The Efficacy of Attentional Bias Modification for Anxiety: A Registered Replication

Nathan Pond[a*], Frances Meeten[ab], Patrick Clarke[c], Lies Notebaert[d], & Ryan Scott[a]

[a] School of Psychology, University of Sussex, Brighton, BN1 9RH, UK

[b] Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, SE5 8AB, UK

[c] Cognition and Emotion Research Group, Faculty of Health Sciences, Curtin University

[d] Centre for the Advancement of Research on Emotion, School of Psychological Science, University of Western Australia

*Corresponding author: np286@sussex.ac.uk

Acknowledgements

Nathan Pond is in receipt of a University of Sussex, School of Psychology funded PhD studentship.

Dr Frances Meeten receives salary support from a Medical Research Council (MRC) New Investigator Grant (MR/W005077/1).

Abstract

Generalised anxiety disorder (GAD) is a prevalent condition that has been linked to the presence of certain cognitive biases, including attention bias. Attention bias is the tendency to attend preferentially to threat-related stimuli and has been consistently observed in high anxious samples. Naturally, interventions aiming to modify these biases have been developed with the hopes of alleviating anxiety symptoms. However, while initial studies were promising, over time the reported efficacy of these attention bias modification (ABM) procedures in alleviating symptoms has become mixed, with some studies reporting moderate to large effect sizes, and others reporting non-significant effects. Furthermore, concerns have been raised regarding the potential for demand effects to be underlying previous significant findings. Therefore, we revisit the efficacy of ABM as a method for alleviating both attention bias, and in turn anxiety symptomology. As our primary objective we seek to conduct a direct replication of one of the seminal studies showing successful alleviation of anxiety symptoms using multi-session ABM training (Hazen et al., 2009), while adopting a Bayesian approach to analyses. As a secondary goal, we aim to quantify the potential influence of demand effects in the paradigm.

Keywords: anxiety, worry, GAD, attention bias, CBM-A, cognitive bias modification, ABM

Supplementary material available here: https://osf.io/mw8p6/

Generalised anxiety disorder (GAD) is a common and debilitating condition characterised by excessive and uncontrollable worry, as well as somatic symptoms including restlessness, muscle tension and fatigue (DSM-V; APA, 2013). Furthermore, GAD has been associated with impaired social and occupational functioning and increased suicide risk (Wittchen et al., 2002). GAD is estimated to have a lifetime prevalence of 3.7% according to a cross-nationally representative survey by Ruscio et al. (2017), with this figure rising to 5% in high-income countries. However, the prevalence rate has likely risen significantly since this estimate was produced, with global events such as the COVID-19 pandemic having led to a sharp increase in GAD symptoms in the general population (Jia et al., 2020). Despite the devastating impact of this condition, the leading treatment, Cognitive Behavioural Therapy (CBT), has a variable success rate, with only 50% of individuals receiving CBT for GAD achieving full remission (Loerinc et al., 2015). Given the increasing prevalence of GAD, it is essential that we identify new treatment approaches to reduce GAD symptoms.

In the model outlined by Hirsch and Mathews (2012), pathological worry (the key cognitive component of GAD) is associated with two key cognitive biases, interpretation bias and attention bias. Interpretation bias can be defined as the tendency of anxious individuals to interpret ambiguous information as threatening, whereas attention bias is the tendency to attend preferentially to threat-related stimuli (Hirsch & Mathews, 2012). Given the diverse nature of worry topics present in GAD samples, it is difficult to describe one 'archetypal' example of an attention bias present in GAD, as it is characterised by a generalised attentional bias to a variety of minor emotional/threatening stimuli that is not present in non-anxious populations (Mogg & Bradley, 2005). One such example would be an attentional bias towards

negative and/or threatening words, or to negative faces (Mogg & Bradley, 2005). However, attention biases are also cross-modal and alongside visual attentional bias, there is also evidence to support auditory attentional bias (e.g., Wang et al. (2019)). Hirsch and Mathews (2012) argue that these cognitive biases lead to the more frequent intrusion of negative threat representations into conscious perception, and thus increase the occurrence of prolonged worry periods characteristic of GAD. Though both biases play an important role in this process, whereas modification of interpretation bias and its effects on anxious symptomatology are robustly reported in the literature, the evidence is mixed regarding the malleability of attention bias compared to interpretation bias (Cristea et al., 2015; Liu et al., 2017), thus this paper focuses on attention bias.[1] In addition to the Hirsch and Mathews (2012) model, a number of theories have been proposed suggesting an important role for attention bias in both the development and maintenance of fear and anxiety (for a review, see Van Bockstaele et al., 2014), suggesting that the influence of this cognitive bias extends robustly beyond pathological worry, and into the context of anxiety. Indeed, an in-depth assessment based on Hill's (1965) criteria of causality concluded that there is evidence of a causal relationship between attention bias and anxiety (Van Bockstaele et al., 2014). Furthermore, research has shown that attention biases are prominent in GAD samples, with a meta-review by Goodwin et al. (2017) reporting

[1] It is worth noting that combined cognitive bias hypotheses are an important part of the literature, as we know that cognitive biases, such as interpretation and attention, do not necessarily operate in isolation (see Hirsch et al., 2006). For example, Hirsch and Mathews (2012) note in their model that high worriers interpret ambiguous information in a negative manner (interpretation bias), and these negative thoughts may potentially become the focus for attention bias. However, given that this report is a registered replication of a study of attention bias, and that findings regarding the malleability of attention bias have been mixed, the focus of the present report remains exclusively on attention bias.

that 69% of the 29 studies reviewed found a reliable attention bias towards threatening stimuli in GAD populations.

Given the role of these cognitive biases in GAD, it is unsurprising that they are being studied as potential targets for modification in the treatment of GAD and other anxiety disorders. Research has shown that multi-session interpretation bias modification training, in which individuals are presented or asked to generate positive or benign interpretations to emotionally ambiguous situations (Beard & Peckham, 2020), is effective at reducing GAD symptoms (Hirsch et al., 2018; 2020; 2021; Ji et al., 2021). Given this success, naturally the potential efficacy of attention bias modification (ABM) training is also a subject of interest. A common method of ABM training is threat-avoidance training (Mogg et al., 2017). Threat-avoidance training often implements visual-probe tasks to reduce attentional bias towards threat. In these tasks, two cues (one threat-related, one neutral) are presented to participants simultaneously. When the cues disappear, a target-probe is located in place of one of the cues and participants are required to respond to that target as quickly as possible. To train threat-avoidance, the target-probe is more likely to appear where the neutral cue was previously located, thus encouraging participants to orient their attention consistently away from the threat cue.

While there is a sound theoretical rationale for implementing ABM training interventions to reduce anxiety symptoms, the experimentally assessed efficacy of ABM training has produced mixed results (Martinelli et al., 2022). Given the large amount of variability across different ABM training paradigms, including variances in ABM task paradigms (dot-probe; visual probe; etc.), training settings (offline; online) and stimuli used (pictures; words), this is perhaps unsurprising (Hallion & Ruscio, 2011; Martinelli et al., 2022). Seminal studies by Amir et al. (2009) and

Hazen et al. (2009) that were the first to assess the efficacy of an ABM training intervention for GAD delivered promising results, with both studies finding a significant alleviation of reported anxiety in both a clinically diagnosed GAD sample, and a high-worry student sample, respectively. However, subsequent meta-analyses of studies into the effectiveness of ABM training for alleviating anxiety disorders have reported varying levels of efficacy, with some reviews suggesting moderate to large overall effect sizes (Hakamata et al., 2010; Linetzky et al., 2015), while others have reported very small, and in some cases non-significant, overall effect sizes (Cristea et al., 2015; Hallion & Ruscio, 2011). This has led some researchers to question the therapeutic value of ABM interventions for anxiety, with Cristea et al. (2015) arguing that the previously reported efficacy of ABM training interventions may be the result of demand effects. The moderator analysis by Cristea et al. (2015) showed that effect sizes were larger for studies conducted in a laboratory setting in which demand effects are known to be more influential. Furthermore, the authors also reported that control conditions in the studies they reviewed often led to improvements in anxiety symptoms equal to that of the ABM intervention conditions, further suggesting that the presence of expectancy effects may be influencing responding.

Despite the discouraging findings of some reviews, researchers have noted that while the increase of null findings, particularly in ABM interventions targeting social anxiety disorder (SAD) and post-traumatic stress disorder (PTSD), has led some to become wary of ABM training as an anxiety intervention, this wariness may be premature (Clarke et al., 2014; Grafton et al., 2017). Grafton et al. (2017) note that there has been a confusion in the cognitive bias modification literature, in that studies that *attempted* to modify cognitive biases and studies that *actually* managed

to modify cognitive biases are oft referred to using the same label of cognitive bias modification training. Thus, in meta-reviews the efficacy of studies attempting to alleviate anxiety symptoms that both did and did not manage to actually modify the intended cognitive bias are analysed together. The authors argue that this has led to an increase in null findings, which have artificially lowered the apparent efficacy of cognitive bias modification training. For example, in a reappraisal of the data reported by Cristea et al. (2015), Grafton et al. (2017) found that, out of nine ABM training studies that included a measure of attention bias change pre- and post-training, the six studies that failed to change attention bias had a near-zero effect size on reported anxiety (Hedges' g = -0.01). However, the remainder of the studies that managed to change attention bias reported an overall moderate effect size on reported anxiety (g = 0.6). Further, Clarke et al. (2014) compiled a list of 29 studies measuring both the mental health outcomes of ABM training for anxiety disorders and the change in attention bias from pre- to post-training at the group level. The authors reported that 26 of the 29 studies reviewed were consistent with the idea that when attention bias is successfully modified, an improvement in mental health outcomes is observed, and when attention bias is not successfully modified, there is no such improvement in mental health outcomes. Therefore, it was concluded that the null findings reported by some authors may likely be due to a failure of the specific methodology of the ABM training intervention applied to modify attention bias, rather than a failure of ABM training in any form to alleviate anxiety. As such, Clarke et al. (2014) conclude that though further research into the most effective methodology for ABM training interventions is warranted, interventions that do effectively reduce attention bias do lead to improvements in mental health outcomes.

It is worth noting that the ideas expressed in the above studies regarding the finding that when ABM training actually modifies bias, then anxiety is generally reduced and when ABM training does not modify bias anxiety is not reduced, have been criticised by some researchers (Cristea, 2018; Kruijt & Carlbring, 2018). Specifically, Cristea (2018) states that "identifying the trials in which both bias and outcomes were successfully changed is only possible post hoc, as these are both outcomes measured after randomisation; reverse engineering the connection between the two is subject to confounding. Bias and symptom outcomes are usually measured at the same time points in the trial, thus making it impossible to establish temporal precedence" (p. 247). Cristea (2018) goes on to argue that this increases the risk of reverse causality in our arguments (as symptom change may have caused bias change), as well as the risk of conflating demand effects as evidence for ABM procedure effectiveness (the trials in which bias and/or symptom change occurred successfully may simply be due to bias, including experimenter effects). ~~While it is certainly true that to go through on an individual trial basis would be an inappropriate way to analyse such data and would certainly increase the risk of bias, Clarke et al. (2014) observed this effect at the group level, suggesting there is sound evidence for this argument. Furthermore, the~~On balance, the overall argument made by Grafton et al. (2017) and Clarke et al. (2014) that an ABM training that does not modify bias has failed at being an ABM training, and therefore should not be considered an ABM procedure, still makes logical sense. ~~We agree~~Furthermore, in response to the criticisms raised by Cristea (2018), ~~with the response raised by~~ Parsons (2018)~~, who~~ argues that what these studies show is that ABM procedures currently do not robustly modify bias (unsurprising, given the earlier discussed variability among ABM paradigms), and therefore we must develop more effective

procedures. However, it is important to acknowledge that the current pattern of results reported does not provide conclusive evidence for this particular interpretation over other possibilities. Overall, this debate perfectly highlights the importance of performing a registered replication of a seminal ABM training study, in order to ascertain its effectiveness using a rigorous design.

A more encouraging picture arises from a recent meta-analysis by Martinelli et al. (2022), whose study represents the largest review of the efficacy of ABM training interventions for modifying bias in mental health disorders to date. Across 13 studies looking into ABM training in populations of participants with non-specific anxiety, Martinelli et al. (2022) reported an overall moderate effect size (g = 0.53), suggestive that ABM interventions may yet prove to be effective for alleviating attention bias. Importantly, the authors make the distinction of separating out non-specific anxiety from other types of anxiety disorders including SAD and PTSD. Interestingly, Martinelli et al. (2022) reported that while ABM interventions are moderately effective in alleviating attention bias in those with non-specific anxiety, these interventions are far less effective for reducing attention bias in SAD populations (g = 0.25, non-significant), and in the case of PTSD larger decreases in attention bias were observed in control than experimental conditions (g = -0.7). This variation in efficacy will clearly be important in the future targeting of ABM.

A review of the efficacy of ABM interventions in relieving anxiety symptomology by Mogoaşe et al. (2014) found results consistent with Martinelli et al. (2022). They similarly distinguished between sub-groups of anxiety disorders and found that, across 22 studies of ABM training for alleviating the symptoms of anxiety disorders, the overall postintervention effect size was small (g = 0.26). However, the effect size was found to be moderated by the type of anxiety disorder,

with an overall moderate effect size reported by studies looking into generalised anxiety (g = 0.61), a relatively small effect size detected in studies of SAD (g = 0.24), and a non-significant effect in other types of anxiety disorders, a category which included phobias and PTSD (g = 0.16). Therefore, it may be that reviews failing to make the distinction between different types of anxiety disorder may fail to detect stronger effect sizes observed in studies solely targeting general anxiety. Given the findings of Clarke et al. (2014) and Grafton et al. (2017), one may conclude that the lack of experimental evidence of attention bias modifiability, and in turn symptom alleviation, in the context of SAD and PTSD may be due to a failure of the ABM paradigms to modify attention bias, rather than SAD and PTSD populations simply being less responsive to the induced change. Regardless, there is still evidence to suggest that in the case of general anxiety, ABM training may be an effective therapeutic method.

There is a clear need to assess the replicability of previous findings suggesting that ABM training is an effective method of alleviating GAD symptoms, including high worry. Since the publication of the Reproducibility Report: Psychology (Open Science Collaboration, 2015), which estimated a low replication rate of around 36% in psychological literature, it has become clear that it is exceedingly important to establish the replicability of key psychological findings. This replication crisis has been driven largely by the prevalence of false positive results arising from poor research practices, as well as publication bias, among other things (Hu et al., 2016). As introduced by Chambers (2013), registered reports are a new publication format that aim to reduce the likelihood of these issues occurring, and therefore performing a registered replication is a rigorous way of testing the replicability of a seminal ABM training study.

In line with this ethos, the aim of the proposed research is to conduct a pre-registered replication of a study assessing the influence of an ABM training intervention on symptoms of anxiety. A seminal study by Hazen et al. (2009) exploring the effect of ABM training on symptoms of anxiety was identified as an ideal target. This study both found a substantial effect of ABM and, as it utilised a high worry student sample, offers greater feasibility for data collection. The original researchers recruited a sample of psychology undergraduates exhibiting high levels of trait worry (as assessed via the Penn State Worry Questionnaire; Meyer et al., 1990) and randomly assigned them to either an experimental ABM training group, or a sham training control group for a 5-session training course. The researchers then assessed the change in indicators of anxiety and depression, as well as the change in attention bias, from pre- to post-training assessment. The researchers found that, based on a composite measure of mental health symptoms including indicators of general anxiety and depression, ABM training led to a significant decrease in reported symptomology. Furthermore, assessment of attention bias found that the ABM intervention did indeed reduce the observed attention bias as predicted. In both cases, no such improvements were observed in the sham training control group. Consistent with the original study, the present replication predicts that the active versus sham ABM training will lead to a ~~significant~~ reduction in symptoms of general anxiety and depression as measured by a composite symptom index (CSI)[2], and a ~~significant~~ reduction in attention bias as assessed via a dot-probe based task.

[2] A composite mental health measure is being used in the interests of fully replicating Hazen et al.'s (2009) methodology. While this is not a 'standard' choice in the literature, the original authors used the CSI in the interests of reducing the familywise error rate (by analysing one CSI measure, as opposed to three individual measures). However, though the CSI is still included, we will also report Bayesian analyses on the three individual mental health measures that comprise the CSI, to allow for the separation of depression and anxiety as constructs.

In addition to the originally assessed hypotheses, in response to the concerns raised by Cristea et al. (2015) and Cristea (2018), an additional analysis will be conducted to assess the hypothesis that demand effects may be influencing the apparent reduction in symptomology. Demand effects occur when participants behave differently, consciously or unconsciously, in response to any aspect of the design or delivery of an experiment that indicates how they are expected to respond. First discovered by Orne (1962), demand effects have proven exceedingly difficult to mitigate (Nichols & Maner, 2008), and are particularly problematic in a training experiment such as this, in which the aim (in this case to reduce anxiety) is inherently clear. Therefore, quantifying the presence of demand effects in the present study is essential, as their presence would suggest that expectancy effects are driving the clinical outcomes of the ABM intervention. This would have implications for the underlying theory (if ABM treatment outcomes are driven by expectancy, then there may in truth not be a link between attention bias and worry/anxiety).

The relationship between a measure of Phenomenological Control (PC) and the CSI will be assessed. PC describes the ability to create phenomenological experiences that match the expectancies of a given situation. One such example of a phenomenological experience would be the rubber hand illusion, in which the sight of a rubber hand being brushed at the same time brushing is felt on one's real hand leads participants to report subjective ownership of the rubber hand; this is an example of a visual hallucination leading to a subjectively reported experience that is driven by demand effects (Dienes et al., 2022; Lush et al., 2020). The application of PC to create such experiences can occur in the absence of any conscious intention. As such, where there are implicit or explicit expectations put on an individual's response or experience, those with higher PC will be more likely to exhibit the

corresponding outcome; they will have increased susceptibility to demand effects (Lush et al., 2020). This difference in susceptibility can be applied to detect the influence of demand effects in any experimental paradigm with outcome variables related to experienced phenomenology; where demand effects are present there should be a corresponding correlation between PC and the outcome measures. Note, while the relationship between responses and PC has been shown to provide an indication of demand effects in a variety of experimental paradigms (Lush et al., 2020), a null result for this experimental hypothesis does not rule out the presence of demand effects entirely. The absence of a relationship between outcome measures and PC is also consistent with the presence of demand effects that are unrelated to PC. In contrast, the presence of a relationship can be taken as a strong indication of the influence of demand effects. In the present context, if demand effects contribute to the apparent efficacy of the ABM training, then it would be predicted that participants with higher PC will exhibit greater reductions in mental health symptomology.

Final Hypotheses

1) ABM training will lead to a ~~significant~~ reduction in attention bias.
2) ABM training will lead to a ~~significant~~ reduction in symptoms of general anxiety and depression.
3) Demand effects may ~~have an~~ influence ~~on~~ ABM training efficacy, contributing to a reduction in ~~reducing~~ mental health symptomology.

Method

The experiment aims to directly replicate the procedure reported by Hazen et al. (2009) to examine whether the same ABM training paradigm can be shown to both reduce attention bias and successfully alleviate symptoms of anxiety and depression. Therefore, all methods will be in keeping with those used in the original study. The only exception to this is that Hazen et al. (2009) also conducted the Structured Clinical Interview for Axis I DSM-IV Disorders on a subsection of participants. However, this was not part of the inclusion/exclusion criteria for the study, nor was it used for any of the analyses. Therefore, we do not conduct the interview in this replication. The original frequentist analyses will be replaced by Bayesian analyses, to allow for more nuanced interpretation of potentially non-significant results as denoting either evidence for the null hypothesis or indicating insensitive data. These Bayesian analyses will be specified in line with the method outlined by Dienes (2021, p. 5-8), in which one first specifies a rough scale of effect, then compares the evidence for a model predicting this effect to a model predicting no effect using a Bayes factor. The final analytical decision as to whether the data supports or refutes the experimental hypotheses will be made based on these Bayesian analyses.

Statistical Hypotheses

1) H_0: ABM training will not lead to a ~~significant~~ reduction in attention bias from pre- to post-training, as assessed via the change in Probe Discrimination Task score.

H_1: ABM training will lead to a ~~significant~~ reduction in attention bias from pre- to post-training, as assessed via the change in Probe Discrimination Task score.

2) H_0: ABM training will not lead to a ~~significant~~ reduction in symptoms of general anxiety and depression from pre- to post-training, as measured via change in both individual measures (PSWQ, STAI-T and BDI) and via change in the CSI.

   H_1: ABM training will lead to a ~~significant~~ reduction in symptoms of general anxiety and depression from pre- to post-training, as measured via change in both individual measures (PSWQ, STAI-T and BDI) and via change in the CSI.

3) H_0: PC score will have no influence on the change in CSI score from pre- to post-training.

   H_1: The higher the PC score, the greater the reduction in CSI score from pre- to post-training.

Participants

Participants will be gathered from a worry database maintained at the University of Sussex, which contains the PSWQ scores of psychology undergraduates who have indicated they are willing to be contacted for future research. Students with a PSWQ score of 60 or above will be contacted via email and invited to take part in the study.[3] Of those who express an interest, respondents who

[3] A PSWQ score of 60 or above is being used as the cut-off for high worry, in line with Hazen et al. (2009), who reported that scores of 60 or above represent the 90th percentile of normative population values. Furthermore, recent cognitive bias

are currently receiving treatment for an anxiety disorder, not fluent in English or do not have normal or corrected-to-normal vision will be excluded. These exclusions will be self-declared, as the consent form presented in the screening questionnaire asks participants to confirm that they meet the above stated eligibility criteria. Sample size will be based on a Bayesian stopping rule, adopting a sequential design with a maximal *n* of 200 (as described by Schönbrodt and Wagenmakers, 2018). Specifically, data collection will continue until we have obtained a sensitive Bayes factor for the change in CSI ($B > 30$, suggesting that ABM training has led to a reduction in mental health symptomology, or $B < 1/6$, suggesting that ABM training has not led to a reduction in mental health symptomology), or cease collection at 200 participants if the result remains insensitive. The only exception to this rule will be that we will collect a minimum of 40 participants. In using the procedure detailed by Palfi & Dienes (2019, Version 3, p. 15), it was determined that given a long-term relative frequency of good enough evidence of 50%, the proposed sample size allows for a discriminating Bayes factor ($B > 30$ if H1 is true, and a $B < 1/6$ if H0 is true).[4] Participants will be reimbursed in course credit. In line with the original authors' procedure, any participants who fail to attend all seven sessions will be excluded from analysis. This study has been granted ethical approval by the Sciences & Technology Cross-Schools Research Ethics Committee (C-REC), University of Sussex (Application Number: ER/NP286/13).

modification research has used a PSWQ score of 56 or above as the cut-off for high worry (see Feng et al., 2019), suggesting that the cut-off chosen here remains valid.

[4] This analysis was performed on the sensitivity of the main Bayesian analysis (the change in the CSI score). The analysis revealed that the likely *N* needed for a sensitive Bayes factor was 16 for H1, and 56 for H0. Given our maximal *N* of 200, it is apparent that a sensitive Bayes factor will be achievable.

Materials

Penn State Worry Questionnaire (PSWQ). Designed by Meyer et al. (1990) and with established internal consistency, test-retest reliability and convergent validity (Behar et al., 2003; Brown et al., 1992; Stöber, 1998), the PSWQ is a 16-item scale that gives an indication of trait worry, with higher scores indicating higher trait worry. Each item uses a 5-point Likert scale, ranging from 1 (*not at all typical of me*) to 5 (*very typical of me*), in response to statements regarding worry behaviours (for example, "Many situations make me worry"). The PSWQ is widely used across a number of clinical and research settings and is a robust measure of pathological worry that can also reliably indicate whether or not an individual has GAD (Startup & Erickson, 2006).

State-Trait Anxiety Inventory (STAI). Developed by Spielberger et al. (1983) and with established internal consistency, test-retest reliability, convergent validity, and internal validity (Guillen-Riquelme & Buela-Casal, 2011; Oei et al., 1990; Ortuno-Sierra et al., 2016), the STAI is a 40-item measure comprised of two subscales that can be used to measure state anxiety (the temporary feeling of anxiety triggered by a situation perceived as threatening) and trait anxiety (the general disposition to become anxious in situations perceived as threatening), respectively. Higher scores indicate greater levels of anxiety. Each item asks participants how frequently they feel certain things, responding on a 4-point Likert scale ranging from 1 (*almost never*) to 4 (*almost always*), in the case of the trait subscale, and from 1 (*not at all*) to 4 (*very much so*) in the case of the state subscale. The STAI is a psychometrically adequate measure of anxiety, with evidence suggesting that the scale can reliably distinguish between clinical and non-clinical anxiety (Ortuno-Sierra et al., 2016).

Beck Depression Inventory (BDI-II). Developed by Beck, Steer and Brown (1996), the BDI-II has established concurrent, content, and structural validity, as well as strong internal consistency and test-retest reliability across a number of settings, as determined by a review of 118 studies by Wang and Gorenstein (2013). The BDI-II is a 21-item measure that assesses key symptoms of depression, with higher scores indicating more severe depression. Each item asks participants to select a statement pertaining to the severity of a given depressive symptom, responding on a 4-point Likert scale ranging from 0 to 3. The BDI-II is a psychometrically strong measure that can reliably distinguish depressed and non-depressed individuals (Wang & Gorenstein, 2013).

Probe Discrimination Task (PDT). Two slightly variant versions of this task will be presented, one at the pre-training session and the other at the post-training session. The pre-training session version will consist of 72 trials, of which 50% will feature neutral-neutral word pairs, and the other 50% will feature threat-neutral word pairs.[5] All word pairs are matched for length and familiarity with emotional valence having been validated[6], and will be presented supraliminally and in a randomised

[5] We contacted the Hazen et al. (2009) authors to request the original word pairs used but were unable to obtain them. Therefore, we created the word pairs from the subset of words from MacLeod et al. (2002) as used by the original authors and paired them based on the pairing rules described in the paper.

[6] The threat-neutral word pairs had already been paired based on length and familiarity, and validated regarding their emotional valence, by MacLeod et al. (2002). However, the neutral-neutral word pairs needed to be created from the pool of neutral words present in MacLeod et al. Familiarity of the neutral words was determined by gathering frequency of use data from the Corpus of Contemporary American English (COCA; https://www.english-corpora.org/coca/), a database that has been used to determine word familiarity in previous cognitive bias research (see Feng et al., 2019). The words were paired based on their closeness in frequency of use. Then, the frequency scores were compiled into two lists (Frequency of Word A; Frequency of Word B) for the neutral-neutral word pairs and the mean frequency of the word lists compared with a bootstrapped t-test. This confirmed that the frequency did not significantly differ between the word lists (p = .875).

order, see supplementary material. Each trial will commence with a fixation cross (specifically, a plus sign) presented in the centre of the screen for 500ms. Subsequent to its disappearance, the word pair will appear on screen for 500ms, presented vertically with one word just above and the other just below the position of the fixation cross. The word pairs will be presented on a black screen and written in white letters that are .5cm tall, with the words being separated by 3cm and subtending approximately 2° of visual angle. On threat-neutral trials, the threat word will appear in either position with equal probability. Subsequent to the disappearance of the word pairs, a probe will appear in the location of either the upper or lower word. The probe will be either one dot or two dots. Participants are instructed to press one of two buttons to indicate which probe they saw. Following a response, there will be a 1000ms pause before the next trial begins (see Figure 1 for a visual example of a trial). At the post-training session, the task will be almost identical. However, this time the task will consist of 96 trials, of which 36 will feature neutral-neutral words, and the other 60 will feature threat-neutral words. This difference has been implemented in the interests of fully replicating the original paper.[7]

[7] All code for the PDT tasks and the ARTS/Sham-ARTS training is included in the supplementary material.

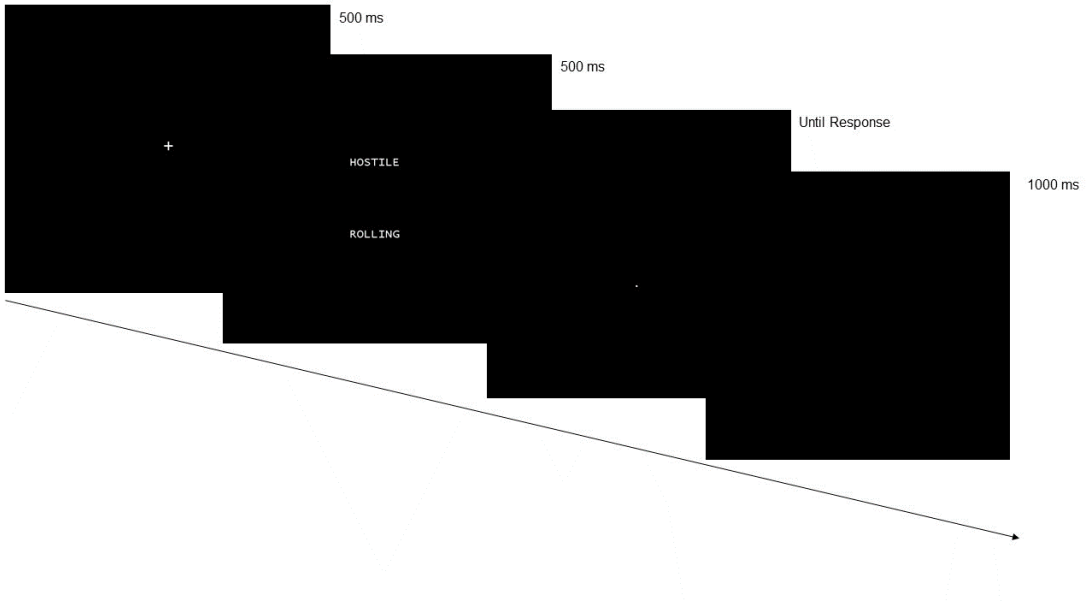

Figure 1. A typical PDT trial.

Attentional Retraining for Threat Stimuli (ARTS). This is an adapted version of the PDT designed to reduce attention to threatening stimuli that will be given to the experimental group during each training session. Each trial is identical to the PDT, except in this version the word pairs are always threat-neutral, and the probe almost always appears behind the neutral word. Each ARTS features 216 trials, in which the probe appears behind the neutral word in 204.

Sham Attentional Retraining for Threat Stimuli (Sham-ARTS). This is an adapted, placebo version of the ARTS that will be administered to the control group during each training session. Sham-ARTS is identical to the ARTS procedure, except in this version the probe appears behind the threat word and the neutral word with equal frequency, thereby not training a bias to either stimulus type.

Phenomenological Control (PC) Scale. Developed by Lush et al. (2021), the PC scale is a measure of one's ability to exercise PC, with higher scores indicating a greater ability. The measure consists of 10-items that capture experiences elicited by different imaginative suggestions and the extent to which

they are felt as real, measured on a scale ranging from 0 to 5. This scale has been found to have good internal consistency (α = .68; Lush et al., 2021).

Procedure

Potential participants who have expressed an interest in the study will be emailed a brief screening survey including an information and consent form, and the PSWQ. If they meet all eligibility requirements, which includes their newly assessed PSWQ score again being 60 or above, participants will be invited to the first session. The study will involve a total of seven sessions, all conducted by experimenters who are blind to the participants' experimental condition. The first session will be a pre-training session, in which the PSWQ, STAI, BDI, and PDT will be administered, to assess anxiety symptoms, depressive symptoms and level of attention bias at baseline. Any participants who's PSWQ score has dropped below 60 since initial screening will be excluded from the study. Before leaving this session, participants will be informed that they will be assigned to either an experimental treatment for worry or a placebo condition, but they will not know which they have received until completion of the study. Following this, sessions two through six will be training sessions, lasting approximately 15 minutes each, with participants being randomly assigned to the control or the experimental group. In line with Hazen et al. (2009), we will aim to run two training sessions a week for each participant, schedule dependent. Each training session will have participants completing either the ARTS (experimental group) or the Sham-ARTS (control group), with both groups simply being instructed to become as fast as possible at discriminating between the two types of probes without making any mistakes. Finally, the post-training session will be conducted one week after completion of the final training session. This session

will be identical to the pre-training session, except in this session the PDT will use different materials, as previously described. Additionally, before leaving participants will be asked whether they believe they received the actual training or the placebo. Then, they will be informed of their experimental condition and debriefed. PC scores will not be collected during the experimental procedure, as most of the participants will already have their PC scores in a PC database maintained by researchers at the University. Therefore, PC scores will be taken from the database and linked to the main dataset prior to analysis being conducted. The PC scale is an adapted version of the SWASH measure of hypnotisability (Lush et al., 2021). Hypnosis is a stable trait (Piccione et al., 1989), and therefore PC is equally stable.

Planned Analysis

Scoring

The PSWQ, STAI-T and BDI from the pre- and post-training sessions will be scored for each of the experimental conditions (ARTS; sham-ARTS), and the means and standard deviations for each of the measures will be calculated at pre- and post-training for each experimental condition. Additionally, a CSI will be calculated. This measure, as calculated by Hazen et al. (2009), represents an overall, standardised indicator of general mental health symptomology, inclusive of measures of anxiety and depression. Firstly, the mean PSWQ, STAI-T and BDI scores for each time point will be standardized according to normative estimates of general population means and standard deviations. These estimates of population means and standard deviations are those used in Hazen et al.'s (2009) original paper (PSWQ: $M = 45.7$, $SD = 13.5$; STAI-T: $M = 37.96$, $SD = 9.42$; BDI: $M = 7.65$, $SD = 5.9$). Then, the

mean of these three standardised scores will be calculated for each time point in each of the experimental conditions. Therefore, each CSI score represents the average level of depression and anxiety symptomology in each group relative to levels of symptomology in the general population.

The PDT from the pre- and post-training sessions will also be scored for each of the experimental conditions. Firstly, in line with Hazen et al.'s (2009) exclusion criteria, RTs from neutral-neutral word pair trials, trials in which participants fail to accurately detect the probe or trials with extreme RT values (RT < 150ms, or RT > 1500ms) will be excluded from analysis. If more than 10% of a participant's pre- or post-training PDT trials are excluded due to these criteria, then that participant's data will be excluded from analysis. In the PDT task, the position of probes is crossed with threat word position to create four within-subject conditions: threat upper\probe upper (TU\PU); threat upper\probe lower (TU\PL); threat lower\probe upper (TL\PU); threat lower\probe lower (TL\PL). Harmonic means will be calculated for each of these four conditions, in line with Hazen et al.'s (2009) method. Then, the four conditions will be combined using the following formula to create an attention bias score: [(TU\PL - TU\PU) + (TL\PU - TL\PL)] / 2. This results in a final attention bias score in which a positive value represents faster discrimination of probes following threat words as opposed to neutral words (biased attention towards threatening words), and a negative value represents faster discrimination of probes following neutral words as opposed to threat words (reduced attention towards threatening words). Finally, the mean and standard deviation for the PDT will be calculated at pre- and post-training for each experimental condition.

Preliminary Analyses

Bayesian Credibility Intervals, assuming a uniform prior and normal approximation, will be calculated for the difference in pre-training scores between conditions (ARTS vs. Sham-ARTS) for each of the four measures (PSWQ, STAI-T, BDI, PDT). Further, for each experimental condition, Bayesian Credibility Intervals, again assuming a uniform prior and normal approximation, will be calculated for the difference between the pre-training scores for the PSWQ and STAI-T and their corresponding general population means (as previously derived by Hazen et al. 2009).

Main Analyses

A series of Bayesian t-tests will be conducted to assess each hypothesis. For all Bayes Factors we will adopt the conventional thresholds of values greater than 3 indicating evidence for the alternate hypothesis and values less than 1/3rd indicating evidence for the null. Robustness regions will be reported as: RRconclusion [x1, x2], where x1 is the smallest and x2 is the largest SD that gives the same conclusion: $B < 1/3$; $1/3 < B < 3$; $B > 3$. All Bayes factors will be calculated using an online Bayes factor calculator (URL: https://harry-tattan-birch.shinyapps.io/bayes-factor-calculator/). For every Bayes factor, we will also report the corresponding t and p-values.

Firstly, the change in attention bias scores will be assessed. A Bayes factor will be computed on the difference between groups in their respective change in attention bias scores from pre- to post-training (ARTS pre-post attention bias minus Sham-ARTS pre-post attention bias). H_1 will be modelled as a half-normal distribution with a mode of 0 and *SD* of 17.49. This *SD* is the raw effect size

originally observed by Hazen et al. (2009; see Table 2, *'Mean Threat Bias Scores by Group'*).

Next, the change in CSI score will be assessed. Bayes factors will be computed on the difference between groups in their respective change in CSI scores from pre- to post-training (ARTS pre-post CSI minus Sham-ARTS pre-post CSI). H_1 will be modelled as a half-normal distribution with a mode of 0 and *SD* of 0.86. This *SD* is the raw effect size originally observed by Hazen et al. (2009; see Table 1, *'Means and SDs for composite symptom index and individual measures by Group and Time'*).

In addition to the comparisons conducted by the original authors, Bayes factors will be computed on the difference between groups in their respective changes in each of the individual mental health indicators (PSWQ, STAI-T and BDI), pre-post training as above. All SDs are the raw effect sizes originally observed by Hazen et al. (2009; see Table 1, *'Means and SDs for composite symptom index and individual measures by Group and Time'*). For the mean difference in PSWQ scores, H_1 will be modelled as a half-normal distribution with a mode of 0 and *SD* of 8.14. For the mean difference in STAI-T scores, H_1 will be modelled as a half-normal distribution with a mode of 0 and *SD* of 6.72. For the mean difference in BDI scores, H_1 will be modelled as a half-normal distribution with mode of 0 and *SD* of 7.25.

Further to these, an additional analysis will be conducted to assess the potential impact of demand characteristics on the outcome of the intervention on symptomology. Specifically, we will test the prediction that higher PC scores will lead to greater reductions in the CSI in the ARTS vs. the Sham Arts condition. This prediction is derived from the hypothesis that individuals with higher PC scores may (consciously or unconsciously) utilise PC to experience the expected change in

symptomology, thus showing a greater response to demand characteristics (Lush et al., 2020). As such, the extent to which this hypothesised relationship is observed will be an indication of the degree to which demand characteristics are influencing the results. This will be assessed by examining the interaction between Condition (ARTS vs. Sham ARTS) and PC in a multiple regression predicting change in CSI. If the interaction term is greater than zero, this will indicate that PC is influencing the apparent efficacy of ABM. The difference from zero will be evaluated by applying a Bayes Factor to the interaction term. To the authors' knowledge, no research has yet investigated the relationship between PC and the efficacy of ABM. Therefore, a theory of the relationship based on the ratio-of-means heuristic (Dienes, 2019) will be used to model H_1 for the Bayes Factor. If we theorise that PC is required for a change in CSI, then there can be no change in CSI without using PC and thus both scales will approach zero together. This is an idealised theory in which change in CSI cannot occur independently of PC. However, if we theorise that PC is one important contributory factor leading to a change in CSI, then the ratio represents the maximum interaction effect one may expect to observe. Therefore, the ratio of the mean CSI change score and the mean PC score will be used as an estimate of the maximum slope. Specifically, H_1 will be modelled as a half-normal distribution with a mode of 0 and *SD* = (Mean CSI Change score / Mean PC score)/2.

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Amir, N., Beard, C., Burns, M., Bomyea, J. (2009). Attention modification program in individuals with generalized anxiety disorder. *Journal of Abnormal Psychology, 118*(1), 28-33. https://doi.org/10.1037/a0012589

Beard, C., & Peckham, A. D. (2020). Interpretation bias modification. In J. S. Abramowitz & S. M. Blakey (Eds.), *Clinical handbook of fear and anxiety: Maintenance processes and treatment mechanisms* (pp. 359–377). American Psychological Association. https://doi.org/10.1037/0000150-020

Beck, A. T., Steer, R. A,, & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II.* San Antonio, TX: Psychological Corporation

Behar, E., Alcaine, O., Zuellig, A. R., & Borkovec, T. D. (2003). Screening for generalized anxiety disorder using the Penn State Worry Questionnaire: A receiver operating characteristic analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 34*(1), 25–43. https://doi.org/10.1016/ S0005-7916(03)00004-1

Brown, T. A., Antony, M. M., & Barlow, D. H. (1992). Psychometric properties of the Penn State Worry Questionnaire in a clinical anxiety disorders sample. *Behavior Research and Therapy, 30*(1), 33–37. https://doi.org/10.1016/0005-7967(92)90093-V

Chambers, C. D. (2013). Registered reports: a new publishing initiative at Cortex. *Cortex*, *49*(3), 609-610. https://doi.org/10.1016/j.cortex.2012.12.016

Clarke, P. J., Notebaert, L., & MacLeod, C. (2014). Absence of evidence or evidence of absence: reflecting on therapeutic implementations of attentional bias modification. *BMC psychiatry, 14*(1), 1-6. https://doi.org/10.1186/1471-244X-14-8

Cristea, I. (2018). Author's reply. *The British Journal of Psychiatry, 212*(4), 247-247. doi:10.1192/bjp.2018.43

Cristea, I. A., Mogoașe, C., David, D., & Cuijpers, P. (2015). Practitioner review: Cognitive bias modification for mental health problems in children and adolescents: A meta‐analysis. *Journal of Child Psychology and Psychiatry, 56*(7), 723-734. https://doi.org/10.1111/jcpp.12383

Dienes, Z. (2019). How do I know what my theory predicts?. *Advances in Methods and Practices in Psychological Science*, *2*(4), 364-377. https://doi.org/10.1177/2515245919876960

Dienes, Z. (2021). Obtaining evidence for no effect. *Collabra: Psychology*, *7*(1), 28202. https://doi.org/10.1525/collabra.28202

Dienes, Z., Lush, P., Palfi, B., Roseboom, W., Scott, R., Parris, B., . . . Lovell, M. (2022). Phenomenological control as cold control. *Psychology of Consciousness: Theory, Research, and Practice, 9*(2), 101-116. doi:https://doi.org/10.1037/cns0000230

Feng, Y. C., Krahe, C., Sumich, A., Meeten, F., Lau, J. Y., & Hirsch, C. R. (2019). Using event-related potential and behavioural evidence to understand interpretation bias in relation to worry. *Biological Psychology*, *148*, 107746. https://doi.org/10.1016/j.biopsycho.2019.107746

Goodwin, H., Yiend, J., & Hirsch, C.R. (2017). Generalized Anxiety Disorder, worry and attention to threat: A systematic review. *Clinical Psychology Review, 54,* 107-122. https://doi.org/10.1016/j.cpr.2017.03.006

Grafton, B., MacLeod, C., Rudaizky, D., Holmes, E. A., Salemink, E., Fox, E., & Notebaert, L. (2017). Confusing procedures with process when appraising the impact of cognitive bias modification on emotional vulnerability. *The British Journal of Psychiatry, 211*(5), 266-271. doi:10.1192/bjp.bp.115.176123

Guillen-Riquelme, A., & Buela-Casal, G. (2011). Psychometric revision and differential item functioning in the State Trait Anxiety Inventory (STAI). *Psicothema*, *23*(3), 510-515.

Hakamata, Y., Lissek, S., Bar-Haim, Y., Britton, J. C., Fox, N. A., Leibenluft, E., ... & Pine, D. S. (2010). Attention bias modification treatment: a meta-analysis

toward the establishment of novel treatment for anxiety. *Biological psychiatry, 68*(11), 982-990. https://doi.org/10.1016/j.biopsych.2010.07.021

Hallion, L. S., & Ruscio, A. M. (2011). A meta-analysis of the effect of cognitive bias modification on anxiety and depression. *Psychological bulletin, 137*(6), 940-958. https://doi.org/10.1037/a0024355

Hazen, R.A., Vasey, M.W., & Schmidt, N.B. (2009). Attentional retraining: A randomized clinical trial for pathological worry. *Journal of psychiatric research, 43*(6), 627-633. https://doi.org/10.1016/j.jpsychires.2008.07.004

Hill, A. B. (1965). The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine*, *58*(5), 295–300. https://doi.org/10.1177/003591576505800503

Hirsch, C. R., & Mathews, A. (2012). A cognitive model of pathological worry. *Behaviour research and therapy, 50*(10), 636-646. https://doi.org/10.1016/j.brat.2012.06.007

Hirsch, C. R., Clark, D. M., & Mathews, A. (2006). Imagery and interpretations in social phobia: Support for the combined cognitive biases hypothesis. *Behavior therapy*, *37*(3), 223-236. https://doi.org/10.1016/j.beth.2006.02.001

Hirsch, C. R., Krahé, C., Whyte, J., Bridge, L., Loizou, S., Norton, S., & Mathews, A. (2020). Effects of modifying interpretation bias on transdiagnostic repetitive negative thinking. *Journal of consulting and clinical psychology*, *88*(3), 226. https://doi.org/10.1037/ccp0000455

Hirsch, C. R., Krahé, C., Whyte, J., Loizou, S., Bridge, L., Norton, S., & Mathews, A. (2018). Interpretation training to target repetitive negative thinking in

generalized anxiety disorder and depression. *Journal of consulting and clinical psychology, 86*(12), 1017. https://doi.org/10.1037/ccp0000310

Hirsch, C.R., Krahé, C., Whyte, J., Krzyzanowski, H., Meeten, F., Norton, S., & Mathews, A. (2021). Internet-Delivered Interpretation Training Reduces Worry and Anxiety in Individuals With Generalized Anxiety Disorder. *Journal of Consulting and Clinical Psychology, 89*(7), 575-589. https://doi.org/10.1037/ccp0000660

Hu, C., Wang, F., Guo, J., Song, M., Sui, J., & Peng, K. (2016). The replication crisis in psychological research. *Advances in Psychological Science*, *24*(9), 1504. doi: 10.3724/SP.J.1042.2016.01504

Ji, J. L., Baee, S., Zhang, D., Calicho-Mamani, C. P., Meyer, M. J., Funk, D., ... & Teachman, B. A. (2021). Multi-session online interpretation bias training for anxiety in a community sample. *Behaviour research and therapy*, *142*, 103864. https://doi.org/10.1016/j.brat.2021.103864

Jia, R., Ayling, K., Chalder, T., et al. (2020). Mental health in the UK during the COVID-19 pandemic. *BMJ Open*;10:e040620. doi:10.1136/bmjopen-2020-040620

Kruijt, A., & Carlbring, P. (2018). Processing confusing procedures in the recent re-analysis of a cognitive bias modification meta-analysis. *The British Journal of Psychiatry, 212*(4), 246-246. doi:10.1192/bjp.2018.41

Linetzky, M., Pergamin‐Hight, L., Pine, D. S., & Bar‐Haim, Y. (2015). Quantitative evaluation of the clinical efficacy of attention bias modification treatment for anxiety disorders. *Depression and anxiety, 32*(6), 383-391. https://doi.org/10.1002/da.22344

Liu, H., Li, X., Han, B., & Liu, X. (2017). Effects of cognitive bias modification on social anxiety: A meta-analysis. *PloS one*, *12*(4), e0175107. https://doi.org/10.1371/journal.pone.0175107

Loerinc, A.G. et al. (2015). Response rates for CBT for anxiety. *Clinical Psychology Review, 42,* 72–82. https://doi.org/10.1016/j.cpr.2015.08.004

Lush, P, Botan, V, Scott, R B, Seth, A K, Ward, J and Dienes, Z. (2020). Trait phenomenological control predicts experience of mirror synaesthesia and the rubber hand illusion. *Nature*, 11. a4853 1-10. ISSN 0028-0836 https://doi.org/10.1038/s41467-020-18591-6

Lush, P., Scott, R. B., Seth, A. K., & Dienes, Z. (2021). The phenomenological control scale: Measuring the capacity for creating illusory nonvolition, hallucination and delusion. *Collabra: Psychology*, *7*(1), 29542. https://doi.org/10.1525/collabra.29542

MacLeod, C., Rutherford, E., Campbell, L., Ebsworthy, G., & Holker, L. (2002). Selective Attention and Emotional Vulnerability: Assessing the Causal Basis of Their Association Through the Experimental Manipulation of Attentional Bias. *Journal of Abnormal Psychology (1965)*, *111*(1), 107–123. https://doi.org/10.1037/0021-843X.111.1.107

Martinelli, A., Grüll, J., & Baum, C. (2022). Attention and interpretation cognitive bias change: A systematic review and meta-analysis of bias modification paradigms. *Behaviour Research and Therapy,* 104180. https://doi.org/10.1016/j.brat.2022.104180

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the penn state worry questionnaire. *Behaviour research and therapy, 28*(6), 487-495. https://doi.org/10.1016/0005-7967(90)90135-6

Mogg, K., & Bradley, B. P. (2005). Attentional bias in generalized anxiety disorder versus depressive disorder. *Cognitive therapy and research*, *29*, 29-45. https://doi.org/10.1007/s10608-005-1646-y

Mogg, K., Waters, A.M., & Bradley, B.P. (2017). Attention bias modification (ABM): Review of effects of multisession ABM training on anxiety. *Clinical Psychological Science, 5*(4), 698-717. https://doi.org/10.1177/2167702617696359

Mogoaşe, C., David, D., & Koster, E. H. (2014). Clinical efficacy of attentional bias modification procedures: An updated meta‐analysis. *Journal of Clinical Psychology, 70*(12), 1133-1157. https://doi.org/10.1002/jclp.22081

Nichols, A. L., & Maner, J. K. (2008). The Good-Subject Effect: Investigating Participant Demand Characteristics. *The Journal of General Psychology*, *135*(2), 151–166. https://doi.org/10.3200/GENP.135.2.151-166

Oei, T. P., Evans, L., & Crook, G. M. (1990). Utility and validity of the STAI with anxiety disorder patients. *British Journal of Clinical Psychology, 29*(4), 429-432. https://doi.org/10.1111/j.2044-8260.1990.tb00906.x

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716. https://doi.org/10.1126/science.aac4716

Orne, M. T. (1962). On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications. *Am. Psychol. 17,* 776–783. https://doi.org/10.1037/h0043424

Ortuno-Sierra, J., Garcia-Velasco, L., Inchausti, F., Debbane, M., & Fonseca-Pedrero, E. (2016). New approaches on the study of the psychometric properties of the STAI. *Actas espanolas de psiquiatria*, *44*(3), 83-92.

Palfi, B., & Dienes, Z. (2019). The role of Bayes factors in testing interactions. https://doi.org/10.31234/osf.io/qjrg4

Parsons, S. (2018). *Moving forward with questions of process and procedure in cognitive bias modification research: Three points of consideration.* OSF preprint. https://dx.doi.org/10.31234/osf.io/k3vxc

Piccione, C., Hilgard, E. R., & Zimbardo, P. G. (1989). On the degree of stability of measured hypnotizability over a 25-year period. *Journal of Personality and Social Psychology, 56*(2), 289–295. https://doi.org/10.1037/0022-3514.56.2.289

Ruscio, A. M., Hallion, L. S., Lim, C. C., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., ... & Scott, K. M. (2017). Cross-sectional comparison of the epidemiology of DSM-5 generalized anxiety disorder across the globe. *JAMA psychiatry*, *74*(5), 465-475. doi:10.1001/jamapsychiatry.2017.0056

Schönbrodt, F. D., & Wagenmakers, E. J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic bulletin & review*, *25*(1), 128-142. https://doi.org/10.3758/s13423-017-1230-y

Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory STAI (Form Y).* Palo Alto, CA: Consulting Psychologists Press.

Startup, E., & Erickson, T. M. (2006). The Penn State Worry Questionnaire (PSWQ). In G. C. L. Davey & A. Wells (Eds.), *Worry and its psychological disorders* (pp. 101–119). Wiley Publishing.

Stöber, J. (1998). Reliability and validity of two widely-used worry questionnaires: Self-report and self-peer convergence. *Personality and Individual differences, 24*(6), 887-890. https://doi.org/10.1016/S0191-8869(97)00232-8

Van Bockstaele, B., Verschuere, B., Tibboel, H., De Houwer, J., Crombez, G., & Koster, E. H. W. (2014). A review of current evidence for the causal impact of attentional bias on fear and anxiety. *Psychological Bulletin, 140*(3), 682-721. doi:https://doi.org/10.1037/a0034834

Wang, Y. P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: a comprehensive review. *Brazilian Journal of Psychiatry, 35,* 416-431. https://doi.org/10.1590/1516-4446-2012-1048

Wang, Y., Xiao, R., Luo, C., & Yang, L. (2019). Attentional disengagement from negative natural sounds for high-anxious individuals. *Anxiety, Stress, & Coping*, *32*(3), 298-311. https://doi.org/10.1080/10615806.2019.1583539

Wittchen, H.U. (2002). Generalized anxiety disorder: Prevalence, burden, and cost to society. *Depression and Anxiety, 16*(4), 162–171. http://doi.org/10.1002/da.10065