Running head: CLINICAL EFFICACY OF ATTENTIONAL RETRAINING

Clinical Efficacy of Attentional Bias Modification Procedures: An Updated Meta-analysis

Cristina Mogoaşe and Daniel David

Babeş-Bolyai University, Cluj-Napoca, Romania

Ernst H.W. Koster

Gent University, Belgium

Reference: Mogoase, C., David, D., & Koster, E.H.W. (in press). Clinical Efficacy of Attentional Bias Modification Procedures: An Updated Meta-analysis. Journal of Clinical Psychology.

Author Note

Cristina Mogoaşe, Department of Clinical Psychology and Psychotherapy, Babeş-Bolyai University, Romania and Department of Medicine, Iuliu Haţieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania; Daniel David, Department of Clinical Psychology and Psychotherapy, Babeş-Bolyai University, Cluj-Napoca, Romania and Mount Sinai School of Medicine, New York; Ernst H.W. Koster, Department of Experimental Clinical and Health Psychology, Ghent University, Belgium.

Correspondence concerning this article should be addressed to: Cristina Mogoaşe, Department of Clinical Psychology and Psychotherapy, Babeş-Bolyai University, No. 37, Republicii Street 400015, Cluj-Napoca, Cluj, Romania. Tel/Fax: + 40 264434141; E-mail: cristina.mogoase@ubbcluj.ro

Abstract

Context: Attentional bias modification (ABM) procedures is a promising intervention tool for a variety of clinical conditions. **Objectives:** This study provides an updated review of the clinical impact of ABM by employing standard meta-analytic procedures to: (a) estimate the average effect size of ABM in reducing both attention bias (AB) and symptoms; (b) estimate the average effect size for different conditions (e.g., anxiety, depression, substance abuse); (c) test possible variables that may moderate the effect sizes, and (d) investigate the relationship between pre-existent AB and the reduction in AB and symptoms. **Method:** We included 43 controlled trials with a total of 2268 participants providing 47 group comparisons (i.e., training vs. control condition). Inclusion criteria were: AB was specifically targeted to reduce symptomatology and/or emotional vulnerability; participants were randomized to the experimental conditions; a control condition (defined as sham training) existed; symptoms were assessed at least post-intervention; sufficient data were provided to allow effect size estimation. **Results:** We obtained a small overall effect size on symptoms post intervention, g = 0.160, 95% CI = [0.055; 0.265], driven by anxiety studies, g = 0.260, 95% CI = [0.132; 0.388], and studies conducted in healthy participants, g = 0.211, 95% CI = [0.046; 0.375]; no significant effect sizes were found post intervention for other symptom categories. **Conclusion:** The therapeutic benefit of ABM is rather small for anxiety, while the amount of data for other symptom categories is limited. We argue that more efficient, psychometrically sound procedures are needed for assessing and modifying AB.

Key words: attentional bias modification, meta-analysis, anxiety, depression, addictive behaviors.

Clinical Efficacy of Attentional Bias Modification Procedures: An Updated Meta-analysis

Introduction

In recent years, an extensive body of research on attentional bias modification (ABM) procedures has accumulated. The ABM procedures build upon cognitive theories of psychopathology, which assume that attentional biases (AB) are critical cognitive mechanisms underlying a wide range of clinical conditions. The clinical purpose of ABM procedures is to reduce excessive allocation of attention to disorder-relevant information (e.g., threat, negative information or addiction-related material); by targeting this critical cognitive process, ABM is expected to reduce symptoms, emotional reactivity or craving (MacLeod & Mathews, 2012).

To date, three experimental paradigms have been used for ABM: visual probe (or dot-probe) task, emotional spatial (or visual) cueing task, and visual search task. A common feature of these procedures is that they manipulate attention allocation to disorder-relevant stimuli when these compete for attention with disorder-incompatible stimuli in such a way that performance on the task benefits from allocating attention towards the disorder-incompatible stimuli (Koster, Fox, & MacLeod, 2009).

Most research so far employed the modified visual probe task (MacLeod, Rutherford, Campbell, Ebsworthy, & Holker, 2002). In this task, two stimuli are simultaneously presented for brief durations (e.g., 500ms) either at one side of the fixation cross (i.e., left/right or up/down). The valence of the stimuli is manipulated in that one stimulus has a disorder-congruent valence (e.g., angry facial expression) whereas the other has a disorder incongruent valence (e.g., happy or neutral facial expression). Next, a neutral probe (e.g., a letter, an asterix or a dot) appears at the location previously occupied by one of the two stimuli (i.e., right/left or up/down). Participants either indicate the location of the probe or discriminate between different probes (e.g., “E” or “F”) as quickly and accurately as possible. To train attention away from disorder-relevant stimuli and towards disorder-incompatible stimuli, the target consistently replaces the disorder-incompatible stimulus during the training phase.

Other studies used a modified version of the emotional spatial cueing task (Fox, Russo, Bowles, & Dutton, 2001; Posner, 1980) to train attention. This task presents an emotional cue in one of two possible

locations, followed by a target stimulus either at the cued location or at the alternative location. AB is revealed by facilitated target responding when the target is presented at the location of the disorder relevant cue or delayed target responding when the target is presented at the opposite location of the disorder-relevant cue. In the training variant, a benign AB is obtained by having the target stimulus never appear at the location previously occupied by a disorder-relevant stimulus (Baert, De Raedt, Schacht, & Koster, 2010; Bar-Haim, Morag, & Glickman, 2011).

Finally, other studies (e.g., Dandeneau, Baldwin, Baccus, Sakellaropoulo, & Pruessner, 2007) used the visual search task to manipulate attention. This task involves repeatedly asking the participant to find a target disorder-incompatible stimulus (e.g., a smiling face) among distracting disorder-relevant stimuli (e.g., frowning faces). The participants are encouraged to find the target stimulus as soon as possible. It is assumed that, through repetitive practice, they implicitly learn to overcome their tendency to preferentially process disorder-relevant stimuli. Instead, they learn a pattern of attentional selectivity favoring positive information.

Seminal studies suggested that ABM procedures have considerable clinical potential in that training efficiently leads to clinical improvement with little time investment of the patient at minimal costs (Amir, Beard, Burns, & Bomyea, 2009; Amir, Beard, Taylor, et al., 2009; Schmidt, Richey, Buckner, & Timpano, 2009). Because of such findings and the fact that ABM can be easily disseminated, several narrative reviews have pointed out that ABM has the potential to become an important clinical intervention for the treatment and prevention of psychopathology (Bar-Haim, 2010; Browning, Holmes, & Harmer, 2010; Hertel & Mathews, 2011; MacLeod & Mathews, 2012). Moreover, three meta-analyses (Beard, Sawyer, & Hofmann, 2012; Hakamata et al., 2010; Hallion & Ruscio, 2011) examined ABM effects on symptoms and AB change. However, two main issues of these syntheses of the ABM literature should be noted: namely, (1) their reports are inconsistent, both in terms of change in AB and symptoms, and (2) none of them included any of the recent negative findings reported with the ABM, although all of them reported publication bias. We briefly discuss each of these limitations below.

Limitations of the existing reviews of the ABM clinical efficacy

1. Inconsistent findings

Some of the previous meta-analyses indicated large effect sizes for change in AB (Beard, Sawyer, et al., 2012; Hakamata et al., 2010), while others indicated small effect sizes (Hallion & Ruscio, 2011). Similarly, for change in symptoms, some reported a medium effect size (Hakamata et al., 2010), while others indicated a small effect size (Beard, Sawyer, et al., 2012; Hallion & Ruscio, 2011). These incongruent findings may be explained in terms of the methodological approach of these meta-analyses. A first issue concerns the range of studies included in each of them. For instance, Hakamata et al. (2010) included only anxiety studies employing ABM, while Hallion and Ruscio (2011) included only anxiety and depression studies employing ABM and/or interpretation bias modification, but investigated the effects of ABM and interpretation bias modification mainly together (e.g., without reporting separate effect sizes for ABM's versus interpretation bias modification's impact on symptoms). This is unfortunate as the different biases (AB, interpretation bias) may operate in different ways in psychopathology and there may also be differences in the plasticity of these biases by training.

Beard et al. (2012) included a wide range of samples and symptoms. However, they did not report different estimates of ABM effect on symptoms by diagnostic status of participants or symptom category. This is problematic because it may be easier to change AB in healthy compared with clinical or analogue samples. In addition, although a diversity of clinical conditions have been found to be characterized by AB (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007; Beard & Amir, 2010; Brooks, Prince, Stahl, Campbell, & Treasure, 2011; Chapman & Martin, 2011; Field & Cox, 2008; Hertel & Mathews, 2011), they present considerable differences in symptoms, etiology, and prognosis, and AB has been found to display different particularities as a function of disorder. Therefore, it may play a different functional role in different disorders (e.g., Mathews & MacLeod, 2005), meaning that it can be causally involved in symptoms, may contribute to the maintenance of symptoms, or may be merely a characteristic of the symptoms. Thus, ABM might be more effective for some psychological conditions (e.g., anxiety) than

for others. Such information is crucial for clinical applications and may help orienting subsequent research. In addition, as AB is thought to be the ABM mechanism of change, it would make sense to consider the pre-existent AB level when investigating how successful is ABM in inducing the desired changes. However, none of the existing quantitative syntheses of ABM literature considered the role of pre-existing AB in relationship with change in AB and/or change in symptoms. This could have contributed also to their inconsistent findings.

2. Publication bias

Previous ABM meta-analyses reported evidence of publication bias for ABM findings. Arguably, this limits our understanding of the current state of research on ABM. No wonder that current views on the clinical utility of ABM vary widely among researchers and clinicians. Some hold positive views, advocating the ABM clinical potential (e.g., Bar-Haim, 2010; Hakamata et al., 2010; MacLeod & Mathews, 2012), while others have negative and skeptical views (Emmelkamp, 2012), pointing out that ABM is efficient only in laboratory-based studies and there is no robust evidence of ABM efficiency in clinical samples. Several additional controlled clinical studies have been reported since the last meta-analyses, many of which reported negative results with ABM (e.g., Boettcher, Berger, & Renneberg, 2012; Carlbring et al., 2012; Julian, Beard, Schmidt, Powers, & Smits, 2012; Neubauer et al., 2013). Notably, none of the previous meta-analyses included these recent negative findings. Arguably, including such failures to replicate promising findings is critical in assessing the clinical relevance of ABM. Therefore, an updated quantitative review with a clinical focus is timely.

The present study

This paper is aimed to comprehensively examine the clinical efficacy of ABM, both in terms of outcome (i.e., reducing and/or preventing symptoms of subjective distress, dysfunctional behaviors, and biological markers of psychopathology) and in terms of the presumed mechanism of change (i.e., AB). To this end, we considered studies that trained attention *away* from disorder relevant stimuli (congruent with the theory-specified direction of clinical improvement) and compared this intervention with an adequate control

group (i.e., no training of attention). This allowed us to provide a global estimate of ABM effect size on symptoms and AB. Second, we aimed to investigate the degree to which ABM yields therapeutic benefits for different symptom categories. Third, we aimed to test possible moderators of the ABM effect. Finally, we were interested to investigate the relationship between pre-existent AB and the reduction in AB and symptoms.

Based on the potential moderators considered in the previous reviews and on the theoretically-derived assumptions about the factors affecting ABM efficacy, we considered the following potential moderators:

1. Type of psychopathology. Hallion and Ruscio (2011) reported a small ABM effect size for anxiety studies, and non-significant effect for depression studies. However, their results reflected the combined effect of ABM and interpretation bias modification on symptom reduction. As the number of ABM studies has doubled since that time, it is important to examine whether their results can be replicated and extended to other types of symptoms (e.g., substance abuse symptoms). In addition, we aim to refine the previous reported results (Hakamata et al.; 2010; Hallion & Ruscio, 2011) by examining whether, within anxiety studies, the ABM effect varies as function of the different types of anxiety problems.

2. Clinical status of the sample was not found to be a significant moderator of the ABM effect in previous meta-analyses (Hakamata et al., 2010; Hallion & Ruscio, 2011). However, as our sample of ABM studies is considerably larger, we aimed to replicate these results. We considered three sample categories: diagnosed participants (clinical samples), undiagnosed participants with elevated symptoms (analogue samples), and healthy participants who were exposed to a stressor following ABM (as we were interested in the buffering role that ABM could play in relation to negative emotional reactivity). Although any protective effect of the ABM in healthy participants would be important, it is essential to demonstrate ABM is capable to modify AB and symptoms in clinical samples and/or to reduce relapse in recovered samples.

3. Training methodology. Previous meta-analyses (Beard et al., 2012; Hakamata et al., 2010) reported that type of stimuli used for training (linguistic versus pictorial stimuli) and stimuli position during training (top-bottom versus left-right) significantly moderated the ABM effect on symptoms, with studies using words

and top-bottom orientation of the stimuli yielding larger effects. Hallion and Ruscio (2011) as well as Beard et al. (2012) reported the number of training sessions to be a significant moderator of change in symptoms assessed post-intervention, but not as assessed following an experimental stressor. However, Hakamata et al. (2010) did not find the number of session to moderate the ABM impact on symptoms. To clarify the potential moderators related to training methodology, we re-examined as moderators the type of stimuli used for training as well as number of sessions in a larger sample of studies and with different symptom categories. In addition, we considered several other potential moderators: the number of trials per session, type of training task, nature of the ABM training (i.e., towards positive versus towards neutral stimuli), training setting (i.e., laboratory versus home), and temporal separation of ABM sessions.

4. Type of outcomes. We considered two clinically important aspects of the ABM outcomes, namely (1) primary versus secondary interest variables, and (2) self-report, clinician-rated, and bio-behavioral measures. If training attention away from negative information broadly reduces stress reactivity (i.e., secondary outcomes) it presumably impacts on primary as well as secondary interest outcomes. Conversely, there may be a specific effect depending on the relevance of the training stimuli relative to the specific investigated condition: the effect of ABM designed for lowering depressive symptoms, for example, may be larger on depression measures (primary outcome) than on general distress or anxiety measures (secondary outcome). In addition, treatment change would be more reliable if it is observed on all type of the outcome measures (self-report, clinician-rated, and bio-behavioral measures). If the improvement is evident only on self-report measures, it could be more liable to demand effects or be less robust.

5. Participants' age. Hakamata et al. (2010) reported that participants' age did not significantly moderate the ABM effect. However, there was relatively little variability among studies included in their meta-analysis, as most participants were students. The participants' age in ABM studies has expanded lately considerably, with studies conducted on children (e.g., Bar-Haim et al., 2011; Eldar et al., 2012) or older adults (e.g., Schoorl, Putman, & Van Der Does, 2013; Sharpe et al., 2012) being added to the field. We treated participants' age as a continuous moderator.

We investigated the relationship between AB and symptom change in two ways. First, we examined the relation between the pre-existent AB and the reduction in AB and symptoms, respectively, given that the reduction in AB is the presumed mechanism of change of ABM interventions. Yet, none of the previous meta-analyses considered the role of the pre-existent AB in relation with ABM efficacy. Second, we examine the relation between AB change and symptom changes. Previous meta-analyses failed to find a statistically significant relationship between reduction in AB and reduction in symptoms following ABM (Hakamata et al., 2010; Hallion & Ruscio, 2011). This may be due to the possibility that ABM lowers symptoms via AB reduction only in persons with a pre-existing AB, where this could be an important selection criteria to enroll in training (see Eldar et al., 2012).

This meta-analysis has both theoretical and practical implications. From a theoretical point of view, as compared to previous meta-analyses in the field, it brings the following new innovations: it includes newer studies that failed to replicate the original positive findings with ABM; it assesses the pre-existing bias levels; and it looks for differential ABM effect within anxiety studies. All of these have potential to further inform the research work in the field. From a practical point of view, this meta-analysis is timely because ABM is already marketed to anxiety patients. Therefore, to rigorously guide the clinical practice of psychology we need a comprehensive meta-analysis, organized from a theoretical point of view and which tries to delineate clear practical implications and future directions of development the clinical utility of ABM procedures.

Method

Literature search

Potential relevant studies were identified through a systematic search of the ISI Web of Science, Scopus, and Medline databases through July 2013, using the following search terms: “attentional bias modification”, and “attention bias” combined with “attentional (re)training” and “experimental manipulation”. We also systematically searched the references from the empirical papers, meta-analyses, and reviews on the topic.

We identified 1231 records (for details see Figure 1). Duplicates were removed and the remaining records were screened based on title and abstract. After excluding clearly irrelevant publications, a total of 86 potential relevant articles were retained for detailed inspection. Their full-texts were reviewed for inclusion in the meta-analysis.

(Insert Figure 1 about here)

Inclusion criteria

The following criteria were applied for inclusion in the meta-analysis: (a) the study was designed to manipulate AB to reduce symptoms and/or lower emotional vulnerability (in the latter case, to be included a study should have included at least one measure of distress); (b) the study assessed clinically-relevant symptoms; (c) participants were randomized to training conditions; (d) a control condition (defined as sham training) existed; (e) the study was written in English and published/accepted for publication in a peer-reviewed journal; (f) sufficient data to compute effect size were available. We did not include studies that: investigated the effect of other types of cognitive bias modification (e.g., interpretation bias; see Amir, Bomyea, & Beard, 2010; Brosan, Hoppitt, Shelfer, Sillence, & Mackintosh, 2011) or used attention (re)training to improve attention functioning and not for modifying AB (Dvorkin et al., 2013) or symptoms (e.g., Van Bockstaele, Koster, Verschuere, Crombez, & De Houwer, 2012).

Selection of the comparison groups

Some studies included a third group (trained to attend to threat stimuli; e.g., Heeren, Reese, McNally, & Philippot, 2012; Klumpp & Amir, 2010), but we focus exclusively on comparing the ABM group trained towards the neutral/positive stimuli to the control group. We chose to do this because we were specifically interested in the clinical implications of AB reduction rather than in the consequences of experimental induction of AB.

There were two studies that investigated ABM effect using four groups, two experimental and two control groups (Browning, Holmes, Charles, Cowen, & Harmer, 2012; Julian et al., 2012). Browning and

colleagues (2012) investigated the efficacy of ABM in (1) a training group using pictorial stimuli compared with a control group using the same pictorial stimuli, and (2) a training group using linguistic stimuli compared with a control group using the same linguistic stimuli. Julian and colleagues (2012) compared a typical ABM intervention (involving one experimental and one control group) with an enhanced ABM intervention, in which participants in both the training and the control groups did exercise before ABM. We extracted two sets of data from each of these studies, contrasting the pictorial ABM group with the pictorial control group and the linguistic ABM group with the linguistic control group for Browning et al. (2012), while for Julian et al. (2012) we compared typical ABM with the control group, and the enhanced ABM with the enhanced control group. Therefore, these two studies were analyzed at the level of their four different conditions (which is warranted as each condition contains new participants).

Two other studies deserve specific consideration here. Eldar and Bar-Haim (2010) investigated ABM in a sample of low and high anxious participants randomized to experimental and control conditions. We extracted two sets of data, related to the ABM impact on high anxious persons, and low anxious persons, respectively. Amir, Taylor, & Donohue (2011) reported findings on an extended sample: the authors had reported previously findings from a subsample of it (Amir, Beard, Taylor, et al., 2009). In order to avoid data overlap, we used only the data reported for LSAS from Amir et al. (2011). We chose to do this because the sample size of Amir et al. (2011) was larger than the sample of their previous study (Amir, Beard, Taylor, et al., 2009).

Coding procedures

For every eligible study we retained the following variables: study identification data (author, year of publication), symptoms category, clinical status of the sample, sample size, participants' mean age, type of ABM procedure, type of stimuli, position of stimuli during training, number of training trials, number of ABM sessions, temporal separation of ABM sessions, ABM treatment duration, follow-up interval, and outcome measures (see Table 1).

(Insert Table 1 about here)

The dependent variables were classified as follows:

1. *Primary outcomes* (or measures of the core symptoms related to the investigated condition) versus *secondary outcomes* (or measures of general distress/nonspecific symptoms). Primary and secondary outcome for each symptom category are listed in Appendix A.
2. *Self-report, clinician-rated, and bio-behavioral measures*. The same outcomes classified previously as being primary or secondary were classified here based on how they were measured (for example, for bio-behavioral measures we considered the cortisol level or indicators of heart rate variability). Measures included in each of these three categories care listed in Appendix B.

Statistical analyses

We calculated Hedges's *g* effect sizes for every outcome measure for which sufficient data were reported. Hedges's *g* in an alternative to Cohen's *d*, that corrects for the bias in the estimation of the population effect size (Hedges & Olkin, 1985). All the effect sizes were coded such that a positive value of Hedges's *g* indicated greater improvement in the experimental group compared with the control group. In order to investigate the relationship between change in AB and change in symptoms, separate sets of effect sizes were compiled to index change in AB and change in symptoms, respectively. In addition, as we were interested to investigate the extent to which the pre-existent AB influenced ABM impact on AB and symptoms, we computed a third set of effect sizes, reflecting baseline differences in AB between groups[1]. For this third set of data, the effect sizes were coded such that a positive value of AB indicated larger AB in the experimental group compared with the control group.

To compute effect size for change in symptoms, we used the following data: means and standard deviations, when available; *t* values from between group analyses, and sample size; Chi-squared values from between group analyses; precise *p* values, and degrees of freedom from between group analyses. Some of

[1] Although participants were randomized to the experimental groups, some baseline differences in AB existed between groups even if they were not always statistically significant. To quantify these differences, we chose to use standardized measures of AB (i.e., effect size estimates) because: (1) many studies did not report means and standard deviations for AB in baseline, but only a *t* value or just a *p* value indicating no statistically significant differences; and (2) different studies used different procedures to estimate AB.

the studies reported outcome measures post-intervention, following a stressor, and/or at follow-up. We calculated separate sets of effect sizes for post-intervention, post-stressor, and follow-up measures (see Table 1). For studies reporting multiple outcomes at a given time point, we calculated an average effect size based on all the measures reported. We computed an overall estimate of the ABM effect, for which data from all symptom categories were taken together. We also computed effect sizes for each symptom category separately.

To compute effect sizes reflecting baseline group differences in AB or change in AB following ABM, we used means and standard deviations reported in the original study, if these data were available. Otherwise, we used the reported statistical information (i.e., independent *t* tests or *p* values and sample size).

For all sets of the calculated effect sizes we used the random effects model (based on the assumption that studies come from populations where the effect sizes varies). To test the assumption that the effect sizes included in each data set estimate the same population mean, we used the *Q* statistic and the I^2 statistic (Borenstein, Hedges, Higgins, & Rothstein, 2009). *Q* statistic relates the true heterogeneity in effect sizes to random error, with statistically significant *Q* suggesting true heterogeneity in the effect sizes, beyond that attributable to error. I^2 statistic indicates the proportion of the observed variance reflecting real differences in the effect sizes. We chose to use both *Q* and I^2 given their complementary nature: *Q* is an inferential statistic, but it is sensitive to the number of the studies, while I^2 is a descriptive statistic, not affected by the number of the studies (Borenstein et al., 2009).

To address publication bias, we calculated a fail-safe *N* for all the effect size subsets. Fail-safe *N* estimates the number of unpublished studies with effect sizes of zero needed to reduce the computed effect size below significance (Rosenthal, 1991). In addition, we generated and visually examined funnel plots, which plot standard error for each study (determined by sample size) against the effect size computed for that study. Studies with larger samples are expected to cluster toward the top of the plot, as they will yield larger and more reliable effect sizes, while studies with smaller sample sizes are expected to be scattered more widely around the mean and towards the bottom of the plot, as they will yield smaller effect size and

are more susceptible to be affected by error. When publication bias is present, the funnel plot will be asymmetrical, with fewer small sample-sized studies that would be predicted falling below the mean effect size. In case of an asymmetrical funnel plot, Duval and Tweedie's trim-and-fill procedure (Duval & Tweedie, 2000) estimates the likely number of the missing studies that would correct for the publication bias. We used this procedure to adjust the computed effect size and the confidence interval according to the estimated number of missing studies, to estimate the corrected effect size. All analyses were run using Comprehensive Meta-Analysis, Version 2.2.046 (Borenstein, Hedges, Higgins, & Rothstein, 2005).

Results

ABM effect on AB

Overall change in AB. We computed the overall ABM effect size on bias from data reported in 38 group comparisons (N = 1685 participants), considering only the data reported *post intervention*. The results showed a medium and statistically significant effect size, g = 0.453, p = 0.000, 95% CI = [0.284; 0.623], $Q(37)$ = 111.536, p = 0.000, I^2 = 66.827. Three outliers were identified, with effect sizes exceeding two standard deviations from the average effect size (Heeren et al., 2012; See, MacLeod, & Bridle, 2009; Wells & Beevers, 2010). After excluding them, the mean effect size was 0.312, p = 0.003, 95% CI = [0.216; 0.409], $Q(34)$ = 33.966, p = 0.469, I^2 = 0.000. For *follow-up measurements*, the average effect size was non-significant, g = 0.553, p = 0.137, 95% CI = [-0.177; 1.282], $Q(4)$ = 26.349, p = 0.000, I^2 = 84.819.

Change in AB across disorders. The average effect size computed for *anxiety studies* (18 group comparisons, 839 participants) was 0.329, p = 0.000, 95% CI = [0.183; 0.474], $Q(17)$ = 19.304, p = 0.311, I^2 = 11.937. For *depression* (6 group comparisons, 237 participants), the average effect size was non-significant, g = 0.217, p = 0.099, 95% CI = [-0.040; 0.475], $Q(5)$ = 2.188, p = 0.823, I^2 = 0.000. Similarly, for pain (2 group comparisons, 72 participants) and for substance abuse (5 group comparisons, 288 participants) the effect sizes were non-significant (*pain*: g = 0.202, p = 0.379, 95% CI = [-0.248; 0.651], $Q(1)$ = 0.372, p = 0.542, I^2 = 0.000; *substance abuse*: g = 0.340, p = 0.070, 95% CI = [-0.027; 0.707], $Q(4)$ = 8.569, p = 0.70, I^2 = 53.806). However, the effect size for AB change was significant for *healthy*

participants facing a stressful situation (4 group comparisons, 251 participants), $g = 0.378$, $p = 0.003$, 95% CI = [0.125; 0.631], $Q(3) = 2.370$, $p = 0.499$, $I^2 = 0.000$.

Moderators of the ABM effect on change in AB. Only training setting was found to significantly moderate the ABM effect for change in AB in the overall data set, $Q(1) = 4.770$, $p = 0.029$, with studies conducted in laboratory yielding significant lager effects ($g = 0.371$, $p = 0.000$, 95% CI = [0.261; 0.480], $Q(25) = 24.720$, $p = 0.478$, $I^2 = 0.000$) than studies conducted out of the laboratory ($g = 0.116$, $p = 0.259$, 95% CI = [-0.085; 0.317], $Q(8) = 4.476$, $p = 0.812$, $I^2 = 0.000$). The same was true for *anxiety studies*, $Q(1) = 5.202$, $p = 0.023$, studies conducted in laboratory being found to yield significant larger effect size ($g = 0.407$, $p = 0.000$, 95% CI = [0.254; 0.561], $Q(14) = 13.281$, $p = 0.505$, $I^2 = 0.000$) compared with studies conducted out of the laboratory ($g = 0.032$, $p = 0.824$, 95% CI = [-0.251; 0.315], $Q(2) = 0.821$, $p = 0.663$, $I^2 = 0.000$). In addition, in anxiety study subsample, participants' age significantly moderated the ABM effect on bias, with younger participants benefiting more from the intervention (slope = - 0.021, $p = 0.01$).

ABM effect on symptoms

***Overall change in symptoms*.** We computed the overall ABM effect size on symptoms from data reported in 42 group comparisons ($N = 1979$ participants), considering only the data reported *post intervention*. The results showed a small, yet statistically significant effect size, $g = 0.196$, $p = 0.001$, 95% CI = [0.085; 0.308], $Q(41) = 85.190$, $p = 0.000$, $I^2 = 51.173$. Two outliers were identified, with effect sizes exceeding two standard deviations from the average effect size (Mathews & MacLeod, 2002; Waters, Pittaway, Mogg, Bradley, & Pine, 2013). After excluding them, the mean effect size was 0.160, $p = 0.003$, 95% CI = [0.055; 0.265], $Q(39) = 70.079$, $p = 0.002$, $I^2 = 44.349$. For studies reporting outcome measures *following a stressor* (15 group comparisons, 680 participants), the average effect size was 0.404, $p = 0.000$, 95% CI = [0.278; 0.531], $Q(14) = 13.789$, $p = 0.466$, $I^2 = 0.000$. Two outliers were identified, with effect sizes exceeding two standard deviations from the average effect size (Dandeneau et al., 2007, Study 3a & 3b). After excluding these group comparisons, the average effect size was reduced to 0.375, $p = 0.000$, 95% CI = [0.246; 0.504], $Q(12) = 8.794$, $p = 0.720$, $I^2 = 0.000$. For studies reporting *follow-up measures* (12

group comparisons, 597 participants), the average effect size was 0.454, p = 0.003, 95% CI = [0.171; 0.738], $Q(11) = 49.828$, $p = 0.000$, $I^2 = 77.968$. However, after excluding three outliers (Browning et al., 2012; Heeren et al., 2012; Schoenmakers et al., 2010), the average effect size was reduced to non-significance, $g = 0.227$, $p = 0.087$, 95% CI = [-0.033; 0.488], $Q(8) = 19.190$, $p = 0.014$, $I^2 = 58.313$. There were not statistically significant differences between the overall effect sizes computed for different time points, $Q(2) = 5.191$, p = 0.075.

Change in symptoms across disorders. We computed separate effect sizes for different disorder categories and different time points. Results are shown in Table 2. As shown, ABM yielded reliable effects at post-intervention and following a stressor only for *healthy* and *anxious* participants. No statistically significant effect sizes were obtained for follow-up measures, except for one study conducted in healthy participants that had a 2-week follow-up period.

(Insert Table 2 about here)

Moderators of the ABM effect on symptom reduction. We ran the moderation analyses based on data collected *post intervention*, and *post-stressor*, respectively, considering only *anxiety studies*. We decided to limit moderation analysis to these data because (1) the overall ABM effect seemed to be driven by anxiety studies and studies conducted in healthy participants, and (2) the low number of studies using ABM to reduce symptoms/emotional vulnerability in other participants than anxious individuals precluded us for running moderation analyses.

Categorical moderations are shown in Table 3. When symptoms were measured post intervention, ABM yielded a small but significant effect on symptoms in social anxiety, and a medium effect in generalized anxiety. The effect sizes for other anxiety disorders (e.g., phobias, post-traumatic stress disorder) were not significant. Similarly, studies using the modified dot-probe task as well as those conducted in laboratory yielded significant larger effects compared with studies using spatial cueing task, or conducted via Internet, out of laboratory (i.e., at home). The effect sizes were significant only for primary

outcomes, regardless of the time of measurement (i.e., post intervention or post stressor). No other significant categorical moderators were identified.

(Insert Table 3 about here)

In terms of continuous moderators, we used meta-regression to test the number of trials per session, total number of training sessions, the temporal frequency of training sessions, and participants' age as possible moderators. Only participants' age was found to significantly moderate the ABM effect when symptoms were measured post-intervention, with younger individuals benefiting most from ABM (slope = -0.034, $p = 0.000$).

Relationship between the pre-existent AB, change in AB, and change in symptoms

In the overall dataset, the pre-existent AB was significantly related to the change in AB, $r(34) = .519$, $p = 0.002$, and the change in AB correlated significantly with the change in symptoms, $r(34) = .342$, $p = 0.048$. However, there was no relationship between the pre-existent AB and change in symptoms, $r(32) = -.005$, $p = 0.977$. We tested in a mediation model the assumption that the pre-existent AB (predictor) influenced symptom reduction (outcome) via change in AB (mediator). For mediation analysis we used bootstrapping tests with 1000 re-samples and corrected confidence interval (Preacher & Hayes, 2008). The results indicated no significant direct or indirect effects of the pre-existent AB on change in symptoms (for the direct effect, 95% CI = [-0.384; 0.181]; for the indirect effect, 95% CI = [-0.039; 0.234]).

The same results pattern was observed within the anxiety sample of studies, where the pre-existent AB correlated strongly with the change in AB, $r(18) = .745$, $p = 0.000$, but failed statistical significance when related to change in symptoms, $r(17) = .313$, $p = 0.221$. Similarly, the correlation between change in bias and change in symptoms was non-significant, $r(17) = .255$, $p = 0.324$. We did not find any evidence of direct or indirect effect of the pre-existent AB on change in symptoms (for the direct effect, 95% CI = [-0.403; 0.578]; for the indirect effect, 95% CI = [-0.659; 0.949]).

Publication bias

Publication bias for change in symptoms. There was evidence of publication bias in the overall data set for change in symptoms as measured post-intervention: fail-safe *N* was 133, smaller than 5*K*+10 [2]. In addition, Duval and Tweedie's trim-and-fill procedure (Duval & Tweedie, 2000) estimated 10 missing studies with effect sizes smaller than the mean effect size, which would have reduced the mean effect size to non-significance, g = 0.031, 95% CI = [-0.080, 0.147]. However, there was no evidence of publication bias in the overall data set for change in symptoms as measured post-stressor: fail-safe *N* was 127, larger than 5*K*+10, and trim-and-fill procedure estimated no missing studies.

Similarly, for anxiety studies, there was some evidence of publication bias for change in symptoms when post intervention data were considered: fail-safe *N* was 119, and trim-and-fill procedure estimated 3 missing studies, which would have reduced the effect size to 0.205, 95% CI = [0.072, 0.339]. However, when post-stressor data were considered, there was no evidence of publication bias (fail-safe *N* = 63; no missing studies were estimated).

Publication bias for change in AB. No evidence of publication bias was found for change in AB as measured post intervention in the overall data set (fail-safe *N* was 740, no missing studies were estimated). However, in the anxiety studies subset there was some evidence of publication bias: fail-safe *N* was 101, and trim-and-fill procedure 4 missing studies which would have reduced the effect size to 0.229, 95% CI = [0.066, 0.393].

Discussion

This meta-analysis aimed to assess the clinical utility of ABM procedures in the context of the growing literature, as the previous reviews on ABM procedures indicated both inconsistent findings and publication bias. For this purpose we performed a quantitative review of studies that aimed to reduce AB via an ABM procedure and examined effects on symptoms or features of psychopathology. In an attempt to provide a better estimation of the clinical utility of ABM, we included recent studies reporting mixed results on ABM. As at the publication time of the previous reviews on the topic such studies were not available,

[2] According to Rosenthal (1991), the computed fail-safe *N* should be larger than 5*K*+10, where *K* is the number of studies included in meta-analysis

considering all the available data could offer a better view on ABM, possibly correcting for publication bias reported in previous meta-analyses. In addition, we aimed to organize the extant empirical evidence to delineate clear practical implications and future research directions in ABM field.

The obtained results indicated that: (1) ABM successfully reduces AB and symptoms/emotional vulnerability in anxious individuals and healthy participants; and (2) although the pre-existent AB was significantly related to change in AB, and change in AB was positively related with change in symptoms in the overall data set (but not in the anxiety subsample of studies), no direct or indirect effect of the pre-existent AB on the change in symptoms was observed. Therefore, with regard to the clinical utility of ABM, existing data indicated a small clinical impact (mainly driven by anxiety studies), with about 58%-62% of participants in the control group having more symptoms compared with the average participant in the experimental group (McGough & Faraone, 2009). In addition, as change in AB did not mediate the relationship between the pre-existent level of AB and change in symptoms, the ABM mechanism of change is not fully clear. With respect to publication bias, our results suggested that the small effect size could be instable, i.e., several additional studies with negative findings would have reduced the effect size to (almost) non-significance.

These results have two main research and clinical implications. First, although ABM reduced AB significantly, we need more powerful ABM procedures, to strengthen the clinical effects of ABM: we should strive to improve the existing ABM procedures, and/or to develop and test new theory-driven training procedures. Future ABM procedures should be aimed to reliably modify AB in a way that allows and promotes generalization of the effect. In addition, they should have greater ecological validity. In this sense, researchers should seek for modalities of increasing the requirement of selective attention to be oriented away from threat and towards neutral or positive material, for instance by providing more visual stimuli. Furthermore, we need training task that are more captivating and engaging for participants. Data from a recent study (Beard, Weisberg, & Primack, 2012) showed that participants reported a lower engagement with ABM programs compared with interpretation modification programs. The authors pointed out the need of improving cognitive bias modification programs' rationale and credibility.

Related to this first implication, the fact that ABM conducted out of the laboratory yielded non-significant effect seriously threatens the clinical utility of ABM, as this intervention was thought to be an efficient, independently implemented, and easily disseminated treatment. It is true that the available evidence regarding ABM clinical utility out the laboratory is currently limited, precluding firm conclusions. However, future studies should investigate the critical factors involved in ABM efficiency in the laboratory, so that ABM efficiency out of the laboratory might be improved.

The second implication of these results concerns the need to clarify the ABM mechanism of change. ABM is thought to work by counteracting AB, a dysfunctional automatic way of allocating attentional resources. However, the empirical data seem to not support this assumption. Here we need to be cautious provided that the amount of data is still limited. A possible explanation could lie in the way that AB is conceptualized and measured. We need to know which component of AB is modified (see Cisler & Koster, 2010) and we need reliable AB measurements. In terms of the modified AB components, it was suggested that ABM works through facilitating disengagement from disorder-relevant material (Heeren, Lievens, & Philippot, 2011). However, ABM may also act through a mechanism of attentional avoidance, which could later contribute to the long-term maintenance of symptoms (Koster, Baert, Bockstaele, & De Raedt, 2010). Future studies should clarify this aspect. In terms of AB measurement, most AB tasks are questionable in terms of reliability. For instance, the dot-probe task, one of the most used tasks both for AB assessment, has been criticized for poor reliability (Cisler, Bacon, & Williams, 2009; Schmukle, 2005). Therefore, relying on it for the purpose of measuring a mechanism of change is less than ideal. Future studies should develop and test alternative ways to measure different components of AB in a psychometrically sound manner.

Several other findings deserve discussion here. First, post-stressor measurements seemed to yield somewhat stronger results compared with measurements collected right after training, suggesting that AB contributes to the vulnerability for developing psychopathology only in the presence of critical stimuli or events. Despite this finding, the follow-up effects of ABM did not yield a significant effect size. However, the latter conclusion is still preliminary as fewer studies investigated the long-term effects of ABM. Future studies should investigate the extent to which ABM contributes to stable resilience to stressors.

Second, the possibility that the small effect size is due to non-specific factors (e.g., demand effects) should be kept in mind, especially since the effect sizes were significantly larger for studies conducted in the laboratory (compared with studies conducted out of laboratory) and for primary outcomes (compared with secondary ones). Although we included only randomized trials and most of them reported that participants were blind to randomization and to the purposes of the study, a large part of them did not report if experimenters were blind as well, and how the participants and/or experimenters were kept blind throughout the study. Future studies should clarify these aspects and researchers should take care to demonstrate that ABM effects on AB and symptoms are attributable to the experimental manipulation and not to other possible factors by adhering more stringently to standards for randomized clinical trials.

Interestingly, for anxiety studies, the effect size was statistically significant at post-intervention only for clinician-rated measures (although the effect size indicators computed for self-report measures and for bio-behavioral measures were significant post-stressor). These results might be explained in terms of the participant's ability to consciously report on the benefits of the ABM post-intervention. Indeed, ABM is thought traditionally to reduce anxiety vulnerability (not anxiety per se). In support of this view, ABM effects on bio-behavioral measures were non-significant post intervention (when no immediate challenge was faced), but significant following a stressor. Arguably, participants might be unaware of the beneficial ABM effects until they face critical events. However, skilled clinicians might be able to assess change in anxiety vulnerability.

Third, the small effect size we obtained is in line with previous reported findings (Beard et al., 2012; Hallion & Ruscio, 2011) and are mostly driven by anxiety studies. However, this result should be interpreted with caution, as there are far fewer studies which investigated ABM in other symptom categories than studies which employed ABM to reduce anxiety symptoms. There was also evidence of publication bias when measurement collected post intervention were considered, both in the overall data set and in anxiety subsample effect sizes set. When the trim-and-fill procedure was applied, the estimated ABM effect size became non-significant in the overall data set and approached non-significance for anxiety studies, suggesting that the small computed effect size may be instable. As such, more randomized clinical trials

should be published before meaningful conclusions can be derived. Given the potential problem of publication bias, the contribution from studies reporting negative findings/small effect sizes should be especially encouraged.

Fourth, participants' age significantly moderated the ABM's impact both on AB and symptoms, with younger participants benefiting the most. This may be due to younger participants being more accustomed to technology and believe more in the therapeutic potential of a computerized task. Alternatively, AB at younger age may be more easily changed compared with adults. Future studies should investigate factors that may contribute to ABM efficacy in different age categories.

Limits and directions for future studies

The results of this meta-analysis are limited in several ways. First of all, one should notice that many of the included studies were anxiety studies conducted in analogue samples and having small sample size. Second, the same group of investigators was involved in conducting more than one study. Third, not every study included an AB measure, the purported mechanism of change. Even when a measure of AB change was included, most times the task used for AB assessment was a modified version of the task used for training. As noted elsewhere (e.g., Koster et al., 2009), this does not allow to examine real-world transfer of training effects.

This is the first meta-analytical work that considered specifically the role of the pre-existent AB in relation to ABM efficacy. Our mediational analysis indicated no significant involvement of pre-existing bias in the response to ABM. However, two main limitations should be made clear. First, because our approach of estimating the pre-existing AB relied on contrasting the experimental group with the control group, we cannot say for sure that there was a real pre-existing AB in group receiving the intervention compared with the control group. To do that, we should have considered the pre-existent AB in both groups relative to zero (i.e., the absence of the AB). Second, change in bias was measured at the same time with change in symptoms. Consequently, the causal relationship between change in bias and change in symptoms cannot be empirically proved. Therefore, our results should be interpreted cautiously. Future experimental studies are

needed to specifically investigate the role of the pre-existent AB in relation with change in AB and symptoms following ABM, by providing intermediate measurements of AB.

Conclusions

As ABM therapeutic benefits are rather limited, we believe it is clearly premature to speak of ABM treatment (see Bar-Haim, 2010; Hakamata et al., 2010). The results of the present work urges for more adequately powered, randomized controlled clinical trials, conducted by different research groups, and aimed to rigorously assess ABM impact on both AB and symptoms.

References

Note: References with an asterix (*) were included in the meta-analysis; a complete reference list for the articles included in the meta-analysis can be found in Appendix C.

*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. doi:10.1037/a0012589

*Amir, N., Beard, C., Taylor, C. T., Klumpp, H., Elias, J., Burns, M., & Chen, X. (2009). Attention Training in Individuals with Generalized Social Phobia: A Randomized Controlled Trial. *Journal of Consulting and Clinical Psychology*, *77*(5), 961–973. doi:10.1037/a0016685

Amir, N., Bomyea, J., & Beard, C. (2010). The effect of single-session interpretation modification on attention bias in socially anxious individuals. *Journal of Anxiety Disorders*, *24*(2), 178–182. doi:10.1016/j.janxdis.2009.10.005

*Baert, S., De Raedt, R., Schacht, R., & Koster, E. H. W. (2010). Attentional bias training in depression: therapeutic effects depend on depression severity. *Journal of Behavior Therapy and Experimental Psychiatry*, *41*(3), 265–274. doi:10.1016/j.jbtep.2010.02.004

Bar-Haim, Y. (2010). Research Review: attention bias modification (ABM): a novel treatment for anxiety disorders. *Journal of Child Psychology and Psychiatry*, *51*(8), 859–870. doi:10.1111/j.1469-7610.2010.02251.x

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study. *Psychological Bulletin*, *133*(1), 1–24. doi:10.1037/0033-2909.133.1.1

*Bar-Haim, Y., Morag, I., & Glickman, S. (2011). Training anxious children to disengage attention from threat: a randomized controlled trial. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *52*(8), 861–869. doi:10.1111/j.1469-7610.2011.02368.x

Beard, C., & Amir, N. (2010). Attention bias for sexual words in female sexual dysfunction. *Journal of Sex & Marital Therapy*, *36*(3), 216–226. doi:10.1080/00926231003719616

Beard, C., Sawyer, A. T., & Hofmann, S. G. (2012). Efficacy of Attention Bias Modification Using Threat and Appetitive Stimuli: A Meta-Analytic Review. *Behavior Therapy*, *43*(4), 724-740. doi:10.1016/j.beth.2012.01.002

Beard, C., Weisberg, R. B., & Primack, J. (2012). Socially Anxious Primary Care Patients' Attitudes Toward Cognitive Bias Modification (CBM): A Qualitative Study. *Behavioural and Cognitive Psychotherapy*, *40*(05), 618–633. doi:10.1017/S1352465811000671

*Boettcher, J., Berger, T., & Renneberg, B. (2012). Internet-Based Attention Training for Social Anxiety: A Randomized Controlled Trial. *Cognitive Therapy and Research*, *36*(5), 522–536. doi:10.1007/s10608-011-9374-y

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis* John Wiley & Sons, Ltd.

Brooks, S., Prince, A., Stahl, D., Campbell, I. C., & Treasure, J. (2011). A systematic review and meta-analysis of cognitive bias to food stimuli in people with disordered eating behaviour. *Clinical Psychology Review*, *31*(1), 37–51. doi:10.1016/j.cpr.2010.09.006

Brosan, L., Hoppitt, L., Shelfer, L., Sillence, A., & Mackintosh, B. (2011). Cognitive bias modification for attention and interpretation reduces trait and state anxiety in anxious patients referred to an out-patient service: results from a pilot study. *Journal of Behavior Therapy and Experimental Psychiatry*, *42*(3), 258–264. doi:10.1016/j.jbtep.2010.12.006

*Browning, M., Holmes, E. A., Charles, M., Cowen, P. J., & Harmer, C. J. (2012). Using Attentional Bias Modification as a Cognitive Vaccine Against Depression. *Biological Psychiatry*, *72*(7), 572-579. doi:10.1016/j.biopsych.2012.04.014

Browning, M., Holmes, E. A., & Harmer, C. J. (2010). The modification of attentional bias to emotional information: A review of the techniques, mechanisms, and relevance to emotional disorders. *Cognitive, Affective & Behavioral Neuroscience*, *10*(1), 8–20. doi:10.3758/CABN.10.1.8

*Carlbring, P., Apelstrand, M., Sehlin, H., Amir, N., Rousseau, A., Hofmann, S. G., & Andersson, G. (2012). Internet-delivered attention bias modification training in individuals with social anxiety

disorder - a double blind randomized controlled trial. *BMC Psychiatry*, *12*(1), 66. doi:10.1186/1471-244X-12-66

Chapman, S., & Martin, M. (2011). Attention to pain words in irritable bowel syndrome: increased orienting and speeded engagement. *British Journal of Health Psychology*, *16*(Pt 1), 47–60. doi:10.1348/135910710X505887

Cisler, J. M., Bacon, A. K., & Williams, N. L. (2009). Phenomenological Characteristics of Attentional Biases Towards Threat: A Critical Review. *Cognitive Therapy and Research*, *33*(2), 221–234. doi:10.1007/s10608-007-9161-y

Cisler, J. M., & Koster, E. H. W. (2010). Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. *Clinical Psychology Review*, *30*(2), 203–216. doi:10.1016/j.cpr.2009.11.003

*Dandeneau, S. D., Baldwin, M. W., Baccus, J. R., Sakellaropoulo, M., & Pruessner, J. C. (2007). Cutting stress off at the pass: reducing vigilance and responsiveness to social threat by manipulating attention. *Journal of Personality and Social Psychology*, *93*(4), 651–666. doi:10.1037/0022-3514.93.4.651

Duval, S., & Tweedie, R. (2000). Trim and Fill: A Simple Funnel-Plot–Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics*, *56*(2), 455–463. doi:10.1111/j.0006-341X.2000.00455.x

Dvorkin, A. Y., Ramaiya, M., Larson, E. B., Zollman, F. S., Hsu, N., Pacini, S., … Patton, J. L. (2013). A "virtually minimal" visuo-haptic training of attention in severe traumatic brain injury. *Journal of Neuroengineering and Rehabilitation*, *10*, 92. doi:10.1186/1743-0003-10-92

*Eldar, S., Apter, A., Lotan, D., Edgar, K. P., Naim, R., Fox, N. A., … Bar-Haim, Y. (2012). Attention bias modification treatment for pediatric anxiety disorders: a randomized controlled trial. *The American Journal of Psychiatry*, *169*(2), 213–220.

Emmelkamp, P. M. (2012). Attention bias modification: the Emperor's new suit? *BMC Medicine*, *10*(1), 63. doi:10.1186/1741-7015-10-63

Field, M., & Cox, W. M. (2008). Attentional bias in addictive behaviors: a review of its development, causes, and consequences. *Drug and Alcohol Dependence*, *97*(1-2), 1–20. doi:10.1016/j.drugalcdep.2008.03.030

Fox, E., Russo, R., Bowles, R., & Dutton, K. (2001). Do Threatening Stimuli Draw or Hold Visual Attention in Subclinical Anxiety? *Journal of Experimental Psychology. General*, *130*(4), 681–700.

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. doi: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. doi:10.1037/a0024355

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press.

*Heeren, A., Lievens, L., & Philippot, P. (2011). How does attention training work in social phobia: Disengagement from threat or re-engagement to non-threat? *Journal of Anxiety Disorders*, *25*(8), 1108–1115. doi:10.1016/j.janxdis.2011.08.001

*Heeren, A., Reese, H. E., McNally, R. J., & Philippot, P. (2012). Attention training toward and away from threat in social phobia: effects on subjective, behavioral, and physiological measures of anxiety. *Behaviour Research and Therapy*, *50*(1), 30–39. doi:10.1016/j.brat.2011.10.005

Hertel, P. T., & Mathews, A. (2011). Cognitive Bias Modification Past Perspectives, Current Findings, and Future Applications. *Perspectives on Psychological Science*, *6*(6), 521–536. doi:10.1177/1745691611421205

*Julian, K., Beard, C., Schmidt, N. B., Powers, M. B., & Smits, J. A. J. (2012). Attention training to reduce attention bias and social stressor reactivity: an attempt to replicate and extend previous findings. *Behaviour Research and Therapy*, *50*(5), 350–358. doi:10.1016/j.brat.2012.02.015

*Klumpp, H., & Amir, N. (2010). Preliminary Study of Attention Training to Threat and Neutral Faces on Anxious Reactivity to a Social Stressor in Social Anxiety. *Cognitive Therapy and Research*, *34*(3), 263–271. doi:10.1007/s10608-009-9251-0

Koster, E. H. W., Baert, S., Bockstaele, M., & De Raedt, R. (2010). Attentional retraining procedures: manipulating early or late components of attentional bias? *Emotion*, *10*(2), 230–236. doi:10.1037/a0018424

Koster, E. H. W., Fox, E., & MacLeod, C. (2009). Introduction to the special section on cognitive bias modification in emotional disorders. *Journal of Abnormal Psychology*, *118*(1), 1–4. doi:10.1037/a0014379

MacLeod, C., & Mathews, A. (2012). Cognitive Bias Modification Approaches to Anxiety. *Annual Review of Clinical Psychology*, *8*(1), 189–217. doi:10.1146/annurev-clinpsy-032511-143052

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*, *111*(1), 107–123.

*Mathews, A., & MacLeod, C. (2002). Induced processing biases have causal effects on anxiety. *Cognition & Emotion*, *16*(3), 331–354. doi:10.1080/02699930143000518

Mathews, A., & MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. *Annual Review of Clinical Psychology*, *1*, 167–195. doi:10.1146/annurev.clinpsy.1.102803.143916

McGough, J. J., & Faraone, S. V. (2009). Estimating the size of treatment effects: moving beyond p values. *Psychiatry*, *6*(10), 21–29.

*Neubauer, K., von Auer, M., Murray, E., Petermann, F., Helbig-Lang, S., & Gerlach, A. L. (2013). Internet-delivered attention modification training as a treatment for social phobia: a randomized controlled trial. *Behaviour Research and Therapy*, *51*(2), 87–97. doi:10.1016/j.brat.2012.10.006

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, *32*(1), 3–25.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891. doi:10.3758/BRM.40.3.879

Rosenthal, R. (1991). *Meta-Analytic Procedures for Social Research*. SAGE.

*Schmidt, N. B., Richey, J. A., Buckner, J. D., & Timpano, K. R. (2009). Attention training for generalized social anxiety disorder. *Journal of Abnormal Psychology*, *118*(1), 5–14. doi:10.1037/a0013643

Schmukle, S. C. (2005). Unreliability of the dot probe task. *European Journal of Personality*, *19*(7), 595–605. doi:10.1002/per.554

*Schoenmakers, T. M., de Bruin, M., Lux, I. F. M., Goertz, A. G., Van Kerkhof, D. H. A. T., & Wiers, R. W. (2010). Clinical effectiveness of attentional bias modification training in abstinent alcoholic patients. *Drug and Alcohol Dependence*, *109*(1–3), 30–36. doi:10.1016/j.drugalcdep.2009.11.022

*Schoorl, M., Putman, P., & Van Der Does, W. (2013). Attentional bias modification in posttraumatic stress disorder: a randomized controlled trial. *Psychotherapy and Psychosomatics*, *82*(2), 99–105. doi:10.1159/000341920

*See, J., MacLeod, C., & Bridle, R. (2009). The reduction of anxiety vulnerability through the modification of attentional bias: a real-world study using a home-based cognitive bias modification procedure. *Journal of Abnormal Psychology*, *118*(1), 65–75. doi:10.1037/a0014377

*Sharpe, L., Ianiello, M., Dear, B. F., Nicholson Perry, K., Refshauge, K., & Nicholas, M. K. (2012). Is there a potential role for attention bias modification in pain patients? Results of 2 randomised, controlled trials. *Pain*, *153*(3), 722–731. doi:10.1016/j.pain.2011.12.014

Van Bockstaele, B., Koster, E. H. W., Verschuere, B., Crombez, G., & De Houwer, J. (2012). Limited transfer of threat bias following attentional retraining. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*(2), 794–800. doi:10.1016/j.jbtep.2011.11.001

*Waters, A. M., Pittaway, M., Mogg, K., Bradley, B. P., & Pine, D. S. (2013). Attention training towards positive stimuli in clinically anxious children. *Developmental Cognitive Neuroscience*, *4*, 77–84. doi:10.1016/j.dcn.2012.09.004

*Wells, T. T., & Beevers, C. G. (2010). Biased attention and dysphoria: Manipulating selective attention reduces subsequent depressive symptoms. *Cognition & Emotion*, *24*(4), 719–728. doi:10.1080/02699930802652388

Running head: CLINICAL EFFICACY OF ATTENTIONAL RETRAINING

Table 1. *Characteristics of the studies included in meta-analysis*

Study	Disorder	Sample type	Sample size		Mean age	ABM Condition								Outcome measures	Hedges' g for change in symptoms		
			ABM Group	ACC Group		Training task	Type of stimuli	Type of training	Stimuli presentation during training	No. of training sessions	No. of trials per session	Temporal separation of ABM session (in days)	Training setting		Post intervention	Post-stressor	Follow-up
Amir, Beard, Burns, et al. (2009)	Anxiety	Clinical	14	15	26	dot-probe	pictures	neutral	vertical	8	240	2	lab	BDI-II, HAM-D, HRSA, PSWQ, SCID, STAI-S, STAI-T, WDQ	0.68	-	-
Amir, Beard, Taylor, et al. (2009)	Anxiety	Clinical	22	22	29	dot-probe	words	neutral	vertical	8	160	2	lab	BDI-II, HAM-D, SCID, SDS, SPAI, STAI-T	0.39	-	-
Amir, et al. (2011)	Anxiety	Clinical	55	57	31	dot-probe	pictures	neutral	vertical	8	160	2	lab	LSAS	0.61	-	-
Amir, et al. (2008)	Anxiety	Analogue / subclinical	47	47	19	dot-probe	pictures	neutral	vertical	1	160	not applicable	home	speech performance, STAI-S	0.17	0.48	-
Baert, et al. (2010) S1	Depression	Analogue / subclinical	25	23	20	visual cueing	words	positive	horizontal	10	220	1	home	BDI-II, MASQ, POMS, RRS	-0.27	-	-
Baert, et al. (2010) S2	Depression	Clinical	15	20	42	visual cueing	words	positive	horizontal	10	220	1	home	BDI-II, MASQ, POMS, RRS	0.27	-	-
Baert et al. (2012)	-	Healthy participants	16	16	27	dot-probe	pictures	positive	horizontal	6	210	1	home	HF, LF, LF/HF ratio, RMSSD	0.19	0.51	-
Bar-Haim, et	Anxiety	Analogue /	18	16	10	visual	picture	neutral	horizontal	2	384	2	lab	CDI, STAI-C, VAS anxiety,	-0.2	0.04	-

al. (2011)		subclinical				cueing	s							VAS depression			
Boettcher, et al. (2012)	Anxiety	Clinical	33	35	38	dot-probe	pictures	neutral	vertical	10	160	2	home	BDI-II, BSI, LSAS, SCID, SIAS, SPS	-0.21	-	0.09
Browning, et al. (2012) (pictures)	Depression	Clinical	16	14	36	dot-probe	pictures	positive	vertical	28	96	0.5	home	BDI-II, CAR, HAM-D, STAI-T	-0.27	-	1.32
Browning, et al. (2012) (words)	Depression	Clinical	16	15	41	dot-probe	words	positive	vertical	28	96	0.5	home	BDI-II, CAR, HAMD, STAI-T	-0.61	-	-0.48
Bunnell, et al. (2013)	Anxiety	Clinical	15	16	24	dot-probe	pictures	neutral	vertical	8	160	0.5	lab	IST, UCT, BARS, BSPS, CGI-I, CGI-S, BDI-II, LSAS-SR, SPAI	0.17	0.04	
Carlbring, et al. (2012)	Anxiety	Clinical	40	39	37	dot-probe	pictures	neutral	vertical	8	160	2	home	BAI, CGI, LSAS, SCID, SIAS, SPSQ, QOLI	0.09	-	-0.01
Dandeneau & Baldwin (2009)	Low self-esteem	Healthy participants	72	72	22	visual search	pictures	positive	matrix	1	112	not applicable	lab	VAS feelings of rejection, SES, CIQ	0.16	-	-
Dandeneau, et al. (2007) S3a	Low self-esteem	Healthy participants	12	13	20	visual search	pictures	positive	matrix	5	80	1	home	School abilities self-esteem, STAI-S	-	-	-
Dandeneau, et al. (2007) S3b	Low self-esteem	Healthy participants	11	12	unspecified	visual search	pictures	positive	matrix	5	80	1	home	Cortisol index, cortisol reactivity, self-confidence rated by other	0.53	1.24	-
Eldar & Bar-Haim (2010)	Anxiety	Analogue / subclinical	15	15	23	dot-probe	pictures	neutral	horizontal	1	480	not applicable	lab	ERP-N2, ERP-P2, ERP-	0.82	-	-

(anxious)													P3			
Eldar & Bar-Haim (2010) (non-anxious)	-	Healthy participants	15	15	23	dot-probe	pictures	neutral	horizontal	1	480	not applicable	lab	ERP-P3	0.04	-	-
Eldar, et al. (2012)	Anxiety	Clinical	15	15	10	dot-probe	pictures	neutral	horizontal	4	480	unspecified	lab	Anxiety severity scale, anxiety symptoms count, SCID	0.43	-	-
Field, et al. (2007)	Substance abuse	Analogue / subclinical	20	20	23	dot-probe	pictures	neutral	horizontal	1	960	not applicable	lab	Amount of beer consumed, DAQ, VAS urge to drink	0.02	-	-
Field, et al. (2009)	Substance abuse	Analogue / subclinical	24	24	23	dot-probe	pictures	neutral	horizontal	1	896	not applicable	lab	Choose to smoke, amount willing to pay, delay discounting, VAS urge to smoke, QSU	0.25	-	-
Hayes, et al. (2010)	Anxiety	Analogue / subclinical	24	24	27	dot-probe	words	neutral	vertical	1	480		lab	Negative thoughts intrusions, VAS depression, difficulty to worry, positive thoughts during worry period	-	0.48	-
Hazen, et al. (2009)	Anxiety	Analogue / subclinical	12	12	19	dot-probe	words	neutral	vertical	5	216	6	lab	BDI-II, PSQW, STAI-T	0.58	-	-
Heeren, et al.	Anxiety	Clinical	20	19	22	visual	words	neutral	horizontal	1	560	not	lab	BASA, VAS anxiety, VAS	-0.08	0.63	-

(2011)						cueing						applicable		mood			
Heeren, et al. (2012)	Anxiety	Clinical	20	18	22	dot-probe	pictures	positive	horizontal	4	744	1	lab	BASA, FNES, LSAS, SUDS	0.51	-	1.13
Julian, et al. (2012) (exercise)	Anxiety	Analogue / subclinical	28	28	20	dot-probe	pictures	neutral	vertical	1	160	not applicable	lab	STAI-S	0.07	0.07	
Julian, et al. (2012) (rest)	Anxiety	Analogue / subclinical	28	28	20	dot-probe	pictures	neutral	vertical	1	160	not applicable	lab	STAI-S	0.41	0.26	-
Klumpp & Amir (2010)	Anxiety	Analogue / subclinical	31	22	20	dot-probe	pictures	neutral	vertical	1	160	not applicable	lab	STAI-S	-	0.61	-
Kruijt, et al. (2013)	Depression	Analogue / subclinical	20	20		visual search	pictures	positive	matrix	1	256	not applicable	lab	PANAS	-0.31	-	-
Li, et al. (2008)	Anxiety	Analogue / subclinical	12	10	20	dot-probe	pictures	neutral	horizontal	7	480	1	lab	FNES, SIAS, SPS	0.24	-	-
Mathews & MacLeod (2002) E7	Anxiety	Analogue / subclinical	15	14	unspecified	dot-probe	words	neutral	vertical	10	750	unspecified	lab	STAI-T	1	-	-
Mathews & MacLeod (2002) E8	Anxiety	Analogue / subclinical	16	14	unspecified	dot-probe	words	neutral	vertical	8	750	unspecified	lab	STAI-T	0.72	-	-
McHugh, et al. (2010)	Substance abuse	Analogue / subclinical	25	26	38	dot-probe	pictures	neutral	vertical	1	560		lab	QSU	0.06	0.05	-
Najmi & Amir (2010)	Anxiety	Analogue / subclinical	26	26	19	dot-probe	words	neutral	vertical	1	288		lab	BAT, STAI-S, VAS anxiety	0.13	0.4	-
Neubauer et al., (2013)	Anxiety	Clinical	24	24	40	dot-probe	pictures	neutral	vertical	8	160	0.5	home	SCID, LSAS, SIAS, SPS, BDI-II,	0.05	-	0.003
Reese, et al. (2010)	Anxiety	Analogue / subclinical	21	21	26	dot-probe	pictures	neutral	vertical	1	768		lab	BAT, SPQ, SUDS, VAS mood	-0.02	-	-
Schmidt, et al. (2009)	Anxiety	Clinical	18	18	23	dot-probe	pictures	neutral	vertical	8	160	2	lab	BDI-II, BSPS, LSAS, SCID,	0.71	-	0.9

														SPAI, STAI-T			
Shoenmarkers, et al. (2007)	Substance abuse	Analogue / subclinical	53	53	21	dot-probe	pictures	neutral	horizontal	1	624		lab	Preference task, VAS craving	-0.23	-	-
Schoenmarkers, et al. (2010)	Substance abuse	Clinical	21	22	45	dot-probe	pictures	neutral	horizontal	5	528	2	lab	DAQ, time to discharge, time to relapse	0.04	-	1.41
Schoorl, et al. (2013)	Anxiety	Clinical	38	38	37	dot-probe	pictures	neutral	vertical	8	200	0.5	home	CAPS, HADS-A, HADS-D, SRIP	0.13	-	0.11
See, et al. (2009)	Anxiety	Analogue / subclinical	22	18	22	dot-probe	words	neutral	vertical	15	192	1	home	STAI-S, STAI-T	-	0.64	-
Sharpe, et al. (2012) S1	Pain	Clinical	23	23	42	dot-probe	words	neutral	vertical	1	320		lab	DASS, Roland-Morris Disability Questionnaire, Tampa Scale for Kinesiophobia, VAS pain	-	-	0.27
Sharpe, et al. (2012) S2	Pain	Clinical	17	9	47	dot-probe	words	neutral	vertical	4	320	2	home	ASI, DASS, Roland-Morris Disability Questionnaire, Tampa Scale for Kinesiophobia, VAS pain	-0.15	-	0.02
Taylor et al. (2011)	Anxiety	Analogue / subclinical	43	34	19	dot-probe	words	positive	vertical	1	384		lab	PA, STAI-S		0.16	-
Tsumura, et al. (2012)	Depression	Analogue / subclinical	27	26	22	dot-probe	words	neutral	horizontal	1	510	not applicable	lab	Cortisol level, VAS depressive mood	0.28	-	-
Verwoerd, et	-	Healthy	22	23	20	visual	picture	neutral	horizontal	1	384		lab	Impact of Movie Scale,	0.82	-	0.66

al. (2012)		participants				cueing	s							Intrusion monitoring task, no. of intrusions (diary)			
Waters, et al. (2013)	Anxiety	Clinical	12	12	10	visual search	picture s	positive	matrix	12	160	5.3	home	ADIS-C-IV-C/P, SCAS-P, CES-DC, SCAS-C	0.97	-	-
Wells & Beevers (2010)	Depressio n	Analogue / subclinical	14	17	19	dot-probe	picture s	neutral	horizontal	4	196	2	lab	BDI-II	0.31	-	0.96

Notes: Hedge's g indexes change on symptoms, and combines all the computed effect sizes, irrespective of outcome measures. E = experiment; S = study; ADIS = Anxiety Disorders Interview Shedule (Brown, Dinardo, & Barlow, 1994); ASI = Anxiety Sensitivity Index (Reiss et al., 1986); BAI = Beck Anxiety Inventory (Beck, Epstein, Brown, & Steer, 1988); BARS = Behavioral Avoidance rating Scale (Beidel et al., 2007); BASA = Behavioral Assessment of Speech Anxiety (Mulac & Sherman, 1974); BAT = Behavioral Approach Test; BDI-II = Beck Depression Inventory-II (Beck, Steer, & Brown, 1996); BSI = Brief Symptom Inventory (Derogatis & Melisaratos, 1983); BSPS = Brief Social Phobia Scale (Davidson et al., 1991); CAR = cortisol awaking response; CAPS = Clinician-Administered PTSD Scale (Blake et al., 1990); CDI = Children's Depression Inventory (Kovacs, 1985); CES-DC = Center for Epidemiologic Depression Studies for Children (Weissman et al., 1980); CGI = Clinical Global Impression of Improvement (Guy, 1976); CIQ = Cognitive Interference Questionnaire (Sarason & Stroops, 1978); DAQ = Desires for Alcohol Questionnaire (Love et al., 1998); DASS = Depression, Anxiety, and Stress Scale (Lovibond & Lovibond, 1995); ERP-N2, ERP-P2, ERP-P3 = event-related potential N2, P2, P3; FNES = Fear of Negative Evaluation Scale (Watson & Friend, 1969); HADS = Hospital Anxiety and Depression Scale (Zigmond & Shaith, 1983); HAM-D = Hamilton Rating Scale for Depression (Hamilton, 1960); HRSA = Hamilton Rating Scale for Anxiety (Hamilton, 1959); IST = Impromptu Speech Task (Beidel et al., 2010); LSAS = Liebowitz Social Anxiety Scale (Liebowitz, 1987); MASQ = Mood and Anxiety Symptoms Questionnaire (Watson & Clacrk, 1991); PA = Positive Affect;POMS-SF = Short-form of the Profile of Mood States (Curran, Andykowsky, & Studts, 1995); PSQW = Penn State Worry Questionnaire (Meyer et al., 1990); QSU = Questionnaire on smoking urges (Cox et al., 2001); RRS = Ruminative Responses Scale (Nolen-Hoeksema & Morrow, 1991); QOLI = Quality of Life Inventory (Frish et al., 1992); SCID = Structured Clinical Interview for DSM-IV (First, Spitzer, Gibbons, & Williams, 2001); SES = State Self-esteem Scale (Heatherton & Polivy, 1991); SDS = Sheehan Disability Scale (Leon, Olfson, Portera, Faber, & Sheenan, 1997); SCAS = Spence Childre Anxiety Scale (Spence, 1998); SIAS = Social Interaction Anxiety Scale (Mattick & Clarke, 1998); SPQ = Spider Questionnaire (Klorman et al., 1974); SPS = Social Phobia Scale (Mattick & Clarke, 1998); SPSQ = Social Pshobia Screening Questionnaire (Furmak et al., 1999); SRIP = Self-Rating Inventory for Posttraumatic Stress Disorder (Hovens et al., 2005); STAI-C = State-Trait Anxiety Inventory for Children (Spielberger, Edwards, Lushene, Montuori, & Platzek, 1973); STAI-S/ STAI-T = State-Trait Anxiety Inventory-State/State-Trait Anxiety Inventory-Trait (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983); SUDS = Subjective Units of Discomfort Scale (Wolpe, 1958); UCT = Unstructured Conversation Task (Turner et al., 1994); VAS = visual analogue scale; WDQ = Worry Domains Questionnaire (Tallis, Eysenck, & Mathews, 1992).

Table 2. *ABM effect on symptoms across symptom categories*

	Post intervention					Post-stressor					Follow-up				
Symptom category	*k*	Hedges's *g*	95% CI	*Q within*	I^2	*k*	Hedges's *g*	95% CI	*Q within*	I^2	*k*	Hedges's *g*	95% CI	*Q within*	I^2
Anxiety	22	0.26	[0.132; 0.388]	95.678, $p = 0.000$	78.051	11	0.337	[0.206; 0.468]	14.119, $p = 0.168$	29.171	5	0.216	[-0.108; 0.539]	40.079, $p = 0.000$	90.02
Depression	7	-0.106	[0.382; 0.169]	32.519, $p = 0.000$	81.549	0	-	-	-	-	2	0.161	[-1.245; 1.567]	6.881, $p = 0.009$	85.466
Distress (measured in healthy participants)*	4	0.211	[0.046; 0.375]	1.420, $p = 0.701$	0	1	0.511	[0.152; 0.870]	-	-	1	0.661	[0.235; 1.086]	-	-
Pain	1	-0.149	[0.421; 0.122]	-	-	0	-	-	-	-	1	0.273	[0.033; 0.513]	-	-
Substance Abuse	5	0.003	[0.155; 0.161]	4.853, $p = 0.303$	17.569	1	0.055	[-0.494; 0.605]	-	-	0	-	-	-	-

Notes: k = number of studies; *Verwoerd et al. (2012) was eliminated as outlier

Table 3. *Categorical moderators of the ABM effect on symtoms in anxiety studies*

		Post intervention				Post stressor			
Moderator variable		*g*	95 % CI	*Qwithin*	*Qbetween*	*g*	95 % CI	*Qwithin*	*Qbetween*
Type of anxiety symptoms*	Generalized anxiety	0.607	[0.426, 0.789]	$Q(3) = 0.472$, $p = 0.925$	$Q(2) = 10.753$, $p = 0.005$	0.484	[-0.091; 1.060]	$Q(0) = 0.000$, $p = 1.000$	$Q(2) = 0.123$, $p = 0.940$
	Social anxiety	0.24	[0.052; 0.428]	$Q(12) = 61.739$, $p = 0.000$		0.378	[0.206; 0.550]	$Q(6) = 5.575$, $p = 0.472$	
	Others	0.162	[-0.104, 0.428]	$Q(4) = 12.608$, $p = 0.013$		0.395	[0.159; 0.631]	$Q(2) = 1.609$, $p = 0.447$	
Clinical status	Diagnosed	0.291	[0.095, 0.488]	$Q(11) = 75.224$, $p = 0.000$	$Q(1) = 0.033$, $p = 0.855$	0.399	[-0.168, 0.966]	$Q(1) = 1.879$, $p = 0.170$	$Q(1) = 0.004$, $p = 0.951$
	Analogue sample	0.265	[0.062, 0.469]	$Q(9) = 11.479$, $p = 0.021$		0.381	[0.236, 0.525]	$Q(8) = 5.306$, $p = 0.724$	
Type of training	Towards neutral	0.282	[0.129, 0.435]	$Q(19) = 94.652$, $p = 0.000$	$Q(1) = 0.133$, $p = 0.715$	0.412	[0.270, 0.554]	$Q(9) = 6.237$ $p = 0.716$	$Q(1) = 1.070$, $p = 0.301$
	Towards positive	0.337	[0.083, 0.592]	$Q(1) = 273$, $p = 0.601$		0.163	[-0.288, 0.613]	$Q(0) = 0.000$, $p = 1.000$	
Type of training stimuli	Pictures	0.254	[0.105, 0.403]	$Q(18) = 82.829$, $p = 0.000$	$Q(1) = 2.940$, $p = 0.086$	0.389	[0.210, 0.568]	$Q(6) = 5.618$, $p = 0.467$	$Q(1) = 0.000$, $p = 0.993$
	Words	0.53	[0.252, 0.807]	$Q(2) = 2.825$, $p = 0.244$		0.39	[0.185, 0.596]	$Q(3) = 1.689$, $p = 0.639$	
Training task	Dot-probe	0.319	[0.171, 0.467]	$Q(19) = 90.602$, $p = 0.000$	$Q(1) = 6.179$, $p = 0.013$	0.382	[0.237, 0.526]	$Q(8) = 5.247$, $p = 0.731$	$Q(1) = 0.001$, $p = 0.973$
	Visual cueing	-0.135	[-0.462, 0.191]	$Q(1) = 0.133$, $p = 0.715$		0.392	[-0.176, 0.959]	$Q(1) = 1.968$, $p = 0.161$	
Training setting	Lab	0.369	[0.213, 0.524]	$Q(17) = 59.922$, $p = 0.000$	$Q(1) = 13.461$, $p = 0.000$	0.378	[0.240, 0.516]	$Q(9) = 6.684$, $p = 0.670$	$Q(1) = 0.623$, $p = 0.430$
	Home	-0.015	[-0.148, 0.119]	$Q(3) = 4.545$, $p = 0.208$		0.642	[0.001, 1.282]	$Q(0) = 0.000$, $p = 1.000$	

Stimuli presentation during training	Horizontal	0.3	[0.022, 0.577]	$Q(5) = 14.630$, $p = 0.012$	$Q(1) = 0.015$, $p = 0.902$	0.392	[-0.176, 0.959]	$Q(1) = 1.968$, $p = 0.161$	$Q(1) = 0.001$, $p = 0.973$
	Vertical	0.285	[0.142, 0.428]	$Q(15) = 78.738$, $p = 0.000$		0.382	[0.237, 0.526]	$Q(8) = 5.247$, $p = 0.731$	
Nature of dependent variable	Self-report measures	0.131	[-0.000, 0.263]	$Q(18) = 11.384$, $p = 0.877$	$Q(2) = 4.573$, $p = 0.102$	0.311	[0.141, 0.480]	$Q(9) = 5.541$, $p = 0.785$	$Q(1) = 1.941$, $p = 0.164$
	Clinician-rated measures	0.385	[0.152, 0.617]	$Q(9) = 14.046$, $p = 0.121$		-	-	-	
	Bio-behavioral measures	0.634	[-0.173, 1.441]	$Q(2) = 8.327$, $p = 0.016$		0.527	[0.274, 0.780]	$Q(3) = 3.545$, $p = 0.315$	
Primary versus secondary outcome**	Primary outcome	0.314	[0.188, 0.441]	$Q(20) = 19.451$, $p = 0.493$	$Q(1) = 7.259$, $p = 0.007$	0.369	[0.201, 0.538]	$Q(9) = 7.233$, $p = 0.613$	$Q(1) = 6.214$, $p = 0.013$
	Secondary outcome	0.033	[-0.128, 0.194]	$Q(12) = 11.691$, $p = 0.471$		-0.256	[-0.718, 0.206]	$Q(1) = 7.383$, $p = 0.689$	

Notes: *"Generalized anxiety" symptom category included studies conducted with diagnosed participants, as well as with high anxiety individuals and high worriers; 'Others" symptoms category included phobias and post-traumatic stress disorder, as well as undifferentiated/comorbid symptoms of anxiety (see Eldar et al. 2012; Bar-Haim et al., 2011) ;**For this moderator we used outcomes subgroup within the study as the unit of analysis. For the other categorical moderators, study was used as the unit of analysis.

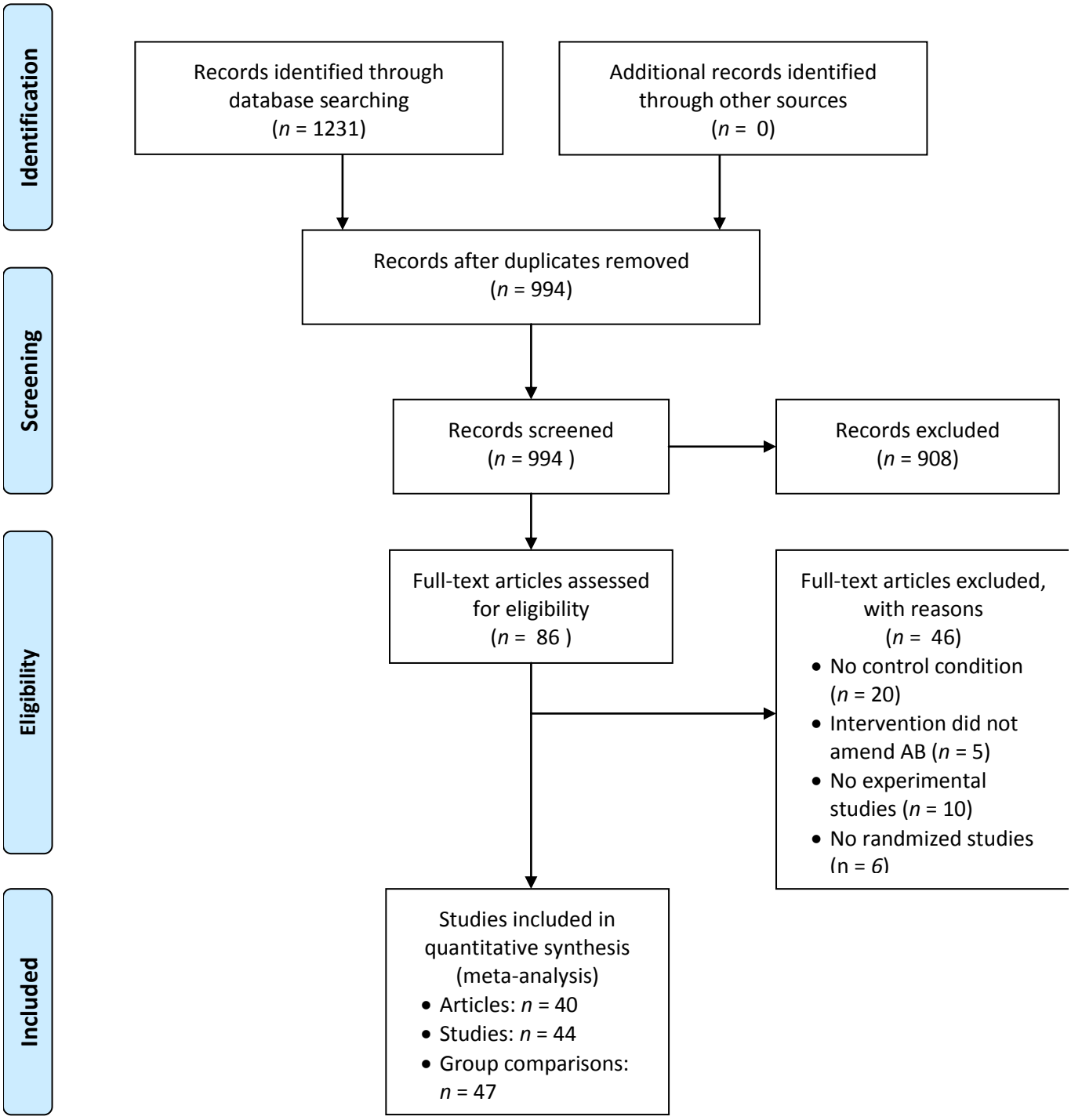

Figure 1. Studies' flow chart

Appendix A
Coding Categories for Dependent Variables: Primary versus Secondary Outcomes

Type of outcome	Symptom category	Measures
Primary outcomes	Generalized Anxiety/high worriers/high anxiety	Anxiety severity scale, anxiety symptoms count, SCAS, SCID, HRSA, PSWQ, WDQ, STAI-S, STAI-T, STAIC, negative thoughts intrusions
	Social anxiety	ADIS, BA, BARS, BASA, BSI, BSPS, CGI, FNES, IST, LSAS, SCID, SDS, SIAS, SPAI, SPS, SPSQ, SUDS, UCT
	Spider Phobia	BAT, SPQ, SUDS
	PTSD	CAPS, SRIP, Impact of Movie Scale, Intrusion monitoring task, no. of intrusions (daily diary)
	OCD symptoms	BAT
	Depression	BDI-II, CAR, HAM-D, RRS, POMS depression, MASQ-GDD, MASQ-AD, MASQ-GDM
	Alcohol consumption	Amount of beer consumed, DAQ, VAS urge to drink, preference task, time to discharge, time to relapse
	Smoking	QSU, VAS craving, choose to smoke, amount willing to pay, delay discounting, VAS urge to smoke
	Pain	Roland-Morris Disability Questionnaire, Tampa Scale for Kinesiophobia, VAS
Secondary outcomes	ASI, BAI, BDI-II, STAI, CDI, CES-DC, HADS, MASQ-GD, MASQ-GDA, MASQ-AA, POMS anger, POMS fatigue, POMS vigor, POMS tension, QOLI, VAS anxiety, VAS mood, DASS, PA	

Notes: ADIS = Anxiety Disorders Interview Shedule (Brown, Dinardo, & Barlow, 1994); ASI = Anxiety Sensitivity Index (Reiss et al., 1986); BAI = Beck Anxiety Inventory (Beck, Epstein, Brown, & Steer, 1988); BARS = Behavioral Avoidance rating Scale (Beidel et al., 2007); BASA = Behavioral Assessment of Speech Anxiety (Mulac & Sherman, 1974); BAT = Behavioral Approach Test; BDI-II = Beck Depression Inventory-II (Beck, Steer, & Brown, 1996); BSI = Brief Symptom Inventory (Derogatis & Melisaratos, 1983); BSPS = Brief Social Phobia Scale (Davidson et al., 1991); CAR = cortisol awaking response; CAPS = Clinician-Administered PTSD Scale (Blake et al., 1990); CDI = Children's Depression Inventory (Kovacs, 1985); CES-DC = Center for Epidemiologic Depression Studies for Children (Weissman et al., 1980); CGI = Clinical Global Impression of Improvement (Guy, 1976); CIQ = Cognitive Interference Questionnaire (Sarason & Stroops, 1978); DAQ = Desires for Alcohol Questionnaire (Love et al., 1998); DASS = Depression, Anxiety, and Stress Scale (Lovibond & Lovibond, 1995); ERP-N2, ERP-P2, ERP-P3 = event-related potential N2, P2, P3; FNES = Fear of Negative Evaluation Scale (Watson &

Friend, 1969); HADS = Hospital Anxiety and Depression Scale (Zigmond & Shaith, 1983); HAM-D = Hamilton Rating Scale for Depression (Hamilton, 1960); HRSA = Hamilton Rating Scale for Anxiety (Hamilton, 1959); IST = Impromptu Speech Task (Beidel et al., 2010); LSAS = Liebowitz Social Anxiety Scale (Liebowitz, 1987); MASQ = Mood and Anxiety Symptoms Questionnaire (Watson & Clacrk, 1991); PA = Positive Affect;POMS-SF = Short-form of the Profile of Mood States (Curran, Andykowsky, & Studts, 1995); PSQW = Penn State Worry Questionnaire (Meyer et al., 1990); QSU = Questionnaire on smoking urges (Cox et al., 2001); RRS = Ruminative Responses Scale (Nolen-Hoeksema & Morrow, 1991); QOLI = Quality of Life Inventory (Frish et al., 1992); SCID = Structured Clinical Interview for DSM-IV (First, Spitzer, Gibbons, & Williams, 2001); SES = State Self-esteem Scale (Heatherton & Polivy, 1991); SDS = Sheehan Disability Scale (Leon, Olfson, Portera, Faber, & Sheenan, 1997); SCAS = Spence Childre Anxiety Scale (Spence, 1998); SIAS = Social Interaction Anxiety Scale (Mattick & Clarke, 1998); SPQ = Spider Questionnaire (Klorman et al., 1974); SPS = Social Phobia Scale (Mattick & Clarke, 1998); SPSQ = Social Pshobia Screening Questionnaire (Furmak et al., 1999); SRIP = Self-Rating Inventory for Posttraumatic Stress Disorder (Hovens et al., 2005); STAI-C = State-Trait Anxiety Inventory for Children (Spielberger, Edwards, Lushene, Montuori, & Platzek, 1973); STAI-S/ STAI-T = State-Trait Anxiety Inventory-State/State-Trait Anxiety Inventory-Trait (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983); SUDS = Subjective Units of Discomfort Scale (Wolpe, 1958); UCT = Unstructured Conversation Task (Turner et al., 1994); VAS = visual analogue scale; WDQ = Worry Domains Questionnaire (Tallis, Eysenck, & Mathews, 1992).

Appendix B

Coding Categories for Dependent Measures: Self-report, Clinician-rated, and Bio-behavioral Measures

Type of outcome	Measures
Bio-behavioral	BA, BAT, BASA, cbehavioral preferences, confidence rated by other, ERP-N2, ERP-P2, ERP-P3, CAR, choose to smoke, cortisol index, cortisol reactivity, delay discounting, HF, LF, HF/LF ratio, IST, mean bear consumption, RMSSD, speech performance, time to relapse, UCT
Clinician-rated	ADIS, BSPS, Anxiety severity scale, anxiety symptoms count, CAPS, CGI, LSAS (clinician-rated), HRSA, HAM-D, SCID, SDS, time to discharge
Self-report	ASI, BAI, LSAS (self-report), STAI-S, STAI-T, BDI-II, SPAI, WDQ, MASQ, PSWQ, POMS-SF, RRS, QSU, VAS craving, VAS urge to smoke/drink, VAS anxiety, VAS deprression, SIAS, FNES, SPS, SPQ, DAQ, CDI, STAI-C, amount willing to pay for a cigarette, VAS feelings of rejection, SES, CIQ, QOLI, school abilities self-esteem, negative thoughts intrusions, difficulty to worry, positie thoughts during worry perios, SUDS, VAS pain, DASS, Roland-Morris Disability Questionnaire, Tampa Scale for Kinesiophobia, PANAS, Impact of Movie Scale, intrusions monitoring task, number of intrusions, SRIP, CES-DC, SCAS

Notes: ADIS = Anxiety Disorders Interview Shedule (Brown, Dinardo, & Barlow, 1994); ASI = Anxiety Sensitivity Index (Reiss et al., 1986); BAI = Beck Anxiety Inventory (Beck, Epstein, Brown, & Steer, 1988); BARS = Behavioral Avoidance rating Scale (Beidel et al., 2007); BASA = Behavioral Assessment of Speech Anxiety (Mulac & Sherman, 1974); BAT = Behavioral Approach Test; BDI-II = Beck Depression Inventory-II (Beck, Steer, & Brown, 1996); BSI = Brief Symptom Inventory (Derogatis & Melisaratos, 1983); BSPS = Brief Social Phobia Scale (Davidson et al., 1991); CAR = cortisol awaking response; CAPS = Clinician-Administered PTSD Scale (Blake et al., 1990); CDI = Children's Depression Inventory (Kovacs, 1985); CES-DC = Center for Epidemiologic Depression Studies for Children (Weissman et al., 1980); CGI = Clinical Global Impression of Improvement (Guy, 1976); CIQ = Cognitive Interference Questionnaire (Sarason & Stroops, 1978); DAQ = Desires for Alcohol Questionnaire (Love et al., 1998); DASS = Depression, Anxiety, and Stress Scale (Lovibond & Lovibond, 1995); ERP-N2, ERP-P2, ERP-P3 = event-related potential N2, P2, P3; FNES = Fear of Negative Evaluation Scale (Watson & Friend, 1969); HADS = Hospital Anxiety and Depression Scale (Zigmond & Shaith, 1983); HAM-D = Hamilton Rating Scale for Depression (Hamilton, 1960); HF = high frequency (HRV indicator); HRSA = Hamilton Rating Scale for Anxiety (Hamilton, 1959); IST = Impromptu Speech Task (Beidel et al., 2010); LF = low frequency (HRV indicator); LSAS = Liebowitz Social Anxiety Scale (Liebowitz, 1987); MASQ = Mood and Anxiety Symptoms Questionnaire (Watson & Clacrk, 1991); PA = Positive Affect;POMS-SF = Short-form of the Profile of Mood States (Curran, Andykowsky, & Studts, 1995); PSQW = Penn State Worry Questionnaire (Meyer et al., 1990); QSU = Questionnaire on smoking urges (Cox et al., 2001); RRS = Ruminative Responses Scale (Nolen-Hoeksema & Morrow, 1991); QOLI = Quality of Life Inventory (Frish et al., 1992); RMSSD = root mean square of the differences of succesive intervals (heart rate

variability - HRV indicator); SCID = Structured Clinical Interview for DSM-IV (First, Spitzer, Gibbons, & Williams, 2001); SES = State Self-esteem Scale (Heatherton & Polivy, 1991); SDS = Sheehan Disability Scale (Leon, Olfson, Portera, Faber, & Sheenan, 1997); SCAS = Spence Childre Anxiety Scale (Spence, 1998); SIAS = Social Interaction Anxiety Scale (Mattick & Clarke, 1998); SPQ = Spider Questionnaire (Klorman et al., 1974); SPS = Social Phobia Scale (Mattick & Clarke, 1998); SPSQ = Social Pshobia Screening Questionnaire (Furmak et al., 1999); SRIP = Self-Rating Inventory for Posttraumatic Stress Disorder (Hovens et al., 2005); STAI-C = State-Trait Anxiety Inventory for Children (Spielberger, Edwards, Lushene, Montuori, & Platzek, 1973); STAI-S/ STAI-T = State-Trait Anxiety Inventory-State/State-Trait Anxiety Inventory-Trait (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983); SUDS = Subjective Units of Discomfort Scale (Wolpe, 1958); UCT = Unstructured Conversation Task (Turner et al., 1994); VAS = visual analogue scale; WDQ = Worry Domains Questionnaire (Tallis, Eysenck, & Mathews, 1992).