King's Research Portal

[Link to publication record in King's Research Portal](Link)

Interpretation biases in clinical paranoia

George Savulich[1]*, Sukhwinder Shergill[2], and Jenny Yiend[2]*

[1]Department of Psychiatry and Behavioural and Clinical Neuroscience Institute, University of Cambridge

[2]Institute of Psychiatry, Psychology and Neuroscience, King's College London

*Corresponding authors:

Dr. George Savulich, Herchel Smith Building for Brain and Mind Sciences, Department of Psychiatry, University of Cambridge, Forvie Site, Robinson Way, CB2 0SZ. Email: gjs46@medschl.cam.ac.uk; Dr. Jenny Yiend, Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, PO 63, De Crespigny Park, London SE5 8AF. Email: jenny.yiend@kcl.ac.uk

Abstract

Interpretation biases matching the concerns of a psychopathology have been implicated in the etiology of psychological disorders, but little research has investigated their presence in psychosis. Here we investigated negative, and specifically paranoia-relevant, interpretation biases in patients with a diagnosis of schizophrenia, with ($n = 32$) and without ($n = 29$) paranoid symptoms and matched healthy controls ($n = 29$). Results revealed negatively biased interpretations of emotional ambiguity in both patient groups compared to controls; paranoid patients showing the stronger biases on material permitting paranoid interpretations, than on other types of ambiguous material; but mixed evidence that this content specific effect applied uniquely to the paranoid patient group. These data support models of psychopathology, including psychosis, which implicate cognitive biases in the formation and maintenance of core symptoms. We conclude that biased interpretation specifically related to paranoia deserves further detailed empirical investigation as a possible causal and maintaining factor for psychosis symptoms.

Word count: Abstract: 136; Main text: 6,XXX

Key words: paranoia, cognitive bias, interpretation, content specificity, psychosis

Psychosis is one of the most disabling mental health conditions, associated with distress and impairment in work, family and social functioning (Schizophrenia Commission, 2012). Lifetime rates are around 3.06% (Perala, Suvisaari, & Saarni, 2007) and persecutory delusions, the most common form of psychotic delusion, are also prevalent in other disorders and present in around 10-15% of the general population (Freeman, 2007). In recent years cognitive treatments for psychosis, such as Cognitive Behavioral Therapy (CBT) for psychosis (Hutton & Taylor, 2014), Metacognitive Training (Moritz et al., 2014; Moritz & Woodward, 2010) and Cognitive Remediation Therapy (Wykes, Huddy, Cellard, McGurk, & Czobor, 2011), have been developed with some success. Most cognitive treatments are intended to work by changing underlying beliefs and maladaptive behaviors so that patients can instead process and respond to information in a manner that promotes wellbeing (McManus, Van Doorn, & Yiend, 2011). However, CBT for psychosis has shown only moderate effect sizes for delusions (van der Gaag, Valmaggia, & Smit, 2014), potentially because treatment is too generic and other cognitive treatments have focused primarily on improving general cognitive impairments, such as attention span and data gathering.

Much psychosis research has focused on cognitive deficits at the global level. Patients are typically characterised by 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 contrast, cognitive biases refer to information-processing biases across specific cognitive domains, such as interpretation, judgment, decision making and reasoning (Blanchette & Richards, 2010). The most widely

researched cognitive bias in psychosis is reasoning bias. Research has shown that deluded individuals draw inferences on the basis of a smaller amount of information (i.e. 'jump to conclusions'; JTC) compared to healthy and psychiatric controls (Garety & Freeman, 1999). Other data-gathering biases associated with psychosis include a bias against disconfirmatory evidence (Moritz & Woodward, 2006) and a liberal acceptance bias of implausible information (Moritz et al., 2009).

Some research in psychosis has focused on biases that occur irrespective of the emotional content of the material being processed. In contrast, we use the term 'pathology-congruent' cognitive bias to refer to the selective processing of information that *matches the content of the core pathology of a disorder* (also known as *content specificity*). These more specific kind of biases have been seen as particularly important to our understanding of the etiology of psychological disorders for several reasons. Firstly, there is a clearly articulated, plausible mechanism through which pathology-congruent cognitive biases can act in a causal manner. The suggested mechanism is that an enhanced tendency to select threatening (or paranoid) items for further processing (be it via attention or interpretation bias) is likely to lead to an artificially increased perception of threat (or paranoia) in the environment, which will enhance and maintain the matching mood and symptoms (e.g. interpreting a stranger's stare as malicious is likely to support paranoid beliefs and increase distress about being at risk of observation by others). This in turn will promote further biased processing and a cycle of reciprocal causation has been suggested (Mathews, 1990; Teasdale, 1988).

Secondly, the empirical evidence actively supports the above suggested causal relationship across a range of pathologies, most notably anxiety and depression.

Considerable research has documented the etiological importance of negatively biased processing in both psychological disorders and high levels of personality traits that act as vulnerability factors to those disorders (e.g. Yiend, 2010). For example, anxious patients, compared to non-anxious controls, have long been known to selectively attend to threatening information matching their personal concerns (MacLeod, Mathews, & Tata, 1986) and to interpret ambiguous information in a manner that supports negative, symptom-related beliefs (Eysenck, Mogg, May, Richards, & Mathews, 1991). The more these biases act on emotional information matching the core symptoms of a psychopathology, the more potent they are in maintaining that pathology and associated symptoms. Recent research suggests that, in depression at least, this relationship is linear, with greater symptom severity being associated with stronger negatively biased interpretation (Lee, Mathews, Shergill, & Yiend, 2016). In addition, biases are known to resolve after recovery from a disorder, precede the onset of disorder and predict emotional response at a subsequent point in time (see MacLeod, Campbell, Rutherford and Wilson, 2004 for a narrative review of this evidence). These findings fail to *disconfirm* the causal hypothesis; active support for the causal role of biases comes from studies directly and experimentally manipulating biases and observing related effects on a range of relevant variables (for example, mood, proxy or actual symptoms and response to emotional stressors; MacLeod, 2012; but see also Cristea et al., 2015).

In short, pathology-congruent cognitive biases are one key mechanism underlying pathological beliefs that are clearly implicated in the cause of the relevant psychopathology (Yiend, 2010). Pathology-congruent *interpretation* bias is an example of one such phenomenon (Savulich, Shergill, & Yiend, 2012). Interpretation bias has

been defined as '…a consistent tendency to interpret emotionally ambiguous stimuli, situations, or events in a negative (or positive) manner…' (Lee et al., 2016) and is implicated in the etiology of psychological disorders, including anxiety and depression, as indicated above (Yiend, 2004). Cognitive experimental studies of interpretation bias typically employ tasks such as Similarity Ratings and Scrambled Sentences, which assess the degree to which pathology-congruent information is endorsed when interpreting ambiguity. Interpretation bias is recognized as a reliable phenomenon and one that is targeted by cognitive therapies across a range of disorders (Mathews, 2012). Furthermore, modification of interpretation bias specifically has led to some therapeutic benefits such as reduced anxiety (Mathews, Ridgeway, Cook, & Yiend, 2007), reduced vulnerability to an external stressor (Mackintosh, Mathews, Yiend, Ridgeway, & Cook, 2006) and improved mood (Holmes, Mathews, Dalgleish, & Mackintosh, 2006). However, cognitive therapies will only be effective to the extent that they target pre-existing biased mechanisms. In this investigation we sought to test for pathology-congruent biases in interpretation associated with clinical symptoms of paranoia and to establish their level of content specificity. In doing so we aimed to highlight the important role that these biases might play in the maintenance of symptoms and associated distress in clinical paranoia.

In psychosis there have been occasional reports of effects similar to those we investigate here. For example, pathology-congruent cognitive biases in attention have been reported using emotional Stroop (Bentall & Kaney, 1989), cue-target pairs (Moritz & Laudan, 2007), visual search (Phillips, Senior, & David, 2000) and eye tracking (Green, Williams, & Davidson, 2003). Bias has been examined in patients with

persecutory delusions who, in one study, showed preferential recall of threat-related stories (Kaney, Wolfenden, Dewey, & Bentall, 1992) and, in another, an attributional bias blaming others for negative events (Kinderman & Bentall, 1996). Questionnaire (Combs et al., 2009; Combs, Penn, Wicher, & Waldheter, 2007; An, Kang, Park, Kim, Lee, & Lee, 2010) and virtual reality methods (Freeman et al., 2005; Freeman et al., 2008) have been used to investigate perceived hostility about the intentions of others and jumping to conclusions has been examined using emotional as well as specifically delusion-relevant information (Menon, Mizrahi, & Kapur, 2008; Lincoln, Salzmann, Ziegler, & Westermann, 2011). However, none of the above has investigated the phenomenon of biased interpretation in the processing of emotional ambiguity using the established, reliable methodologies from the field of cognitive experimental psychology.

In a non-clinical analogue to the current study, our own laboratory has investigated pathology-congruent interpretation bias in individuals with high levels of trait paranoia (Savulich, Freeman, Shergill, & Yiend, 2015). Individuals with high trait paranoia interpreted ambiguous information more negatively than those with low trait paranoia, and crucially, this effect was more pronounced for information directly related to paranoid concerns. This suggests that interpretation of paranoia-specific (and thus more highly pathological) information might underlie paranoid symptoms in individuals with elevated vulnerability. We set out to test this same hypothesis in the current study by conducting a comprehensive and methodologically rigorous investigation of interpretation bias in a clinical sample of patients with psychosis. We sought to pinpoint biases, which are likely to directly precipitate negative paranoid cognitions, sustain persecutory delusions and further entrench paranoid beliefs. This in turn could

help to better inform the best mechanisms to target with cognitive treatments.

The specific aim of the present study was to investigate the degree to which paranoia-relevant and more generally valenced (positive/negative) interpretation biases differed between three groups with varying levels of paranoia: patients with a diagnosis of schizophrenia with paranoid symptoms (mild paranoid or beyond), patients with a diagnosis of schizophrenia without paranoid symptoms (absent or in the upper limits of normal) and matched healthy controls.

We included a measure of the data-gathering bias discussed earlier, specifically the jumping to conclusions (JTC) task, for two reasons. Firstly, we wished to illustrate the important conceptual differences between data-gathering biases (which have been widely investigated in psychosis populations), and the biases in emotion processing that were the focus of our current investigation (see Savulich et al., 2012). Secondly, we wished to benchmark the cognitive performance of our sample against previous results reported in similar samples in the literature (see Garety & Freeman, 2013 and Freeman, 2007 for reviews). We considered it important to evidence that any emotionally relevant interpretation biases that we observed were not due to some idiosyncrasy of our sample selection.

We hypothesized:

1) that both patient groups would be more negatively biased in their interpretations of emotionally ambiguous information than controls, and

2) that content specificity would be shown, whereby paranoid patients would be more biased on material permitting paranoid interpretations, than on other types of ambiguous material, and

3) this specificity would be less apparent in patients with fewer paranoid symptoms ('non-paranoid' group) compared to those with more paranoid symptoms ('paranoid' group).

Method

Participants

Sixty-one patients were recruited to participate in the study from: the South London and Maudsley NHS Foundation Trust; the Psychological Interventions Clinic for Outpatients with Psychosis; the Lewisham Early Intervention Service and the Cognition, Schizophrenia & Imaging laboratory at the Institute of Psychiatry, Psychology & Neuroscience, King's College London. Inclusion criteria were fluency in English; age 18 to 65; a diagnosis of schizophrenia according to DSM-IV-TR criteria (American Psychiatric Association, 2000), either with mild paranoia or beyond (paranoid patient group, $n = 32$) or without paranoia or in the upper limits of normal (non-paranoid patient group, $n = 29$); prescribed medication and dose unchanged in the last three months; and not having received CBT or any talking therapy in the two weeks prior to participation. Exclusion criteria were previous head injury resulting in a loss of consciousness for more than three minutes and current physical illness.

Healthy control participants ($n = 29$) were recruited from within the staff and student population of King's College London and the local community of South East London. Inclusion criteria were fluency in English; age 18 to 65; not currently taking any psychological or psychiatric medication; and not having a past or current psychological or psychiatric diagnosis. Exclusion criteria were as for patients.

Clinical Measures

The Positive and Negative Symptoms Scale (PANSS). The PANSS (Kay, Fiszbein, & Lewis, 1987) is a 30-item clinical tool that measures symptom severity in schizophrenia. It consists of three subscales with seven items for positive symptoms, seven items for negative symptoms and 16 items for general symptoms. Each symptom is rated on a 7-point Likert scale from 0 (symptom absent) to 7 (symptom extreme) (with a minimum score of 7 and a maximum score of 40 for the positive and negative symptoms subscales; and a minimum score of 16 and a maximum score of 112 for the general symptoms subscale). This tool was administered at the beginning of the experimental session to establish the presence and severity of paranoid symptoms. Patients were allocated to either the paranoid or non-paranoid patient group based on their score on the P6 Suspiciousness/Persecution item of this instrument. This item measures unrealistic or exaggerated ideas of persecution using a scale where 1 = absent; 2 = minimal, may be at upper extreme of normal limits; 3 = mildly symptomatic, then incrementally upwards to 7 = extreme persecutory delusions. Accordingly patients scoring 2 or lower were allocated to the non-paranoid group and those scoring 3 or over were allocated to the paranoid group. Thus, the non-paranoid group was characterised by absent to 'normal' paranoia and the paranoid group was characterised by mild to severe paranoia.

The Mini International Neuropsychiatric Interview (M.I.N.I). The M.I.N.I (Sheehan, Lecrubier, & Sheehan, 2011) is a structured diagnostic interview used to screen for major Axis 1 psychiatric disorders. Participants answer 'yes/no' questions corresponding to 15

diagnostic categories. Participants answering 'yes' to questions are then probed further with a series of questions that indicate whether diagnostic criteria are met. This interview was used to confirm the diagnosis of schizophrenia and to identify comorbid psychiatric disorders in the patient groups and to screen for psychiatric disorders in controls; only those screening negative for all disorders were invited to participate in the study as controls.

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, delusions, anxiety and depression.

The Wechsler Test of Adult Reading (WTAR). The WTAR (Wechsler, 2001) is an estimate of premorbid intelligence. Participants are instructed to read a 50-item word list aloud. The total score is the number of words pronounced correctly, with higher scores indicating higher intelligence (range is 0 – 50).

The Green et al. Paranoid Thoughts Scale (GPTS). The GPTS (Green et al., 2008) is a 32-item multidimensional measure of paranoid thinking. This scale is comprised of two 16-item subscales that assess ideas of reference and thoughts of persecution. Participants indicate thoughts that they might have had about others over the past month by rating statements using a 5-point Likert scale (1 = not at all to 5 = totally). Higher scores

indicate more paranoid thoughts (range is 0 – 160).

The Paranoia Scale. The Paranoia Scale (Fenigstein & Vanable, 1992) is a 20-item measure of paranoia. Participants indicate thoughts about themselves and others by rating statements using a 5-point Likert scale (1 = not at all to 5 = totally). Higher scores indicate a higher degree of paranoia (range is 0 – 100).

Peters' Delusion Inventory (PDI-21). The PDI-21 (Peters, Joseph, & Garety, 1999) is a 21-multidimensional measure of delusions. Participants answer 'yes/no' questions about their beliefs. Participants answering 'yes' to any question then rate the level of distress, preoccupation and conviction of their belief using a 5-point Likert scale. Higher scores indicate a greater degree of delusional ideation (range is 0 – 336).

The Hospital Anxiety and Depression Scale (HADS). The HADS (Zigmond & Snaith, 1983) is a 14-item measure of anxiety and depression. Each item is scored from 0-3, with a range from 0 to 21 for each subscale. Higher total score indicates higher level of anxiety and depression.

Experimental Tasks

Similarity Rating Task (SRT). The SRT (Eysenck et al., 1991; Mathews & Mackintosh, 2000) is a widely used measure of interpretation bias that presents participants with emotionally ambiguous passages (for a detailed description of this task see Yiend, Mackintosh & Mathews, 2005). Disambiguated sentences are subsequently rated for

similarity to the corresponding passage. Individuals with a tendency towards negative interpretation bias endorse negatively disambiguated sentences more strongly than positively disambiguated sentences. Sentences denoted as 'targets' measure interpretation bias; sentences denoted as 'foils' measure response bias (the tendency to endorse any negative/positive sentence irrespective of whether it is related to the passage previously read). The task was comprised of 15 paranoia relevant and 15 generally valenced passages. Each passage consisted of three sentences that described a variety of social situations. First, participants encoded the emotionally ambiguous passages by reading and completing a word fragment (i.e. 'fill in the missing letter') and answering a question at the end of a text passage. For example: 'After a long morning you enter the canteen for lunch. While waiting in the queue, you look for your colleagues and hope to join them. Across the room, two girls begin to …' followed by the word fragment wh-sp-r (whisper) and comprehension question 'Did you enter the canteen for lunch?' (Correct answer: yes).

Subsequently, in the so-called 'Recognition Test' participants rated the similarity to the corresponding original text passage, of individual disambiguated sentences. Four sentences were randomly presented, one at a time, beneath the title of their corresponding passage. Participants were asked to rate 'how similar is this sentence to the original passage?', on a scale from 1 (= very different) to 4 (= very similar). 'Target' sentences reflected possible meanings of the original text passage; 'foil' sentences were not directly relevant to the passage. Foil sentences retained the same degree of paranoia/non-paranoia or negativity/positivity as target sentences but were factually unrelated to the content of their associated passage. Corresponding disambiguating target and foil sentences for the

previous example item were 'The girls are plotting against you' (target, paranoid interpretation), 'The girls are talking about their friend' (target, non-paranoid interpretation), 'Your doctor gives you suspicious medication' (foil, unrelated paranoid interpretation), and 'Your doctor gives you beneficial medication' (foil, unrelated non-paranoid interpretation). Interpretation bias is inferred from higher average similarity ratings that favour one type of meaning over another (paranoid/non-paranoid; negative/positive). In this study, paranoid relevant items reflected intention of harm by a persecutor (Freeman & Garety, 2000). Five independent clinicians with expertise in psychosis had previously validated these items by rating each item's level of paranoia/non-paranoia (pilot analyses reported by Savulich et al., 2015). Generally valenced items had been validated in numerous other studies with different non-clinical analogue and patient groups (e.g. Yiend et al., 2014; Yiend, Parnes, Shepherd, Roche, & Cooper, 2014; Yiend, Savulich, Coughtrey, & Shafran, 2011) and were taken from those originally reported by Eysenck et al. (1991). Bias scores for paranoid and negative target sentences (mean paranoid/negative target rating − mean non-paranoid/positive target rating) were moderately correlated ($r = .40$, $p = .001$) and showed good Spearman-Brown's split-half reliability (.53).

Scrambled Sentences Task (SST). The SST (Wenzlaff, 1993) is a widely used measure of interpretation bias, requiring participants to reorder strings of words to construct grammatically correct statements. Each word string can be reordered into one of two possible statements, with positive and negative meanings, respectively. The proportion of negative statements constructed (out of the total number of items completed in the time

allowed) gives an index of the degree of negative bias. Here, participants were instructed to unscramble two blocks of 20 sentences (each block contained 10 sentences with paranoia-relevant meanings and 10 sentences with generally emotional meanings) by reordering five out of six jumbled words to create grammatically correct sentences. For each set of unscrambled sentences, either a paranoid/non-paranoid or a negative/positive sentence could be made. For example, the paranoia relevant scrambled sentence 'aggressive someone me friendly toward was' could be unscrambled to be either 'Someone was aggressive toward me' (paranoid interpretation) or 'Someone was friendly toward me' (non-paranoid interpretation); the valenced scrambled sentence 'winner am born I loser a' could be unscrambled to be either 'I am a born loser' (negative interpretation) or 'I am a born winner' (positive interpretation). Participants were given four minutes to complete this task and were asked to recall a six-digit number before and after unscrambling to create a cognitive load. This load was included to interfere with any tendency for participants to suppress or control their bias. The paranoia items of the SST, also validated by five practicing clinicians (as above), have shown sensitivity to detect symptoms relevant to interpretation biases in high levels of trait paranoia (Savulich et al., 2015). Valenced items have shown sensitivity to detect symptom-relevant interpretation biases in depression (Yiend et al., 2014; Rude, Valdez, Odom, & Ebrahimi, 2003; Lee et al., 2016). Percentage of paranoid and negative sentences were strongly correlated ($r = .70$, $p < .001$) and showed good Spearman-Brown's split-half reliability (.82).

Jumping to Conclusions Task (JTC). On this task, participants were instructed to decide which was the predominant colour (black or white) in a bag of 100 beads (Huq et al.,

1988). Participants were informed that the ratio of coloured beads was 60:40 (difficult version of task: 60:40; easy version of task: 85:15; Garety et al., 2013), but were not told which colour was the dominant one. The experimenter reached into a bag without looking and removed one bead at a time, continuing to take out as many beads as needed before the participant could confidently decide which colour ratio the bag contained (more black or more white). Participants were permitted as many beads as needed before making a decision. A fewer number of beads needed before making a decision indicates a greater 'jumping to conclusions' reasoning bias. Typically, a third of patients with delusions will make a decision after choosing around two beads on this version of the task.

Procedure

This study received full ethical approval from the South East London Research Ethics Committee (11/LO/0070). All referrals were outpatients at NHS mental health services at the time of taking part in the study. Patients were identified by a research nurse at the Institute of Psychiatry, Psychology & Neuroscience, a participant research panel provided by the Psychological Interventions Clinic for Outpatients with Psychosis and lead clinicians at the Lewisham Early Intervention Service. Control participants were recruited from internal mailing of King's College London students and staff and advertisements in the local community. Control participants were screened for psychiatric disorders using the M.I.N.I. and selected to match patients for premorbid IQ, gender and age distribution. All participants completed the experimental and clinical measures described above. Experimental measures were counterbalanced using a three-factor Latin-square design. Clinical measures were administered after the experimental measures to reduce demand

characteristics.

Results

Participant Characteristics

The mean scores and standard deviations of all demographic, clinical symptoms (patients only), and personality measures for the paranoid patient group ($n = 32$), non-paranoid patient group ($n = 29$) and control group ($n = 29$) are presented in Table 1.

Table 1

Groups were matched on age, premorbid IQ and gender ratio. Any differences in cognitive test measures are therefore not attributable to differences in these demographic variables. All patients were in receipt of medication but with dose unchanged in the last three months. By design, the paranoid patient group scored significantly higher on the PANSS P6 Suspiciousness/Persecution item than the non-paranoid patient group and also on the positive, general and total symptoms subscales of the PANSS (Table 1). Scores on item P6 ranged from 1 to 6 in the entire sample at the following percentages: scoring 1 (16.75%); scoring 2 (15.6%); scoring 3 (23.3%); scoring 4 (7.8%); scoring 5 (3.3%); and scoring 6 (1.1%).

As expected, the paranoid and non-paranoid patient groups significantly differed on measures of paranoia (GPTS: $t(52.12) = 3.76$, $p < .001$, d = .95, 95% CI, 14.74, 48.94; Paranoia Scale: $t(59) = 2.14$, $p = .036$, d = .55, 95% CI, .67, 20.07), delusions (PDI-21: $t(52) = 2.60$, $p = .012$, d = .72, 95% CI, 10.40, 80.30) and anxiety and depression (HADS:

($t(48) = 2.16$, $p = .036$, d $= .62$, 95% CI, .37, 10.09). Due to the potential confound with anxiety and depression levels, the analyses reported below were repeated using HADS as a covariate. This did not change the pattern of results reported; key interactions remained significant. Furthermore, the two patient groups were well matched diagnostically for co-morbid depression ($X^2 = .01$, $p = .94$), generalized anxiety disorder ($X^2 = 1.66$, $p = .20$), mania/hypomania ($X^2 = 1.87$, $p = .17$), obsessive-compulsive disorder ($X^2 = .01$, $p = .94$), posttraumatic stress disorder ($X^2 = .01$, $p = .94$) and substance abuse in the last 12 months ($X^2 = 1.87$, $p = .17$).

Table 2 shows correlations between bias measures for each task (SRT, SST, JTC) across all participants and correlations between bias measures for each task and the Paranoia Scale, Paranoid Thoughts Scale and Peter's Delusion Inventory.

Table 2

Similarity Ratings Task

An interpretation bias on this task is indicated by the presence of a Sentence Type (target, foil) x Stimulus Direction (pathological congruent, non-pathological congruent) interaction, indicating interpretation bias effects specific to targets rather than control items (foils). Alternatively, an effect involving Stimulus Direction alone (here, a Stimulus Direction x Group interaction) would be expected where an interpretation bias exists but is indistinguishable from response bias. In the present study we additionally expected the factor Content to interact with the above terms to demonstrate bias that was specific to paranoid items for the relevant groups (i.e. pathology-congruent bias).

A four-way mixed design ANOVA was therefore conducted on mean similarity ratings, with factors Content (paranoid, valenced) x Sentence Type (target, foil) x Stimulus Direction (pathological congruent, non-pathological congruent) x Group (paranoid patient, non-paranoid patient, control). This revealed a significant four-way interaction, $F(2,78) = 4.23$, $p = .018$, partial $\eta_p^2 = .10$, $\varepsilon = 1$ with main effects of Content, $F(1,78) = 30.86$, $p < .001$, $\eta_p^2 = .28$, $\varepsilon = 1$, Sentence Type, $F(1,78) = 136.00$, $p < .001$, $\eta_p^2 = .64$, $\varepsilon = 1$ and Stimulus Direction, $F(1,78) = 144.13$, $p < .001$, $\eta_p^2 = .65$, $\varepsilon = 1$. The Sentence Type main effect reflected endorsement of target sentences as more similar in meaning to the original passages than foil sentences (M = 2.30, SD = .32 vs. M = 1.85, SD = .39), as might be expected and is commonly found on this task. Mean participant ratings and standard deviations of each sentence type (targets and foils) for paranoia relevant and valenced items are indicated in Table 3[1].

---

[1] For comparability with studies using alternative participant groupings, we repeated analyses allocating participants to groups based on the PANSS P1 Delusions item (i.e. beliefs which are unfounded, unrealistic and idiosyncratic) where Group 1 comprised those scoring two or lower and Group 2 comprised those scoring 3 or over where 1 = absent, 2 = minimal ('normal'), 3 = mild and up to 7 = extreme. Relevant interactions changed little and were: Content x Sentence x Stimulus x Group, $F = 4.28$, $p = .017$ and Content x Group, $F = 4.56$, $p = .013$ for the SRT and SST tasks respectively. Analyses were also repeated using a more stringent PANSS P6 Suspiciousness/Persecution group allocation score (i.e. those scoring 3 or lower were allocated to the non-paranoid group and those scoring 4 or higher were allocated to the paranoid group). Again relevant

Table 3

To interpret the four-way interaction, follow up mixed ANOVAs (Sentence Type x Stimulus Direction x Group) were conducted for each type of content separately (paranoid, valenced).

Valenced material: For valenced items, the three-way interaction was not significant, $F(2,78) = 2.55$, p = .084, $\eta_p^2 = .06$, $\varepsilon = 1$. However, the Stimulus Direction x Group interaction was significant, $F(2,78) = 12.45$, p < .001, $\eta_p^2 = .24$, $\varepsilon = 1$ (negative direction: paranoid patient: 1.97, SD = .37, non-paranoid patient: 1.96, SD = .48, control = 1.65, SD = .23; positive direction: paranoid patient: 2.37, SD = .45, non-paranoid patient: 2.44, SD = .43, control = 2.52, SD = .44). Follow-up independent samples t-tests indicated that both patient groups were significantly more negatively biased than the healthy control group, both in terms of specific interpretation of the ambiguous text presented (target items) and in terms of a more general response bias (foil items) (paranoid vs. control, $t(45.24) = 3.89$, $p < .001$, d = 1.04, 95% CI, .16, .48; non-paranoid vs. control, $t(31.87) = 2.91$, $p = .007$, d = .81, 95% CI, .09., .53). The patient groups did not differ significantly from healthy controls on positive interpretations (paranoid vs. control, $t(54.85) = -1.28$, $p = .21$, d = .34, 95% CI, -.39, .09; non-paranoid vs. control, $t(51) = -.68$, $p = .50$, d = .18, 95% CI, -.32., .15). Thus in support of hypothesis 1, both patient groups were more biased, both in their responses to and interpretations of

interactions changed little and were: Content x Sentence x Stimulus x Group, $F = 4.51$, $p = .014$ and Content x Group, $F = 4.04$, $p = .021$, for the SRT and SST tasks, respectively.

emotionally ambiguous information, compared to controls. Results on valenced items alone, could not, however speak to our other hypotheses.

Paranoid material: For paranoid items, the corresponding three-way interaction was significant, $F(2,78) = 17.30$, $p <.001$, $\eta_p^2 = .31$, $\varepsilon = 1$. To interpret this we conducted separate two-way ANOVAs (Stimulus Direction x Group) for each Sentence Type (target, foil) on paranoid items only, in line with previous studies (Yiend et al., 2011; Yiend et al., 2005; Savulich et al., 2015). These revealed a significant Stimulus Direction x Group interaction for paranoid targets, $F(2,78) = 16.68$, $p < .001$, $\eta_p^2 = .30$, $\varepsilon = 1$, but not foils, $F(2,78) = .60$, $p = .55$, $\eta^2 = .02$, $\varepsilon = 1$. The pattern of results for paranoia relevant target items across groups is shown in Figure 1a.

Figure 1a

Follow-up independent samples t-tests showed that both patient groups made significantly more paranoid interpretations compared to controls (non-paranoid vs. control, $t(51) = 2.79$, $p = .007$, d = .75, 95% CI, .07, .52; paranoid vs. control, $t(43.62) = 4.68$, $p < .001$, d = 1.23, 95% CI, .27, .68 and significantly fewer non-paranoid interpretations compared to controls, (non-paranoid vs. control, $t(51) = -2.86$, $p = .006$, d = .77, 95% CI, -.58, -.10; paranoid vs. control, $t(55) = -3.27$, $p = .002$, d = .84, 95% CI -.58, -.14). The two patient groups did not differ significantly from each other in their biased interpretations of paranoia relevant material (on paranoid sentences: paranoid vs. non-paranoid patients, $t(50) = 1.38$, $p = .175$, d = .40, 95% CI -.08, .44, paranoid group mean = 2.21, SD= .46, non-paranoid group mean = 2.02, SD = .48); on non-paranoid

sentences: paranoid vs. non-paranoid patients, *t*(50) = -.12, *p* = .91, d = .75, 95% CI -.27, .24, paranoid group mean = 2.37, SD = .43, non-paranoid group mean = 2.38, SD = .48). Figure 1a illustrates *the specific finding* from this task, reflecting the above two-way interaction (*F* = 16.68, *p* < .001), namely significant group differences in the specific interpretation of ambiguous material (i.e. target items), only when that material permits interpretations relevant to paranoid thinking (i.e. paranoia relevant content only). These findings supported hypothesis two by revealing a specific bias in the interpretation of ambiguity related to potentially paranoid content in paranoid patients. In contrast, hypothesis three was not supported in that the degree of content specificity did not differ significantly between our two patient groups.

Figure 1b summarizes *the overall pattern* of the four-way interaction (*F* = 4.23, *p* = .018) using 'interpretation bias scores', now including both types of material (paranoia relevant and general negative/positive valence) for target items only. Unlike raw similarity rating scores (Figure 1a), interpretation bias scores reflect the overall degree of bias by taking into account both directions of endorsement (paranoid or negative and non-paranoid or positive) and are, arguably, a more accurate indicator of bias (see Method for calculation of interpretation bias scores).

Figure 1b

Scrambled Sentences Task

Following Yiend and colleagues (2014), interpretation bias was calculated as the percentage of sentences unscrambled to create a paranoid (or negative) meaning. The

total number of sentences unscrambled to create a paranoid (or negative) meaning was divided by the total number of paranoid (or valenced) sentences attempted, multiplied by 100. Only sentences that were exact matches to unscrambled paranoid/negative stimuli were included in the count of the numerator. Any sentences containing errors (e.g. sentences that were grammatically incorrect; sentences unscrambled using fewer than five words) were excluded from the numerator, but included in the denominator. Thus a higher percentage would indicate evidence of a more paranoid (or negative) interpretation bias. Percentages and standard deviations of paranoia relevant and valenced sentences for each group are indicated in Table 3.

A two-way mixed design ANOVA was conducted on the percentage of negative sentences created, with factors Content (paranoid, valenced) x Group (paranoid patient, non-paranoid patient, control). A main effect of Group, $F(2,81) = 22.20$, $p < .001$, $\eta_p^2 = .35$, revealed that both patient groups were negatively biased compared to healthy controls (paranoid patients: 42.68%, SD = 22.88; non-paranoid patients: 22.56%, SD = 19.95; controls: 10.94%, SD = 10.35). Both the paranoid and non-paranoid patient groups unscrambled significantly more negative sentences than the control group (paranoid vs. control, $t(40.68) = 6.90$, $p < .001$, d = .07, 95% CI, 22.45, 41.03; non-paranoid vs. control, $t(35.82) = 2.63$, $p = .013$, d = .73, 95% CI, 2.64, 20.62). Again this pattern of results supported hypothesis 1, in suggesting negative biases in both patient groups, compared to controls.

The analysis also revealed a significant Content x Group interaction, $F(2,81) = 5.26$, $p = .007$, $\eta_p^2 = .12$, $\varepsilon = 1$, which was broken down by the factor Group. Paranoid patients made significantly more paranoid interpretations (i.e. created more sentences

with a paranoid meaning; 48.96%, SD = 24.84 see Table 2) than negative interpretations

(36.40%, SD =26.47, $t(29)$ = 2.96, p = .006, d = .49, 95% CI, 3.87, 21.24), lending

support for hypothesis 2. Regarding hypothesis 3, neither the non-paranoid patients, nor

control groups, differed in the percentage of paranoid and negative interpretations made

(non-paranoid patients: $t(24)$ = 1.26, p = .218, d = .22, 95% CI, -3.19, 13.37; controls:

$t(28)$ = 1.67, p = .106, d = .28, 95% CI, -7.14, .72). This suggested that the content

specific effect observed in paranoid patients was not apparent in non-paranoid patients or

healthy controls on this task, meaning hypothesis three was also supported because

specificity was less evident in these groups. Independent samples t-tests showed that

paranoid patients were significantly more negatively biased than non-paranoid patients,

both in terms of a paranoia specific and generally valenced interpretation bias, $t(53)$ =

3.52, p = .001, d = .95, 95% CI 10.28, 37.47 and $t(53)$ = 2.59, p = .012, d = .71, 95% CI

3.68, 29.03, respectively. Figure 2 illustrates the main findings on this task.

Figure 2

Jumping to Conclusions Task

The paranoid patient group requested the least number of beads before making a decision

(M = 6.88, SD = 5.60), followed by the non-paranoid patient (M = 13.10, SD = 15.79)

and control groups (M = 16.21, SD = 17.70). A one-way ANOVA revealed these

differences were significant, $F(2,87)$ = 3.61, p = .031, η_{p2} = .08. Follow-up tests indicated

that the patient groups did not differ significantly from each other (p = 0.25), and that the

paranoid group differed significantly from controls, $t(59) = -2.83$, $p = .031$, d = .74, 95% CI, -18.01, -.65, whereas the non-paranoid group did not ($p = 1.0$).

Discussion

The aim of the present study was to use robust cognitive experimental methods to investigate interpretation biases in psychosis and to observe variations in these biases between clinical groups, with greater and fewer symptoms of paranoia, and matched healthy controls. ==We hypothesized, firstly, that negatively biased interpretations would be more evident in patient groups compared to controls. Secondly, we hypothesized that there would be evidence of content specificity, with paranoid patients being more biased on material permitting paranoid interpretations, compared to other ambiguous material. Thirdly, we hypothesized that this content specificity would be less apparent in non-paranoid than paranoid patients==.

In support of our first hypothesis, evidence of negative interpretation bias was found in both patient groups compared to healthy controls, on two independent measures of interpretation. Using ambiguous text passages, and measuring interpretations using similarity ratings of disambiguating sentences, both patient groups were more negatively biased than healthy controls. This bias was evident both in terms of the specific interpretation of the ambiguity presented (target items) and in terms of a more general response bias (foil items). Using scrambled sentences, which could be unscrambled to make either negative or positive meanings, both patient groups created a significantly higher percentage of negative meanings (both valenced and paranoid) than did healthy controls.

There was also support for our second hypothesis. On the Scrambled Sentences task, paranoid interpretation bias was stronger than negative interpretation bias for paranoid patients only, thus showing the *selective* processing of information matching their core symptoms. Results from the Similarity Ratings task further showed that consistent with hypothesis two, interpretation bias was only found for material relevant to paranoid thinking (i.e. target sentences). In line with our third hypothesis, neither non-paranoid patients nor controls showed evidence of content specificity on the Scrambled Sentences task. However, no clear differences were found in paranoia relevant interpretation bias between the patient groups on the Similarity Ratings task.

Evidence of a pathology-congruent interpretation bias on the Scrambled Sentences task, but not Similarity Ratings task, could be because paranoid material is relevant to all patients with psychosis to some degree, or that we did not have enough power to detect differences between patient groups given the more complex design of the latter. Lack of content specificity could also be due to a materials effect, in which materials on the Scrambled Sentences task are more sensitive, or better matched, to participants' specific concerns than the Similarity Ratings task, or because the former may tap into additional biased cognitive mechanisms beyond interpretation (e.g. selective attention) and thus produce additive effects. A further difference between the tasks is the use of cognitive load that may have prevented top-down inhibition of cognitive biases. Nonetheless, correlational analyses revealed that measures of paranoia and delusions were, for the most part, more strongly associated with indices of paranoid interpretation bias than with indices of negative interpretation bias.

Results from the 'jumping to conclusions' task revealed a pattern similar to that

reported in the previous literature (Garety et al., 2013), suggesting the cognitive performance of our sample was broadly as expected. Correlational analyses further revealed that interpretation bias measures were highly associated with each other, but not with the JTC task, reinforcing that biases in interpretation and reasoning can co-occur, but are conceptually different phenomenon in psychosis.

The present results confirm and extend the findings of our previous study, conducted in an analogue, non-clinical sample (Savulich et al., 2015). In that study we reported pathology-congruent interpretation bias in individuals with high trait paranoia, compared to those with low trait paranoia. As replicated here, content specific effects in that study were most pronounced for information directly related to paranoid concerns on the Scrambled Sentences task. In other work we have shown how important it is to replicate findings from subclinical samples in the corresponding patient group (Yiend et al., 2015) and the current study provides this. Another strength of the current investigation, which enhances validity, was that we found a consistent pattern of negatively biased interpretations across multiple tests of interpretation bias (ambiguous passages and scrambled sentences). This convergence of findings, both within the current study, and across subclinical and clinical groups, lends more weight to our results and increases the confidence which can be placed in the conclusions. Together these data suggest that people with clinical symptoms of paranoia, and those with elevated vulnerability to clinical paranoia, make interpretations of emotionally ambiguous information in a manner that could maintain paranoid beliefs. The content specificity of interpretation biases associated with paranoia has been relatively under researched (Savulich et al., 2012), but the present data go some way to redress this.

Based on our findings, we propose a specific cognitive mechanism exists in subclinical and clinical paranoia, whereby content specific interpretation biases lead to increasingly selective processing of information as it more closely matches an individuals' paranoid concerns. Additional research involving either manipulation of biases or longitudinal designs will be needed to test this hypothesis, as the present data are merely correlational. The proposed mechanism could work by selective processing acting to enhance both the *subjective perception* and the *actual exposure* to information consistent with paranoid thinking. Indeed this suggestion in consistent with existing general cognitive models of selective processing biases in psychopathology (e.g. Mathews & Mackintosh, 1998) and adds more specificity, including additional testable predictions including causality, to specific cognitive models of psychosis. For example, the 'threat anticipation cognitive model of persecutory delusions' (Freeman, Garety, Kuipers, Fowler, & Bebbington, 2002) proposes that the cognitive biases associated with psychosis are one route to an anomalous experience. According to this model, search for the meaning of an anomalous experience leads to the formation of a threat belief, which is maintained, if confirmed, by cognitive processes. Interpretation biases could be one such process, and their etiological role deserves further detailed empirical investigation in clinical paranoia.

Our study suffered from several limitations. Firstly, we did not have a patient control group without psychosis, only a matched healthy sample who inevitably differed markedly from patients on a wide range of variables. Although this omission does not undermine our conclusions, as outlined above, it nevertheless precludes the possibility to demonstrate a dissociation in which the reverse form of content specificity could have

been demonstrated. For example showing a pattern of content-specific (i.e. paranoid) interpretation bias in clinical psychosis, which was simultaneously absent in another psychological disorder, such as clinical depression, would have made our results even more compelling. Secondly, we did not assess other cognitive mechanisms, such as attentional biases, only interpretation. It is possible that similar content-specific effects could operate during selective attentional processes and might make additional important contributions to the maintenance of symptoms involving paranoid beliefs. Future work should investigate this. Finally, materials used for the jumping to conclusions task were not emotional and therefore could not capture differences in pathology-congruent biases. Future work could include materials that use social/emotional stimuli (e.g. Lincoln et al., 2011; Menon et al., 2008).

There are some clinical implications of our study. Our data do not demonstrate causality, but it may be possible in future work to manipulate the biases that we have measured here in an initial proof of principle. If so, it would then be possible to go on to assess whether clinical benefits might follow from manipulations specifically designed to reduce negative biases related to paranoid thoughts and beliefs. Cognitive Bias Modification (CBM) is a relatively new theory-driven treatment approach that uses a computerised task to manipulate pathological biases toward more adaptive information processing. CBM has been applied with varying levels of efficacy across a range of disorders (Cristea et al., 2015). The data presented here suggest that the version of CBM designed to manipulate interpretation (e.g. Lee et al., 2015) might be suitable for adaptation and testing as a possible technique to modify interpretation biases associated with paranoia. Subject to proof of principle of this sort, further development could lead to

interventions designed to specifically target unhelpful paranoid beliefs. Such work could complement and enhance other training packages, such as Metacognitive training of cognitive biases (Mortiz & Woodward, 2010), Social Cognition and Interaction Training (SCIT; Penn, Roberts, Combs, & Sterne, 2007) and reasoning training in delusions (Ross, Freeman, Dunn, & Garety, 2011).

In summary, this study extends previous work in subclinical paranoia and provides evidence of interpretation biases directly related to core content of symptomatic paranoid beliefs in psychosis patients. Two convergent measures of interpretation bias used carefully controlled and matched stimuli to experimentally measure both valenced and paranoia-specific biases in interpretation of emotional ambiguity. Negative biases were observed in both patient groups compared to healthy controls and paranoid patients showed stronger biases on material permitting specifically paranoid interpretations, than on other types of ambiguous material, although evidence was mixed concerning whether this content specific effect applied uniquely to the paranoid patient group. Biased interpretation is already established as an underlying cognitive mechanism with etiological importance in a range of other disorders (Yiend, 2010) and this study is an important preliminary step toward advancing our understanding of similar mechanisms in psychosis. We conclude that biased interpretation specifically related to paranoia deserves further detailed empirical investigation as a putative causal and maintaining mechanism in psychosis and may be a suitable target for developing new interventions such as bias modification.

Authors Contributions

J. Yiend, S. Shergill and G. Savulich conceived and designed the study; G. Savulich recruited participants and collected data; J. Yiend and G. Savulich analysed and interpreted data and wrote the manuscript.

Declarations of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

References

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

An, S. K., Kang, J. I., Park, J. Y., Kim, K. R., Lee, S. Y., & Lee, E. (2010). Attribution bias in ultra-high risk for psychosis and first-episode schizophrenia. *Schizophrenia Research, 118*, 54-61.

Bentall, R. P., & Kaney, S. (1989). Emotional Stroop performance predicts disorganization in schizophrenia. *Schizophrenia Research, 77*, 141-149.

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 and Emotion, 24*, 561-595. doi:10.1080/02699930903132496

Combs, D. R., Penn, D. L., Michael, C. O., Basso, M. R., Wiedeman, R., Siebenmorgan, M., ... & Chapman, D. (2009). Perceptions of hostility by persons with and without persecutory delusions. *Cognitive Neuropsychiatry, 14*, 30-52. doi: 10.1080/13546800902732970.

Combs, D. R., Penn, D. L., Wicher, M., & Waldheter, E. (2007). The Ambiguous Intentions Hostility Questionnaire (AIHQ): a new measure for evaluating hostile social-cognitive biases in paranoia. *Cognitive Neuropsychiatry, 12*, 128-143.

Cristea, I.A., Kok, R.N., & Cuijpers, P. (2015). The efficacy of cognitive bias modification interventions in anxiety and depression: Meta-analysis. *British Journal of Psychiatry, 206*, 7–16. doi:10.1192/bjp.bp.114.146761

Eysenck, M. W., Mogg K., May J., Richards A., & Mathews A. (1991). Bias in interpretation of ambiguous sentences related to threat in anxiety. *Journal of*

Abnormal Psychology, 100, 144-150. doi:10.1037/0021-843X.100.2.144

Fenigstein, A., & Vanable, P. A. (1992). Paranoia and Self-Consciousness. *Journal of Personality and Social Psychology, 62*, 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*: 64. doi:10.1186/1471-244X-12-64

Freeman, D. (2007). Suspicious minds: The psychology of persecutory delusions. *Clinical Psychology Review, 27*, 425-457. doi:10.1016/j.cpr.2006.10.004

Freeman, D., Dunn, G., Garety, P. A., Bebbington, P., Slater, M.,Kuipers, E., . . . Ray, K. (2005). The psychology of persecutory ideation. II—A virtual reality experimental study. *Journal of Nervous and Mental Disease, 193*, 309–315.

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

Freeman, D., Pugh, K., Slater, M., Gittins, M., Dunn, G., Kuipers, E., . . . Garety, P. A. (2008). Virtual reality study of paranoid thinking in the general population. *British Journal of Psychiatry, 192*, 258–263. doi: 10.1192/bjp.bp.107.044677

Garety, P.A. & Freeman, D. (2013). The past and future of delusions research: from the inexplicable to the treatable. *The British Journal of Psychiatry, 203,* 327-333. doi: 10.1192/bjp.bp.113.126953

Garety, P.A. & Freeman, D. (1999). Cognitive approaches to delusions: A critical review of theories and evidence. *British Journal of Clinical Psychology, 38*, 113-154.

doi:10.1348/014466599162700

Garety, P., Joyce, E., Jolley, S., Emsley, R., Waller, H., Kuipers, E., … Freeman, D.

(2013). Neuropsychological functioning and jumping to conclusions in delusions.

Schizophrenia Research, 150, 570-574. doi:10.1016/j.schres.2013.08.035

Green, C. E. L., Freeman, D., Kuipers, E., Bebbington, P., Fowler, D., Dunn, G., &

Garety, P. A. (2008). Measuring ideas of persecution and social reference: the

Green et al. Paranoid Thought Scales (GPTS). *Psychological Medicine, 38*, 101-

111. doi:10.1017/S0033291707001638

Green, M. J., Williams, L. M., & Davidson, D. (2003). Visual scanpaths to threat-related

faces in deluded schizophrenia. *Psychiatry Research, 119*, 271-285.

doi:10.1016/S0165-1781(03)00129-X

Holmes, E. A., Mathews, A., Dalgleish, T., & Mackintosh B. (2006). Positive

interpretation training: effects of mental imagery versus verbal training on

positive mood. *Behaviour Therapy, 37*, 237-247. doi:10.1016/j.beth.2006.02.002

Hutton, P., & Taylor, P.J. (2014). Cognitive behavioural therapy for psychosis

prevention: A systematic review and meta-analysis. *Psychological Medicine, 44*,

449-468. doi: 10.1017/S0033291713000354

Huq, S., Garety P., & Hemsley D. (1988). Probabilistic judgments in deluded and non-

deluded subjects. *Quarterly Journal of Experimental Psychology: Human*

Experimental Psychology, 40, 801-812. http://dx.doi.org/

doi:10.1080/14640748808402300

Kaney, S., Wolfenden, M., Dewey, M.E., & Bentall, R.P. (1992). Persecutory delusions

and the recall of threatening and non-threatening propositions. *British Journal of*

Clinical Psychology, 31, 85-87.

Kay, S.R., Fiszbein A., & Opler L.A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bulletin, 13,* 261-276. doi:10.1093/schbul/13.2.261

Kinderman, P., & Bentall, R.P. (1996). Self-discrepancies and persecutory delusions: Evidence for a defensive model of paranoid ideation. *Journal of Abnormal Psychology, 105*, 106-114.

Lee, J.-S., Mathews, A., Shergill, S., Yiu 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

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

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

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. *Behaviour Therapy, 37*, 209-222. doi:10.1016/j.beth.2006.03.001

MacLeod, C., Campbell, E., Rutherford, L., & Wilson, E. (2004). The causal status of

anxiety-linked attentional and interpretive bias. In J. Yiend (Ed.), Cognition,

emotion and psychopathology: Theoretical, empirical and clinical directions (pp.

172–189). New York, NY: Cambridge University Press.

doi:10.1017/CBO9780511521263.010

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders.

Journal of Abnormal Psychology, 95, 15-20. doi:10.1037/0021-843X.95.1.15

MacLeod, C. (2012). Cognitive bias modification procedures in the management of

mental disorders. *Current Opinion in Psychiatry, 25*, 114–120.

doi:10.1097/YCO.0b013e32834fda4a.

Mathews, A. (2012). Effects of modifying the interpretation of emotional ambiguity.

Journal of Cognitive Psychology, 24, 92-105.

doi:10.1080/20445911.2011.584527

Mathews, A. (1990). Why worry? The cognitive function of anxiety. *Behavior Research

and Therapy, 28*, 455-468. doi:10.1016/0005-7967(90)90132-3

Mathews, A. & Mackintosh, B. (2000). Induced emotional interpretation bias and

anxiety. *Journal of Abnormal Psychology, 109*, 602-615. doi:10.1037/0021-

843X.109.4.602

Mathews, A. & Mackintosh, B. (1998). A cognitive model of selective processing in

anxiety. *Cognitive Therapy and Research, 22*, 539-560.

doi:10.1023/A:1018738019346

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

interpretational bias reduces trait anxiety. *Journal of Behaviour Therapy and

Experimental Psychiatry, 38*, 225-236. doi:10.1016/j.jbtep.2006.10.011

McManus, F., Van Doorn, K., & Yiend, J. (2011). Examining the effects of thought

    records and behavioral experiments in instigating belief change. Journal of

    Behavior Therapy and Experimental Psychiatry, 43, 540-547. doi:

    10.1016/j.jbtep.2011.07.003

Menon, M, Mizrahi, R., & Kapur, S., (2008). 'Jumping to conclusions' and delusions in

    psychosis: Relationship and response to treatment. *Schizophrenia Research, 98*,

    225-231. doi:10.1016/j.schres.2007.08.021

Moritz, S., Andreou C., Schneider B.C., Wittekind, C.E., Menon, M., Balzan, R.P., &

    Woodward, T.S. (2014). Sowing the seeds of doubt: A narrative review on

    metacognitive training in schizophrenia. *Clinical Psychology Review, 34*, 358-

    366. doi: 10.1016/j.cpr.2014.04.004

Moritz, S., & Laudan, A. (2007). Attention bias for paranoia-relevant visual stimuli in

    schizophrenia. *Cognitive Neuropsychiatry, 12*, 381-390.

Moritz, S., & Woodward, T. S. (2010). Metacognition Training for Schizophrenia (MCT)

    Manual (4th ed). Metacognition Study Group. Hamburg: VanHam Campus.

Moritz, S., & Woodward, T. S. (2006). Specificity of a generalized bias against

    disconfirmatory evidence (BADE) to schizophrenia. *Psychiatry Research, 142*,

    157–165. doi:10.1016/j.psychres.2005.08.016

Moritz, S., Veckenstedt, R., Randjbar, S., Hottenrott, B., Woodward, T. S., Eckstaedt, F.

    V., ... & Lincoln, T. M. (2009). Decision making under uncertainty and mood

    induction: further evidence for liberal acceptance in schizophrenia. *Psychological

    medicine, 39*, 1821-1829. doi: 10.1017/S0033291709005923.

O'Carroll, R. (2000). Cognitive impairment in schizophrenia. *Advances in Psychiatric*

Treatment, 6, 161-168.

Penn, D.L., Roberts, D.L., Combs, D., & Sterne, A. (2007). Best practices: The

development of the Social Cognition and Interation Traiing program for

schizophrenia specturm disorders. *Psychiatric services (Washington, DC), 58*,

449-451.

Perala, J., Suvisaari, J. & Saarni, S.I. (2007). Lifetime prevalence of psychotic and

bipolar I disorders in a general population. *Archives of General Psychiatry*, *64*, 19-

28. doi:10.1192/apt.6.3.161

Peters, E., Joseph, S. A., & Garety, P. A. (1999). Measurement of delusional ideation in

the normal population: Introducing the PDI (Peters et al. Delusions Inventory).

Schizophrenia Bulletin, 25, 553-576.

Phillips, M. L., Senior, C., & David, A. S. (2000). Perception of threat in schizophrenics

with persecutory delusions: an investigation using visual scan paths. *Psychological

Medicine, 30*, 157-167. doi:10.1017/S0033291799001397

Ross, K., Freeman, D., Dunn G., & Garety, P. A. (2011). A randomized experimental

investigation of reasoning training for people with delusions. *Schizophrenia

Bulletin Advance Access, 37*, 324-333. doi:10.1093/schbul/sbn165

Rude, S. S., Valdez, C. R., Odom, S., & Ebrahimi, A. (2003). Negative cognitive biases

predict subsequent depression. *Cognitive Therapy and Research, 27*, 415-429.

doi:10.1023/A:1025472413805

Savulich, G., Edwards, A., Guven, H., Shergill, S. & Yiend, J. (2015). Cognitive bias

modification of interpretations in paranoia. Manuscript in preparation.

Savulich, G., Freeman, D., Shergill, S., & Yiend, J. (2015). Interpretation biases in

paranoia. *Behavior Therapy, 46*, 110-124. doi:10.1016/j.beth.2014.08.002

Savulich, G., Shergill, S., & Yiend, J. (2012). Biased cognition in psychosis. *Journal of Experimental Psychopathology, 3*, 514-536. http://dx.doi.org/10.5127/jep.016711

Schizophrenia Commission. *The Abandoned Illness: A Report by the Schizophrenia Commission*. Rethink Mental Illness, 2012.

Sheehan, D.V., Lecrubier, Y., & Sheehan, K.H., Amorim, P., Janavas, J., Weller, E., … Dunbar, G.C. (2006). The Mini-International Neuropsychiatric Interview (M.IN.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry, 59,* 22-33. doi:10.1016/S0924-9338(97)83296-8

Teasdale, J. D. (1988). Cognitive vulnerability to persistent depression. *Cognition and Emotion, 2*, 247-274. doi:10.1080/02699938808410927

van der Gaag, M., 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*, 30-37. doi:10.1016/j.schres.2014.03.016

Wechsler, D. (2001). *Wechsler test of adult reading (WTAR).* New York: Psychological Incorporation.

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

Wykes, T., Huddy, V., Cellard, C., McGurk, S. R. & Czobor, P. (2011). A meta-analysis of cognitive remediation for schizophrenia: Methodology and effect Sizes.

American Journal of Psychiatry, 168, 472-485.

doi:10.1176/appi.ajp.2010.10060855

Yiend, J. (2010). The effects of emotion on attention: A review of attentional processing

of emotional information. *Cognitive and Emotion, 24*, 3-47.

doi:10.1080/02699930903205698

Yiend, J., Lee, J-S., Mathews, A., Tekes, S., Atkins. L., Ferragamo, C., … Shergill, S.

(2014). Modifying interpretation in a clinically depressed sample using 'cognitive

bias modification-errors': A double blind randomised controlled trial. *Cognitive

Therapy and Research*, *38*, 146-159. doi:10.1007/s10608-013-9571-y

Yiend, J., & Mackintosh, B. (2004). The experimental modification of processing biases.

In J. Yiend. (Ed.) Cognition, Emotion and Psychopathology (pp. 190-210).

Cambridge, UK: Cambridge University Press. Pp 190-210.

doi:10.1017/CBO9780511521263.011

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

experimentally induced biases in interpretation. *Behavior Research and Therapy,

43,* 779-797. doi:10.1016/j.brat.2004.06.007

Yiend, J., Mathews, A., Burns, T., Dutton, K., Fernandez-Martin, A., Georgiou, G., …

Fox, E. (2015). Mechanisms of selective attention in generalized anxiety disorder.

Clinical Psychological Science, 3, 758-771. doi:10.1177/2167702614545216

Yiend, J., Parnes, C., Shepherd, K., Roche, M., & Copper, M. J. (2014). Negative self-

beliefs in eating disorders: A cognitive-bias-modification study. *Clinical

Psychological Science*, *26*, 756-766. doi:10.1177/2167702614528163

Yiend, J., Savulich, G., Coughtrey, A., & Shafran, R. (2011). Biased interpretation in

perfectionism and its modification. *Behavior Research and Therapy, 49*, 892–900.

doi:10.1016/j.brat.2011.10.004

Zigmond, A.S., & Snaith, R.P. (1983). The Hospital Anxiety and Depression Scale. *Acta*

Psychiatrica Scandanavia, 67, 361-70. doi:10.1111/j.1600-0447.1983.tb09716.x

Table 1. Group demographics, clinical profile, mood state and personality trait measures

Measures	Paranoid patients $n = 32$	Non-paranoid patients $n = 29$	Healthy controls $n = 29$	Statistic, p value
Age	42.75 (7.72)	39.59 (8.80)	37.41 (15.51)	$F(2,87) = 1.79, .17$
Gender (% male)	59.38	65.51	44.82	$X^2 = 2.68, .26$
WTAR	40.40 (10.16)	40.93 (8.95)	42.79 (6.16)	$F(2,85) = .63, .54$
PANSS				
P6 Persecution/Suspiciousness	3.45 (0.88)	1.55 (0.63)		$t(58) = 9.49, < .001$
Positive Symptoms	15.90 (4.05)	11.90 (4.69)		$t(58) = 3.55, .001$
Negative Symptoms	15.19 (6.32)	12.59 (4.10)		$t(58) = 1.88, .065$
General Symptoms	30.45 (7.37)	25.62 (6.49)		$t(58) = 2.69, .009$
Total Symptoms	61.55 (15.18)	50.10 (12.15)		$t(58) = 3.21, .002$
Green Paranoid Thoughts	90.03 (40.00)	58.41 (24.52)	38.90 (7.33)	$F(2,87) = 26.14, < .001$
Paranoia Scale	52.41 (20.36)	42.03 (17.13)	27.45 (7.21)	$F(2,87) = 18.39, < .001$
PDI-21total	109.53 (61.84)	64.18 (64.41)	21.38 (23.56)	$F(2,80) = 21.41, < .001$
HADS total	18.27 (9.54)	13.04 (7.30)	5.07 (3.67)	$F(2,76) = 23.89, p < .001$

Notes: WTAR: Welscher Test of Adult Reading; PANSS: Positive and Negative Syndrome Scale, GPTS: Green Paranoid Thoughts

Scale; PDI: Peter's Delusions Inventory; HADS: Hospital Anxiety/Depression Scale

Table 2. (a) Correlations between bias measures for each task (SRT: Similarity Ratings task, SST: Scrambled Sentences task, JTC: Jumping to Conclusions tasks) across all participants and (b) correlations between bias measures for each task and the Paranoia Scale, Green Paranoid Thoughts Scale and Peter's Delusion Inventory across patients.

(a) *n* = 90	SRT Negative Index	SST Paranoid %	SST Negative %	JTC Beads
SRT Paranoid Index	.36, .001*	.58, < .001*	.42, < .001*	-.07, .56
SRT Negative Index		.23, .046*	.13, .26	-1.76, .12
SST Paranoid %			.70, < .001*	-.10, .35
SST Negative %				.02, .89

| (b) *n* = 61	Paranoia Scale	Green Paranoid Thoughts	PDI-21
SRT Paranoid Index	.39, < .001*	.34, .002*	.38, .001*
SRT Negative Index	.30, .006*	.22, .046*	.27, .02*
SST Paranoid %	.55, < .001*	.59, < .001*	.74, < .001*
SST Negative %	.55, < .001*	.61, < .001	.60, < .001*
JTC Beads	-.17, .11	-.15, .16	-.22, .046*

Notes: Values are Pearson's *r*, *p* value

Table 3. Mean similarity ratings for each sentence type on the Similarity Ratings task and percentage of unscrambled sentences for each sentence type on the Scrambled Sentences task

		Paranoid Content				**Valenced Content**			
		Targets		Foils		Targets		Foils	
		Paranoid direction	Non-paranoid direction	Paranoid direction	Non-paranoid direction	Negative direction	Positive direction	Negative direction	Positive direction
Similarity Ratings (mean item rating on 1-4 scale)	Paranoid patients	2.21 (.46)	2.37 (.43)	1.81 (.47)	1.82 (.47)	2.16 (.47)	2.51 (.54)	1.78 (.42)	2.23 (.61)
	Non-paranoid patients	2.02 (.48)	2.38 (.48)	1.80 (.49)	1.93 (.46)	2.18 (.49)	2.45 (.50)	1.75 (.55)	2.43 (.54)
	Healthy controls	1.73 (.27)	2.72 (.40)	1.55 (.23)	1.59 (.43)	1.97 (.39)	2.84 (.40)	1.33 (.26)	2.21 (.65)
		Paranoid meaning				Negative meaning			
Scrambled Sentences (% of total attempted sentences)	Paranoid patients	48.96% (24.84)				36.40% (26.47)			
	Non-paranoid patients	25.09% (25.26)				20.04% (18.87)			
	Healthy controls	9.33% (11.46)				12.54% (11.67)			

Figure 1. Patterns of interpretation bias on the Similarity Ratings task across groups

a) Pattern of results for material relevant to paranoid concerns only. Values are mean ratings of similarity between disambiguating target sentences (reflecting one or other possible interpretation) and the original ambiguous passage and reflect the significant Stimulus Direction x Group interaction reported in the text. Higher ratings reflect more paranoid/non-paranoid interpretations. * denotes significant difference at $p<.05$
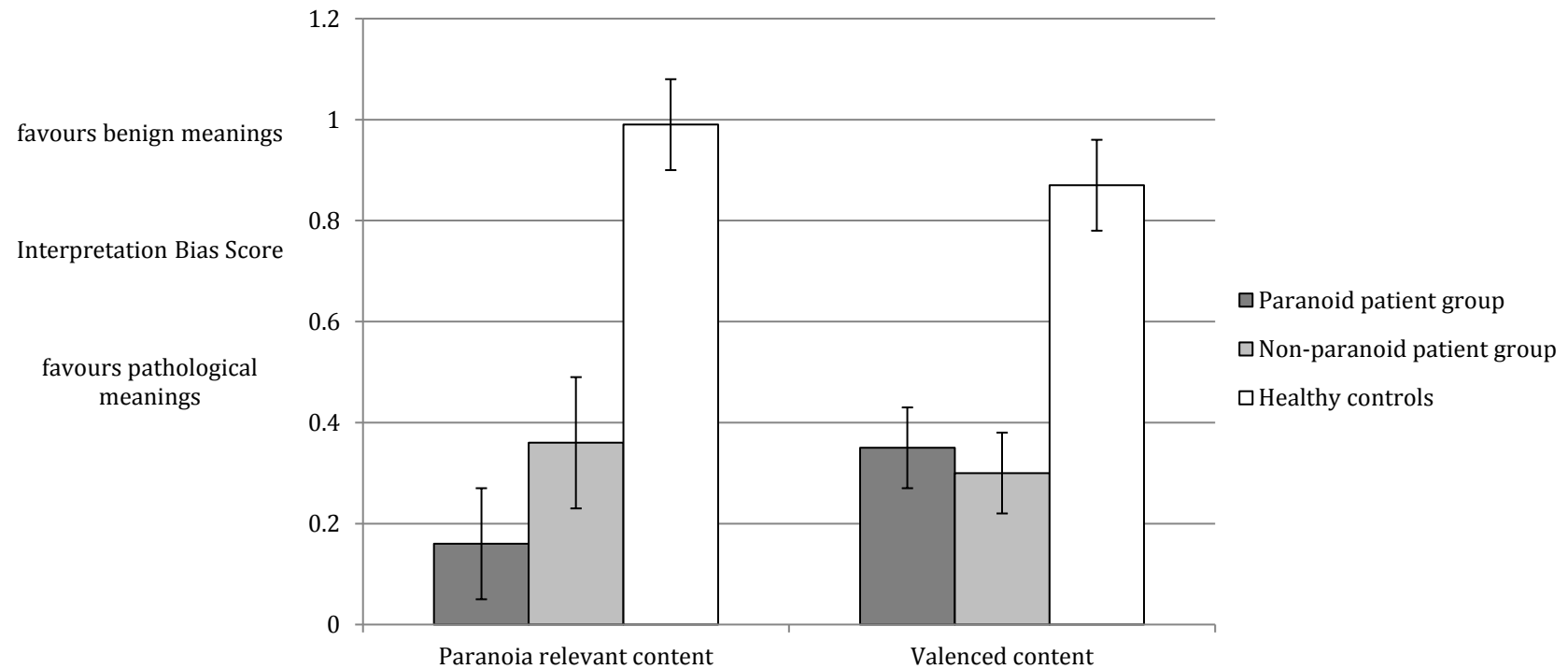
b) Pattern of results across both types of material (paranoia relevant and general negative/positive valence) for target items only. Values are 'interpretation bias scores' which, unlike raw similarity rating scores, reflect the overall degree of bias by taking into account both directions of endorsement (paranoid or negative and non-paranoid or positive). Index score = magnitude of (mean paranoid/negative target rating) – (mean non-paranoid/positive target rating)
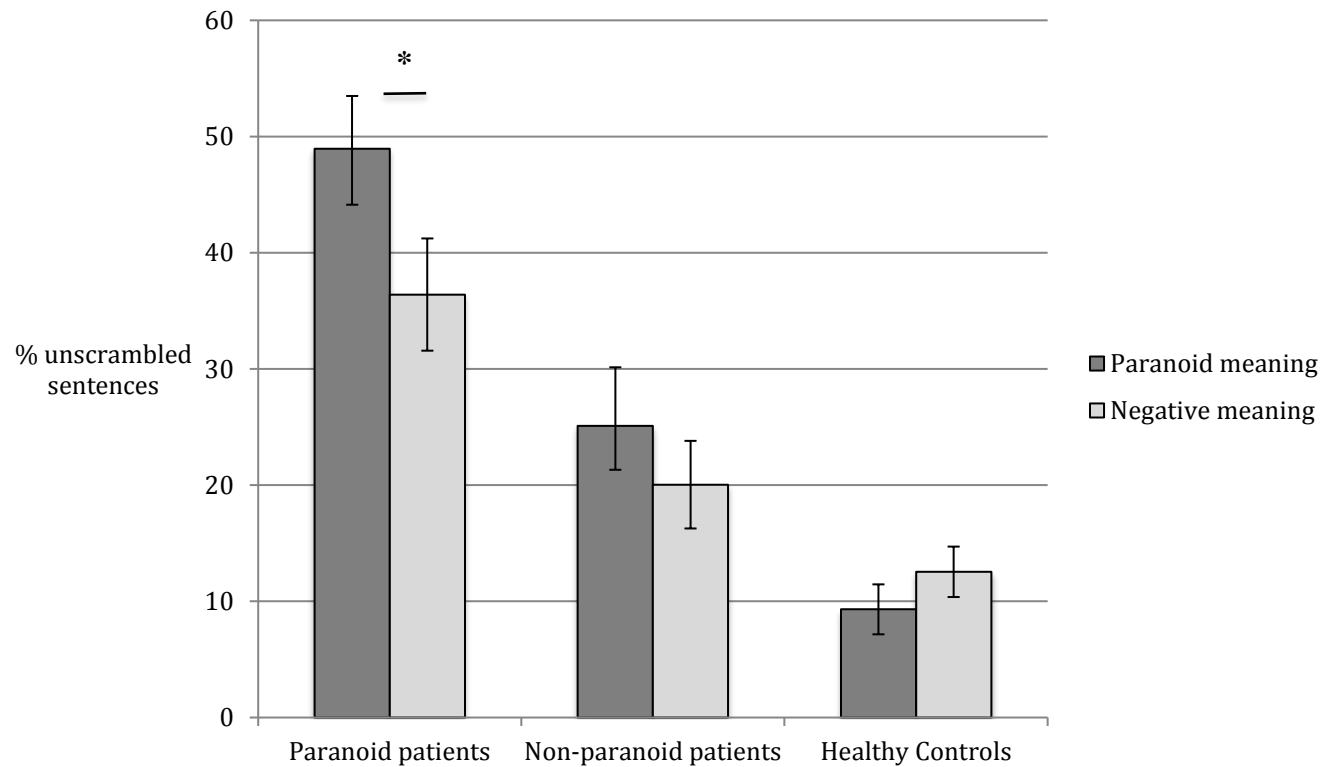
Figure 2 Patterns of interpretation bias on the Scrambled Sentences task across groups. Values are percentages of sentences unscrambled as paranoid/negative. Paranoid patients made significantly more paranoid interpretations than negative interpretations, reflecting content specificity for paranoia relevant information; Neither the non-paranoid patients, nor control groups, differed in the percentage of paranoid and negative interpretations made