King's Research Portal

DOI:

[10.1016/j.jbtep.2020.101575](https://doi.org/10.1016/j.jbtep.2020.101575)

Document Version

Peer reviewed version

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

Citation for published version (APA):

Savulich, G., Edwards, A., Assadi, S., Guven, H., Leathers-smith, E., Shergill, S., & Yiend, J. (2020). Biased interpretation in paranoia and its modification. *Journal of Behavior Therapy and Experimental Psychiatry*, 69, 101575. [101575]. <https://doi.org/10.1016/j.jbtep.2020.101575>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Biased interpretation in paranoia and its modification

George Savulich^{1*}, Annabel Edwards², Sara Assadi², Husniye Guven², Emily
Leathers-Smith², Sukhi Shergill² and Jenny Yiend^{2*}

¹ Department of Psychiatry and Behavioural and Clinical Neuroscience Institute,
University of Cambridge, School of Clinical Medicine

² Department of Psychosis Studies, Institute of Psychiatry, Psychology, &
Neuroscience, King's College London

Corresponding authors:

*Dr Jenny Yiend, Department of Psychosis Studies, Institute of Psychiatry,
Psychology & Neuroscience, PO 63, De Crespigny Park, London SE5 8AF, UK;
jenny.yiend@kcl.ac.uk, Tel: +44 020 7848 5119; Dr George Savulich, Department of
Psychiatry, Herchel Smith Building for Brain and Mind Sciences, Forvie Site,
Robinson Way, Cambridge, CB2 0SZ; gjs46@medschl.cam.ac.uk; Tel: 01223 336798

Abstract

Background and objectives: Cognitive models of psychosis implicate interpretation biases as one of the mechanisms involved in the formation and maintenance of symptoms. First we measured the strength of association between interpretation biases and psychosis-relevant traits. Next we manipulated these biases and quantified the effects of doing so on psychosis-relevant outcomes.

Methods: Experiment 1 used two measures of interpretation bias in a healthy sample ($n = 70$). Experiment 2 used a novel cognitive bias modification procedure (CBM-pa) in individuals with moderate trait paranoia ($n = 60$).

Results: Experiment 1 revealed that over a third of the variance in interpretation bias could be explained by the combined effect of trait measures of paranoia/psychosis. In Experiment 2, CBM-pa produced training-congruent changes in the interpretation of new ambiguous information and influenced the interpretation, attribution and distress associated with a real-life social event.

Limitations: The potentially confounding effects of elevated anxiety and depression on interpretation bias and the restricted range of outcome measures to assess the wider effects of CBM-pa.

Conclusions: These studies are consistent with interpretation biases contributing to the maintenance of paranoia. CBM-pa could next be adapted and evaluated to test its efficacy as a therapeutic intervention.

Word count: Abstract: 197; **Main text:** 6,261

Key words: paranoia; interpretation bias; cognitive bias modification; information processing

1. Introduction

Interpretation bias can be defined as a consistent tendency to interpret emotionally ambiguous stimuli, situations, or events in a negative (or positive) manner (Blanchette & Richards, 2010). In two experiments we attempted to, first, characterise and, then, manipulate interpretation bias, one of the putative cognitive mechanisms underlying paranoid thoughts and beliefs (Savulich, Shergill, & Yiend, 2012). Interpretation bias represents an explicit plausible mechanism for symptom maintenance across a range of disorders. Individuals with psychopathology or elevated trait vulnerability to psychopathology tend to make more negative interpretations of emotionally ambiguous information than controls or those with lower vulnerability. For example, upon reading the ambiguous sentence ‘the doctor examined little Emily’s growth’ anxious people are more likely to interpret this to mean that Emily had cancer, rather than that she was having a routine measurement of her height. Repeatedly interpreting emotional ambiguity in this manner can result in overexposure to negative information which in turn reinforces the corresponding mood and can contribute to maintaining symptoms and or/symptomatic beliefs. However, ambiguous information permitting broadly negative interpretations (such as in the example above) has less direct relevance to disorders like psychosis, where a predominant symptom is often paranoia.

For paranoid symptoms a related form of emotional ambiguity involving different and more directly relevant content is likely to be more important. For example, a stranger’s stare is emotionally ambiguous and could be interpreted in a paranoid manner, as malicious, or a benign manner, as curious. Likewise the whispers of others could be interpreted as conspiratorial (a paranoid interpretation) or playful (a benign interpretation). Paranoid interpretations such as these examples are likely to

lead to an artificially increased perception of danger of harm to the self, which in turn will reinforce paranoid beliefs, maintain symptoms and increase distress about being at risk of observation by or harm from others. This in turn will promote further biased processing and a cycle of reciprocal causation would ensue (Yiend & Mackintosh, 2004).

Theoretical models of psychosis already implicate information-processing biases in the formation and maintenance of positive symptoms (Garety, Kuipers, Fowler, Freeman, & Bebbington, 2001; Freeman, Garety, Kuipers, Fowler, & Bebbington, 2002). These models propose that persecutory delusions are supported by the interaction between emotional changes, anomalous experiences and cognitive biases associated with psychosis that confirm threat beliefs. Much psychosis research has focused on reasoning biases, whereby patients with delusions draw inferences on the basis of little information (e.g. ‘jumping to conclusions’; JTC) (Garety & Freeman, 1999), biases against disconfirmatory information (Moritz & Woodward, 2006) and attributional biases (Bentall, Corcoran, Howard, Blackwood, & Kinderman, 2001).

Fewer studies have investigated the interpretation of emotional ambiguity in the context of paranoia. A previous study from our laboratory has shown that healthy individuals with high trait paranoia interpret ambiguous information in a negative manner, an effect that was more pronounced for items specifically matching paranoid concerns (Savulich, Freeman, Shergill, & Yiend, 2015). Another study revealed negatively biased interpretations of emotional ambiguity in patients with schizophrenia compared to healthy controls, with some evidence showing that paranoid interpretation bias was stronger than negative interpretation bias in patients with symptoms of paranoia (Savulich, Shergill, & Yiend, 2017). Most recently, negative interpretation bias has been shown to precede the onset of psychosis (Yiend

et al., 2018). Finally, a meta-analysis found that a stronger negative bias is associated with both the presence and a greater severity of paranoia in clinical and subclinical groups (see Trotta et al., 2019).

This body of work is now sufficient to warrant investigating, in psychosis, interventions specifically designed to target pre-existing biased interpretation mechanisms, such as cognitive bias modification for interpretation (CBM-I; Mathews & Mackintosh, 2000). When applied to the affective disorders, CBM-I has established beneficial effects, such as symptom reduction, although further work is needed to understand and improve variations in efficacy (Jones & Sharpe, 2017). Here we aimed to, first, provide additional evidence of paranoia-relevant biases in interpretation (Experiment 1) and, second, to extend these by developing and testing a new version of CBM-I designed to be suitable to the content of paranoid interpretations (Experiment 2).

1a. Experiment 1

The aim of Experiment 1 was to examine paranoia-relevant interpretation bias in individuals with elevated vulnerability to trait paranoia in an attempt to further establish the reliability of these biases in a subclinical population. We used two well established, reliable and valid cognitive experimental measures of interpretation bias using ambiguous materials which had the potential to elicit global paranoid interpretations. Similar to our previous study (Savulich et al., 2015), we hypothesized that the combined effect of trait paranoia/psychosis measures (i.e. paranoid thinking, delusions, anomalous perceptions) would predict around a third of the variance in paranoid interpretation bias.

2a. Method

2a.1. Participants

Participants were recruited from advertisements within King's College London and the local community. Inclusion criteria were fluency in English, not currently taking any psychiatric medication or receiving psychological treatment and not having a past or current psychiatric diagnosis. Participants were screened on the latter criteria using the Mini-International Neuropsychiatric Interview (Lecrubier et al., 1997). Those meeting criteria were then asked to complete the Green Paranoid Thoughts Scale (GPTS; Green et al., 2008). Participants were selected based on their total GPTS score; frequency plots were examined after every ten participants to ensure a wide distribution of scores. Selected participants were invited to attend one experimental session at the Institute of Psychiatry, Psychology and Neuroscience within three days of returning the GPTS.

2a.2. Trait Measures

Measures were selected from our previous study (Savulich et al., 2015) to replicate sample characteristics and improve construct validity when capturing paranoia in the general population. These are described in detail elsewhere (Savulich et al., 2015; Savulich et al., 2018); therefore only outline information is repeated here. Measures of paranoia were the GPTS (Green et al., 2008) and the Paranoia Scale (PS; Fenigstein & Vanable, 1992). Measures of delusions and anomalous perceptions were Peter's Delusions Inventory (PDI; Peters, Joseph, & Garety, 1999) and the Cardiff Anomalous Perceptions Scale (CAPS; Bell et al., 2006). Measures of anxiety and depression included Spielberger's Trait Anxiety Inventory (STAI-Trait; Spielberger, Gorsuch, & Lushene, 1979) and Beck's Depression Inventory (BDI-II; Beck et al.,

1996). Cognitive flexibility was assessed using the Cognitive Flexibility Scale (CFS; Martin & Rubin, 1995).

2a.3. Interpretation Bias Measures

2a.3.1. Similarity Rating Task (SRT)

The SRT was adapted from Mathews & Mackintosh (2000) and Eysenck, Mogg, May, Richards, & Mathews (1991) and developed with paranoia-relevant content. The task is comprised of 20 items reflecting thoughts of persecution (e.g. intent of harm by another person; ‘*someone wants to hurt me*’) and 20 items reflecting ideas of reference (i.e. relating ordinary actions/events to the self in a hostile or personally meaningful manner; ‘*people are watching me*’)¹. The task first involved an ‘encoding’ stage in which participants read three-sentence passages of emotionally ambiguous text, interpretable in either a paranoid or non-paranoid way. In a later ‘test’ phase interpretations were measured (indirectly to reduce the influence of any strategic control strategies) by showing sentences that participants rated according to

¹For both tasks (SRT and SST, below), items were generated by the research team and revised with clinician feedback. Persecutory items for both tasks were previously reported and validated by Savulich and colleagues (2015). Reference items were developed for this study; for the SRT, pilot work (independent volunteers rated each target/foil sentence on a 7-point Likert scale [7 being the most paranoid] indicated that the main effect of Sentence type was not significant, $p = .7$ (target and foil sentences obtained similar ratings), but that the main effect of Stimulus Direction was, $p < .001$ (i.e. paranoia and non-paranoia were well differentiated).

how similar they were to the earlier passage. Sentences were unambiguous and reflected either paranoid or non-paranoid meanings of the passage. Interpretation bias was inferred from systematically higher similarity ratings for one type of sentence over another. At encoding participants were instructed to read each passage and complete a word fragment in the final sentence by indicating the first missing letter, followed by a neutral comprehension question (yes/no response, with feedback). For example, a passage entitled ‘Canteen Lunch’ read: ‘After a long morning you enter the canteen for the 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 the comprehension question ‘Did you enter the canteen for lunch?’ (Correct answer: yes).

During the test phase two target (i.e. one paranoid and one non-paranoid disambiguation) and two foil (i.e. one unrelated paranoid and one unrelated non-paranoid sentence to act as control items) sentences were individually randomly presented beneath the title of the corresponding passage. For the example above test items 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). Participants rated similarity of each sentence to the original passage using a four-point scale (1 = very different in meaning to 4 = very similar in meaning). Forty encoding passages were presented in random order, followed by 160 test sentences. For each passage, the associated title was followed by the appropriate ‘package’ of test sentences (two targets and two foils), and the test sentences appeared at random within this package. Duration of the task was approximately 25 minutes. This task

showed moderate reliability in the present sample (Cronbach's alpha = .63).

2a.3.2. Scrambled Sentences Task (SST)

The SST (Wenzlaff, 1993) is a robust measure of interpretation bias that requires participants to reorder five out of six mixed up (scrambled) words to create grammatically correct sentences. Participants were instructed to re-order two blocks of 20-word strings (ten of persecutory and ten of referent content²). Each word string can be unscrambled to create sentences of either paranoid or non-paranoid meanings. For example, the 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). Participants were given four minutes to complete as many items as possible. They performed the task under cognitive load (recalling a six-digit number after the task) to reduce the influence of any effortful control or suppression of interpretation biases (Rude, Valdez, Odom, & Ebrahimi, 2003). Duration of the task was approximately eight minutes. This task showed moderate reliability in the present sample (Cronbach's alpha = .64).

2a.4. Procedure

This study received ethical approval from King's College London Psychiatry, Nursing, and Midwifery Research Ethics Committee (Ref: PNM 09/10-34). Participants completed the SRT and SST (delivered in counterbalanced order and prior to questionnaires to minimise paranoia-related priming effects) followed by the

²A paired samples t-test confirmed that ratings of reference sentences significantly differed from ratings of non-reference sentences (i.e. were more paranoid), $p = .04$.

trait measures. All participants provided written informed consent.

3a. Results

3a.1. Regression Analyses

3a.1.1. Participants

Participant characteristics for the entire sample ($n = 70$) are presented in Table 1.

3a.1.2. Analytical approach

Experiment 1 was intended to provide additional conceptual support of Savulich et al. (2015), who examined the prediction of trait variables, including measures of paranoia, on variance in interpretation bias. In doing this we were taking the first step in the investigation of a new, putative bias - psychopathology association between paranoia and interpretation of ambiguous (potentially paranoid) material. Typically this first step involves using cross sectional data to establish a statistical association by examining whether bias varies according to different levels of psychopathology. This approach deliberately avoids making any assumptions about the causal direction of the association. Index of paranoid interpretation bias (calculations for each task below) that significantly correlated with psychosis-relevant trait measures ($p < .05$) were entered as predictor variables in a multiple regression analysis with bias index as the dependent variable. In line with the recommendations of Miller and Chapman (2011), we did not control for anxiety and depression in order not to remove variance associated with paranoia. As Miller and Chapman (2001) point out, attempts to ‘control for’ such variables statistically are invalid because doing so removes too much of the independent variable of interest (here paranoia)

because the confounders (here anxiety and depression) are conceptually and empirically too closely associated with it. According to their influential paper, removing variance associated with conceptually similar confounders would, in effect, ‘corrupt the variable [of interest] itself.’ (p 44). Correlational analyses confirmed that these variables were associated (STAI-Trait: with PS: $r = .59, p < .001$; with GPTS: $r = .45, p < .001$; with PDI: $r = .31, p = .009$; and BDI: with PS: $r = .70, p < .001$; with GPTS: $r = .60, p < .001$; with PDI: $r = .41, p = .001$).

3a.1.3. SRT

Following previous studies, an index of paranoid interpretation bias was calculated by subtracting the mean non-paranoid target rating from the mean paranoid target rating (Yiend et al., 2008; Casey, Rogers, Burns, & Yiend, 2013; Mathews, Yiend, & Lawrence, 2004), with a more positive score indicating more paranoid bias (range +3 to -3). Bias index was significantly associated with the GPTS ($r = .52, p < .001$), PS ($r = .57, p < .001$), PDI ($r = .45, p < .001$) and CAPS ($r = .33, p = .005$). Regression results are shown in Table 2. The model accounted for 38.0% of the variance in paranoid interpretation bias and was significant³ $F(4,65) = 9.95, p < .001$. Trait paranoia (PS) made a significant independent contribution to predicting paranoid interpretation bias (Table 2).

3a.1.4. SST

The total number of sentences unscrambled as paranoid was divided by the total number of sentences attempted in the time allowed and multiplied by 100 to

³ In both models (SRT, SST), task order had no effect on index of paranoid interpretation bias (p 's $> .50$).

calculate a percentage bias score (Lee, Mathews, Shergill, & Yiend 2016). Sentences containing errors were included only in the denominator (total sentences attempted)⁴. A higher percentage indicates a more paranoid interpretation bias. Bias score was significantly associated with the GPTS ($r = .47, p < .001$), PS ($r = .62, p < .001$) and PDI ($r = .28, p = .018$). Regression results are shown in Table 2. The model accounted for 38.2% of the variance in paranoid interpretation bias and was significant, $F(3,66) = 13.58, p < .001$. Again, trait paranoia (PS) made an independent significant contribution to predicting paranoid interpretation bias (Table 2).

1b. Experiment 2

The aim of Experiment 2 was, firstly, to manipulate paranoid interpretation biases in order to provide a test of their aetiological significance and, secondly, to act as a proof of principle that these biases were amenable to manipulation in vulnerable individuals and therefore an appropriate target for further development as a possible clinical intervention. ‘Cognitive bias modification’ (CBM) techniques that ‘train’ participants to interpret emotionally ambiguous information in either a positive or negative direction (Mathews & Mackintosh, 2000; MacLeod, 2012; MacLeod & Holmes, 2012) are now well established in a number of disorders (see Jones & Sharpe, 2017 and Hallion & Ruscio, 2011 for reviews), but are not without criticism (Cristea, Kok, & Cuijpers, 2015). Significant outcomes following training have included

⁴Including errors and grammatically incorrect sentences that suggested a paranoid interpretation (in the numerator) did not change results from the regression analysis; the model accounted for 38.4% of the variance and was significant, $F(3,66) = 15.35, p < .001$, with trait paranoia making an independent prediction.

reductions in trait anxiety (Salemink, van den Hout, & Kindt, 2009; Mathews Ridgeway, Cook, & Yiend, 2007), social anxiety (Klein et al., 2015), anxiety sensitivity (Capron & Schmidt, 2016), negative intrusions (Grol et al., 2018) and depressed mood/symptoms (Holmes, Lang, & Shah, 2009; Watkins, Baeyens, & Read, 2009); decreased vulnerability to external stressors (Mackintosh, Mathews, Yiend, Ridgeway, & Cook, 2006; Lester, Mathews, Davison, Burgess, & Yiend, 2011); improvements in risk factors associated with body dysmorphic/eating disorders (Dietel et al., 2018; Yiend, Parnes, Shephard, Roche, & Cooper, 2014; Matheson, Wade, & Yiend, 2018; Brockmeyer et al., 2018) and behaviour change (Yiend, Savulich, Coughtrey, & Shafron, 2011; Wong & Grisham, 2017). Interpretation bias training has also been shown to change mood and to survive changes in environment and context (Salemink & van den Hout, 2010; Mackintosh et al., 2006).

We developed a novel CBM for paranoia (CBM-pa) programme that could induce interpretations identified in Experiment 1 in either a paranoid or non-paranoid direction in a healthy sample. Training in both a paranoid and non-paranoid direction is needed to avoid ceiling effects typically observed in non-clinical samples that are already positively biased at baseline. We also tested the influence of training on the experience of two stressful, ambiguous, situational events ('Stooge-laughter task'). We hypothesized that:

- 1) CBM-pa would produce training-congruent group differences in interpretation of newly presented ambiguous information; and
- 2) the paranoid direction group would report greater stress and more paranoid explanations of both events than the non-paranoid direction group, consistent with their training.

2b. Method

2b.1. Participants

Sixty participants were recruited from King's College London and the local community. Inclusion criteria were as in Experiment 1 and also included not having participated in any CBM study within the last six months. Participants were selectively recruited using the GPTS to have moderate levels of trait paranoia (scoring > 40).

2b.2. Trait and state mood measures

Paranoia and psychosis measures included the GPTS, PS, PDI and CAPS. Mood state measures of paranoia, anxiety and depression were self-reported using 10 cm visual analogue scales. Participants were asked to mark the point along the scale (from not at all to totally) the extent to which they felt each state measure.

2b.3 Experimental Tasks

2b.3.1 SRT

The SRT was as described in Experiment 1. Two parallel versions were used so that spontaneous interpretations of ambiguous text could be compared pre and post CBM-pa. Items across both versions showed moderate reliability (Cronbach's alpha = .62).

2b.3.2 Cognitive Bias Modification for Paranoia (CBM-pa)

Participants were instructed to read a total of 90 ambiguous scenarios at a self-paced rate. For each scenario, a disambiguating final word resolved the passage with either a paranoid or a non-paranoid meaning. Participants were required to indicate

the first missing letter in the final sentence of the passage and then answer a ‘yes/no’ comprehension question, with feedback.

An example of a CBM-pa item is: ‘You enter your office after returning from lunch. You immediately notice a group of colleagues crowding around one computer and talking in low voices. You think that your colleagues are being... consp-rat-rial (conspiratorial, paranoid direction training)/c-ns-derate (considerate, non-paranoid direction training, followed by the comprehension question ‘Did you just return to the office from lunch?’ (Correct answer: yes). To ensure validity of the programme’s training content, CBM-pa items were rated for their level of implied paranoia by six independent raters who were clinicians or academic researchers working in the area of paranoia and psychosis, unaware of the study aims. All items were presented to raters in one of two fixed orders, with half of the raters receiving each order. Scenarios were given with the two alternative end words (i.e. paranoid/non-paranoid interpretation). Ratings were made on a 7-point Likert scale from 1 (= ‘not paranoid at all’) to 7 (= ‘very paranoid’). A paired samples t-test showed that paranoid endings were rated significantly more paranoid than non-paranoid endings (paranoid: $M = 5.34$, $SD = .78$; non-paranoid: $M = 1.33$, $SD = .33$; $t(89) = 43.82$, $p < .001$, $d = 7.23$). The main difference between CBM-pa and the SRT is that the CBM-pa training programme consistently resolves the ambiguity of each scenario in *either* a paranoid or non-paranoid direction, whereas the SRT maintains the same level of ambiguity throughout the encoding phase, but assesses interpretation bias for *both* directions during the recognition phase. Changes in SRT target sentences from pre to post training are therefore essential to demonstrate whether or not interpretation biases have been effectively manipulated. An induced interpretation bias would be reflected by higher recognition ratings for sentences resolved in the same direction as training,

rather than the opposite direction (Yiend, Mackintosh, & Mathews, 2005).

2b.3.3 Stooge-Laughter task

The Stooge-Laughter task is a ‘real life’ situational measure known to elicit (unfounded) paranoid interpretations in the general population (Green et al., 2011). During a five minute ‘filler’ task in which participants were asked to provide a written description of their journey to the laboratory, participants were exposed to two ambiguous situational events.

2b.3.3.1 Event 1 Interruption event

Participants were asked to provide a written description of their journey to the laboratory. After beginning, the experimenter explained that they would leave the room to file the participant’s questionnaires. This allowed the experimenter to cue the stooge. After two minutes, the experimenter re-entered the room. The stooge waited for 15 seconds after the experimenter had re-entered the room and then knocked on the door. The experimenter announced ‘come in’ at which point the stooge opened the door and said, ‘Sorry to interrupt, may I have a word?’ The two then exited the room.

2b.3.3.2 Event 2 Laughter event

With both the experimenter and the stooge outside of the testing room (not visible), a recording of male laughter was then played in front of the door for 35 seconds. The experimenter then waited a further 15 seconds before re-entering the room. The laughter recording was played using a portable speaker and was identical for each participant. The laughter was piloted in four individuals to ensure that it was audible to those inside of the testing room, without sounding artificial.

After the two events, participant's interpretations were first assessed using a semi-structured interview to check if the manipulation was successful (Green et al., 2011). The experimenter apologised for the interruption and framed questions of the interview as standard procedure to ensure that the participant's performance would not be affected. Participants were then asked to rate each of three possible reasons for the interruption on a scale from 0 (= not at all the reason') to 10 (= 'completely the reason'). The three reasons given (outcomes) were it was caused by '*something about you*' (internal personal), '*something about the other person*' (external personal) or '*something about circumstances or chance*' (external situational). Conviction ('How certain are you that [selection] is the reason?') and distress ('How distressing is it that [selection] the reason?') were assessed on scales from 0 to 10 (with higher scores indicating higher levels of conviction and distress) as additional outcome measures. The same questions were repeated for laughter event, for those participants who noticed it.

2b.3.4. Procedure

Ethical approval was received from King's College London Psychiatry, Nursing, and Midwifery Research Ethics Committee (Ref: PNM 10/11-157). Selected participants provided informed written consent and completed baseline visual analogue scales. Baseline SRT was administered followed by the trait measures. Participants were assigned to either paranoid (n = 29) or non-paranoid (n = 31) CBM-pa using a computer-generated randomisation formula. Immediately following training, the second SRT, counterbalanced from baseline, was given, followed by the Stooge-Laughter task. State mood was assessed at three additional time points: immediately after CBM-pa, before the filler task and immediately after the laughter event.

3b. Results

Hypothesis-driven group contrasts were conditional on the significance of preceding relevant interactions, thereby protecting against Type 1 errors (Howell, 2010). Dependent measures met assumptions for parametric testing. A sensitivity analysis showed that our sample size gave us 80% power to detect effect sizes of $d = 0.65$ or larger between paranoid and non-paranoid trained groups, at $\alpha = .05$, suggesting large effect sizes ($d = .8$) should be detectable, whereas small ones ($d = .2$) might not be. The SRT typically produces large effect sizes (Hallion & Ruscio, 2011). Fisher's exact test was used for the Stooze-Laughter task given the low cell counts.

3b.1. Group Characteristics

Participant characteristics are shown in Table 1. Groups did not differ significantly on any of the variables measured at baseline (all p 's < 1).

3b.2 SRT

Table 3 shows the condition means for this task. A four-way mixed design ANOVA was conducted on mean similarity ratings, with factors Time (baseline, post-training) x Sentence Type (target, foil) x Stimulus Direction⁵ (pathological congruent,

⁵The factor stimulus direction refers to the *direction* of the interpretation (paranoid or non-paranoid). For example, for targets in the scenario ending '*...Across the room, two girls begin to whisper*', the paranoid direction would be '*The girls are plotting against you*' and the non-paranoid direction would be '*The girls are talking about their friend.*' For foils the paranoid direction would be '*Your doctor gives you*

non-pathological congruent) x Group (paranoid direction, non-paranoid direction). The four-way interaction did not reach significance, $F(1,58) = 1.06, p = .30$.

However, the three way Time x Stimulus Direction x Group interaction was significant, $F(1,58) = 5.03, p = .029, \eta_p^2 = .08$. Interactions like this one, *without* the Sentence Type (targets versus foils) factor indicate effects which are indistinguishable across interpretation and response bias, but may nevertheless be consistent with hypotheses. There is a well-justified practice within the interpretation bias training literature using this task, which has discussed this issue and established this practice (e.g. LeMoult et al., 2018; Hoppitt et al., 2014; Yiend et al., 2005). We therefore followed up this significant three-way interaction, as recommended by Howell (2010) using two-way ANOVAs (Stimulus Direction x Group) conducted at each time point separately. At baseline, the Stimulus Direction x Group interaction was not significant, $F(1,58) = .20, p = .65$, suggesting groups did not differ in their interpretation or response biases. In contrast, after CBM-pa training the Stimulus Direction x Group remained significant, $F(1,58) = 8.65, p = .005, \eta_p^2 = .13$. Hypothesis-driven group contrasts (Howell, 2010) revealed that, for non-paranoid material, there was no significant difference between groups after training, $t(58) = -1.12, p = .27, d = -.29$. However, for paranoid material, there was a significant difference, $t(42.95) = 2.36, p = .023, d = .64$, with those receiving paranoid training giving significantly higher similarity ratings than those receiving non-paranoid training, see Figure 1.

3b.3 Influence of CBM-pa on Situational Events

suspicious medication' and the non-paranoid direction would be *'Your doctor gives you beneficial medication'*.

Event 1 manipulation check: One participant (1.67%) gave a paranoid interpretation of the interruption event (*'There was something I had done wrong'*). All other participants made non-paranoid interpretations (98.33%). *Outcome data:* There was no significant difference between training groups in their interpretation or attributions of the interruption, level of conviction or distress associated with the interruption event (all p 's $> .06$) (Table 3).

Event 2 manipulation check: Around half of the participants (55%) noticed the laughter and provided an explanation for it. Of the 33 participants noticing the laughter, 16 had received paranoid training and 17 non-paranoid, which did not represent a significant difference (Fisher's Exact $p = .59$ two-sided). 8 participants (24.24%) made paranoid interpretations (7 from the paranoid direction group and 1 from the non-paranoid direction group) and 25 (75.75%) made non-paranoid interpretations (9 from the paranoid direction group and 16 from the non-paranoid direction group). *Outcome data:* Interpretations of the laughter significantly differed between groups, Fisher's Exact $p = .017$ (two-sided). Examples of paranoid interpretations from the paranoid direction group were: *'I thought you might be laughing at my answers'*; *'You were laughing at something to do with me'*. Examples of non-paranoid interpretations from the non-paranoid direction group were: *'You must have heard a joke'*; *'I didn't really think about it, some random people in the corridor'*. 15% of the sample reported a paranoid interpretation for at least one of the ambiguous events ($n = 9$, equivalent to Green et al., 2011 with a similar sample size), of whom 13% received paranoid training.

Independent samples t -tests revealed that the paranoid direction group believed that the laughter was more likely to do with them (i.e. made higher internal personal attribution ratings) $t(58) = 2.22$, $p = .034$, $d = .77$ and were significantly

more distressed by it, $t(24.30) = 2.55$, $p = .02$, $d = .92$, compared with the non-paranoid direction group, see Table 3.

4. Discussion

In a two Experiment paper we demonstrated, firstly, an interpretation bias relevant to paranoid thinking in the general population and, secondly, that information processing could be induced in a more paranoid direction using a novel CBM-pa programme. In Experiment 1, trait paranoia significantly predicted paranoid interpretation bias across two different measures. In Experiment 2, after training (but not before) those trained in a paranoid direction endorsed more previously unseen ambiguous text passages in a paranoid manner compared to those trained in the opposite direction (although response bias could not be ruled out and groups did not differ on non-paranoid material following training). A subsequent situational laughter event produced training congruent differences in the pattern of participants' interpretations and personal attributions of the laughter and in their level of distress experienced in response to it, thus showing preliminary evidence of wider effects of the training procedure.

A number of findings are worthy of more detailed discussion. Experiment 1 showed that the combined effect of measures of paranoia and psychosis accounted for over a third of the variance in paranoid interpretation bias on the SRT, with trait paranoia making an independent prediction. Similarly on the SST, over a third of the variance was explained by measures of paranoia and psychosis, with trait paranoia again making an independent prediction. Across both tasks, paranoid interpretation bias was predicted by trait paranoia when measured by the PS, but not the GPTS. This might be due to differences in the content of these instruments, in which items of the

PS reflect milder paranoid thinking (e.g. mistrust, resentment) more characteristic of the general population, with only one item including intent of harm. Similarly, trait paranoia and delusions (PDI) were associated with bias indices on both the SRT and SST, whereas the CAPS was only associated with the former, suggesting items on this task may have been more sensitive for capturing perceptual experiences. Overall, Experiment 1 demonstrated that biased interpretation associated with paranoia in the subclinical population is a recurring phenomenon (Savulich et al., 2015) and, together with recent clinical sample results (Savulich et al., 2017; Yiend et al., 2018), is consistent with a continuum of biased processing across the spectrum of paranoia.

Experiment 2 produced differential effects of CBM-pa on subsequent interpretations of new ambiguous information in individuals with moderate levels of trait paranoia. However, the lack of interaction with the factor ‘Sentence Type’ limits the interpretation of the type of information that has been modified. It suggests our manipulation altered a wider range of paranoia-related information processing that included response bias (i.e. training effects were found for both target and foil sentences rather than target sentences alone). Nonetheless significant group differences were found in the endorsement of paranoid, but not non-paranoid meanings after CBM-pa, suggesting modest effects of training in the present sample. For instance, those receiving paranoid training were more likely than those receiving non-paranoid training to interpret our example (seeing two girls whispering on the other side of the room) to mean that the girls were ‘plotting against you’. However when these participants also rated harmless (i.e. non-paranoid) interpretations of the same situation (that the girls were ‘talking about a friend’) there was no statistical evidence of any effect of training. Therefore training appeared to affect overtly paranoid material by making people more or less likely to endorse paranoid

interpretations, but had little observable effect on harmless interpretations. This is consistent with the cognitive bias modification literature (Hirsch & Mathews, 2000; Eysenck et al., 1991), in which manipulation of biases in the pathological domain is usually more successful in healthy samples given their typically pre-existing adaptive biases (Yiend et al., 2019). Training in a clinical sample, or one with higher levels of trait paranoia, may increase specificity of the training programme and be more likely to elicit changes on both types of material beyond response bias. Both increasing the accessibility of non-paranoid meanings *as well as* reducing the paranoid interpretations made would be the optimum outcome of any clinically-motivated CBM intervention. Future work should focus on clinical participants or individuals with clinical levels of trait paranoia to gauge the effectiveness of increasing the endorsement of non-paranoid meanings as well as decreasing the endorsement of paranoid meanings, as both of these would be of potential clinical utility.

Experiment 2 also showed that individuals in the paranoid training group were more likely to believe that the laughter they heard had something to do with them and were significantly more distressed by it than the non-paranoid direction group, consistent with a putative etiological role for biased processing in individuals with moderate levels of trait paranoia. It also identifies one specific cognitive mechanism that could be targeted by suitably designed CMB interventions for psychosis. In addition to replicating the current findings, future work could also assess whether deluded conviction only differs at more severe levels of paranoia or in clinical populations.

Our data open another route for future CBM-I studies explicitly targeting paranoia. They suggest that paranoid interpretation bias is a reliable phenomenon and that CBM-pa could be further developed and evaluated as a potential clinical

intervention (see Yiend et al., 2017). Few studies to date have used CBM programmes for modifying interpretation in psychosis, and with limited success. Steel and colleagues (2010) targeted comorbid anxiety, but were not able to improve interpretation bias or state anxiety, whereas others targeted negative interpretation biases and only reported significant improvements in social anxiety (e.g. Turner et al., 2011). Their study manipulated interpretation bias by inducing interpretation of information of generally emotional content using versions of CBM designed for other disorders and did not attempt to modify the paranoia relevant interpretation bias that is directly related to the paranoid beliefs underlying psychosis. In contrast the present data, together with our previous studies in high trait (Savulich et al., 2015) and patient (Savulich et al., 2017; Yiend et al., 2018) samples, provide evidence of biased interpretations *specific* to paranoia and we suggest that it is these biases that CBM should seek to target when applied to paranoia and psychosis (Yiend et al., 2017). Not only would this further assess the etiological significance of paranoid interpretation biases in related disorders, but it would also help establish a new technique by which paranoid symptoms, related beliefs and mood might be clinically improved.

There are several limitations to the studies presented here. Firstly, it is not possible within the current design, to preclude the potential confounding effects of elevated anxiety and depression (or other potential confounders) on interpretation bias, especially in Experiment 1. Instead this question would need to be addressed using specifically adapted experimental designs, such as recruiting a sample of individuals with a fixed level of paranoia, but who vary in their level of depression (and/or vice versa) and measure both paranoid and negative interpretation biases. Using a design which holds one variable constant and varies only the other, as discussed above, would guarantee the independence (albeit artificial) of the constructs to be teased

apart, and permit meaningful analysis and conclusions about disorder specificity and bias. Further limitations include the sampling mainly from internal University mailing and advertisement in the South East London area in both studies and the restricted range of outcome measures used to assess the wider effects of CBM-pa.

Although Experiment 2 may have been underpowered to achieve the critical four-way interaction, the sample sizes of both of our studies were similar to those included in recent meta-analyses reporting overall medium effect sizes on interpretation biases following CBM (.5 to .8; Jones et al., 2017; Krebs et al., 2018). Furthermore, results from the Stoope-Laughter task were not adjusted for multiple comparisons; nonetheless, the percentage of participants making a paranoid interpretation for one of the ambiguous events was similar to the sample in which the task was originally developed (Green et al., 2011).

Importantly, we did not include a ‘no-training’ control condition (e.g. reading/resolving neutral passages of text) in Experiment 2, which was an experimental study designed to provide a ‘proof of principle’ that interpretation bias specifically relevant to paranoia could be manipulated. In common with most pre-clinical bias training studies, we were not so much interested in whether one group changed more or less than the other, or in the direction of change, but rather in the principle that change was simply possible at all. In any future work with patients, as opposed to healthy participants, it will be important to directly compare non-paranoid training with a ‘no-training’ control condition. The types of material needed to induce non-paranoid interpretations (and appropriate control measures) in patients requires further development and is already forming the basis of new work (Yiend et al., 2017). Lastly, these tasks were administered prior to the questionnaire measures in both

experiments, which may have increased participant's ratings of themselves as paranoid.

The present studies extend recent work on biased interpretation processes in paranoia and demonstrate that these specific biases can be manipulated in a limited context. In sum, CBM-pa only induced training effects in a maladaptive direction, and did not act solely to influence interpretation bias, but rather a wider range of paranoia-related information processing. This paper highlights an additional and more specific mechanism through which therapeutic changes in cognition might occur, in addition to those already widely investigated in psychosis (e.g. meta-cognition, data gathering). Unlike other interventions, CBM-pa is a self-administered procedure, and might eventually be developed as an adjunct to CBTp or other cognitive and social-cognitive interventions, designed to reduce susceptibility to biases in psychosis (e.g. meta-cognitive training; Moritz et al., 2014). Shifting maladaptive interpretation with CBM-pa has the potential to instil more adaptive biases, which in turn may help resolve symptoms and promote wellbeing (MacLeod, 2012; Lester, Mathews, Davison, Burgess, & Yiend, 2011). However, further research and refinement of the CBM-pa programme presented here would be essential prior to use in a clinical population in order to ensure generalizability of the intervention. In addition replication will be essential to validate the present findings and explore the possibility of achieving stronger training effects across a wider range of material and in suitable samples.

References

- Beck, A.T., Steer, R. A. & Brown, G. K. (1996) *Manual for the Beck Depression Inventory-II* (Vol. 1). San Antonio, TX: Psychological Corporation.
- Bell, V., Halligan, P. W., & Ellis, H.D. (2006). The Cardiff Anomalous Perceptions Scale (CAPS): A New Validated Measure of Anomalous Perceptual Experience. *Schizophrenia Bulletin*, 32, 366-377. doi:10.1093/schbul/sbj014
- Bentall, R. P., & Kaney, S. (1996). Abnormalities of self-representation and persecutory delusions. *Psychological Medicine*, 26, 1231–1237.
- Brockmeyer, T., Anderle, A., Schmidt, H., Ferby, S., Wunsch-Leiteritz, W., Leiteritz, A., & Friederich, H-C. (2018). Body image related negative interpretation bias in anorexia nervosa. *Behavior Research and Therapy*, 104, 69-73. doi: 10.1016/j.brat.2018.03.003
- Capron, D.W., & Schmidt, N.B. (2016). Development and randomized trial evaluation of a novel computer-delivered anxiety sensitivity intervention. *Behaviour Research and Therapy*, 81, 47-55. doi: 10.1016/j.brat.2016.04.001
- Casey, H., Rogers, R. D., Burns, T., & Yiend, J. (2013). Emotion regulation in psychopathy. *Biological Psychology*, 92, 541-548. doi:10.1016/j.biopsycho.2012.06.011
- Combs, D. R., Michael, C. O., & Penn, D. L. (2006). Paranoia and emotion perception across the continuum. *British Journal of Clinical Psychology*, 45, 19–31.
- 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
- Dietel, F.A., Mobius, M., Steinbach, L., Dusend, C., Wilhelm, S., & Buhlmann, U.

- (2018). Effects of induced appearance-related interpretation bias: A test of the cognitive-behavioral model of body dysmorphic disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, 61, 180-187. doi: 10.1016/j.jbtep.2018.07.003
- 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
- Freeman, D. (2007). Suspicious minds: The psychology of persecutory delusions. *Clinical Psychology Review*, 27, 425-457. doi:10.1016/j.cpr.2006.10.004
- Freeman, D. & Garety, P. A. (2000). Comments on the content of persecutory delusions: Does the definition need clarification? *British Journal of Clinical Psychology*, 39, 407-414. doi: 10.1348/014466500163400
- 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.
- 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, C. E. L., Freeman, D., Kuipers, E., Bebbington, P., Fowler, D., Dunn, G., & Garety, P. A. (2011). Paranoid Explanations of Experience: A Novel Experimental Study. *Behavioural and Cognitive Psychotherapy*, 39, 21-34.

doi:10.1017/S1352465810000457

- Grol, M., Schwenzfeier, A.K., Stricker, J., Booth, C., Temple-McCune, A., Derakshan, N., Hirsch, C., Becker, E., & Fox, E. (2018). The worrying mind in control: An investigation of adaptive working memory training and cognitive bias modification in worry-prone individuals. *Behaviour Therapy and Research*, 103, 1-11. doi: 10.1016/j.brat.2018.01.005
- Hallion, L.S., & Ruscio, A.M. (2011). A meta-analysis of the effect of cognitive bias modification on anxiety and depression. *Psychological Bulletin*, 136, 940-958. Doi:10.1037/a0024355
- Holmes, E.A., Lang, T.J., & Shah, D.M. (2009). Developing interpretation bias modification as a 'cognitive vaccine' for depressed mood: Imagining positive events make you feel better than thinking about them verbally. *Journal of Abnormal Psychology*, 118, 76-88. doi:10.1037/a0012590
- Hoppit, L., Illingworth, J.L., MacLeod, C., Hampshire, A., Dunn, B.D., & Mackintosh, B. (2014). Modifying social anxiety related to a real-life stressor using online Cognitive Bias Modification for interpretation. *Behaviour Research & Therapy*, 52, 45-52. doi: 10.1016/j.brat.2013.10.008
- Howell, D.C. *Statistical Methods for Psychology* (7th Ed.). 2010: Wadsworth, Cengage Learning.
- 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.
- Klein, A.M., Rapee, R.M., Hudson, J.L., Schniering, C.A., Wulthrich, V.M., ... Rinck, M. (2015). Interpretation bias modification training reduces social anxiety in clinically anxious children. *Behaviour Research and Therapy*, 75,

78-84. doi: 10.1016/j.brat.2015.10.006

Lecrubier, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, I., Harnett Sheehan,

K., . . . Dunbar, G. C. (1997). The Mini International Neuropsychiatric

Interview (MINI). A Short diagnostic structured interview: reliability and

validity according to the CIDI. *European Psychiatry*, 12(5), 224-231. doi:

10.1016/S0924-9338(97)83296-8

LeMoult, J., Colich, N., Joorman, J., Singh, M.K., Eggeslton, C., Gotlib, I.H. (2018).

Interpretation bias training in depressed adolescents: Near- and far-transfer

effects. *Journal of Abnormal Child Psychology*, 46, 159-167. doi:

10.1007/s10802-017-0285-6

Lee, J-S., 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

Lester, K. J., Mathews, A., Davison, P. S., Burgess, J. L., and Yiend, J. (2011).

Modifying cognitive errors promotes cognitive well being: A new approach to

bias modification. *Journal of Behavior Therapy and Experimental Psychiatry*,

42, 298-308. doi: 10.1016/j.jbtep.2011.01.001

Martin, M., & Rubin, R. (1995). A new measure of cognitive flexibility.

Psychological Reports, 76, 623-626.

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, 209-222.

doi:10.1016/j.beth.2006.03.001

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

- MacLeod, C., & Holmes, E.A. (2012). Cognitive bias modification: An intervention approach worth attending to. *American Journal of Psychiatry*, 169, 118-120. doi: 10.1176/appi.ajp.2011.11111682
- Matheson, E., Wade, T.D., & Yiend, J. (2018). A new cognitive bias modification technique to influence risk factors for eating disorders. *The International Journal of Eating Disorders*, 10.1002/eat.22938
- 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., Ridgeway, V., Cook, E., & Yiend, J. (2007). Inducing a benign interpretational bias reduces trait anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, 38, 225-236. doi:10.1016/j.jbtep.2006.10.011
- Mathews, A., Yiend, J., & Lawrence, A. D. (2004). Individual differences in the modulation of fear-related brain activation by attentional control. *Journal of Cognitive Neuroscience*, 16, 1683-1694.
- 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., & 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
- 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.
doi:10.1093/oxfordjournals.schbul.a033401
- 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
- Salemink, E., & van den Hout, M. (2010). Trained interpretive bias survives mood change. *Journal of Behavior Therapy & Experimental Psychiatry*, 41, 310-315. doi: 10.1016/j.jbtep.2010.02.010
- Salemink, E., van den Hout, M., & Kindt, M. (2009). Effects of positive interpretive bias modification in highly anxious individuals. *Journal of Anxiety Disorders*, 23, 676-683. doi: 10.1016/j.janxdis.2009.02.006
- 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., Jeanes, H., Rossides, N., Kaur, S., Zacharia, A., Robbins, T.W., & Sahakian, B.J. (2018). Moral emotions and social economic games in paranoia. *Frontiers in Psychology*, 9:615. doi: 10.3389/fpsy.2018.00615
- Savulich, G., Shergill, S., & Yiend, J. (2017). Interpretation biases in clinical paranoia. *Clinical Psychological Science*, 5, 985-1000. doi:10.1177/2167702617717180
- Savulich, G., Shergill, S., & Yiend, J. (2012). Biased cognition in psychosis. *Journal of Experimental Psychopathology*, 3, 514-536. doi:10.5127/jep.016711
- Spielberger, C.D., Gorsuch, R. L., Lushene, P. R., Vagg, P. R., & Jacobs, A. G. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y)*. Palo Alto, CA: Consulting Psychologists Press.
- Steel, C., Wykes, T., Ruddle, A., Smith, G., Shah, D., & Holmes, E. A. (2010). Can we harness computerised cognitive bias modification to treat anxiety in

- schizophrenia? A first step highlighting the role of mental imagery. *Psychiatry Research*, 178, 451-455. doi:10.1016/j.psychres.2010.04.042
- Turner, R., Hoppitt, L., Hodgekings, J., Wilkinson, J., Mackintosh, B., & Fowler D. (2001). Cognitive bias modification in the treatment of social anxiety in early psychosis: A single case series. *Behavioral and Cognitive Psychotherapy*, 39, 341-347. doi:10.1017/S1352465810000706
- Trotta, A., Kang, J., Stahl, D., & Yiend, J. (submitted). Interpretation bias in paranoia: a systematic review and meta-analysis.
- Watkins, E.R., Baeyens, C.B., & Read, R. (2009). Concreteness training reduces dysphoria: Proof-of-principle for repeated cognitive bias modification in depression. *Journal of Abnormal Psychology*, 118, 55-64. doi: 10.1037/a0013642
- Wilson, E., MacLeod C., Mathews A., & Rutherford E. (2006). The causal role of interpretative bias in anxiety reactivity. *Journal of Abnormal Psychology*, 115, 103-111. doi:10.1037/0021-843X.115.1.103
- 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.
- Wong, S.F & Grisham, J.R. (2017). Causal role for inverse reasoning on obsessive-compulsive symptoms: Preliminary evidence from a cognitive bias modification for interpretation bias study. *Journal of Behavior Therapy and Experimental Psychiatry*, 57, 143-155. doi: 10.1016/j.jbtep.2017.06.001
- 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., Allen, P., Lopez, N., Falkenburg, I., Tseng, H-H, & McGuire, P. (2018). Negative interpretation biases precede the onset of psychosis. *Behaviour Therapy*, 46, 110-124.
- Yiend, J., Lee, J-S., Mathews, A., Tekes, S., Atkins, L., Ferragamo, C., ... Shergill, S. (2014a). 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. *Behaviour Research and Therapy*, 43, 779-797. doi:10.1016/j.brat.2004.06.007
- Yiend, J., Mathews, A., Weston, B., Dunn, B. D., Cusack, R., & Mackintosh, B. (2008). An investigation of the implicit control of the processing of negative pictures. *Emotion*, 8, 828-837. doi:10.1037/a0014146
- Yiend, J., Parnes, C., Shepherd, K., Roche, M., & Copper, M. J. (2014b). 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. *Behaviour Research and Therapy*, 49, 892–900. doi:10.1016/j.brat.2011.10.004
- Yiend, J., Trotta, A., Meek, C., Dzafic I., Baldus, N., Crane, B., ... Peters, E. (2017). Cognitive Bias Modification for paranoia (CBM-pa): study protocol for a

randomised controlled trial. *Trials*, 18, 298. doi:10.1186/s13063-017-2037-x

Table 1. *Participant demographic information and questionnaires measures by experiment*

	Experiment 1		Experiment 2	
	Whole sample (<i>n</i> =70)	Paranoid Direction (<i>n</i> =29)	Non-Paranoid Direction (<i>n</i> =31)	Statistic, p-value
Demographic				
Age	26.24 ±7.73	26.14 ±8.14	25.97 ±9.09	<i>t</i> (58)=.08, .94
Gender	17 M, 53 F	11 M, 18 F	11 M, 20 F	<i>X</i> ² =.04, .84
Questionnaires				
GPTS total	54.54 (23.50)	54.45 ±12.38	51.39 ±10.42	<i>t</i> (58)=1.04, .30
PS	40.11 (13.10)	38.69 ±10.36	36.10 ±10.46	<i>t</i> (58)=.96, .34
PDI total	40.11 (28.73)	42.17 ±28.19	39.45 ±36.56	<i>t</i> (58)=.32, .75
CAPS	6.23 (6.65)	7.07 ±5.15	5.42 ±5.76	<i>t</i> (58)=1.17, .25
CF	53.34 (9.55)			
BDI-II	8.57 (7.63)			
STAI-Trait	40.26 (11.41)			
Mood state				
VAS Paranoia		.89 ±.74	.82 ±1.34	<i>t</i> (58)=.27, .79
VAS Anxiety		1.20 ±2.31	2.00 ±1.92	<i>t</i> (58)=-.003, .10
VAS Social Anxiety		2.19 ±1.85	1.94 ±1.84	<i>t</i> (58)=.54, .59
VAS Depression		1.33 ±1.11	1.54 ±1.72	<i>t</i> (58)=-.57, .57

Notes: GPTS: Green Paranoid Thoughts Scale; PS: Paranoia Scale; PDI: Peter's Delusions Inventory; CAPS: Cardiff Anomalous Perceptions Scale; CF: Cognitive Flexibility Scale; BDI-II: Beck's Depression Inventory; STAI-Trait: Spielberger Trait Anxiety Inventory; VAS: visual

analogue scales; Bold indicates significant group differences where $p < .01$.

Table 2. *Experiment 1 prediction of paranoia/psychosis measures on indices of paranoid interpretation bias by task (overall model)*

Model	Predictor	β	t	p	95% CI	Partial Correlations	R^2 , F , p -value
SRT	GPTS	.16	1.02	.31	-.004, .01	.13	.38, 9.95, <.001
	PS	.36	2.27	.027	.002, .03	.27	
	PDI	.10	.74	.46	-.004, .01	.09	
	CAPS	.15	1.31	.20	-.008, .04	.16	
SST	GPTS	.01	.08	.94	-.20, .21	.009	.38, 13.58, <.001
	PS	.64	4.12	<.001	.40, 1.15	.45	
	PDI	-.07	-.60	.55	-.17, .09	-.07	

Notes: SRT: Similarity Ratings task; SST: Scrambled Sentences task; CI: Confidence interval for β ; GPTS: Green Paranoid Thoughts Scale; PS:

Paranoia Scale; PDI: Peter's Delusions Inventory; CAPS: Cardiff Anomalous Perceptions Scale; Bold indicates $p < .01$.

Table 3. Experiment 2 SRT ratings at baseline and post-training; Stooge-Laughter task outcomes by training group (counts for paranoid/non-paranoid interpretations of the events and mean ratings and standard deviations for attributions, conviction and distress); Bold indicates $p < 0.01$.

	Paranoid Training Group ($n=29$)	Non-Paranoid Training Group ($n=31$)	Statistic, p -value, Cohen's d
<i>SRT Baseline</i>			
Paranoid material	1.66 (.43)	1.61 (.32)	$t(58)=.46, .65, .13$
Non-paranoid material	2.15 (.37)	2.17 (.49)	$t(58)=-.12, .90, .04$
<i>SRT Post-training</i>			
Paranoid material	1.79 (.57)	1.51 (.32)	$t(42.95)=2.63, .02, .61$
Non-paranoid material	2.09 (.43)	2.21 (.45)	$t(58)=-1.12, .27, .27$
<i>Event 1 Interruption</i>			
Paranoid interpretations	1	0	Fisher's exact $p=.48$
Non-paranoid interpretations	28	31	
Internal-personal	1.62 $\pm$ 2.01	1.27 $\pm$ 2.77	$t(58)=.55, .58, .14$
External-other	6.66 $\pm$ 3.22	7.47 $\pm$ 2.76	$t(58)=-1.05, .30, .27$
External-situational	6.21 $\pm$ 2.84	5.82 $\pm$ 2.95	$t(58)=.51, .61, .13$
Conviction	5.98 $\pm$ 2.67	6.74 $\pm$ 2.54	$t(58)=-1.30, .26, .29$
Distress	1.83 $\pm$ 2.65	.81 $\pm$ 1.14	$t(37.47)=1.92, .06, .50$
<i>Event 2 Laughter</i>			
Paranoid interpretations	7	1	Fisher's exact $p=.017$
Non-paranoid interpretations	9	16	
Internal-personal	3.94 $\pm$3.49	1.56 $\pm$2.66	$t(32)=2.22, .034, .77$
External-other	5.69 $\pm$ 2.72	6.13 $\pm$ 3.34	$t(32)=-.41, .68, .14$
External-situational	5.08 $\pm$ 2.77	5.81 $\pm$ 3.12	$t(32)=-.72, .48, .25$
Conviction	6.50 $\pm$ 1.98	6.06 $\pm$ 2.84	$t(32)=.53, .60, .18$
Distress	3.33 $\pm$3.31	1.13 $\pm$1.50	$t(24.30)=2.55, .017, .86$

Figure 1. *Experiment 2 mean similarity ratings for the paranoid and non-paranoid directions by group at baseline and post-CBM-pa*