Clinical Psychology Review 40 (2015) 76–90

Contents lists available at ScienceDirect

Clinical Psychology Review

Attention bias modification for social anxiety: A systematic review and meta-analysis

Alexandre Heeren [a,*], Cristina Mogoaşe [b], Pierre Philippot [a], Richard J. McNally [c]

[a] Psychological Sciences Research Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium
[b] Department of Clinical Psychology and Psychotherapy, Babeş-Bolyai University, Cluj-Napoca, Romania
[c] Department of Psychology, Harvard University, Cambridge, MA, USA

HIGHLIGHTS

- We examined the effects of attention bias modification (ABM) for social anxiety (SA).
- ABM had small effects on SA symptoms, attentional bias, and reactivity to speech challenge.
- ABM's characteristics, study design, and trait anxiety moderated effect sizes.
- Effects on secondary symptoms and SA symptoms at 4-month follow-up were nonsignificant.
- The quality of the studies was substandard and wedged the effect sizes.

ARTICLE INFO

Article history:
Received 12 November 2014
Received in revised form 31 May 2015
Accepted 4 June 2015
Available online 6 June 2015

Keywords:
Attention bias modification
Social anxiety disorder
Speech performance
Cognitive bias modification
Attentional bias
Meta-analysis
Systematic review

ABSTRACT

Research on attention bias modification (ABM) for social anxiety disorder (SAD) is inconclusive, with some studies finding clear positive effects and other studies finding no significant benefit relative to control training procedures. In this meta-analysis, we assessed the efficacy of ABM for SAD on symptoms, reactivity to speech challenge, attentional bias (AB) toward threat, and secondary symptoms at posttraining as well as SAD symptoms at 4-month follow-up. A systematic search in bibliographical databases uncovered 15 randomized studies involving 1043 individuals that compared ABM to a control training procedure. Data were extracted independently by two raters. The *Q* statistic was used to assess homogeneity across trials. All analyses were conducted on intent-to-treat data. ABM produced a small but significant reduction in SAD symptoms ($g = 0.27$), reactivity to speech challenge ($g = 0.46$), and AB ($g = 0.30$). These effects were moderated by characteristics of the ABM procedure, the design of the study, and trait anxiety at baseline. However, effects on secondary symptoms ($g = 0.09$) and SAD symptoms at 4-month follow-up ($g = 0.09$) were not significant. Although there was no indication of significant publication bias, the quality of the studies was substandard and wedged the effect sizes. From a clinical point of view, these findings imply that ABM is not yet ready for wide-scale dissemination as a treatment for SAD in routine care. Theoretical implications for the integration of AB in the conceptualization of SAD are discussed.

Contents

1. Introduction . . . 77
2. Previous comprehensive evidence . . . 77
3. Overview of the present meta-analysis . . . 78
4. Method . . . 79
   4.1. Literature search . . . 79
   4.2. Selection of the studies . . . 79
   4.3. Quality assessment and data extraction . . . 79
   4.4. Meta-analytic procedure . . . 80

* Corresponding author at: Laboratory for Experimental Psychopathology, Institute of Psychological Science, Université catholique de Louvain, Place du Cardinal Mercier, 10, B-1348 Louvain-la-Neuve, Belgium.
E-mail addresses: heeren.alexandre@gmail.com, Alexandre.Heeren@uclouvain.be (A. Heeren).

http://dx.doi.org/10.1016/j.cpr.2015.06.001
0272-7358/

5. Results . . . 81
5.1. Characteristics of the studies . . . 81
5.2. Quality assessment . . . 81
5.3. Quantitative data synthesis . . . 83
5.3.1. SAD symptoms at posttest . . . 83
5.3.2. Speech challenge at posttest . . . 85
5.3.3. AB at posttest . . . 85
5.3.4. Secondary symptoms at posttest . . . 85
5.3.5. SAD symptoms at 4-month follow-up . . . 86
5.3.6. Complementary analyses . . . 86
6. Discussion . . . 86
6.1. Main effects . . . 86
6.2. Moderator effects . . . 86
6.3. Risk of bias . . . 88
6.4. Theoretical and clinical implications . . . 88
6.5. Limitations . . . 88
Role of funding sources . . . 88
Contributors . . . 89
Conflict of interest . . . 89
Appendix A. Supplementary data . . . 89
References . . . 89

1. Introduction

Social anxiety disorder (SAD) is the most common anxiety disorder with a lifetime prevalence of more than 12% (e.g., Stein & Stein, 2008). SAD is characterized by intense fear in social situations, causing considerable distress and impaired daily functioning. Although there are several empirically supported psychological (for a meta-analysis, see Acarturk, Cuijpers, van Straten, & de Graaf, 2009) and pharmacological treatments for SAD (for a meta-analysis, see Blanco et al., 2003), many patients with this condition do not access treatment for a number of reasons (e.g., inability to afford treatment, concern about what others might think, concern over side effects; Gunter & Whittal, 2010; Lovell & Richards, 2000; Olfson et al., 2000; Weisberg, Dyck, Culpepper, & Keller, 2007). Moreover, even when they inquire about treatment, only about 15% initiate it (e.g., Olfson et al., 2000). These findings highlight the importance of developing effective treatments that are widely accessible and acceptable for individuals with SAD.

Recently, a growing body of research has accumulated on a new treatment for reducing anxiety, called attention bias modification (ABM). ABM builds upon cognitive theories of psychopathology that implicate attentional bias for threat (AB) in the maintenance, and perhaps the etiology, of SAD (Morrison & Heimberg, 2013). The clinical purpose of ABM is to reduce AB, thereby diminishing anxiety proneness and symptoms (MacLeod & Mathews, 2012). The most common ABM procedure is a modification of the visual dot-probe task (MacLeod, Rutherford, Campbell, Ebsworthy, & Holker, 2002) based on the classic work of MacLeod, Mathews, and Tata (1986). In early versions of the dot-probe task (e.g., MacLeod et al., 1986), participants viewed two stimuli (e.g., a pair of threatening-neutral words or photographs) presented in two distinct locations (presented either horizontally or vertically) of a computer screen for a brief duration (usually 500 ms). Immediately thereafter, a dot appeared in the location previously occupied by one of the two stimuli. In different versions, participants had to indicate the location of the probe (right or left versus up or down) or to indicate its identity (e.g., "E" or "F") as quickly as possible. An AB occurred when participants responded faster to the probe when it replaced a threatening stimulus than when it replaced a nonthreatening stimulus, indicating that their attention was directed to the location occupied by the threatening stimulus.

In ABM, researchers typically modify the original task so that the probe nearly always (e.g., 95% of the trials) replaces the neutral or positive stimulus, thereby redirecting subjects' attention to non-threatening cues. In the control condition, there is no contingency between cues and probes. Relative to the control condition, ABM often reduces symptoms in people with SAD (e.g., Amir, Weber, Beard, Bomyea, & Taylor, 2008; Amir et al., 2009; Heeren, Reese, McNally, & Philippot, 2012b; Li, Tan, Qian, & Liu, 2008; Schmidt, Richey, Buckner, & Timpano, 2009). These findings suggest that ABM could have important clinical potential for treating SAD, as it entails a very simple protocol, little effort and motivation from the patient, little contact with a mental health professional, and can be easily disseminated (e.g., Amir, Taylor, & Donohue, 2011; Heeren, Maurage, & Philippot, 2013). However, over the past two years, other studies have reported mixed findings (e.g., Boettcher, Berger, & Renneberg, 2012; Boettcher et al., 2013; Carlbring et al., 2012; Heeren, Lievens, & Philippot, 2011; Julian, Beard, Schmidt, Powers, & Smits, 2012; McNally, Enock, Tsai, & Tousian, 2013). More specifically, these studies have shown that ABM and the control condition did not differ significantly at posttraining in reducing AB or SAD symptoms. That is, although the AB condition often attenuated anxiety symptoms, the control condition performed just as well. These failures to replicate initial results with ABM have prompted a dismissive appraisal of ABM's prospects as a viable clinical intervention from some commentators (e.g., Emmelkamp, 2012).

2. Previous comprehensive evidence

Over the last four years, several systematic reviews have affirmed the clinical potential of ABM across a variety of clinical conditions. Indeed, to date, five meta-analyses have been published on the effects of ABM (Beard, Sawyer, & Hofmann, 2012a; Cristea, Kok, & Cuijpers, 2015; Hakamata et al., 2010; Hallion & Ruscio, 2011; Mogoaşe, David, & Koster, 2014).

The first meta-analysis (Hakamata et al., 2010) summarized the findings of 12 studies that used ABM to reduce AB and anxiety (all anxiety disorders included). It revealed that ABM has a small-to-medium effect ($d = 0.51$) on symptom reduction and a large effect on AB ($d = 1.16$). Moreover, the correlation between the effect sizes (ES) on AB and anxiety was large ($r = .75$) and nearly significant ($p = 0.052$). Exploring potential moderators, the investigators found that the ES for symptom reduction was larger for trait anxiety than for state anxiety, words versus face stimuli, and top-bottom versus left–right presentation of stimuli during training. Regarding ABM's effect on AB reduction, the ES was larger for multiple sessions than for a single session of training. However, the inclusion of diverse anxiety disorders and the small number of studies rendered it impossible to ascertain the effects on SAD alone.

The meta-analysis of Hallion and Ruscio (2011) had a broader focus, including a larger number of studies (45, and 24 concerning ABM). However, their meta-analysis examined *both* ABM and another cognitive bias modification (CBM) technique (i.e., interpretation training), and it combined anxious and depressive symptoms. Results revealed a small, but reliable effect on AB ($g = 0.29$) and anxiety ($g = 0.13$ for post-test, $g = 0.28$ for poststressor), and a non-significant effect on depression ($g = 0.12$). However, unlike Hakamata et al. (2010), the effects did not vary as a function of number of training sessions. As pointed out by Clarke, Browning, Hammond, Notebaert, and MacLeod (2014b), the Hallion and Ruscio meta-analysis had several serious limitations. First, it pooled the results of both ABM and interpretive bias modification studies *without reporting separate ES* for each method. This is unfortunate as the mechanisms of attention and interpretive bias may differ, including in their responsiveness to training interventions. Second, like Hakamata et al. (2010), Hallion and Ruscio (2011) did not differentiate among anxiety disorders and did not assess whether the benefits of ABM may be moderated by the type of anxiety disorder (e.g., social anxiety, specific phobia, generalized anxiety disorder). Accordingly, uncertainty abounds regarding the benefits of ABM for individuals suffering from SAD.

Beard et al. (2012a) published a more comprehensive meta-analysis, including 37 studies examining ABM effects on anxiety, depression, and addictive behavior symptoms. Their results showed a large ES for AB when training toward neutral was compared with a control condition ($g = 0.80$) and small changes when training toward positive was compared with a control condition ($g = 0.24$). They found a small and non-significant ES for symptoms after a single session of training ($g = 0.01$ for training toward neutral stimuli versus control condition, and $g = 0.09$ for training toward positive stimuli versus control condition). Small-to-medium ESs were obtained at poststressor (0.22 for training toward neutral stimuli, and 0.60 for training toward positive stimuli versus control condition) and at posttreatment (i.e., multiple ABM sessions; $g = 0.41$ for training toward neutral stimuli versus control condition, and $g = 0.09$ for training toward positive stimuli versus control condition). Consistent with Hakamata et al. (2010), they reported that training with top-bottom orientation was more effective than left-right orientation. However, unlike Hakamata et al. (2010) who found that words were superior to pictures, Beard et al. (2012a) reported that training with pictures was more effective in reducing AB than training with words. Consistent with Hakamata et al. (2010), they also found that number of sessions moderated effects, with greater number of sessions yielding larger ES. However, unlike Hakamata et al. (2010), who included exclusively anxiety studies, Beard et al. (2012a) included a wide range of disorders and symptoms, *without providing different estimates of ABM effect on symptoms by disorder type or symptom category*. This is problematic as the plasticity of AB may operate differently in clinical, subclinical (analogue), and healthy samples. Given the therapeutic nature of ABM, it is critical to ensure the clinical efficacy of ABM in clinical samples.

Recently, Mogoaşe et al. (2014) provided an updated meta-analysis examining the clinical efficacy of ABM in reducing both AB and clinical symptoms in people with anxiety, depression, and substance abuse problems as well as in healthy participants. Their analysis comprised 43 controlled trials. Interestingly, they were the first to include studies with negative results in their ES estimates. Their results showed a small overall ES on symptoms at posttraining ($g = 0.16$) and medium ES ($g = 0.45$) on AB at posttraining, both driven by anxiety studies ($g = 0.26$) and healthy participants ($g = 0.21$). ES for both symptoms and AB were larger for studies conducted in the laboratory than online. They also found, in the anxiety study subsample, that participants' age significantly moderated the ABM effect on both AB and symptoms at posttraining, with younger participants benefiting more from intervention than older ones. More importantly, irrespective of clinical status, they also found that the preexisting AB was significantly related to change in AB, and the change in AB correlated significantly with change in symptoms. However, like Beard et al. (2012a), Mogoaşe et al. (2014) did not report separate overall and moderator ES analyses for SAD studies. This is unfortunate as clinical efficacy and the effect of moderators (e.g., participant's age, number of session, change in AB, the type of the threatening stimuli during training) may operate differently across anxiety disorders, and the disorders may differ in their plasticity or responsiveness to training.

Cristea et al. (2015) meta-analyzed all the RCTs on CBM interventions across a diversity of disorders. Their analysis comprised 49 studies, grouped by anxiety and depression outcomes. ES for both anxiety (for all measures, $g = 0.37$; for general anxiety, $g = 0.38$; for social anxiety, $g = 0.40$) and depression ($g = 0.43$) were small but significant. Consistent with Mogoaşe et al. (2014), ESs were larger for studies conducted in the laboratory than online. For anxiety outcomes, ESs were larger when participants were compensated for their participation than when not compensated. However, contrary to previous meta-analyses, ES was negatively correlated with the number of sessions. Importantly, they assessed the quality of the studies, finding that most had methodological limitations, and that the quality of the study varied inversely with the ES for symptom reduction. Strong evidence of a time-lag publication bias emerged such that studies published earlier had larger ESs than those published later (Ioannidis, 1998, 2008). In the same vein, they found a near-significant positive correlation between ES and the impact factor of the journal publishing the study. Taken together, this meta-analysis suggests that early studies with large ESs published in top-tier journals reinforced the impression of CBM as a powerful potential treatment, methodological limitations of the studies notwithstanding.

However, like previous meta-analyses, this study has several limitations. First, although they computed distinct ESs for anxiety and depression outcomes, they did not differentiate among different disorders and did not assess whether the benefits of CBM are moderated by type of disorder. Second, they pooled the results of ABM and other CBM studies. Third, they computed ESs for clinically relevant outcomes (standardized symptom or distress measures) without examining the effect of CBM on AB, reactivity to stressor, or follow-up data.

3. Overview of the present meta-analysis

Despite the five previous meta-analyses on the clinical efficacy of ABM, none focused on a single disorder. Moreover, this is especially unfortunate for SAD — the chief target in most studies and the disorder for whom ABM may be most appropriate as either a stand-alone treatment (e.g., Amir et al., 2011; Heeren et al., 2013) or as integrated into a standard cognitive-behavioral treatment package (e.g., Rapee et al., 2013).

Furthermore, only two meta-analyses included studies with negative findings (i.e., Cristea et al., 2015; Mogoaşe et al., 2014). Yet during the past two years, many such studies have appeared for SAD (e.g., Boettcher et al., 2012, 2013; Carlbring et al., 2012; Heeren et al., 2011; Julian et al., 2012). Including published "failure to replicate" studies is essential as all previous meta-analyses found evidence of publication bias.

In the same vein, only one previous meta-analysis considered the quality of the studies, finding that it moderated ESs. The quality of RCTs in meta-analyses can bias the conclusions (e.g., Moher et al., 1998; Wood et al., 2008), especially for trials with small sample sizes (Kjaergard, Villumsen, & Gluud, 2001). This point is critical, as Cristea et al. (2015) emphasized, because CBM researchers are overly reliant on small studies of substandard quality. Substandard quality is often associated with artificial inflation of ESs (Ioannidis, 2008). Indeed, Cristea et al. (2015) found that the ES of CBM is inversely related to the quality of the study.

Moreover, previous meta-analyses yielded inconsistent findings, partly because they covered different studies and included diverse disorders. For example, Hakamata et al. (2010) reported that ABM has a medium ES for symptom reduction and a large ES for AB reduction, whereas Mogoaşe et al. (2014) found a small ES for symptom reduction

and medium ES for AB reduction. In the same vein, there were several discrepancies regarding potential moderators of ABM efficacy. For instance, although both Hakamata et al. (2010) and Mogoaşe et al. (2014) revealed that the number of training sessions moderated the effect of ABM, this moderator was nonsignificant in the Hallion and Ruscio's (2011) meta-analysis. Combining outcomes for different diagnostic groups may be one cause of such inconsistencies. Indeed, as we already pointed out, the clinical efficacy and the different moderators (e.g., participant's age, sessions' number, change in AB, the type of the threatening stimuli during training) may operate in different ways according to the psychopathological condition.

In contrast to most previous meta-analyses, ours examined moderators that may account for inconsistencies in the effects of ABM across SAD studies. For example, our moderators included baseline severity of SAD symptoms as measured by the Liebowitz Social Anxiety Scale (LSAS; Liebowitz, 1987); whether a study involved a stressful speech challenge; and baseline trait anxiety, a variable often associated with AB (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007). Furthermore, we examined how investigators presented their studies to potential participants during recruitment. Did investigators describe the study as an experiment or as a potentially therapeutic intervention? Finally, following Mogoaşe et al. (2014), we examined baseline AB as a moderator. Indeed, studies failing to find a therapeutic advantage of ABM over the control procedures also failed to show an effect on AB (e.g., Carlbring et al., 2012; McNally et al., 2013; Neubauer et al., 2013), and two studies revealed that reduction in AB predicted reduction in symptoms (Amir et al., 2011; Kuckertz et al., 2014). Indeed, as MacLeod and Clarke (2015) emphasized, ABM presupposes that one must reduce AB for ABM to reduce symptoms.

To summarize, we conducted an updated quantitative review with a clinical focus, aimed to assess the efficacy of ABM for SAD, testing crucial moderators. Our first goal was to investigate the degree to which ABM reduces SAD symptoms and AB. Our second goal was to test for possible moderators of ABM effects in SAD. Finally, our third goal was to gauge the quality of these studies and examine potential publication bias.

4. Method

4.1. Literature search

We performed the meta-analysis in accordance with the PRISMA guidelines (Liberati et al., 2009; see Appendix A). Potentially relevant studies were identified following a systematic search of the Scopus, PubMed, and PsycInfo database through October 2014, using the following keywords: “attentional bias modification”, “attentional training”, “attentional retraining”, combined with “social anxiety”. We also systematically searched the references within the most recent articles (De Voogt, Wiers, Prins, & Salemink, 2014; McNally et al., 2013), and recent reviews of ABM for anxiety disorders (Cristea et al., 2015; Mogoaşe et al., 2014; Van Bockstaele et al., 2014). We did not search for unpublished studies as failures to replicate the beneficial effects of ABM for SAD have been appearing regularly in the literature (e.g., Boettcher et al., 2012, 2013; Carlbring et al., 2012; Heeren et al., 2011; Julian et al., 2012).

4.2. Selection of the studies

The flow of information from identification to inclusion of studies is summarized in the PRISMA diagram depicted in Fig. 1 (Moher, Liberati, Tetzlaff, Altman, & the PRISMA Group, 2009). Our search strategy identified 189 publications. Two additional records (i.e. online first publication papers) were identified through other sources. Duplicates were removed, and the abstracts from the remaining 97 publications were screened. We excluded review articles, qualitative studies, case studies, dissertation abstracts, study protocols, and non-English articles (N = 77; in this article, N refers to number of studies, n to number of participants). The remaining 20 articles were selected for further screening, and we excluded articles for the following reasons: (a) the study was not designed specifically to manipulate AB in order to reduce SAD symptoms, emotional vulnerability among SAD individuals, or both; (b) the study did not assess clinically-relevant symptoms of SAD (e.g., self-reported symptoms of SAD, reactivity to a speech challenge, or symptoms assessed through clinical interview); (c) participants were not randomly allocated to training conditions or a control condition (defined as sham training) was not used; (d) the studies investigated the effects of a subliminal version of ABM training that likely targets different mechanisms; (e) the studies investigated the effects of ABM combined with other treatments (e.g., cognitive-behavioral therapy). After applying these exclusions, we found that 15 studies satisfied the inclusion criteria (see Fig. 1).

4.3. Quality assessment and data extraction

The quality of included studies was assessed with the six criteria of the “Risk of Bias” assessment tool, developed by the Cochrane Collaboration (Higgins & Green, 2011), to assess possible sources of bias in randomized trials: (1) Adequate generation of random allocation sequence, (2) concealment of allocation to conditions, (3) prevention of knowledge of the allocated intervention to assessors (blinding of assessors), (4) prevention of knowledge of the allocated intervention to participants (blinding of participants), (5) dealing with incomplete outcome data, (6) and selective outcome reporting. Following Cristea et al. (2015), criterion 5 (dealing with incomplete outcome data) was rated as positive if there were no missing data or if data were analyzed in an intent-to-treat approach (i.e., a post-treatment score was analyzed for every patient even if the last observation prior to attrition had to be carried forward or that score was estimated from earlier response trajectories). Assessment of the quality of the studies was conducted by two of the authors (AH and CM) and disagreements were solved by discussion.

We coded several aspects of the included studies. Each study's characteristics were extracted by one author (AH), and checked independently by a second author (CM). For each included study, we retained the following characteristics (when available): study identification data (author, year of publication), number of participants per comparison, percentage of female participants per study, clinical status of the sample [presence of SAD diagnostic criteria according to the Diagnostic and Statistical Manual of Mental Disorders (4th ed., text rev.; DSM-IV-TR; American Psychiatric Association, 2000). versus highly socially anxious], severity of social anxiety symptoms (i.e., LSAS score at baseline), depressive symptoms at baseline, trait anxiety score at baseline, mean age of the participants, nature of stimuli used during training (i.e., type of threat and nonthreat cues), the way the study was presented to the participants during the recruitment (i.e., as a treatment versus not as a treatment), the number of different pairs of stimuli used during training, the orientation of the stimuli (vertical versus horizontal), participants' compensation (coded as *yes* if participants received money, course credit, or both for their participation), number of training trials, number of sessions, the percentage of trials in the ABM condition when cues predicted probes (i.e., the contingency was operative), the setting of training (laboratory versus Internet), the presence of a speech challenge during the experiment, follow-up length (in weeks), the temporal separation of training sessions (in days), and impact factor (through Web of Science) of the journal in which the study was published for the year of its publication. To compute the magnitude of AB at baseline for studies using a dot-probe task, we tested whether the bias score at baseline significantly differed from 0 (absence of bias) by using one-sample t-tests. For those studies using a spatial cueing task, we tested whether the mean RT for invalid threat trials significantly differed from invalid neutral trials by using a paired t-test. All these analyses were performed on means and standard deviations at baseline for the whole sample. To obtain a metric-independent measure of AB at baseline, we used the ESs of these analyses. Following Mogoaşe et al.

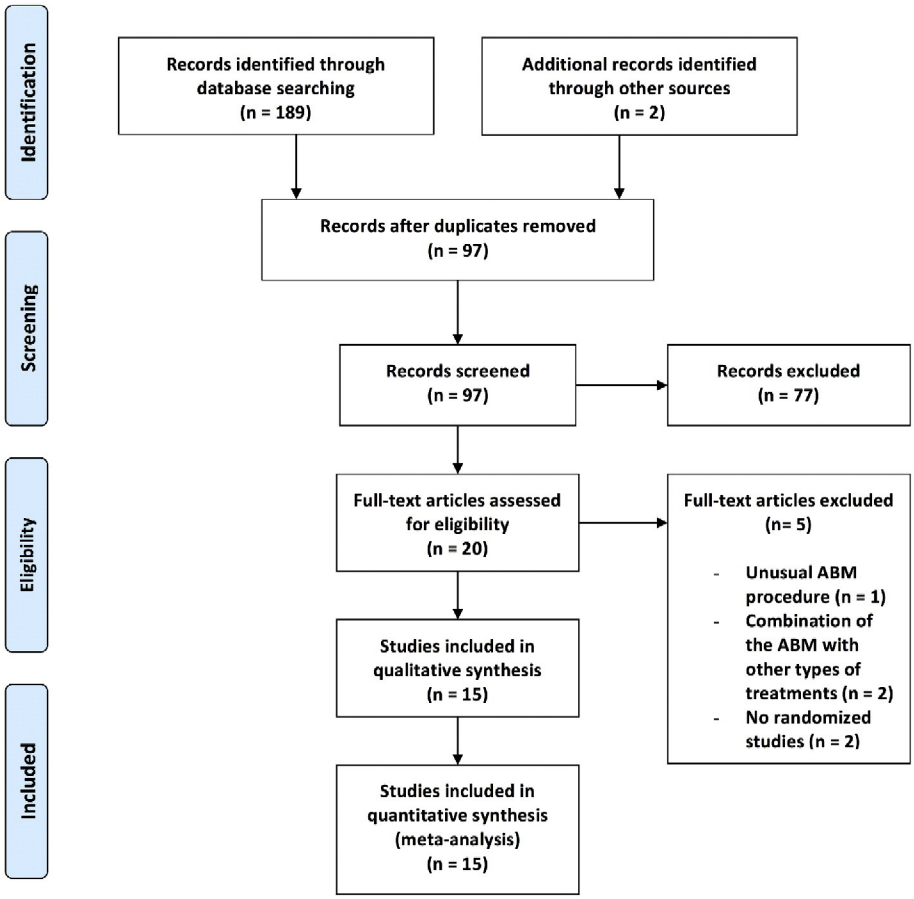

Fig. 1. Flowchart of selection and inclusion of process, following the PRISMA statement.

(2014), we coded these ESs such that a positive value indicated the magnitude of AB at baseline.

Some studies included a third group trained to attend to threat stimuli (e.g., Heeren et al., 2012b; Klumpp & Amir, 2010; McNally et al., 2013). However, as we aimed to compare exclusively the ABM condition to a sham condition, we did not include the data from these third conditions. Two studies reported a four-group design, two experimental and two control groups (Heeren et al., 2011; Julian et al., 2012). For each of these studies, we only extracted data related to standard ABM and control (no-contingency) conditions. Finally, two studies (i.e., Amir et al., 2011; Kuckertz et al., 2014) reported findings on an extended sample of a previous study (i.e., Amir et al., 2009; Carlbring et al., 2012). To prevent non-independence of observations, we used only the data reported in Amir et al. (2009) and Carlbring et al. (2012) because these reports provided more information about potential moderators.

4.4. Meta-analytic procedure

For each comparison between the ABM condition and the sham condition, we computed the ES indicating the difference between the two groups at posttraining (Cohen's *d*). ESs were calculated by subtracting (at posttest) the average score of the ABM group from the average score of the control group, and dividing the result by the pooled standard deviations of the two groups. A value between 0.2–0.5 indicates a small ES, a value of 0.5–0.8 signifies a medium one, and values of 0.8 or larger signify a large ES (Cohen, 1988). Because several studies had small sample sizes, we adjusted the ES for small sample bias according to the procedure developed by Hedges and Olkin (1985) (Hedges's *g*). Hedges's *g* is interpreted like Cohen's *d*. All the ESs were coded such that a positive value of Hedges's *g* indicates greater improvement in the training condition compared to the sham control group. If provided, intention-to-treat data, using a method such as "last observation carried forward" (Ferguson, Aaron, Guyatt, & Herbert, 2002), were preferred over data from completers.

In the calculation of ES, we distinguished between five categories of outcomes measures: (a) self-report measures of social anxiety at posttraining; (b) AB for threat at post-test; (c) reactivity to speech task at posttraining; (d) self-report measures of social anxiety at 4-month follow-up; (e) self-report measures of secondary symptoms at posttraining (i.e., depression, trait anxiety, level of psychopathology, general distress). We did not include clinician-rated semi-structured interview as only two studies (i.e., Amir et al., 2009; Schmidt et al., 2009) provided these data at post-test. Details of the measures included in these categories are provided in Table 1.

We used the Comprehensive Meta-Analyses software (Version 2.2.046; Biostat, Englewood, NJ) to calculate the pooled mean ES. When means and standard deviations were not reported, we used other statistics (Cohen's *d* reported in the study, *t* values and sample sizes, *p* values and degrees of freedom) provided for between-group comparison at posttraining. In addition, when a study reported more than one measure to assess a specific construct (e.g., self-reported measure of SAD symptoms), we computed an average ES of those outcomes at a given point in time (posttraining, follow-up, or both). As we expected considerable heterogeneity among the studies, we computed the mean ES by using a random-effects model, which assumes that studies come from a population of studies where ES varies (Riley, Higgins, & Deeks, 2011). In this model, the ES resulting from the included studies not only differ because of the random error within the studies (as in the fixed effect model), but also because of true variation in ES from one study to another.

Table 1
Coding of the five categories for dependent measures.

Outcome categories	Measures
1. SAD symptoms at post-test	BSPS; BARS; FNE; LSAS; PRCS; SIAS; SPAI; SPS; SPSQ
2. AB at post-test	Dot-probe task; Spatial cueing task
3. Reactivity to speech at posttest	BASA; DBP; IST; HR, SCR; STAI-S; UCT; SUDS; SPRS; SBP; VAS-anxiety; VAS-mood
4. Secondary symptoms at post-test	BDI-II; CGI; DASS; HAM-D; MADRS; PSWQ; STAI-T
5. SAD symptoms at 4-month follow-up	BSPS; FNE; LSAS; SIAS; SPAI; SPS; SPSQ

Notes. AB = attentional bias for threat, BARS = Behavioral Avoidance Rating Scale; BASA = Behavioral Assessment of Speech Anxiety; BDI-II = Beck Depression Inventory-II; BSPS = Brief Social Phobia Scale; CGI = Clinical Global Impression of Improvement; DASS = Depression, Anxiety, and Stress Scale; DBP = diastolic blood pressure; FNE = Fear of Negative Evaluation; HAM-D = Hamilton Rating Scale for Depression; HR = Heart Rate; IST = Impromptu Speech Task; LSAS = Liebowitz Social Anxiety Scale; MADRS = Montgomery–Asberg Depression Rating Scale; PRCS = Personal Report of Confidence as a Speaker; PSQW = Penn State Worry Questionnaire; SAD = Social Anxiety Disorder; SCR = Skin conductance reactivity; SIAS = Social Interaction Anxiety Scale; SPAI = Social Phobia and Anxiety Inventory; SPRS = Social Performance Rating Scale; SPS = Social Phobia Scale; SPSQ = Social Phobia Screening Questionnaire; STAI-S/STAI-T = State-Trait Anxiety Inventory-State/State-Trait Anxiety Inventory-Trait; SUDS = Subjective Units of Discomfort Scale; SBP = systolic blood pressure; UCT = Unstructured Conversation Task; VAS-anxiety = visual analogue anxiety scale; VAS-mood = visual analogue mood scale.

Homogeneity of ES was assessed with the *Q* statistic and the I^2 index (Borenstein, Hedges, Higgins, & Rothstein, 2009; Cooper, Hedges, & Valentine, 2009). The *Q* statistic reflects heterogeneity in ES, comparing true heterogeneity to random error. A statistically significant *Q* value indicates true heterogeneity in ES beyond random error. I^2 indicates the percentage of observed heterogeneity, and unlike *Q*, is not sensitive to the number of studies (Borenstein et al., 2009). A value of 0% indicates no heterogeneity, and larger values signify increasing heterogeneity, with 25% as low and 50% as moderate, and 75% as high heterogeneity (Higgins, Thompson, Deeks, & Altman, 2003).

We assessed potential publication bias in three ways. First, a funnel plot, which plots the standard error for each study (determined by sample size) against the study's ES, was created and visually inspected for each data set (Light & Pillemer, 1984). The assumption of the funnel plot is that studies with larger samples yield larger and more reliable ESs, thereby clustering toward the top of the plot, whereas studies with smaller sample sizes yield smaller ESs that are more susceptible to error. Accordingly, studies with smaller samples should scatter widely about the mean and cluster near the bottom of the plot (Borenstein et al., 2009). Second, we applied Duval and Tweedies's (2000) trim-and-fill procedure to each data set. This procedure calculates the likely number of missing studies on the basis of asymmetry in the funnel plot; it yields corrected ESs and confidence intervals adjusted to account for these missing studies (Borenstein et al., 2009; Cooper et al., 2009; Duval & Tweedies, 2000). We also conducted Egger's test of the intercept to quantify the bias captured by the funnel plot and test its significance (Egger, Davey Smith, Schneider, & Minder, 1997; Sterne, Becker, & Egger, 2005). This procedure relies on a linear regression of normalized effect estimates (estimate divided by its standard error) against precision (the reciprocal of the standard error of the estimate). The intercept in this regression corresponds to the slope in a weighted linear regression of the ES on the standard error. When there is no evidence of funnel plot asymmetry, the intercept should not significantly differ from zero (Egger et al., 1997; Sterne et al., 2005).

Finally, we conducted moderator analyses testing the following variables: the intensity of AB at baseline; severity of social anxiety symptoms at baseline; depressive symptoms at baseline; trait anxiety score at baseline; percentage of female participants per study; mean age of the participants; total number of training trials, number of sessions; the percentage of contingency between cues and probes during ABM; the number of distinct pairs of stimuli; the study's year of publication; the impact factor of the journal in which the study was published; the quality of the study; the clinical status of the sample; the way the study was presented to the participants during the recruitment (treatment versus not a treatment); the nature of stimuli used during the training (i.e., nature of the nonthreat and threatening cues); the orientation of the stimuli; the type of training delivery (laboratory versus Internet); the presence of a speech challenge during the experiment; the presence of compensation for participation.

Categorical variables were tested with a mixed-effects meta-analytic categorical test. In this model, studies within subgroups are pooled with the random-effects model, whereas tests for significant differences between subgroups are conducted with the fixed-effects model. For continuous variables, we used unrestricted maximum likelihood meta-regression analyses to test whether there was a significant relation between the continuous variable and the ESs, as indicated with a *Z* value and an associated *p* value.

5. Results

5.1. Characteristics of the studies

Table 2 summarizes the characteristics of the 15 studies. Study sample size ranged from 24 to 299 with a total of 1043 randomized participants (ABM = 537; Control = 506), 53.63% were women (range 36.75 to 74.45%), and the mean age was 28.47 (range of the means 19.01–39.54). The mean baseline of the LSAS total score was 71.09 (range of the means 39.90–81.81). Five studies included one training session, whereas ten had multiple sessions (mean numbers of sessions = 9.01; *SD* = 15.27). Six studies included a four-month follow-up assessment.

5.2. Quality assessment

The quality of the studies varied. Only five studies reported an adequate random sequence generation. Although 13 studies reported that knowledge of the allocated intervention was adequately hidden to participants, only five studies had assessors adequately blinded to allocation to condition. Six studies conducted intent-to-treat analyses and four did not have any missing data. Regarding selective outcome reporting, only three trials could be traced to a trial registration, and selective outcome reporting was detected in all those that did.

With the exception of the criterion related to selective outcome reporting, only two studies met all five of the remaining criteria (i.e., Boettcher et al., 2013; Carlbring et al., 2012). Although two studies met four criteria and one met three criteria, ten studies only met two ($N = 7$) or one ($N = 3$) criteria. However, with the exception of the criteria related to incomplete data outcome, it should be noted that twelve studies (9/15 for random sequence generation, 8/15 for allocation concealment, 9/15 for blinding of assessors, 2/15 for blinding of participants, 11/15 for selective outcome reporting) did not provide the necessary information for assessing whether the criteria were met. Fig. 2 presents the percentage of studies with a low, unclear (not enough information), and high risk of bias for each of the quality criteria.

Table 2
Characteristics of the studies included in meta-analysis.

Study	Sample type	N	Mean age	% Female	N sessions	N of Training trials per session	Type of training material (threat/nonthreat)	Training stimuli orientation	N distinct stimulus-pairs	Training setting	Compensation	Presence of a speech challenge?	Assessment of AB change?	Presence of 4-month follow-up?	Outcomes
Amir et al. (2008)	Highly socially anxious	94	19	51	1	160	Disgust/neutral faces	Top-bottom	8	Lab	Y	Y	Y	N	Spatial cueing task, STAI-S
Amir et al. (2009)	DSM-IV criteria of SAD	44	29	59	8	160	Disgust/neutral faces	Top-bottom	8	Lab	Y	N	Y	Y	BDI, BSPS, CGI, HAM-D, LSAS, SPAI, Spatial cueing task STAI-T
Boettcher et al. (2012)	DSM-IV criteria of SAD	68	38	37	8	160	Disgust/neutral faces	Top-bottom	8	Internet	N	N	Y	Y	BDI, BSI, LSAS, SIAS, Spatial cueing task, SPS
Boettcher et al. (2013)	DSM-IV criteria of SAD	86	39	57	11	192	Socially threatening words and disgust faces / positive words and happy faces	Top-bottom	173	Internet	N	N	Y	Y	BAI, Dot-probe task, LSAS, MADR, SIAS, SPS,
Bunnell et al. (2013)	DSM-IV criteria of SAD	31	24	45	8	160	Disgust/neutral faces	Top-bottom	8	Lab	N	Y	N	N	BARS, BSPS, BDI, CGI, IST, LSAS, SPAI, UCT
Carlbring et al. (2012)	DSM-IV criteria of SAD	79	37	68	8	160	Disgust/neutral faces	Top-bottom	8	Internet	N	N	N	Y	BDI, LSAS, SIAS, SPS, SPSQ
Enock et al. (2014)	Highly socially anxious	299	35	48	65	80	Disgust/neutral faces	Top-bottom	8	Internet	N	N	Y	N	DASS, Dot-probe task, LSAS, PSWQ, SIAS, Spatial cueing task[a]
Heeren et al. (2011)	DSM-IV criteria of SAD	41	22	74	1	560	Disgust/neutral faces	Left–right	70	Lab	Y	Y	Y	N	BASA, Spatial cueing task, VAS-anxiety, VAS-mood
Heeren et al. (2012b)	DSM-IV criteria of SAD	38	22	56	4	744	Angry/happy faces	Left–right	62	Lab	Y	Y	Y	N	BASA, Dot-probe task, LSAS, FNE, SCR, SUDS
Julian et al. (2012)	Highly socially anxiou	56	20	79	1	160	Disgust/happy faces	Top-bottom	8	Lab	Y	Y	Y	N	Spatial cueing task, STAI-S
Klumpp and Amir (2010)	Moderately socially anxious	53	20		1	160	Disgust/happy faces	Top-bottom	8	Lab	Y	Y	N	N	STAI-S
Li et al. (2008)	Highly socially anxious	24	20	42	7	480	Disgust/happy faces	Left–right	60	Lab	Y	N	Y	N	Dot-probe task, FNE, SIAS, SPS
McNally et al. (2013)	Highly socially anxious	38	39	37	4	384	Disgust/happy faces	Top-bottom	8	Lab	Y	Y	Y	N	DASS, DBP, Dot-probe task, HR, LSAS, PRCS, SBP, SPRS, SUDS
Neubauer et al. (2013)	DSM-IV criteria of SAD	56	40	66	8	160	Disgust/neutral faces	Top-bottom	30	Internet	N	N	Y	N	BDI, Dot-probe task, LSAS, SIAS, SPS
Schmidt et al. (2009)	DSM-IV criteria of SAD	36	23	44	8	160	Disgust/neutral faces	Top-bottom	8	Lab	N	N	N	N	BDI, BSPS, LSAS, SPAI, STAI-T

Notes. AB = attentional bias for threat, BARS = Behavioral Avoidance Rating Scale; BASA = Behavioral Assessment of Speech Anxiety; BDI-II = Beck Depression Inventory-II; BSPS = Brief Social Phobia Scale; CGI = Clinical Global Impression of Improvement; DASS = Depression, Anxiety, and Stress Scale; DBP = diastolic blood pressure; DSM-IV = Diagnostic and Statistical Manual of Mental Disorders (4th ed., text rev.); FNE = Fear of Negative Evaluation; HAM-D = Hamilton Rating Scale for Depression; HR = Heart Rate; IST = Impromptu Speech Task; LSAS = Liebowitz Social Anxiety Scale; MADRS = Montgomery–Asberg Depression Rating Scale; PRCS = Personal Report of Confidence as a Speaker; PSQW = Penn State Worry Questionnaire; SAD = Social Anxiety Disorder; SCR = skin conductance reactivity; SIAS = Social Interaction Anxiety Scale; SPAI = Social Phobia and Anxiety Inventory; SPRS = Social Performance Rating Scale; SPS = Social Phobia Scale; SPSQ = Social Phobia Screening Questionnaire; STAI-S/STAI-T = State-Trait Anxiety Inventory-State/State-Trait Anxiety Inventory-Trait; SUDS = Subjective Units of Discomfort Scale; SBP = systolic blood pressure; UCT = Unstructured Conversation Task; VAS-anxiety = visual analogue anxiety scale; VAS-mood = visual analogue mood scale.

[a] Since only 62 of the 299 participants completed the modified Posner Cueing task, we did not included these data into our statistical analyses.

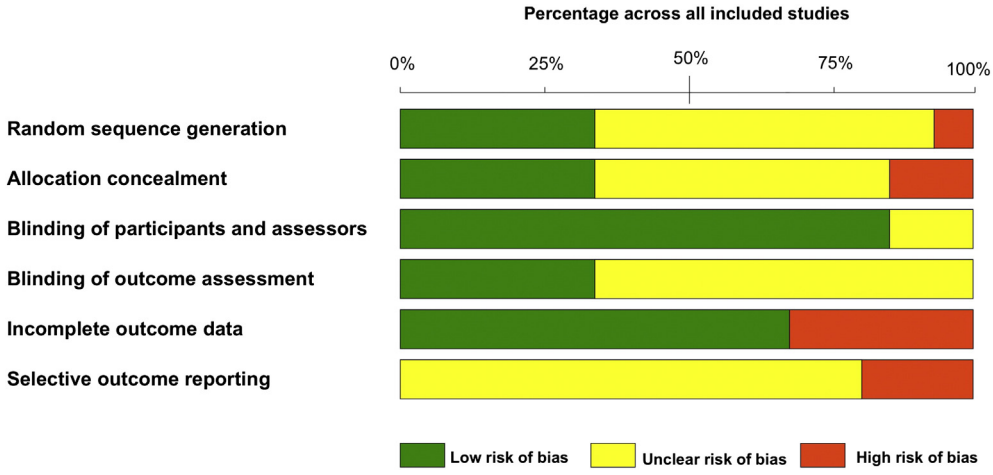

Fig. 2. Risk of bias graph (each risk of bias item plotted as percentage across all included studies).

5.3. Quantitative data synthesis

5.3.1. SAD symptoms at posttest

5.3.1.1. Overall effect. ABM had a small, but significant, effect on self-reported SAD symptoms at posttest ($g = 0.36$, 95% CI = [0.12, 0.61], $N = 11$, $z = 2.97$, $p < 0.004$). There was evidence of heterogeneity among the studies [$Q(10) = 24.14$, $p = .007$, $I^2 = 58.57$]. The study of Amir et al. (2009) was identified as an outlier as its ES was 2.5 *SD* larger than the average ES (Lipsey & Wilson, 2001). When we removed this outlier, the effect remained significant ($g = 0.27$, 95% CI = [0.12, 0.49], $N = 10$, $z = 3.66$, $p < 0.001$) and the heterogeneity reduced to non-significant values [$Q(9) = 5.43$, $p = .78$, $I^2 = 0.01$]. This outlier study was excluded from further analyses. Sensitivity analyses revealed that the results were not by driven by a specific study.

5.3.1.2. Publication bias. Duval and Tweedies's trim-and-fill procedure did not identify any studies to be trimmed. Egger's test ($b_0 = 0.94$, SE = 1.25, $t = 0.75$, one-tailed $p = .24$) corroborated the absence of publication bias. The funnel plot is shown in Fig. 3(a).

5.3.1.3. Moderators. As depicted in Table 3, none of the categorical moderators was significant. However, studies that delivered the training in the laboratory tended to yield marginally significantly larger ESs than those delivered on the Internet. As only one study from this dataset used angry faces as threatening cues during the training (Heeren et al.,

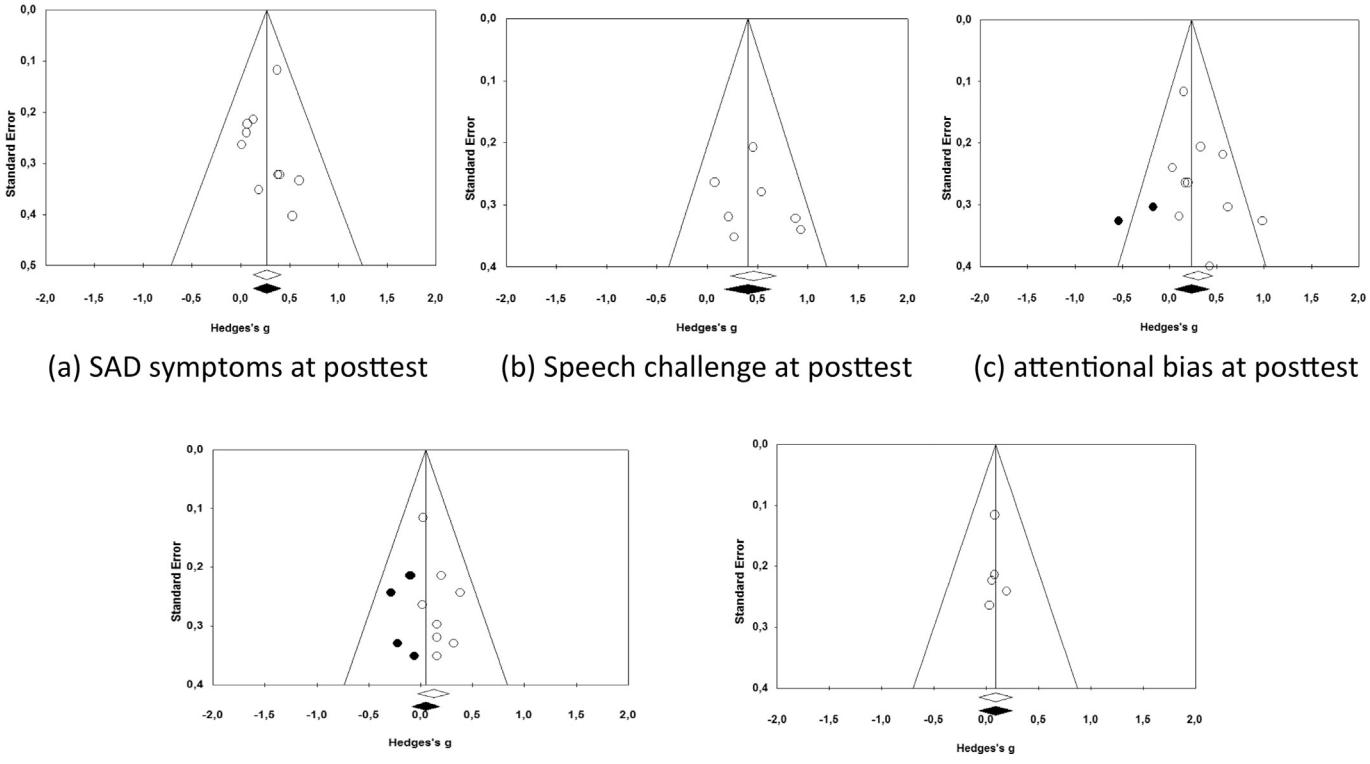

Fig. 3. Funnel plots for the five outcomes (including corrected effect sizes for publication bias). *Note.* Imputed studies were plotted when Duval and Tweedies's trim-and-fill procedure revealed that some studies would need to fall to the left of the mean effect size to make the plot symmetric.

Table 3
Moderation analysis with categorical variables for SAD symptoms at posttest.

Moderator		*N*	*g*	95% CI	Qw	*p*	Qb	*p*
Clinical status	Diagnosed	7	0.16	[−0.04–0.36]	2.89	0.82	2.47	0.12
	Subclinical	3	0.39	[0.18–0.60]	0.14	0.93		
Participants informed about the training nature	Yes	8	0.25	[−0.10–0.40]	4.90	0.67	0.51	0.48
	No	2	0.44	[−0.06–0.93]	0.09	0.77		
Nonthreatening stimuli	Neutral	6	0.26	[0.10–0.43]	4.36	0.49	0.03	0.87
	Happy	4	0.29	[0.01–0.57]	1.11	0.76		
Training orientation	Left–right	2	0.44	[−0.06–0.93]	0.09	0.77	0.51	0.48
	Top-bottom	8	0.25	[0.10–0.40]	4.90	0.67		
Training delivery	**Lab**	**5**	**0.65**	**[0.18–1.12]**	**14.36**	**0.01**	**2.84**	**0.09**
	Internet-based	**5**	**0.22**	**[0.06–0.39]**	**3.45**	**0.49**		
Presence of speech before and after	Yes	4	0.22	[−0.07–0.50]	1.01	0.80	0.14	0.70
	No	6	0.28	[0.12–0.45]	4.34	0.50		
Compensation	No	7	0.24	[0.09–0.40]	4.68	0.59	0.72	0.40
	Yes	3	0.42	[0.04–0.81]	0.10	0.95		

Note. Marginally significant moderating variable is in bold font.

2012b), we were unable to explore the potential moderating role of the threatening material used during training.

Regarding continuous moderators, the meta-regressions revealed that three variables significantly moderated the posttest ES (see Table 4). First, analyses indicated that publication year moderated ES with more recent publications producing significantly smaller ESs. Second, the number of quality criteria met per study showed a significant moderating effect such that studies satisfying a greater number of quality criteria produce smaller ESs than those satisfying fewer criteria. Hence, methodologically stronger studies produced smaller ESs than did methodologically weaker studies. Third, studies published in journals with higher impact factors had larger ESs than did those published in journals with lower impact factors. However, as only two studies from this dataset (after the removal of the outlier) assessed trait-anxiety at

Table 4
Moderation analysis with continuous variables.

Outcomes	Time of measurement	Moderators	*N*	β	*SE*	*Z*	*p*
SAD symptoms	Posttest	AB at baseline	7	−0.72	0.98	0.73	0.46
		LSAS at baseline	9	0.01	0.02	0.33	0.74
		BDI score	7	−0.002	0.03	0.05	0.96
		% Female	10	−0.01	0.01	1.22	0.23
		Age	9	−0.01	0.01	0.53	0.60
		Number of training trials per session	10	0.001	0.001	0.33	0.75
		Number of training sessions	10	0.003	0.001	1.12	0.26
		% Contingency	10	0.001	0.001	0.76	0.45
		Number of distinct stimulus-pairs	10	−0.001	0.003	0.60	0.55
		Publication's year	**10**	**−0.14**	**0.05**	**2.35**	**0.01**
		Quality criteria	**10**	**−0.13**	**0.06**	**2.24**	**0.02**
		Impact factor	**10**	**0.20**	**0.10**	**2.07**	**0.03**
Reactivity to speech	Posttest	AB at baseline	5	1.96	1.86	1.05	0.29
		LSAS at baseline	7	0.01	0.01	0.85	0.40
		BDI at baseline	6	−0.01	0.05	0.14	0.89
		STAI-Trait at baseline	**5**	**−0.07**	**0.03**	**2.00**	**0.04**
		% Female	7	−0.001	0.01	0.05	0.96
		Age	7	−0.01	0.02	0.69	0.49
		Number of training trials per session	**7**	**0.001**	**0.001**	**1.87**	**0.06**
		Number of training sessions	7	−0.01	0.05	0.26	0.79
		% Contingency	7	0.003	0.02	0.20	0.85
		Number of distinct stimulus-pairs	**7**	**0.01**	**0.01**	**2.13**	**0.03**
		Publication's year	7	−0.03	0.05	0.62	0.57
		Quality criteria	7	−0.06	0.12	0.49	0.62
		Impact factor	10	0.04	0.10	0.42	0.67
AB	Posttest	AB at baseline	10	0.01	0.04	0.31	0.75
		LSAS at baseline	9	0.001	0.007	0.03	0.97
		BDI at baseline	7	0.012	0.03	0.48	0.63
		STAI-Trait at baseline	4	−0.003	0.02	0.15	0.88
		% Female	10	0.01	0.01	0.88	0.38
		Age	8	−0.02	0.01	1.61	0.11
		Number of training trials per session	**10**	**0.001**	**0.002**	**1.98**	**0.04**
		Number of training sessions	10	−0.04	0.003	1.35	0.18
		% Contingency	10	0.02	0.01	1.45	0.15
		Number of distinct stimulus-pairs	**10**	**0.003**	**0.01**	**2.00**	**0.04**
		Publication's year	**10**	**−0.09**	**0.03**	**2.70**	**0.007**
		Quality criteria	10	−0.08	0.14	0.59	0.56
		Impact factor	**10**	**0.11**	**0.06**	**2.08**	**0.03**

Note. Significant moderating variables are in bold font.

Table 5
Categorical moderators of the ABM effect on reactivity to speech challenge at posttest.

Moderator		N	g	95% CI	Qw	p	Qb	p
Clinical status	Diagnosed	3	0.71	[0.30–1.12]	2.89	0.32	2.21	0.14
	Subclinical	4	0.34	[0.09–0.60]	1.98	0.58		
Participants informed about the training nature[a]	**Yes**	**3**	**0.17**	**[−0.18–0.51]**	**0.22**	**0.90**	**4.32**	**0.03**
	No	**3**	**0.66**	**[0.35–0.97]**	**2.08**	**0.35**		
Nonthreatening stimuli	Neutral	5	0.43	[0.19–0.67]	4.12	0.39	0.12	0.73
	Happy	2	0.56	[0.09–1.01]	2.40	0.12		
Training orientation	**Left–right**	**2**	**0.91**	**[0.09–0.57]**	**0.01**	**0.91**	**4.70**	**0.03**
	Top-bottom	**5**	**0.34**	**[0.10–0.40]**	**2.02**	**0.73**		
Presence of speech before and after	Yes	4	0.57	[0.25–0.90]	4.07	0.25	0.73	0.39
	No	3	0.37	[0.09–0.65]	1.78	0.41		

Note. Significant moderating variables are in bold font.
[a] Klumpp and Amir (2010) were not included in this analysis as they did not provide information.

baseline (Heeren et al., 2012b; Schmidt et al., 2009), we were unable to explore the potential moderating role of this variable.

5.3.2. Speech challenge at posttest

5.3.2.1. Overall effect. ABM had a small, but significant, effect on speech challenge at posttest ($g = 0.46$, 95% CI = [0.23, 0.68], $N = 7$, $z = 3.98$, $p < 0.001$).[1] There was no evidence of heterogeneity among the studies [$Q(6) = 6.73$, $p = 0.35$, $I^2 = 10.88$]. No ES exceeded ± 2.5 *SD* from the average ES. Sensitivity analyses revealed the results were not driven by a specific study.

5.3.2.2. Publication bias. The funnel plot was symmetric [see Fig. 3(b)]. Duval and Tweedies's trim-and-fill procedure did not identify any studies to be trimmed. Egger's test ($b_0 = 1.32$, SE = 2.33, $t = 0.56$, $p = .30$) corroborated the absence of publication bias.

5.3.2.3. Moderators. Two categorical moderators were significant (see Table 5). First, the information provided to participants significantly moderated posttest ES. Studies in which participants were not informed about the potential therapeutic benefits of ABM yielded significantly larger ESs than did those in which participants were told that ABM may be therapeutic. Second, the location of stimuli during training also significantly moderated posttest ES. Training with a left-right orientation yielded larger ESs than those with a top-down orientation. We were unable to explore the potential moderating role of three categorical variables: the type of threatening stimuli used during the training (as only one study used angry faces as threatening cues, i.e. Heeren et al., 2012b), compensation for participation (as only one study did not compensate participants for their participation, i.e., Bunnell, Beidel, & Mesa, 2013), and the type of training delivery as no Internet studies had speech challenge.

Meta-regressions revealed that the number of distinct pairs of training stimuli, baseline trait anxiety scores, and the number of training trials per session moderated posttest ESs, although the third variable fell short of statistical significance (see Table 4). The more training stimuli, lower baseline trait anxiety, and greater number of trials per training session were associated with larger ESs.

[1] Because only two studies included a physiological measure of reactivity to speech challenge (i.e., Heeren et al., 2012a; McNally et al., 2013), we re-ran the analyses without the ESs for physiological measures to examine the influence of these measurements on the aggregate ES. The elimination of these measures did not alter the size of the aggregate ES ($g = 0.43$, 95% CI = [0.25, 0.60]) nor its significance ($z = 3.80$, $p < 0.01$).

5.3.3. AB at posttest

5.3.3.1. Overall effect. ABM had a significant small-to-medium effect on AB at posttest ($g = 0.48$, 95% CI = [0.19, 0.76], $N = 11$, $z = 3.24$, $p < 0.001$). There was heterogeneity among the studies [$Q(10) = 37.21$, $p = .01$, $I^2 = 73.12$]. The study of Heeren et al. (2012b) was identified as an outlier with an ES 2.5 *SD* larger than the average ES. After we removed this outlier, sensitivity analyses indicated that the ES remained significant ($g = 0.30$, 95% CI = [0.14, 0.46], $N = 10$, $z = 3.68$, $p < 0.001$) and the heterogeneity was reduced to nonsignificance [$Q(9) = 10.59$, $p = .30$, $I^2 = 15.63$]. This outlier study was excluded from further analyses. Complementary sensitivity analyses revealed that the results were not driven by a specific study.

5.3.3.2. Publication bias. Duval and Tweedies's trim-and-fill procedure revealed that two studies would need to fall to the left of the mean ES to make the plot symmetric. If we assume a random effects model, then the new imputed mean ES was $g = 0.23$ (95% CI = [0.05, 0.42], $Q = 18.46$). Likewise, Egger's test ($b_0 = 1.33$, SE = 0.90, $t = 1.47$, one-tailed $p = .08$) showed a marginally significant trend toward publication bias. The adjusted funnel plot with observed and imputed studies is shown in Fig. 3(c). These analyses supported the small, but robust, effect on ABM on AB.

5.3.3.3. Moderators. Regarding the categorical variables, two moderators were significant (see Table 6). First, stimulus location during training moderated posttest ES. Horizontal display of training stimuli yielded significantly larger ESs on anxiety reactivity to the speech task at posttest than did vertical display of training stimuli. Second, participants' compensation also moderated posttest ES. Higher ESs were obtained if participants were compensated than if they were not. As only the outlier study (Heeren et al., 2012b) from this dataset used angry faces as threatening cues during the training, we were unable to explore the potential moderating role of the threatening material used during the training.

Regarding continuous moderators, the meta-regressions revealed that four variables significantly moderated posttest ESs (see Table 4). First, the more distinct pairs of training stimuli, the larger the ESs. Second, the more training trials per session, the larger the ESs. Third, the more recent the publication, the smaller the ESs. Finally, studies appearing in journals with higher impact factors yielded larger ESs than did those appearing in journals with lower impact factors.

5.3.4. Secondary symptoms at posttest

5.3.4.1. Overall effect. ABM had a nonsignificant effect on secondary symptoms at posttest ($g = 0.12$, 95% CI = [−0.03, 0.28], $N = 8$, $z = 1.59$, $p = 0.11$). The study of Boettcher et al. (2012) was identified as an outlier with an ES 2.5 *SD* larger than the average ES. Its removal did not alter the results, as sensitivity analyses indicated ($g = 0.09$, 95% CI = [−0.07, 0.26], $N = 7$, $z = 1.15$, $p = 0.25$). This outlier study was excluded from further analyses. Complementary sensitivity analyses revealed that the results were not driven by a specific study. There was no evidence of heterogeneity among the studies [$Q(8) = 2.54$, $p = 0.92$, $I^2 = 0.00$].

5.3.4.2. Publication bias. Duval and Tweedies's trim-and-fill procedure revealed that four studies would need to fall to the left of the mean ES to make the plot symmetric. If we assume a random effects model, the imputed mean ES would be $g = 0.05$ (95% CI = [−0.09, 0.19], $Q = 2.83$). Egger's test ($b_0 = 0.91$, SE = 0.46, $t = 1.97$, one-tailed $p = .04$) also showed significant a publication bias. The adjusted funnel plot with observed and imputed studies is shown in Fig. 3(d).

5.3.4.3. Moderators. As there was no significant omnibus effect, we did not compute moderator analyses.

Table 6
Categorical moderators of the ABM effects on AB at posttest.

Moderator		*N*	*g*	95% CI	Qw	*p*	Qb	*p*
Clinical status	Diagnosed	5	0.44	[0.12–0.76]	7.43	0.12	1.69	0.19
	Subclinical	5	0.20	[0.03–0.37]	0.97	0.91		
Participants informed about the training nature	Yes	7	0.23	[0.07–0.38]	5.36	0.50	1.97	0.16
	No	3	0.54	[0.13–0.94]	2.93	0.23		
Nonthreatening stimuli	Neutral	5	0.43	[0.19–0.67]	4.12	0.39	0.12	0.73
	Happy	2	0.56	[0.09–1.01]	2.40	0.12		
Training delivery	Internet-based	4	0.21	[0.02–0.41]	3.47	0.33	1.01	0.32
	Lab	7	0.36	[0.15–0.58]	6.32	0.39		
Training orientation	**Left–right**	**2**	**0.77**	**[0.23–1.32]**	**0.01**	**0.91**	**4.70**	**0.03**
	Top-bottom	**8**	**0.24**	**[0.10–0.39]**	**1.13**	**0.28**		
Presence of speech before and after	Yes	2	0.54	[−0.33–1.40]	3.76	0.06	1.35	0.24
	No	8	0.25	[0.11–0.40]	5.56	0.59		
Compensation	**No**	**4**	**0.21**	**[0.02–0.41]**	**27.92**	**0.33**	**5.79**	**0.02**
	Yes	**6**	**0.69**	**[0.20–1.18]**	**5.56**	**0.001**		

Note. Significant moderating variables are in bold font.

5.3.5. SAD symptoms at 4-month follow-up

5.3.5.1. Overall effect. ABM had a nonsignificant effect on SAD symptoms at 4-month follow-up assessment ($g = 0.13$, 95% CI = [−0.03, 0.39], $N = 6, z = 1.60, p = 0.11$). The study of Schmidt et al. (2009) was identified as an outlier with an ES 2.5 *SD* larger than the average ES. Its removal did not alter the results, as sensitivity analyses confirmed ($g = 0.09$, 95% CI = [−0.07, 0.29], $N = 5, z = 1.06, p = 0.29$). This outlier study was excluded from further analyses. Sensitivity analyses revealed that the results were not driven by a specific study. There was no evidence of heterogeneity among the studies [$Q(5) = 0.28$, $p = 0.99$, $I^2 = 0.00$].

5.3.5.2. Publication bias. The funnel plot was symmetric [see Fig. 3(e)]. Duval and Tweedies's trim-and-fill procedure did not identify any studies to be trimmed. Egger's test ($b_0 = 0.06$, SE = 0.42, $t = 0.13$, $p = 0.45$) corroborated this absence of publication bias for this outcome category.

5.3.5.3. Moderators of ABM. As there was no significant omnibus effect, we did not compute moderator analyses.

5.3.6. Complementary analyses

Given the rationale of ABM, we aimed at exploring whether posttest ESs for AB related to posttest ESs for change in emotional reactivity to the speech task. Accordingly, we computed Spearman's correlation coefficient between these variables. The correlation was significant [$r(6) = 0.90$, $p = 0.03$]. We also examined the correlations between ESs of AB and the two other significant outcomes ESs. However, although these correlations reflect a substantial effect, neither was significant the relation between SAD symptoms at posttest and AB at posttest [$r(8) = 0.41$, $p = 0.32$] nor the relation between emotional reactivity to the speech task at posttest and SAD symptoms at posttest [$r(4) = 0.50$, $p = 0.67$]. However, it is likely that the very modest statistical power of these tests explains why these correlations fell far short of significance.

Finally, we used the Benjamini–Hochberg false discovery procedure (Benjamini & Hochberg, 1995) to hold the expected proportion of falsely rejected null hypothesis at 5% for the 3 correlations we computed. The significant correlation between ESs at posttest for AB and emotional reactivity to the speech task remains significant after applying this correction.

6. Discussion

This is the first meta-analysis to assess the efficacy of ABM for SAD on symptoms, reactivity to speech challenge, AB, and secondary symptoms at posttraining as well as on SAD symptoms at 4-month follow-up. We performed a systematic review and meta-analysis on 15 studies with a combined total sample size of 1043 participants. We first examined the degree to which ABM yields therapeutic benefits for the different categories of outcomes measurement depicted above. We then explored possible moderators of these effects. Finally, we also evaluated potential publication bias as well as the potential influence of the quality of the studies by assessing the presence of risk of bias and its potential moderating role in the ES.

6.1. Main effects

Overall, the effects of the ABM for SAD as compared to the sham training were small irrespective of the outcome categories. Although the ESs related to secondary symptoms and SAD symptoms at 4-month follow-up did not reach the significance threshold, those for the other outcomes did. These results remain similar after the exclusion of outlier studies and adjustment for publication bias.

First, we found that ABM had a small ES on self-reported SAD symptoms at posttest. These findings were in line with both Hakamata et al. (2010) and Mogoaşe et al.'s (2014) meta-analyses. Second, ABM had a small effect on speech challenge at posttest, consistent with Beard et al.'s (2012a) meta-analysis. Finally, the ABM had a small ES for AB at posttraining that was much smaller than the large ESs reported by Hakamata et al. (2010) and Beard et al. (2012a). The aforementioned meta-analyses included diverse diagnostic groups and were done before null-findings began to appear in the literature.

Similar to Hakamata et al. (2010) and Mogoaşe et al. (2014), we also observed highly significant correlations between ESs for AB and those for reactivity to stressors (i.e., reactivity to speech challenge). Without confirming causality, these findings are consistent with the possibility that ABM reduces reactivity to stressors via improvement in AB. These findings are likewise consistent with cognitive models of SAD that hold that AB heightens anxiety in social situations, which, in turn, partly maintains SAD (e.g., Rapee & Heimberg, 1997). However, as both factors were measured at the same time point, we can neither confirm a causal account nor rule out third variables that may be driving this correlation (Maurage, Heeren, & Pesenti, 2013). Future studies should thus further explore this issue by assessing these outcomes at distinct time points.

6.2. Moderator effects

Several moderators were significant. First, laboratory studies yielded significantly larger ESs for SAD symptoms at posttraining than did those conducted on the Internet. However, we were unable to examine whether this effect occurs with ESs of AB and reactivity to speech challenge as most Internet-ABM studies did not include these outcomes. Consistent

with previous meta-analyses (Cristea et al., 2015; Mogoaşe et al., 2014), this observation suggest that clinical efficacy of ABM may be harder to achieve using remote training delivery compared to training in the laboratory. As suggested by several authors (e.g., Boettcher et al., 2013; MacLeod & Clarke, 2015), one cannot rule out the possibility that the home setting differs from the laboratory setting in key ways. For instance, individuals are more likely to feel at ease in their own home while still being prone to distractions that may interfere with training. Yet Internet-based cognitive-behavioral treatments have yielded significant effects in reducing symptoms of SAD (e.g., Berger, Boettcher, & Caspar, 2014), perhaps because such interventions are akin to reading a self-help manual. Reading a manual and applying its procedures in one's life does not require the standardized, highly-controlled context of ABM where precise reaction-time measures are recorded. As MacLeod and Clarke (2015) suggested, a conservative approach may advocate in favor of restricting delivery of ABM procedures to the laboratory, given that its efficacy over the Internet is minimal. Alternatively, researchers may endeavor to identify why Internet ABM is less effective than laboratory ABM, and attempt to rectify the deficiencies of the former. For example, Kuckertz et al. (2014) proposed that participants may feel more anxious in the laboratory than they do at home, implying that participants may need to activate anxiety prior to training sessions at home to achieve maximum benefits of ABM.

Second, training with a left–right orientation was more effective than a top-down orientation at odds with Hakamata et al. (2010) and Beard et al. (2012a). There are several possible interpretations for this moderation. First, this effect may partly be attributable to medium-to-large ESs arising from only three studies using the left–right orientation (e.g., Heeren et al., 2011, 2012b; Li et al., 2008). As SAD is characterized by social concerns, a second possibility is that processing two faces presented horizontally is more ecologically relevant than processing two faces presented vertically. A third possibility, not incompatible with the others, is that this effect is due to the use of faces as training material in most of the studies. Indeed, the previous meta-analyses reporting the top-down orientation advantage mostly included studies using words during ABM and did not examine whether the type of training stimuli during ABM moderated this top-down orientation advantage. Because only one study included words during ABM in the present meta-analysis, one cannot exclude the possibility that the top-bottom advantage in previous meta-analyses arises from pooling training with words with that of faces. Future studies must clarify the robustness of this influence.

Third, the number of training trials per session significantly moderates both AB and speech challenge outcomes, with higher number of trials yielding larger ESs. These results are in line with previous meta-analyses (e.g., Beard et al., 2012a; Hakamata et al., 2010) suggesting a dose–response relationship. However, previous meta-analyses reported this effect for the number of training sessions instead of the numbers of trials per session. Because most ABM studies assessing AB and reactivity to speech challenge had only a single session, the number of total trials likely explains the dose–response relationship.

Fourth, we are the first to examine the influence of the number of distinct stimulus-pairs on ABM efficacy. For both AB and speech challenge outcomes, the more distinct stimulus-pairs during training, the more ABM outperformed the control condition. There are several plausible interpretations for this moderation. First, this effect may be partly due to the fact that processing more distinct stimulus-pairs may facilitate stimulus generalization and reduce habituation to training material. A second and – in our opinion – more convincing explanation is that performing a task with many stimulus-pairs diminishes boredom with the task. Indeed, many SAD participants report ABM to be dull and repetitive (Beard, Weisberg, & Primack, 2012b).

Fifth, higher ESs were obtained for AB if participants received compensation than if not, replicating the findings of Cristea et al. (2015) for mood and anxiety disorders in general. Perhaps compensation increases participants' engagement with ABM, bolstering its effects (Beard et al., 2012b). Although Cristea et al. (2015) suggested that this influence may reflect demand effects, it occurred for reductions in AB, an effect seemingly impervious to demand. Moreover, the demand explanation should apply to participants in the control group, yet ABM outperforms sham training when subjects in both groups receive compensation. Because the presence of compensation is likely to mirror the access to research funding and that there is an increasing pressure to be funded throughout science (e.g., Fanelli, 2010; Fang, Steen, & Casadevall, 2012), an alternative explanation might be that the assessors might have been implicitly influenced by this pressure (e.g., leading to more emphasis when delivering instructions to the participants, changes in assessor's beliefs and expectancies). However, several studies that declared funding did not compensate their participants and, conversely, those that compensated did not always report funding. Moreover, the presence of a double-blind procedure for all the studies included in the present meta-analysis tends to run counter to this interpretation since it implies that it ought to occur for control conditions too. Nevertheless, as shown in the quality assessment section, only five studies had assessors adequately blind to allocation to condition. As pointed by Cuijpers, van Straten, Bolhmeijer, Hollon, and Andersson (2010), features like adequate blinding of assessors that have long been required in the testing of medications are only now starting to find their way into the clinical psychological sciences. In part, that reflects the regulatory environment in which pharmacological research is conducted, with studies often larger and better funded than in ABM research, but also are subjected to stricter third-party scrutiny (e.g., governmental regulatory agencies). In contrast, ABM research has often been conducted by academic investigators with absence of third-party external scrutiny.

Sixth, we found that trait anxiety moderated the efficacy of ABM on reactivity to speech challenge, with lower STAI-trait scores yielding larger ESs. The difficulty that people with high trait anxiety have with top-down executive control over their attention (e.g., Bishop, 2009; Pacheco-Unguetti, Acosta, Callejas, & Lupiáñez, 2010) may explain the effect. Indeed, neuroimaging studies have demonstrated that high-trait anxious individuals exhibit a reduced activation of the left dorsolateral part of the prefrontal cortex during tasks assessing such a top-down attention control (e.g., Bishop, 2009; Browning, Holmes, Murphy, Goodwin, & Harmer, 2010). Moreover, recent translational studies show that increasing the activity of this brain region by using neuromodulation may facilitate ABM efficiency (Clarke et al., 2014a; Heeren, Baeken, Vanderhasselt, Philippot, & De Raedt, 2015). Accordingly, if elevated trait anxiety reflects difficulty with top-down attention control, this may explain why SAD participants with lower levels of trait anxiety benefit more from ABM. Moreover, people with higher levels of executive control exhibit less anxiety and better performance during impromptu speeches (Jones, Fazio, & Vasey, 2012), perhaps explaining why this influence only occurs for reactivity to speech challenge. However, this interpretation should be considered cautiously as only five studies included both a speech challenge and the assessment of STAI-trait at baseline. Consequently, researchers need to determine why some people with SAD have relatively low STAI-trait scores. Although one might assume that they also have less severe SAD symptoms, the absence of correlations between LSAS and STAI-trait scores at baseline among the eight studies that provided these data runs counter to this interpretation [$r(8) = 0.13$, $p = 0.76$]. Nevertheless, having a treatment that works significantly less well for individuals who exhibit more severe symptoms poses a serious challenge for the clinical utility of ABM for SAD. Consequently, further explorations of this effect constitute a critical next step for longer tenable clinical directions.

Finally, studies in which participants were not informed about the potential therapeutic nature of the training procedures yielded larger ESs for reactivity to speech challenge than those in which participants were told that the procedures may be therapeutic. Arguably, because one can expect that fostering positive expectancies about possible therapeutic benefits may increase the clinical efficacy, this finding comes as a surprise. However, because a vast majority of the SAD patients from

previous studies who received the ABM condition reported they believed to have been assigned to the sham condition (e.g., Amir et al., 2009; Beard et al., 2012b), one cannot exclude that participants who were informed about the potential therapeutic nature would exhibit lower engagement to the training task as they would believe that they have been assigned to the sham condition. This explanation makes sense in the context of the repetitive and boring nature of the ABM procedure (Beard et al., 2012b). As a consequence, the present findings give rise to a key dilemma. On the one hand, because of the repetitive and boring nature of the ABM procedures, they suggest that active steps should be taken to improve the patients' acceptability of the ABM procedure by communicating the treatment rationale. On the other hand, the results suggest that the benefits of ABM on speech reactivity among SAD individuals may be optimized if active steps were taken to reduce patient insight of the training nature. One may argue that an alternative solution would be to give patients explicit information about the training contingency to increase the treatment acceptability. However, Grafton, Mackintosh, Vujic, and MacLeod (2014) recently reported that when participants are informed about the rationale of ABM (i.e., the contingency), then ABM manipulation continues to exert an impact on AB, but this modification no longer affects anxiety reactivity to subsequent stressor. Grafton et al. (2014) suggest that awareness of the targeted attentional selectivity during ABM may have produced an attentional change that did not generalize and endure beyond ABM training, as would be mandatory to influence anxiety response to subsequent situations. Consequently, as advised by Grafton et al. (2014), an alternative option would be to not give patients explicit information about the training contingency and to instead try to increase the acceptability in alternative ways (e.g., making ABM tasks sufficiently captivating so that motivation to complete them does not essentially depend on the patients' understanding their putative mechanisms).

6.3. Risk of bias

Following Cristea et al. (2015), we examined studies' quality and publication bias. Although the bias favoring positive results was small and mostly nonexistent due to the inclusion of null-findings in the meta-analysis, we found a bias favoring early publications relative to recent ones as did Cristea et al. (2015). Studies published earlier had larger ESs than those published later, a pattern common in science (Ioannidis, 1998, 2008). As typical throughout science (e.g., Bagshaw, McAlister, Manns, & Ghali, 2006; Jennions & Moeller, 2002), studies with larger ESs appeared in journals with higher impact factors relative to studies with smaller ESs. Moreover, as pointed out by Ioannidis (2008), if replication studies were only powered to detect large ESs, as based on early studies, these subsequent replication attempts will be underpowered to detect modest ESs, thereby giving the impression that the original effects were spurious. Therefore, the best way to combat this problem is to improve the methodological quality if replication attempts (Yordanov et al., 2015).

Study quality moderated ES. The higher the quality, the smaller was the effect. Unfortunately, it was difficult to assess quality for many studies as authors often failed to report necessary information, especially how they handled missing data. Very few investigators registered their trials, and selective reporting of data was common, further increasing bias.

6.4. Theoretical and clinical implications

From *a theoretical point of view*, because modifying AB had only a very small effect on reducing stressor reactivity and SAD symptoms, our results seemingly challenge the claim that AB figures prominently in the maintenance, and perhaps the etiology, of this disorder (e.g., Clark & Wells, 1995; Heeren et al., 2012a; Hirsch & Clark, 2004; Rapee & Heimberg, 1997). Indeed, AB is likely one of multiple causal factors in the emergence of SAD. Likewise, the absence of AB at baseline assessment in several recent ABM studies also raises questions about how common AB is among SAD individuals (e.g., Boettcher et al., 2013; Julian et al., 2012; McNally et al., 2013). It is possible that people develop SAD via pathways other than through AB (e.g., Brühl, Delsignore, Komossa, & Weidt, 2014; Gilboa-Schechtman & Shachar-Lavie, 2013; Haller, Cohen Kadosh, & Lau, 2014; Krimbel, 2008). This way, although early ABM publications in top-tier journals have contributed to reinforce the notion that AB does play an acute causal role in the maintenance of SAD, the present findings call for a critical reconsideration of the importance of AB in the conceptualization of this disorder. On the other hand, extant procedures for assessing AB by using reaction-time indices are insufficiently reliable to detect changes in AB (e.g., Dear, Sharpe, Nicholas, & Refshauge, 2011; Schmukle, 2005; Staugaard, 2009; Waechter, Nelson, Wright, Hyatt, & Oakman, 2014). New reaction-time methods for quantifying AB have emerged (Price et al., 2015; Zvielli, Bernstein, & Koster, in press) that conceptualize AB as a dynamic process rather than a static trait. Preliminary research involving these new methods of measuring AB yield modest, but larger, internal reliability estimates vis-à-vis traditional indices of AB that yield near-zero estimates (Zvielli et al., in press). Moreover, highlighting the predictive validity of such a dynamic conceptualization of AB, one study reported that the lack of stability of AB – often assumed to reflect a stable trait-like process – does predict AB plasticity through ABM (Heeren, Philippot, & Koster, in press).

From *a clinical point of view*, the ESs of the present meta-analysis are small relative to those of usual CBT ($g = 0.77$; for a meta-analysis, see Acarturk et al., 2009) or pharmacological treatments for SAD ($g = 0.65–1.02$; for a meta-analysis, see Blanco et al., 2003). This suggests that ABM is not yet ready for dissemination as a stand-alone treatment or as an adjunct to traditional CBT for SAD. Reasonable steps would be to apply new methods for computing reliable RT indices of AB that reflect its dynamic nature (e.g., Price et al., 2015; Zvielli et al., in press), and to use gamified ABM paradigms that are motivationally engaging (e.g., Dennis & O'Toole, 2014; Enock, 2015; Notebaert, Clarke, Grafton, & MacLeod, 2015) and that provide feedback to participants during training (e.g., Bernstein & Zvielli, 2014; Enock, 2015).

6.5. Limitations

Our meta-analysis has several limitations. First, although we had sufficient statistical power to detect small overall ESs, some subgroup analyses involved only a few studies, rendering impossible to compute categorical moderation analyses (e.g., type of threatening material such as disgust versus angry faces, or threatening words versus faces). Second, although exclusion of outliers reduced heterogeneity among studies, I^2 remained large for some outcomes, implicating persistent heterogeneity. Finally, we confined our search to studies published or accepted for publication in English peer-reviewed journals. Although we may have missed some studies, this seems unlikely as the field does not resist publishing failures to replicate.

Role of funding sources

This research was supported by a Grant (Grant # FC 78142) from the Belgian National Science Foundation "F.R.S.—FNRS." (awarded to Alexandre Heeren) and by the Sectorial Operational Program for Human Resources Development 2007–2013 ("Young Successful Researchers"; Grant # POSDRU/159/1.5/S/132400) from the European Social Fund (awarded to Cristina Mogoaşe). This research also received the support from the Belgian Foundation for Vocation ("Vocatio") and the Belgian French Community Grant for Scientific Excellence, both awarded to Alexandre Heeren. These foundations have no role in the study design, collection, analysis or interpretation of the data, writing the manuscript, or the decision to submit the paper for publication.

Contributors

Alexandre Heeren and Richard J. McNally designed the study and wrote the protocol. Alexandre Heeren and Cristina Mogoaşe conducted literature searches, provided summaries of previous research studies, and conducted the statistical analyses. Alexandre Heeren and Richard J. McNally wrote the first draft of the manuscript. All authors revised the manuscript critically and contributed to and have approved the final manuscript.

Conflict of interest

All authors declare that they have no conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.cpr.2015.06.001.

References[2]

* Neubauer, K., von Auer, M., Muray, E., Petermann, F., Helbig-Lang, S., & Gerlach, A.L. (2013). *Behaviour Research and Therapy*, *51*(2), 87–97.

Acarturk, C., Cuijpers, P., van Straten, A., & de Graaf, R. (2009). Psychological treatment of social anxiety disorder: A meta-analysis. *Psychological Medicine*, *39*, 241–254.

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

* 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*, 961–973.

Amir, N., Taylor, C.T., & Donohue, M.C. (2011). Predictors of response to an attention modification program in generalized social phobia. *Journal of Consulting and Clinical Psychology*, *79*, 533–541.

* Amir, N., Weber, G., Beard, C., Bomyea, J., & Taylor, C.T. (2008). The effects of a single-session attention modification program on response to a public-speaking challenge in socially anxious individuals. *Journal of Abnormal Psychology*, *117*, 860–868.

Bagshaw, S.M., McAlister, F.A., Manns, B.J., & Ghali, W.A. (2006). Acetylcysteine in the prevention of contrast-induced nephropathy: A case study of the pitfalls in the evolution of evidence. *Archives of Internal Medicine*, *166*, 161–166.

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, L., & van IJzendoorn, M.H. (2007). Threat-related attentional bias in anxious and non-anxious individuals: A meta-analytic study. *Psychological Bulletin*, *133*, 1–24.

Beard, C., Sawyer, A.T., & Hofmann, S.G. (2012a). Efficacy of attention bias modification using threat and appetitive stimuli: A meta-analytic review. *Behavior Therapy*, *43*, 724–740.

Beard, C., Weisberg, R.B., & Primack, J. (2012b). Socially anxious primary care patients' attitudes toward cognitive bias modification (CBM): A qualitative study. *Behavioural and Cognitive Psychotherapy*, *40*, 618–633.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*, 289–300.

Berger, T., Boettcher, J., & Caspar, F. (2014). Internet-based guided self-help for several anxiety disorders: A randomized controlled trial comparing a tailored with a standardized disorder-specific approach. *Psychotherapy*, *51*, 207–219.

Bernstein, A., & Zvielli, A. (2014). Attention feedback awareness and control training (A-Fact): Experimental test of a novel intervention paradigm targeting attentional bias. *Behaviour Research and Therapy*, *55*, 18–26.

Bishop, S.J. (2009). Trait anxiety and impoverished prefrontal control of attention. *Nature Neuroscience*, *12*, 92–98.

Blanco, C., Schneier, F.R., Schmidt, A., Blanco-Jerez, C. -R., Marshall, R.D., Sanchez-Lacay, A., & Liebowitz, M.R. (2003). Pharmacological treatment of social anxiety disorder: A meta-analysis. *Depression & Anxiety*, *18*, 29–40.

* Boettcher, J., Berger, T., & Renneberg, B. (2012). Internet-based attention training for social anxiety: A randomized controlled trial. *Cognitive Therapy and Research*, *36*, 1–15.

* Boettcher, J., Leek, L., Matson, L., Holmes, E.A., Browning, M., MacLeod, C., Andersson, G., & Carlbring, P. (2013). Internet-based attention bias modification for social anxiety: A randomized controlled comparison of towards negative and training towards positive cues. *PLoS ONE*, *8*, e71760. http://dx.doi.org/10.1371/journal.pone.0071760.

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

Browning, M., Holmes, E.A., Murphy, S.E., Goodwin, G.M., & Harmer, C.J. (2010). Lateral prefrontal cortex mediates the cognitive modification of attentional bias. *Biological Psychiatry*, *67*, 919–925.

Brühl, A.B., Delsignore, A., Komossa, K., & Weidt, S. (2014). Neuroimaging in social anxiety disorder: A meta-analytic review resulting in a new neurofunctional model. *Neuroscience & Biobehavioral Reviews*, *47*, 260–280.

* Bunnell, B.E., Beidel, D.C., & Mesa, F. (2013). A randomized trial of attention training for Generalized Social Phobia: Does attention training change social behavior? *Behavior Therapy*, *44*, 662–673.

* 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*, 66. http://dx.doi.org/10.1186/1471-244X-12-66.

Clarke, P.J.F., Browning, M., Hammond, G., Notebaert, L., & MacLeod, C. (2014a). The causal role of the dorsolateral prefrontal cortex in the modification of attentional bias: Evidence from transcranial direct current stimulation. *Biological Psychiatry*, *76*, 946–952.

Clarke, P.J.F., Notebaert, L., & MacLeod, C. (2014b). Absence of evidence or evidence of absence: Reflecting on therapeutic implementations of attentional bias modification. *BMC Psychiatry*, *14*, 8. http://dx.doi.org/10.1186/1471-244X-14-8.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cooper, H., Hedges, L.V., & Valentine, J.C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY: Russell Sage Foundation.

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

Cuijpers, P., van Straten, A., Bolhmeijer, E., Hollon, S.D., & Andersson, G. (2010). The effect of psychotherapy for adult depression are overestimated: A meta-analysis of study quality and effect size. *Psychological Medicine*, *40*, 211–223.

De Voogt, E.L., Wiers, R.W., Prins, P.J.M., & Salemink, E. (2014). Visual search attentional bias modification reduces social phobia in adolescents. *Journal of Behavior Therapy and Experimental Psychiatry*, *45*, 252–259.

Dear, B.F., Sharpe, L., Nicholas, M.K., & Refshauge, K. (2011). The psychometric properties of the dot-probe paradigm when used in pain-related attentional bias research. *Journal of Pain*, *12*, 1247–1254.

Dennis, T.A., & O'Toole, L.J. (2014). Mental health on the go: Effects of a gamified attention-bias modification mobile application in trait-anxious. *Clinical Psychological Science*, *25*, 576–590.

Duval, S., & Tweedies, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629–634.

Emmelkamp, P.M.G. (2012). Attention bias modification: The Emperor's new suit? *BMC Medicine*, *10*, 63.

Enock, P. M. (2015). *Making an IMPACT: Designing and testing a novel attentional training game to reduce social anxiety*. Unpublished doctoral dissertation, Department of Psychology, Harvard University, Cambridge, MA.

* Enock, P.M., Hofmann, S.G., & McNally, R.J. (2014). Attention bias modification treatment via smartphone to reduce social anxiety: A randomized controlled trial. *Cognitive Therapy and Research*, *38*, 200–216.

Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US States Data. *PLoS ONE*, *5*(4), e10271. http://dx.doi.org/10.1371/journal.pone.0010271.

Fang, F.C., Steen, R.G., & Casadevall, A.C. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 17028–17033.

Ferguson, D., Aaron, S.D., Guyatt, G., & Herbert, P. (2002). Post-randomisation exclusions: The intention to treat principle and excluding patients from analysis. *British Medical Journal*, *325*, 652–654.

Gilboa-Schechtman, E., & Shachar-Lavie, I. (2013). More than a face: A unified theoretical perspective on nonverbal social cue processing in social anxiety. *Frontiers in Human Neuroscience*, *7*, 904. http://dx.doi.org/10.3389/fnhum.2013.00904.

Grafton, B., Mackintosh, B., Vujic, T., & MacLeod, C. (2014). When ignorance is bliss: Explicit instructions and the efficacy of CBM-A for anxiety. *Cognitive Therapy and Research*, *38*, 172–188.

Gunter, R.W., & Whittal, M.L. (2010). Disseminations of cognitive-behavioral treatments for anxiety disorders: Overcoming barriers and improving patients access. *Clinical Psychology Review*, *30*, 194–202.

Hakamata, Y., Lissek, S., Bar-Haim, Y., Britton, J.C., Fox, N.A., Leibenluft, E., Ernst, M., & Pine, D.S. (2010). Attention bias modification treatment: A meta-analysis toward the establishment of novel treatment for anxiety. *Biological Psychiatry*, *68*, 982–990.

Haller, S.P.W., Cohen, Kadosh K., & Lau, J.Y.F. (2014). A developmental angle to understanding the mechanisms of biased cognitions in social anxiety. *Frontiers in Human Neuroscience*, *7*, 846. http://dx.doi.org/10.3389/fnhum.2013.00846.

Hallion, L.S., & Ruscio, A.M. (2011). A meta-analysis of the effect of cognitive bias modification on anxiety and depression. *Psychological Bulletin*, *137*, 940–958.

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

Heeren, A., Baeken, C., Vanderhasselt, M. -A., Philippot, P., & De Raedt, R. (2015a). Impact of anodal and cathodal transcranial direct current stimulation over the left dorsolateral prefrontal cortex during attention bias modification: An eye-tracking study. *PLoS ONE*, *10*(4), e0124182. http://dx.doi.org/10.1371/journal.pone.0124182.

* Heeren, A., Lievens, L., & Philippot, P. (2011). How does attention training work in social phobia: Disengagement from threat or reengagement to nonthreat? *Journal of Anxiety Disorders*, *25*, 1108–1115.

Heeren, A., Maurage, P., & Philippot, P. (2013). A Bayesian case–controls exploration of the malleability of attentional bias for threat in social phobia. *International Journal of Cognitive Therapy*, *6*, 24–39.

[2] References marked with an asterisk indicate studies included in the meta-analysis. The in-text citations to studies selected for meta-analysis are not preceded by asterisks.

Heeren, A., Peschard, V., & Philippot, P. (2012a). The causal role of attentional bias to threat cues in social anxiety: A test on a cyber-ostracism task. *Cognitive Therapy and Research*, *36*, 512–521.

Heeren, A., Philippot, P., & Koster, E.H.W. (2015b). Impact of the temporal stability of pre-existent attentional bias for threat on its alteration through attention bias modification. *Journal of Behavior Therapy and Experimental Psychiatry*. http://dx.doi.org/10.1016/j.jbtep.2014.10.012 (in press).

* Heeren, A., Reese, H.E., McNally, R.J., & Philippot, P. (2012b). Attention training toward and away from treat in social phobia: Effects on behavioural, subjective, and physiological measures of anxiety. *Behaviour Research and Therapy*, *50*, 30–39.

Clark, D. M., & Wells, A. (1995). A cognitive model of social phobia. In R. G. Heimberg, M. R. Liebowitz, D. A. Hope, & F. R. Schneier (Eds.), *Social Phobia: Diagnosis, assessment, and treatment* (pp. 69–93). New York: Guilford Press.

Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. Higgins, J.P., & Green, S. (Eds.). (2011). *The Cochrane Collaboration* (Available from www.cochrane-handbook.org).

Higgins, J.P., Thompson, S.G., Deeks, J.J., & Altman, D.G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*, 557–560.

Hirsch, C.R., & Clark, D.M. (2004). Information-processing bias in social phobia. *Clinical Psychology Review*, *24*, 799–825.

Ioannidis, J.P.A. (1998). Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *The Journal of the American Medical Association*, *279*, 281–286.

Ioannidis, J.P.A. (2008). Why most discovered true associations are inflated. *Epidemiology*, *19*, 640–648.

Jennions, M.D., & Moeller, A.P. (2002). Relationships fade with time: A meta-analysis of temporam trends in publication in ecology and evolution. *Proceedings of the Royal Society B: Biological Sciences*, *269*, 43–48.

Jones, C.R., Fazio, R.H., & Vasey, M.W. (2012). Attentional control buffers the effect of public-speaking anxiety on performance. *Social Psychological and Personality Science*, *3*, 556–561.

* 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 & Therapy*, *50*, 350–358.

Kjaergard, L.L., Villumsen, J., & Gluud, C. (2001). Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Annals of Internal Medicine*, *136*, 982–989.

*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*, 263–271.

Krimbel, A. (2008). A model of the development and maintenance of generalized social phobia. *Clinical Psychological Review*, *28*, 592–612.

Kuckertz, J.M., Gildebrant, E., Liliequist, B., Karström, P., Väppling, C., Bodlund, O., Stenlund, T., Hoffman, S.G., Andersson, G., Amir, N., & Carlbring, P. (2014). Moderation and mediation of the effect of attention training in social anxiety disorder. *Behaviour Research & Therapy*, *53*, 30–40.

* Li, S.W., Tan, J.Q., Qian, M.Y., & Liu, X.H. (2008). Continual training of attentional bias in social anxiety. *Behaviour Research & Therapy*, *46*, 905–912.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gotzche, P.C., Ioannidis, J.P.A., Clarke, M., Devereaux, P.J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, *6*, e1000100. http://dx.doi.org/10.1371/journal.pmed.1000100.

Liebowitz, M.R. (1987). Social phobia. *Modern Problems of Pharmacopsychiatry*, *22*, 141–173.

Light, R.J., & Pillemer, D.B. (1984). *Summing up: The science of reviewing research.* Cambridge, MA: Harvard University Press.

Lipsey, M.W., & Wilson, D.B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Lovell, K., & Richards, D. (2000). Multiple access points and levels of entry (MAPLE): Ensuring choice, accessibility and equity of CBT services. *Behavioural and Cognitive Psychotherapy*, *28*, 379–391.

MacLeod, C., & Clarke, P.J.F. (2015). The attentional bias modification approach to anxiety intervention. *Clinical Psychological Science*, *3*, 58–78.

MacLeod, C., & Mathews, A. (2012). Cognitive bias modification approaches to anxiety. *Annual Review of Clinical Psychology*, *8*, 189–217.

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology*, *95*, 15–20.

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*, 107–123.

Maurage, P., Heeren, A., & Pesenti, M. (2013). Does chocolate consumption really boost Nobel award chances? The peril of over-interpreting correlations in nutrition and health studies. *The Journal of Nutrition*, *143*, 931–933.

* McNally, R.J., Enock, P.M., Tsai, C., & Tousian, M. (2013). Attention bias modification for reducing speech anxiety. *Behaviour Research and Therapy*, *51*, 882–888.

Mogoaşe, C., David, D., & Koster, E.H.W. (2014). Clinical efficacy of attentional bias procedures: An updated meta-analysis. *Journal of Clinical Psychology*, *70*, 1133–1157.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., & the PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*, e1000097. http://dx.doi.org/10.1371/journal.pmed.1000097.

Moher, D., Pham, B., Jones, A., Cook, D.J., Jadad, A.R., Moher, M., Tugwell, P., & Klassen, T.P. (1998). Does quality of reports of randomized trials affect estimated of intervention efficacy reported in meta-analyses. *The Lancet*, *352*, 609–613.

Morrison, A.S., & Heimberg, R.G. (2013). Social anxiety and social anxiety disorder. *Annual Review of Clinical Psychology*, *9*, 249–274.

Notebaert, L., Clarke, P.J.F., Grafton, B., & MacLeod, C. (2015). Validation of a novel attention bias modification task: The future may be in the cards. *Behaviour Research and Therapy*, *65*, 93–100.

Olfson, M., Guardino, M., Struening, E., Schneier, F.R., Hellman, F., & Klein, D.F. (2000). Barriers to the treatment of social anxiety. *American Journal of Psychiatry*, *157*, 521–527.

Pacheco-Unguetti, A.P., Acosta, A., Callejas, A., & Lupiáñez, J. (2010). Attention and anxiety: Different attentional functioning under state and trait anxiety. *Psychological Science*, *21*, 298–304.

Price, R.B., Kuckertz, J.M., Siegle, G., Ladouceur, C., Silk, J.S., Ryan, N.D., Dahl, R.E., & Amir, N. (2015). Empirical recommendations for improving the stability of the dot-probe task in clinical research. *Psychological Assessment*, *27*, 365–376.

Rapee, R.M., & Heimberg, R.G. (1997). A cognitive-behavioral model of anxiety in social phobia. *Behaviour Research and Therapy*, *35*, 741–756.

Rapee, R.M., MacLeod, C., Carpenter, L., Gaston, J.E., Frei, J., Peters, L., & Baillie, A.J. (2013). Integrating cognitive bias modification into a standard cognitive behavioural treatment package for social phobia: A randomized controlled trial. *Behaviour Research and Therapy*, *51*, 207–215.

Riley, R.D., Higgins, J.P.T., & Deeks, J.J. (2011). Interpretation of random effects meta-analyses. *British Medical Journal*, *342*, d549.

* 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*, 5–14.

Schmukle, S.C. (2005). Unreliability of the dot-probe task. *European Journal of Personality*, *19*, 595–605.

Staugaard, S.R. (2009). Reliability of two versions of the dot-probe task using photographic faces. *Psychology Science*, *51*, 339–350.

Stein, M.B., & Stein, D.J. (2008). Social anxiety disorder. *The Lancet*, *371*, 1115–1125.

Sterne, J.A.C., Becker, B.J., & Egger, M. (2005). The funnel plot. In H.R. Rothstein, A.J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 75–98). Chichester: John Willey & Sons, Ltd.

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*, 682–721.

Waechter, S., Nelson, A.L., Wright, C., Hyatt, A., & Oakman, J. (2014). Measuring attentional bias to threat: Reliability of dot probe and eye movement indices. *Cognitive Therapy and Research*, *38*, 313–333.

Weisberg, R.B., Dyck, I., Culpepper, L., & Keller, M.B. (2007). Psychiatric treatment in primary care patients with anxiety disorders: A comparison of care received from primary care providers and psychiatrists. *American Journal of Psychiatry*, *164*, 276–282.

Wood, L., Egger, M., Gluud, L.L., Schulz, K.F., Jüni, P., Altman, D.G., Gluud, C., Martin, R.M., Wood, A.J., & Sterne, J.A. (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: Meta-epidemiological study. *British Medical Journal*, *336*, 601–605.

Yordanov, Y., Dechartres, A., Porcher, R., Boutron, I., Altman, D.G., & Ravaud, P. (2015). Avoidable waste of research related to inadequate methods in clinical trials. *BMJ*, *350*, h809. http://dx.doi.org/10.1136/bmj.h809.

Zvielli, A., Bernstein, A., & Koster, E.H.W. (2015). Temporal dynamics of attentional bias. *Clinical Psychological Science* (in press).