008). Emotion Processing in Persons at Risk for Schizophrenia. *Schizophrenia Bulletin, 34*(5), 888-903. doi:10.1093/schbul/sbn085

Rude, S. S., Wenzlaff, R. M., Gibbs, B., Vane, J., & Whitney, T. (2002). Negative processing biases predict subsequent depressive symptoms. *Cognition & Emotion, 16*(3), 423-440. doi:10.1080/02699930143000554

Sartorius, N., Jablensky, A., Korten, A., Ernberg, G., Anker, M., Cooper, J. E., et al. (1986). Early manifestations and first-contact incidence of schizophrenia in different cultures. *Psychological Medicine, 16*, 909−928.

Savulich, G. J. (2013). *Biased interpretation in paranoia and paranoid psychosis*. (Doctoral dissertation). King's College London, Institute of Psychiatry. Retrieved from https://kclpure.kcl.ac.uk/portal/

Savulich, G., Edwards, A., Assadi, S., Guven, H., Leathers-Smith, E., Shergill, S. & Yiend, J. Biased interpretation in paranoia and its modification. Under

review.

Savulich, G., Freeman, D., Shergill, S., & Yiend, J. (2015). Interpretation Biases in Paranoia. *Behavior Therapy, 46(1),* 110-124. doi:10.1016/j.beth.2014.08.002

Savulich, G., Shergill, S.S & Yiend, J. (2017). Interpretation Biases in Clinical Paranoia. *Clinical Psychological Science*, 5 (6), 985-1000, doi: 10.1177/2167702617718180

Savulich, G., Shergill, S., & Yiend, J. (2012). Biased Cognition in Psychosis. *Journal of Experimental Psychopathology*. doi:10.5127/jep.016711

Smith, L., Leung, W. G., Crane, B., Parkinson, B., Toulopoulou, T., & Yiend, J. (2017). Bilingual comparison of Mandarin and English cognitive bias tasks. *Behavior Research Methods*. doi:10.3758/s13428-017-0871-0

Wechsler, D. (1997). *WAIS-III: Wechsler adult intelligence scale*. San Antonio: Psychological Corporation.

Wenzlaff, R. M., Wegner, D. M., Pennebaker, J. W. (1993). The mental control of depression: Psychological obstacles to emotional well-being *Handbook of mental control.* (pp. 239–257): Prentice-Hall.

Yiend, J. (2010). The effects of emotion on attention: A review of attentional processing of emotional information. *Cognition & Emotion, 24*(1), 3-47. doi:10.1080/02699930903205698

Yiend, J., Lee, J., Tekes, S., Atkins, L., Mathews, A., Vrinten, M., . . . Shergill, S. (2013). Modifying Interpretation in a Clinically Depressed Sample Using 'Cognitive Bias Modification-Errors': A Double Blind Randomised Controlled Trial. *Cognitive Therapy and Research*, *38*(2), 146-159. doi:10.1007/s10608-013-9571-y

Yiend, J., Mackintosh, B., & Mathews, A. (2005). Enduring consequences of

experimentally induced biases in interpretation. *Behaviour Research and Therapy, 43*(6), 779-797. doi:10.1016/j.brat.2004.06.007

Yiend, J., Parnes, C., Shepherd, K., Roche, M-K., & Cooper, M. (2014). Negative Self-Beliefs in Eating Disorders: A Cognitive-Bias-Modification Study. *Clinical Psychological Science, 2*(6), 756-766. 10.1177/2167702614528163

Yiend, J., Savulich, G., Coughtrey, A., & Shafran, R. (2011). Biased interpretation in perfectionism and its modification. *Behaviour Research and Therapy, 49*, 892–900. http://doi.org/10.1016/j.brat.2011.10.004

Yiend, J., Trotta, A., Meek, C., Dzafic, I., Baldus, N., Crane, B., Kabir, T., Stahl, D., Heslin, M., Shergill, S., McGuire, P., & Peters, E. (2017). Cognitive Bias Modification for Paranoia (CBM-pa): study protocol for a randomized controlled trial. *Trials, 18,* 298. doi:10.1186/s13063-017-2037-x

Yung, A. R., Yuen, H. P., McGorry, P. D., Phillips, L. J., Kelly, D., . . . Buckby, J. (2005). Mapping the onset of psychosis: the Comprehensive Assessment of At-Risk Mental States. *Australian and New Zealand Journal of Psychiatry*, *39*(11-12), 964-971. doi:10.1080/j.1440-1614.2005.01714.x

Table 1 Demographic and Clinical Characteristics of the Sample

Group Characteristic	HC (*n*=20)	ARMS (*n*=21)	FEP (*n*=20)	*p*
Age	23.25 (4.67)	24.67 (4.47)	26.85 (4.46)	0.048
Gender: female (male)	13 (7)	11 (10)	6 (14)	0.081
Ethnic Background				0.017
White	15	9	9	
Black	2	11	11	
Other	3	1	0	
Education Level				0.017
GCSE/ O Levels	1	2	5	
A Levels/ Secondary	0	7	3	
Vocational	0	2	1	
Higher Education	19	9	11	
Occupation				0.001
Employed	6	7	7	
Student	14	5	4	
Unemployed	0	7	9	
PANSS Total		49.95 (12.05)	51.10 (12.21)	0.751
CAARMS Total	2.58 (3.93)	35.45 (18.66)		<.001
Estimated Current IQ	116.5 (16.60)	94.00 (13.26)	95.85 (19.34)	<.001
DASS Total	7.85 (11.88)	40.29 (31.70)	32.25 (26.70)	<.001
LSAS Total	24.45 (17.44)	41.76 (32.93)	45.75 (24.02)	0.026
PS	28.10 (11.39)	44.00 (16.70)	46.35 (19.83)	0.001
GPTS Total	39.40 (12.02)	63.62 (28.16)	69.50 (36.64)	0.002

Note. HC = Healthy Controls; ARMS – At Risk Mental State; FEP = First Episode Psychosis GCSE = General Certificate of Secondary Education; PANSS = The Positive and Negative Symptoms Scale; CAARMS = Comprehensive Assessment of At Risk Mental States; DASS = The Depression Anxiety Stress Scale; LSAS = The Liebowitz Social Anxiety Scale; PS = The Paranoia Scale; GPTS = The Green et al. Paranoid Thoughts Scale; Values in parentheses are standard deviations. Cell counts for categorical data may vary which indicates participants declined to provide the information.

Table 2 Means (standard deviation) and confidence intervals for interpretation bias scores on three behavioral tasks

a) *Similarity Rating Task*

Group	Paranoid	95% CI	Non-paranoid	95% CI	Positive	95% CI	Negative	95% CI
	Targets							
HC	1.61 (.47)	[1.39, 1.83]	2.74 (.60)	[2.45, 3.02]	2.81 (.55)	[2.55, 3.08]	2.20 (.34)	[2.03, 2.36]
ARMS	1.72 (.48)	[1.49, 1.94]	2.11 (.45)	[1.90, 2.32]	2.37 (.52)	[2.12, 2.61]	2.04 (.51)	[1.80, 2.28]
FEP	1.89 (.56)	[1.59, 2.17]	2.29 (.61)	[2.00, 2.61]	2.57 (.50)	[2.31, 2.82]	2.07 (.44)	[1.84, 2.29]
	Foils							
HC	1.25 (.33)	[1.09, 1.41]	1.56 (.38)	[1.38, 1.74]	2.11 (.57)	[1.84, 2.39]	1.47 (.39)	[1.28, 1.65]
ARMS	1.39 (.46)	[1.17, 1.60]	1.48 (.47)	[1.26, 1.70]	1.93 (.56)	[1.67, 2.19]	1.58 (.51)	[1.34, 1.82]
FEP	1.51 (.46)	[1.27, 1.75]	1.72 (.57)	[1.43, 2.02]	2.16 (.54)	[1.88, 2.44]	1.72 (.46)	[1.49, 1.96]

b) *Scrambled Sentences Task*

	Maladaptive Bias				Adaptive Bias			
Group	Paranoid	95% CI	Negative	95% CI	Non-paranoid	95% CI	Positive	95% CI
HC	.21 (.24)	[.09, .32]	.20 (.17)	[.12, .28]	.68 (.29)	[.54, .82]	.73 (.21)	[.63, .83]
ARMS	.24 (.18)	[.15, .33]	.30 (.16)	[.23, .38]	.54 (.53)	[.40, .67]	.54 (.24)	[.43, .65]
FEP	.18 (.15)	[.10, .26]	.22 (.18)	[.13, .32]	.48 (.20)	[.37, .58]	.57 (.24)	[.45, .70]

c) *Emotional Identification Task*

Group	Angry	95% CI	Fearful	95% CI	Sad	95% CI	Happy	95% CI
HC	.85 (1.47)	[.16, 1.53]	.97 (1.26)	[.38, 1.56]	.76 (.81)	[.38, 1.14]	.58 (.74)	[.23, .93]
ARMS	-.28 (1.11)	[-.85, .29]	-.36 (1.45)	[-1.10, .39]	-.31 (1.27)	[-.97, .34]	-.41 (1.04)	[-.94, .12]
FEP	-.64 (.84)	[-1.05, -.23]	-.70 (1.45)	[-1.34, .00]	-.52 (1.44)	[-1.21, .17]	-.25 (1.01)	[-.73, .24]

Note. HC = Healthy Controls. ARMS= At Risk Mental State. FEP= First Episode Psychosis.

Table 3 Multiple linear regression of paranoid interpretations predicting level of self-reported composite paranoia symptom score (n = 41)

Predictor	β	t	p	R^2	Partial correlations
SRT-target paranoid	**.635**	**3.279**	**.003**		.534
SRT-target non paranoid	.199	1.325	.196		.247
SST- paranoid (maladaptive)	-.029	-.142	.888	0.47	-.027
SST- non paranoid (adaptive)	.292	1.550	.133		.286
EIT collapsed d prime	.141	.841	.408		.160

Note. SRT = Similarity Rating Task; SST = Scrambled Sentences Task; EIT = Emotion Identification Task. Variables in bold made significant independent predictions to the overall model shown.

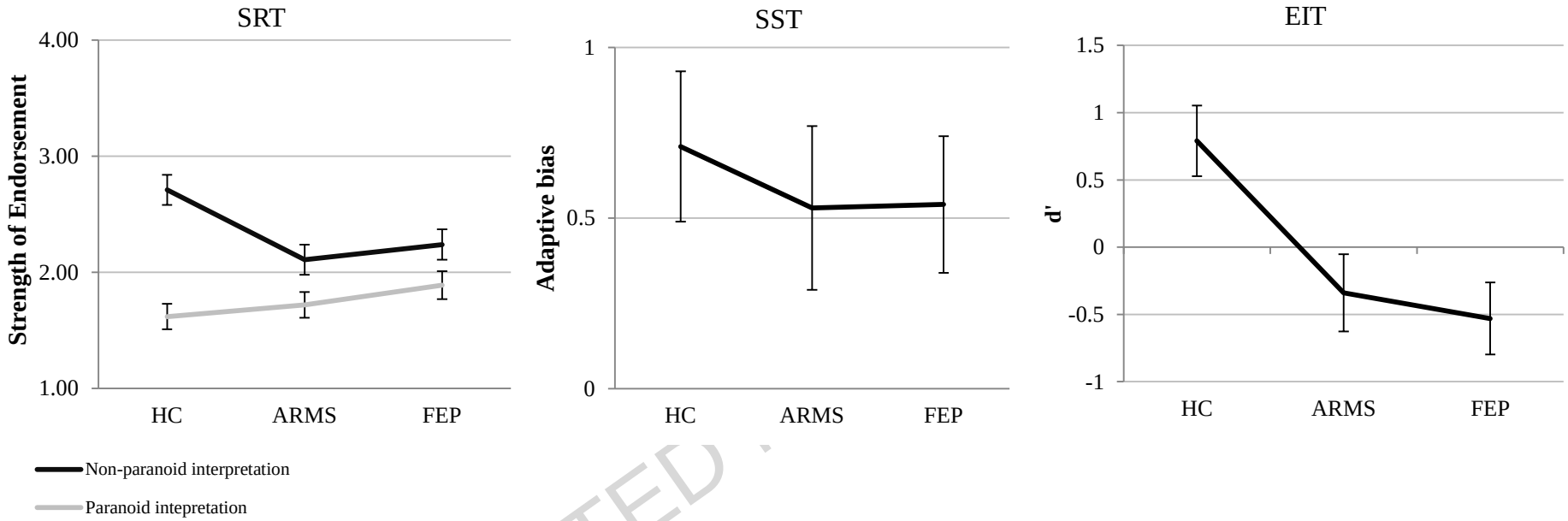


Figure 1: Pattern of findings across three measures of interpretation bias.

Note. SRT= Similarity Rating Task (target items). SST= Scrambled Sentences Task (adaptive bias). EIT= Emotional Identification Task (all emotions). HC = Healthy Controls. ARMS= At Risk Mental State. FEP= First Episode Psychosis. Error bars represent standard error of the mean.

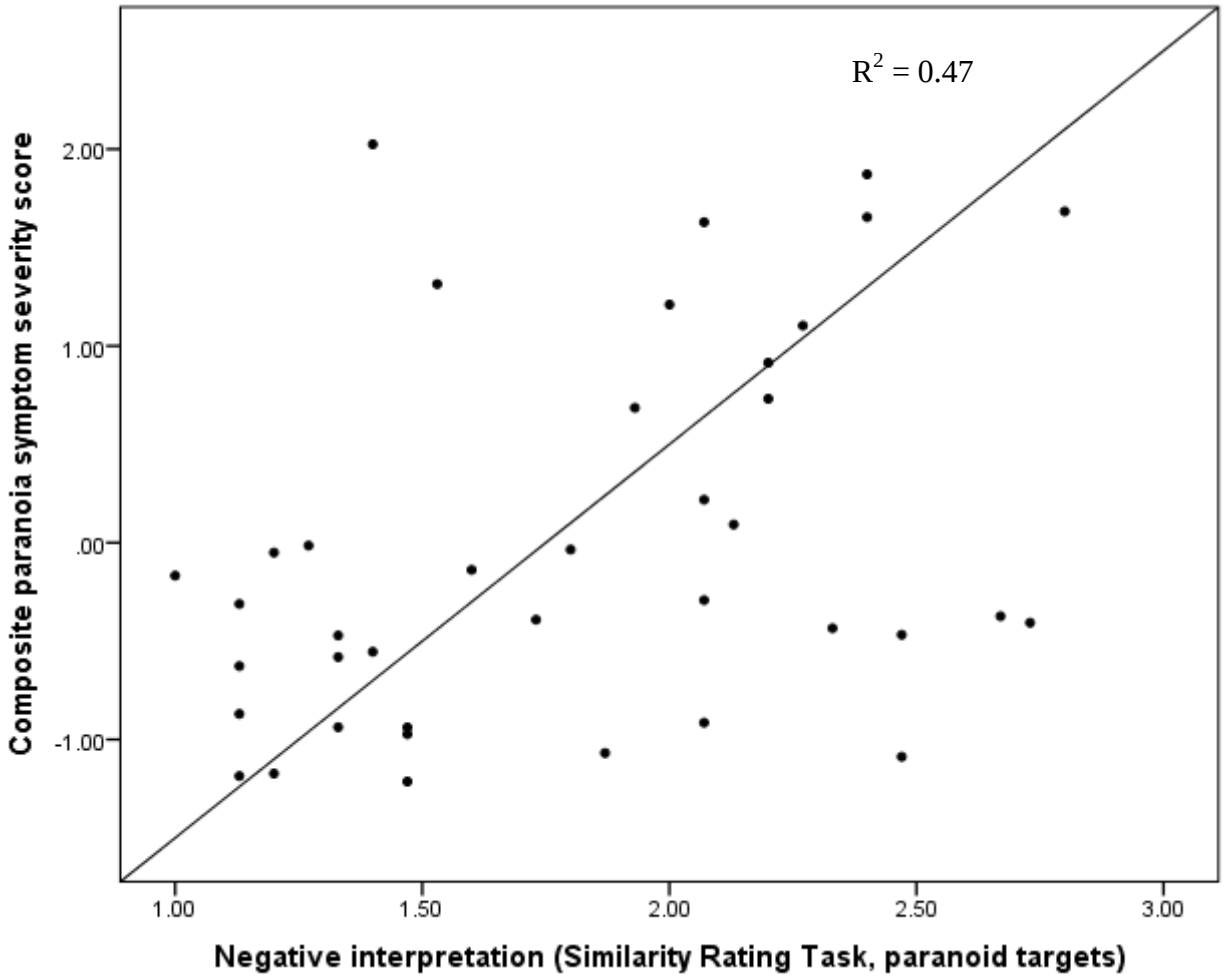


Figure 2: Linear relationship between paranoia symptom severity and interpretation bias

Highlights

- Psychosis patients and people at high risk performed tests of interpretation bias.
- Both showed attenuated positive interpretation bias compared to healthy controls
- Maladaptive interpretative bias could contribute to psychosis development
- This cognitive mechanism could be targeted for early clinical intervention.