Literature Review: Effects of Cognitive Bias Modification - Interpretation (CBM-I) Interventions on Depressive Symptoms: A Systematic Review

Major Research Project: Effects on Depressive Symptoms of a Web-Based Cognitive Bias Modification-Interpretation (CBM-I) Program for Emotion Recognition: A Randomised Controlled Trial

Submitted by Victoria Clare Stephens, to the University of Exeter as a thesis for the degree of Doctor of Clinical Psychology, September 2014.

Signature: ……..

Contents

List of Tables

Literature Review **Page**

Major Research Project

Appendices **Page**

List of Figures

Literature Review **Page**

Major Research Project

LITERATURE REVIEW COVER SHEET

Effects of Cognitive Bias Modification- Interpretation (CBM-I) Interventions on Depressive Symptoms: A Systematic Review

TRAINEE NAME: Victoria Stephens

SUPERVISORS

PRIMARY: Professors Marcus Munafò and Ian Penton-Voak, School of Psychology, University of Bristol.

SECONDARY: Dr Anna Adlam, College of Life and Environmental Science, University of Exeter.

TARGET JOURNAL: British Journal of Psychiatry

WORD COUNT: 3824 + corrections (excluding abstract, tables, titles, references and appendices)

DECLARATION:

“I certify that all material in this literature review which is not my own work has been identified and properly attributed. I have conducted the work in line with the BPS DCP Professional Practice Guidelines.”

SUBMITTED IN PARTIAL FULFILMENT OF REQUIREMENTS FOR THE DOCTORATE DEGREE IN CLINICAL PSYCHOLOGY

UNIVERSITY OF EXETER

Effects of Cognitive Bias Modification- Interpretation (CBM-I) Interventions on Depressive Symptoms: A Systematic Review

Abstract

Depression is a major source of disability and many people do not access treatment. More accessible, cost-effective treatments are required. Cognitive bias modification interventions that target interpretation biases, hypothesised to contribute to the development and maintenance of depression, have been identified as a promising treatment. The current systematic review aimed to examine effects on symptoms of depression, in clinical and analogue samples, of cognitive bias modification interventions that have targeted interpretation biases. Relevant studies were sourced using a systematic search of databases and by contacting researchers involved in cognitive bias modification research. Ten relevant studies were identified, nine of which were classified as randomised controlled trials. Data relating to participants, interventions, comparator groups, outcome measures, study design, study quality, and findings relating to the interventions' effects on symptoms of depression and cognitive bias were extracted and synthesised. The overall picture of the effect of the interventions on symptoms of depression and interpretation bias was mixed, suggesting cognitive bias modification interventions that target interpretation biases are not currently at a stage to be recommended as an evidence-based treatment for depression. Limitations of the studies and review were acknowledged. Identification of the most meaningful outcome measures and effective CBM-I paradigm, perhaps a non-verbal design to maximise accessibility, paired with further rigorous studies are recommendations for further research in this field.

Introduction

Depression

Depression is a major cause of disability worldwide, accounting for 4.3% of the global burden of disease (World Health Organization, 2013). These burdens include increased risk of mortality, especially by suicide, and disruptions to family stability, leading to family breakdown (Lépine & Riley, 2011). Depression contributes to workplace absenteeism, an economic burden for society (Karampampa, Borgström, & Jönsson, 2011). Depression is characterised by persistent low mood and loss of interest in activities, accompanied by changes in psychomotor activity, sleep, and appetite, lack of energy, feelings of worthlessness, concentration difficulties, and suicidal ideation (American Psychiatric Association, 2013). Many cognitive theories of the aetiology, maintenance and risk of relapse of depression emphasise the role of negative cognitive biases, whereby people with depression tend to be drawn to negative stimuli and interpretations.

Negative cognitive bias is central to this review, investigating the effect of cognitive bias modification (CBM) interventions on symptoms of depression. Negative biases have been found to occur at different levels of information-processing including attention, interpretation, and memory (Mathews & MacLeod, 2005). CBM paradigms focusing on these cognitive processes are known as CBM-A, CBM-I, and CBM-M, respectively. Compared to people without depression, people with depression more selectively attend to and recall negative stimuli, and interpret ambiguous situations more negatively. Beck and colleagues (1979) developed a model of depression, suggesting people with depression view the self, world and future more negatively than others; this

model underpins traditional cognitive therapy, which aims to identify and reduce these negative biases.

Evidence-based treatments for depression include cognitive behavioural therapy (CBT) and antidepressants; both treatments are linked to changes in information-processing biases hypothesised to mediate changes in mood. CBT modifies interpretation biases, through thought-challenging and considering alternative interpretations of situations. A recent neurocognitive model of the mechanism of antidepressants (Harmer, Goodwin, & Cohen, 2009) suggests medication leads to neurochemical changes causing early modification of information-processing biases, detectable shortly after administration, and this mediates later improvements in mood. Antidepressants are less acceptable than psychological therapies to many people (e.g., Chabrol, Teissedre, Armitage, Danel, & Walburg, 2004) and psychological therapies are not accessible to everyone. Barriers include lack of skilled therapists, limited clinic opening times, and treatment costs (Titov et al., 2010). It is therefore desirable to develop accessible, acceptable, cost-effective treatments to tackle problems caused by depression.

CBM-I

Development of accessible treatments for depression has included successfully adapting CBT to be computerised (cCBT; Andrews, Cuijpers, Craske, McEvoy, & Titov, 2010). CBM has been proposed as a promising treatment. CBM aims to modify or induce cognitive biases through repeated exposure, without extensive instructions, to simple stimuli via computer software. CBM was initially implemented as an experimental task to test hypotheses that negative cognitive biases contributed causally to mental health problems. For example, anxiety increased in healthy volunteers after a CBM

paradigm reinforced negative interpretation biases (Mathews & Mackintosh, 2000). The success of CBM in modifying cognitive biases in healthy populations highlighted its potential as a mental health intervention. This review considers the potential of CBM to reduce symptoms of depression by a similar mechanism to the models outlined above of the action of CBT and antidepressants.

The majority of CBM studies have targeted negative attentional biases in anxiety. CBM research in depression is in its infancy (Torkan et al., 2014). Nevertheless, two meta-analyses have considered effectiveness of CBM in treating mental health conditions. Hallion and Ruscio (2011) examined effects of CBM-I and CBM-A on anxiety and depression in 45 studies. CBM interventions had a large effect in modifying interpretation biases and a medium effect on attention biases. Regarding depression, they concluded there was no evidence of CBM interventions modifying depression; this was based on a non-significant, small, unreliable effect. Limitations were that only 10 of the 45 studies used a depression measure. Only three investigated an analogue or clinical sample of people with depression, using CBM-A. The other seven measured depression in analogue or clinical samples of people with anxiety, and used a CBM paradigm designed specifically to target biases associated with anxiety, rather than depression. Conclusions of the effectiveness of CBM interventions on depression should be made cautiously in the context of these limitations.

A more recent meta-analysis (Menne-Lothmann et al., 2014) investigated effects of CBM-I on positive interpretations and mood. Different aspects of interpretation bias have been targeted by various CBM-I interventions, using homophones, word-sentence association tasks, ambiguous situations, and emotion recognition. The meta-analysis found CBM-I, which reinforced benign and positive interpretations, decreased negative mood. Bigger effects were

found when CBM-I paradigms encouraged participants to use mental imagery, and also for female participants, or when the number of sessions was greater. These findings do not have a high level of external validity when generalizing to people with symptoms of depression because the meta-analysis combined studies using anxiety and depression paradigms to examine healthy volunteers and analogue and clinical samples of people with anxiety and depression. The current review, in a similar design to that of Menne-Lothmann and colleagues (2014), only examined CBM-I studies, rather than other CBM paradigms.

Objective

The current systematic review aimed to examine effects of CBM-I on depression symptoms. To overcome limitations identified in the previous reviews, only studies that have used CBM-I with analogue or clinical samples with dysphoria or depression were examined. CBM-I, theoretically, has great potential in targeting negative biases that occur in depression to mediate changes in mood. Understanding the effectiveness of CBM-I interventions developed thus far is essential for treatment development and dissemination. The review question was, “What are the effects of CBM-I interventions on symptoms of depression of people with depression or dysphoria?”

Method

Search Strategy

Electronic database searches were carried out in July 2014 using Web of Science, PsycINFO, PubMed, Medline, Psycarticles, and The Cochrane Library to find papers published between 1994 and 2014. The search terms were (depress*, OR dysphori*) AND (cognitive bias modification, OR interpret* bias modification, OR interpret* training, OR bias training OR bias modification).

Additional search strategies included contacting experts in the field of CBM for information about unpublished or other relevant literature, and hand-searching reference lists of key articles.

The researcher carried out a systematic approach to identify appropriate studies (Figure 1). Duplicate studies were removed, titles and abstracts examined, then full-text of the remaining studies was screened to ensure fulfilment of the eligibility criteria. Data were extracted based on the PICOS framework, i.e., considering participants, intervention, comparators, outcomes, and study design. Issues of study quality were examined using guidelines outlined in the PRISMA statement (Moher, Liberati, Tetzlaff, & Altman, 2009) to assess risk of bias for each study.

Eligibility Criteria

Studies were eligible if they met the following criteria. Participants had dysphoria or depression, i.e., analogue or clinical samples. Participants of all ages and from all countries were considered. Outcomes included a depression measure, assessing changes in mood. The intervention used a CBM-I design, as defined by Koster, Fox, and MacLeod (2009). Specifically, the intervention aimed to directly modify an interpretation bias, not through following detailed instructions but instead through intensive practice on a simple, repetitive, cognitive task designed to facilitate the desired cognitive change. Eligible publication dates were 1994 to 2014, since the field of CBM research began around 2000 (Fox, Mackintosh, & Holmes, 2014). There were no publication status restrictions.

Study Synthesis Strategy

Studies were considered in terms of similarities and differences of design, participants, comparators, outcomes, and CBM-I interventions. Issues of study bias were also explored. To examine the objective of the review, investigating the effectiveness of CBM-I on depressive symptoms, the study findings relating to depression and interpretation bias outcomes were evaluated.

Results

Study Characteristics

Ten studies were identified for the review (Table 1).

Study Designs

All studies were described as randomised controlled trials (RCTs) except the earliest dated study, a single case-series (Blackwell & Holmes, 2010). RCTs represent the gold-standard in evaluating healthcare interventions, when well-designed, conducted and reported (Schulz, Altman, & Moher, 2010), therefore they have greater external validity than the case-series. Two studies were described as pilot-studies (Lang, Blackwell, Harmer, Davison, & Holmes, 2012; Torkan et al., 2014). Follow-ups, a component of good quality studies, were implemented by all studies. The follow-up findings from the study by Williams and colleagues (2013), however, will not be considered in this review because the outcomes are confounded by a 10 week web-based CBT intervention implemented after the CBM-I intervention.

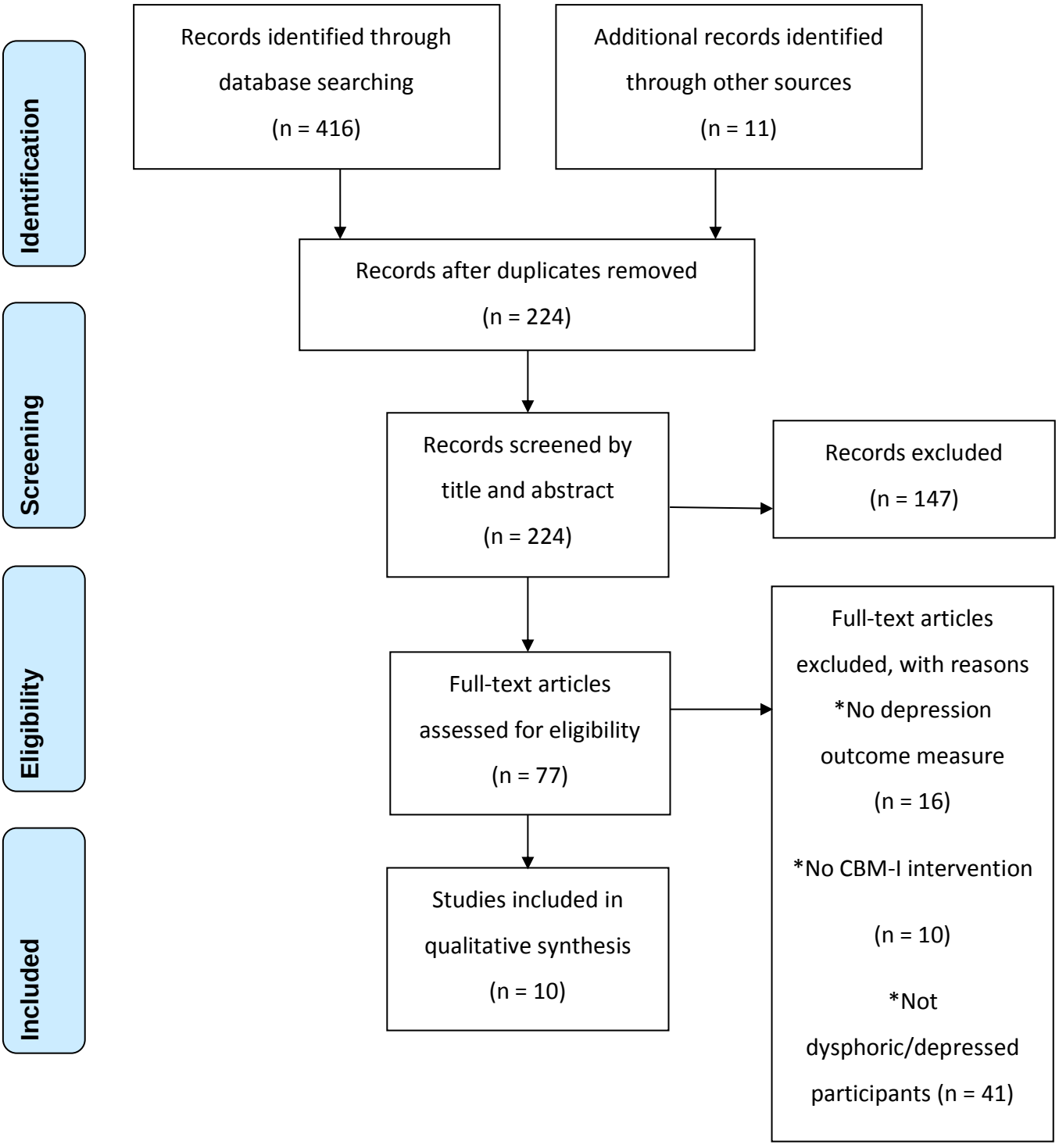

Figure 1. PRISMA flowchart of study selection.

Table 1

Studies investigating effectiveness of CBM-I interventions in reducing depressive symptoms in people with depression or dysphoria

Study Number. Authors: Study Centre	Study Designs	Participants	CBM-I Positive Intervention Paradigms	Comparators	Outcomes	Findings ($p < .05$)
1. Blackwell et al., (2014): Oxford, UK.	RCT. 1, 3, & 6 month follow-up.	MDD (SCID-I). N=150. Average age:35.	Multi-paradigm: auditory (6 sessions) & picture-word (6). 4 weeks at home.	Half negative, half positive stimuli.	(a) BDI-II & SST (b) Clinically significant change on BDI-II (c) Anhedonia (i.e., BDI-II items 4 & 12) (d) Moderating effect of number of episodes of depression on BDI-II (e) Moderating effect of vividness of imagining the CBM-I scenarios on BDI-II (f) Relationship between	(a) No time*condition interaction (b) No group difference (c) Greater reduction for positive v. control group at post intervention (not all time-points) (d) Subgroup with < 5 depressive episodes had a greater reduction on BDI-II in the positive vs control condition, at post-intervention & 1-month follow-up (not other time-points) (e) Greater reduction on BDI-II with higher vividness scores, within the positive v. control group at all time-points (f) Change in SST predicted change in BDI-II

					change in bias (SST) and mood (BDI-II).	score (pre-post). Post-intervention SST score correlated with BDI-II score at 3 & 6 month follow-up (not 1 month):
2. Blackwell & Holmes (2010): Oxford, UK.	Single case series: A-B design. 2 week follow-up	MDD (SCID-I). N=8. Average age:38.	Auditory: 7 days at home.	Week prior to intervention used as control.	(a) BDI-II, PANAS, SST, VAS (depressive statements)	(a) Improved mood & cognitive bias for 4 out of 7 participants that was maintained at follow-up, 3 found difficulties engaging with the CBM-I intervention, 1 excluded.
3. Lang et al., (2012): Oxford, UK.	RCT (pilot). 2 week follow-up	MDD (SCID-I). N=28. Average age:29.	Multi-paradigm: auditory (3 sessions), picture-word (2), word-fragment appraisal (1), mixed session (1). 1 week at home.	Half negative, half positive stimuli.	(a) BDI-II, Ham-D, IES, SST & RIQ, post-intervention (b) BDI-II, at follow-up (c) Clinically significant change on BDI-II & Ham-D post-intervention (d) Clinically significant change	(a) Decreased depressive symptoms & negative cognitive bias for positive group, not control group (b) Time*condition interaction at trend level: decrease for positive group not control (c) More positive group participants than control group, at post-intervention (d) More participants, at trend level, for positive

					on BDI-II at follow-up	group vs control
					(e) Correlation between change in interpretation bias (RIQ & SST) and depression (BDI-II)	(e) BDI-II change positively related with RIQ (at trend level with SST) for positive not control group, at post-intervention.
4. Micco et al., (2014): Boston, USA.	RCT. 2 week follow-up	BDI-II > 13. N = 45. Average age:18.	Word-fragment: scenarios relevant to potential loss, rejection or failure. 4 sessions over 2 weeks in lab.	Unambiguous, emotionally neutral stimuli.	(a) TIB (b) TIB: limited to the subgroup with a baseline negative bias (i.e. a score below 1.0) (c) DAS (d) CANTAB: AGN (e) BDI-II & MDD symptoms (SCID-IV/K-SADS-E).	(a) No time*condition interaction (b) Time*condition interaction mid-treatment at trend level post-treatment (not at follow-up): increased positive bias for positive vs control group (c) Greater reduction for positive vs control group at post-intervention & follow-up (d) No time*condition interaction (e) No time*condition interaction.

5. Newby et al., (2014): Sydney, Australia.	Non-randomised controlled trial. 1 week follow-up.	BDI-II > 12. N=60. Average age:26.	Word-fragment: appraisal of intrusive memories. Single session in lab.	1. CBT: single session. 2. No intervention.	(a) BDI-II, IES, & Appraisals of Intrusive Memories Questionnaire, at follow-up (b) PANAS, post-intervention (c) IMI (d) Intrusive Memory Diary, completed at home over a week (e) Correlation: bias & mood	(a) No time*condition interaction (b) No difference between groups (c) No difference between CBM group vs control or CBT (d) No difference between groups (e) Reduction in negative appraisal (baseline – follow-up) correlated with reduction of distress.
6. Penton-Voak et al., (2012): Bristol, UK.	RCT with 2 week follow-up	BDI-II > 13. N=80. Average age:21.	Morphed faces: 4 sessions in lab.	Sham-training.	(a) PANAS (positive, follow-up) & CBM-I face-morph final threshold (post-intervention) (b) PANAS (negative) & BDI-II, at follow-up.	(a) Increased positive affect & positive bias for positive group vs control (b) No significant difference.

7. Pictet et al., (2011): Oxford, UK.	RCT with 24 hour follow-up.	BDI-II > 10. N=81. Average age:28.	Picture-word: Single session in lab.	1. Half negative, half positive stimuli 2. All negative.	(a) PANAS (positive affect) & Fish Game, post-intervention (b) Homophone Task, at 24-hours follow-up.	(a) Increased positive affect & behavioural motivation for positive CBM-I group vs comparators (b) Increased positive bias for positive CBM-I group vs negative group, not positive vs half negative/half positive group.
8. Torkan et al., (2014): Isfahan, Iran.	RCT (pilot) with 2 week follow-up.	MDD (SCID-I). N=39. Average age:28.	Auditory (Persian version): with use of imagery vs non-imagery. 7 days at home.	No treatment.	(a) BDI-II, at post-intervention (b) SST & RRS, at post-intervention (c) BDI-II, at follow-up (d) SST, at follow-up (e) RRS, at follow-up (f) Clinically significant change on BDI-II (g) Relationship: change in	(a) Decrease in depressive symptoms for both CBM-I groups, not control. Greater decrease for imagery group than non-imagery, not between non-imagery & control (b) Decrease in negative interpretation & ruminative responses in imagery group, not non-imagery CBM-I condition or control (c) Decrease in depressive symptoms from baseline (not post-intervention) for imagery group, not non-imagery CBM-I condition

					bias (SST) & mood (BDI-II)	(d) Decrease in negative interpretation bias for imagery CBM-I group, increase in negative interpretation for non-imagery CBM-I group (e) No time*condition interaction (f) More participants in imagery CBM-I group than non-imagery group, at post-intervention and follow-up (g) Change in BDI-II correlated with change in negativity score on SST (pre-post).
9. Williams et al., (2013): Sydney, Australia.	RCT.	MDD (MINI). N=69. Average age:45.	Auditory: 7 days at home.	Wait list control.	(a) BDI-II, PHQ-9, K10, AST-D (b) SST (c) Clinically significant change on BDI-II (d) Relationship: change in bias (AST-D) & mood (BDI-II).	(a) Increased symptom reduction & more positive bias for positive group vs control (medium effects) (b) No interaction of time*condition (c) More participants, at trend level, for positive group vs control

						(d) AST-D change score (pre-post) was a significant mediator of condition (CBM-I vs control) on BDI-II change score, i.e., change in interpretation, at least in part, mediated reduction in depression symptoms following CBM-I
10. Yiend et al., (2013): London, UK.	RCT. 4 week follow-up.	MDD (MINI). N = 40. Average age:43.	Word-fragment: prompt positive future directed cognition to CBM-errors. A single lab session.	Unambiguous, emotionally neutral stimuli.	(a) SRT, post-intervention (b) SST, post-intervention (c) VAS sad/anxious: change after watching short film, intended to induce stress (d) BDI-II, PANAS, & MDI, at follow-up (e) ATQ-R, at follow-up.	(a) More positive bias for positive vs control group. (b) Significant time*condition interaction when means imputed for missing data, n = 3. Increase in positive bias for positive group not control group. Decrease in negative bias, at trend level (at significance when means imputed for missing data), for positive group not control group. (c) No time*condition interaction. (d) No time*condition interaction.

(e) No difference between groups.

Note. AST-D = Ambiguous Scenarios Test for Depression; ATQ-R = Automatic Thoughts Questionnaire-Revised; BDI-II = Beck Depression Inventory-II; CANTAB: AGN = CANTAB: Affective Go/No-Go Test; CBM-I = cognitive bias modification for interpretation bias; CBT = cognitive behavioural therapy; DAS = Dysfunctional Attitudes Scale ; IES = Impact of Event Scale; K10 = Kessler Psychological Distress Scale; K-SADS-E = Kiddie-Schedule of Affective Disorders and Schizophrenia-Epidemologic Version; IMI = Intrusive Memory Interview; Ham-D = Hamilton Depression Ratings Scale; KSADS-E = ;MDI = Major Depression Inventory; MINI = Mini International Neuropsychiatric Interview Version 5.0.0; PANAS = Positive and Negative Affect Scale; RCT = randomised controlled trial; MDD = major depressive disorder; PHQ-9 = Patient Health Questionnaire – 9 item Depression Scale; RIQ = Response to Intrusions Questionnaire; RRS = Ruminative Responses Scale; SCID-I = Structured Clinical Interview for DSM-IV; SRT = Similarity Ratings Test; SST = Scrambled Sentence Test; TIB = Test of Interpretation Bias; UK = United Kingdom; USA = United States of America; VAS = visual analogue scale.

Participants

The majority of study participants were Caucasian females, limiting generalizability of findings to the global population. One study recruited adolescents and young adults (Study 4), another recruited participants aged 18 to 45 (Study 6). All other studies recruited adults from a wider age range. Six studies (1, 2, 3, 8, 9, & 10) recruited only participants with a major depressive disorder. Four studies recruited participants with dysphoria (4, 5, 6, & 7). All studies reported detailed information of inclusion and exclusion criteria, and similarities in baseline data across conditions, showing good levels of validity in the study samples and transparency in reporting.

Interventions

Four different CBM-I paradigms were implemented. All were computerised interventions aimed at reinforcing, through repeated presentation of simple stimuli, a more positive bias for interpreting ambiguous situations. Heterogeneity of these CBM-I interventions raises the question whether the mechanism by which they work is the same, whether they have similar effects, and whether they are best measured using the same outcomes. Research into CBM-I for depression is relatively new, so investigating multiple paradigms is sensible in the quest to find an effective intervention.

One paradigm, implemented by five studies (1, 2, 3, 8, & 9), involved auditory presentation of initially ambiguous scenarios, which had a positive resolution. Another paradigm, implemented by three studies (1, 3, & 7), involved pictures of ambiguous scenes displayed alongside a positive word. A third paradigm, implemented by four studies (3, 4, 5, & 10), required participants to read short paragraphs, initially ambiguous in content, ending with a word fragment, which participants were asked to amend, creating a forced positive

resolution. The fourth paradigm, implemented by one study (6), involved brief presentations of faces, on a morphed sequence from sad to happy. Participants received feedback to shift interpretations of the ambiguous faces from sad to happy.

Over the ten studies, the number of sessions implemented ranged from one to twelve. Two studies used a multi-paradigm intervention (1 & 3). Five implemented the intervention in a laboratory setting (4, 5, 6, 7, & 10). The other five implemented interventions in participants' homes (1, 2, 3, 8, & 9); four studies initially invited participants to a laboratory setting (1, 2, 3, & 8), only one study (9) involved no face-to-face contact.

Comparators

Six studies included a comparator similar in design to the positive CBM-I paradigm implemented (1, 3, 4, 6, 7, & 10). These comparators consisted of control stimuli that were (a) unambiguous and emotionally neutral (4 & 10) (b) half negative and half positive (1, 3, & 7) or (c) all negative (7); as opposed to the positive CBM-I interventions in which the majority of stimuli were ambiguous and positively biased. Presentation of only negative stimuli as a control condition could be considered ethically questionable, particularly with participants who are depressed. Sham-training was the comparator implemented by one study (6); control group participants received feedback to reinforce their baseline thresholds.

Other comparators were no-treatment control conditions (5, 8, & 9). When examining differences between a positive CBM-I intervention and a no-treatment condition many more confounding factors are present than comparing a positive CBM-I intervention with a CBM-I control condition. In the case-series

study (2), the week before participants began the CBM-I program was used as a control period. One study (5) compared a different treatment paradigm: a therapist-delivered CBT session, aiming to target and modify negative interpretation biases.

Outcomes

All ten studies included at least one outcome measuring depressive mood and one measuring interpretation bias, the relevant study outcomes for this review.

Depressive mood.

All studies, except one (7) used the Beck Depression Inventory-II (BDI-II; Beck, Steer, & Brown, 1996), a gold-standard self-report measure of depression. Use of the same outcome measure facilitates comparisons between studies. The majority of depressive mood outcomes were self-report measures. Two studies implemented clinician-administered assessments of depressive mood (3 & 4). Clinician-administered assessments have been described as more accurate at measuring depressive severity than self-report measures but less able to detect change in mild levels of depression (Cusin, Yang, Yeung & Fava, 2010).

Two other outcomes were implemented, a novel behavioural task using a fishing game to measure motivated behaviour (7), and visual analogue scales (VAS) measuring sadness and anxiety to assess mood changes after participants watched a short video of a life-threatening incident, simulating a stressful experience (10). These outcomes, arguably, have less robust psychometric properties but more ecological validity in measuring symptoms of depression than self-report measures, such as the BDI-II and clinician-

administered assessments, such as the Hamilton Depression Rating Scale (Ham-D; Hamilton, 1960).

Interpretation bias.

Six studies (1, 2, 3, 8, 9, & 10) measured interpretation bias using the Scrambled Sentences Test (SST; Wenzlaff, 1993), a predictor of depressive symptoms (Rude et al., 2002). Six other tests of interpretation bias were implemented (see Table 1). A further measure of interpretation bias utilised visual analogues scales (VAS) associated with depressive statements (2), and three self-report measures were utilised. No single standard measure of interpretation bias has yet emerged. The heterogeneity of measures makes comparing findings difficult.

Study Quality and Issues of Bias

Nine of the reviewed studies claimed to be RCTs (1, 3, 4, 5, 6, 7, 8, 9, &10). One study (5) did not use a true randomisation procedure, alternately allocating participants to different conditions; three studies did not describe randomisation procedures (4, 7 & 8). Four studies (1, 4, 6, & 10) reported that condition allocation was concealed to reduce bias. Only one study collected a full set of data (7). Two studies moderated the risk of bias caused by attrition, clearly reporting flow of participants and appropriately using intention-to-treat analyses (1 & 9). While all study reports clearly described many issues relating to methodology, not all clearly identified the study hypotheses and primary outcome; four studies were transparent on this issue (1, 4, 8, & 9). One study (9), however, identified five measures as primary outcomes. When considering all of these issues of quality and bias, the most rigorous study (1) was implemented by Blackwell and colleagues (2014).

Study Findings

The findings presented are based on statistical tests where the probability of obtaining the results by chance was <.05, some findings are reported that were at trend level. Improvement in scores over time could be caused by regression to the mean; findings are therefore only reported if there was evidence of a greater improvement compared to a control group, not just a main effect of time.

Changes in depressive symptoms.

Reductions in depressive mood for groups allocated to the positive CBM-I intervention were found in eight studies. The most rigorously implemented study (1), using a 12-session multi-component CBM-I home intervention, found no evidence to support the a priori hypothesis regarding reduction in BDI-II score. Reductions in anhedonia were found: Cohen's *d* = 0.41 between groups post-intervention, a 0.54 difference on a 0 to 6 scale. Subgroup analyses revealed improvements on BDI-II score for participants with less than five depressive episodes; *d* = 0.73 between groups, a 6.9 difference post-intervention and 6.6 at 1-month follow-up on the BDI-II scale of 0 to 63. BDI-II score also improved for participants who could vividly imagine the CBM-I stimuli.

Three studies, using different positive CBM-I interventions, found evidence to support improvements on all included depression outcomes (3, 9, & 7). One such study (3) piloted a similar multi-paradigm design to the most rigorous study (1) but with less rigour and statistical power, a shorter intervention, and shorter follow-up; findings included significant reductions in depressive symptoms measured by the BDI-II (*d* = 0.89 within the positive group, a 6.85 difference), Ham-D (*d* = 1.24, a 4.92 difference on a 0 to 54 scale), and Impact of Event Scale (IES; Horowitz, Wilner & Alvarez, 1979; *d* = 1.34, a 8.62 difference on a 0

to 40 scale). Furthermore, more participants in the positive CBM-I group were found to have significant levels of clinical change on the BDI-II and Ham-D, compared with participants allocated to a control CBM-I group presented with half positive and half negative stimuli. The second study (9) compared a 7-day auditory CBM-I home intervention with a wait-list group; significant improvements were found on the BDI-II (a 5.83 difference between groups), Patient Health Questionnaire 9-item depression scale (Kroenke, Spitzer & Williams, 2001; a 1.72 difference on a 0 to 27 scale), and Kessler Psychological Distress Scale (Kessler et al, 2002; a 4.86 difference on a 10 to 50 scale). Medium effect sizes were reported for each improvement. The third study (7) used a single picture-word laboratory-based CBM-I session; significant increases in positive affect were found, measured by the Positive and Negative Affect Scales (PANAS; Watson, Clark & Tellegen, 1988; d = 0.77 between the positive CBI-I group and mixed CBM-I group, a difference of 7.71 on an unspecified PANAS scale, and a difference of 14.89 between the positive and negative CBM-I groups). The study also found increased behavioural motivation and persistence, negatively associated with dysphoria; this was measured by the fishing task, which was a toy requiring participants to use a small plastic rod to catch as many small fish as possible in 2.5 min. The positive CBM-I group caught 3.15 more fish than the mixed CBM-I group (d = 0.75) and 5.78 more than the negative group (d = 1.38).

The three studies that implemented a word-fragment CBM-I intervention, as a single-paradigm design, showed limited evidence for reducing depressive symptoms (4, 5, & 10). One study (4) found no greater reduction in depressive symptoms for the intervention group, following a 4-session laboratory-based CBM-I intervention, compared to the control group, measured by the BDI-II, and

a clinician-administered assessment. However, there was a reduction, post-intervention and at follow-up, on the Dysfunctional Attitudes Scale (DAS; Weissman & Beck, 1978), which measured depressive beliefs relating to perfectionism and the need for social approval (*d* = 0.60 between groups post-intervention, a 17.97 difference on a 40 to 280 scale; *d* = 1.02 at follow-up, a 30.92 difference). The DAS could be more sensitive to specific changes. This finding could also be a Type I error, since many different outcome measures were used. Two studies (5 & 9) used a similar intervention design but different participant inclusion criteria and different comparison conditions; neither found evidence for the superiority of positive CBM-I in reducing depressive symptoms on any mood measure.

The only study that used the face-morph CBM-I paradigm (6) measured depression symptoms 2 weeks following four laboratory-based sessions; evidence was found for improved positive affect, measured by the PANAS (a 3.29 adjusted mean difference on a 10 to 50 scale), but not for reduced negative affect or BDI-II score.

One study (8) compared a no-treatment group with two 7-day home-based positive CBM-I interventions using an auditory paradigm: one group used field perspective imagery; the other did not use imagery. The imagery group was superior to the non-imagery group, which was in turn superior to the no-treatment group in reducing depressive symptoms. The adjusted mean difference in BDI-II score between imagery and non-imagery group was 8.54 post-intervention (*d* = 1.11), 13.0 at follow-up (*d* = 1.75), and 9.85 between imagery and control group post-intervention (*d* = 1.04). The difference in score on the Ruminative Responses Scale (Nolen-Hoeksema & Morrow, 1991) post-intervention between the imagery and non-imagery group was 15.77 on a 22 to

88 scale (d = 1.0), and 17.0 between the imagery and control group (d = 1.13). Finally, the case-series study (2), using a 7-day home-based auditory CBM-I intervention, showed reductions in depressive symptoms for four of seven participants. Qualitative data indicated responsive participants had more fully engaged with the intervention than non-responders.

Changes in interpretation bias.

Increased positive interpretation bias or decreased negative bias for groups allocated to the positive CBM-I intervention, compared to comparator conditions, were found in seven studies. No evidence of superiority for the positive CBM-I intervention in affecting cognitive bias was found in three studies, including the most rigorously implemented study (1).

Improved cognitive bias was found on all implemented outcome measures for three studies. This included the study in which the outcome stimuli were identical to the CBM-I training stimuli (6); therefore generalizability of this finding to other ambiguous situations is uncertain. The adjusted mean difference was 2.72 for shift in bias on a scale from 1 to 15 on an emotion recognition task. The short multi-paradigm study (3) showed a 0.13 difference on a 0 to 1.0 scale for SST between groups, d = 0.60, and a 8.23 difference on a 6 to 42 scale for RIQ, d = 1.34. One study (7) found superiority in increasing positive bias for the positive CBM-I group over the negative CBM-I group (a 6.09 difference on a 0 to 100 scale for a homophone task, d = 0.56) but not the mixed CBM-I group.

Mixed findings occurred in some studies. One study (8) found improved cognitive bias for the positive CBM-I group using imagery, post-intervention (a 28.43 difference compared with the imagery group on a 0 to 100 scale for SST, d = 0.77, and 24.99 compared with the control group, d = 0.73) however, at

follow-up, the non-imagery positive CBM-I group showed increased negative bias (a difference of 42.17, *d* = 2.32). Another study (9) found increased positive bias on the Ambiguous Scenarios Test for Depression (Berna, Lang, Goodwin & Holmes, 2011; a 0.77 difference on an unspecified scale, corresponding to a medium effect size) but no SST difference. The study identified the majority of included measures as primary outcomes but proposed no explanation for why improvements were found on some measures and not others. The case-series study (2) showed improved cognitive bias for four of seven participants.

Association between change in bias and mood.

The relationship between change in bias and mood was examined by five studies (1, 3, 5, 8, & 9); all found a positive association. The importance of this analysis is to investigate the hypothesised mechanism of CBM treatments on mental health symptoms, i.e., change in cognitive bias is hypothesised to mediate changes in mood. These associations are not necessarily causative.

Review Findings

The objective of this review was to examine the effectiveness of CBM-I on depressive symptoms. Theoretical underpinnings of the effect of CBM-I on depression suggest changes in interpretation bias can mediate changes in mood symptoms. Ten studies were systematically selected. All investigated use of a positive CBM-I paradigm with participants experiencing depression or dysphoria. The main findings showed a mixed picture: inconsistent evidence of improvement on a variety of relevant outcomes measuring depressive symptoms and interpretation bias, compared to various control conditions. Significant differences between groups were generally medium sized effects; the clinical significance of these differences is unclear because minimal

clinically important differences were not routinely reported. The mixed picture of findings falls between the negative and optimistic results of the two meta-analyses: no evidence that CBM modifies depression (Hallion & Ruscio, 2011); positive or benign CBM-I paradigms are associated with increased positive interpretations and decreased negative mood states (Menne-Lothmann et al., 2014). Inconsistent findings could relate to issues regarding the underlying theory, implemented interventions, or study methodology. CBM-I might not be an effective treatment for depression; not all CBM-I paradigms might be effective in affecting changes in cognitive bias to mediate changes in mood; limitations in study methodology might have prevented significant positive effects from being found.

Study designs showed different levels of quality. The most rigorous study (1), which incorporated a multi-session, multi-paradigm CBM-I intervention, with the biggest sample, and longest follow-up, found no evidence to support hypotheses of reductions in symptoms of depression and negative bias. Subgroups of responders, moderating factors, and improvements in specific depressive symptoms were identified: participants with less than five episodes of depression, vividly imagining scenarios from a field perspective, and anhedonia. Existing treatments for depression are not effective for everyone (Hollon, Stewart, & Strunk, 2006) so CBM-I could be a viable intervention even if it is only effective for certain groups.

Limitations

Ten studies with heterogeneous designs were reviewed. This small, varied sample reduces the external validity of the findings. A thorough search for relevant papers was implemented but there is a risk of publication bias concealing further null findings. Many outcome measures were included in

studies, increasing the risk of a Type I error. Studies with small samples were underpowered to detect small significant effects (all reported significant effects were more than Cohen's *d* = 0.5). The study with the biggest sample (1; *n*=150) suggested it was still underpowered to detect small effects; the significant change for anhedonia was only a small to medium effect size. CBM researchers have suggested that even if CBM-I interventions produce a change in outcome measure corresponding to a small effect size it would be worth further investigation into CBM-I as a depression treatment because of its low-intensity, cost-effective nature (Blackwell et al., 2014). If a CBM-I intervention were found to lead consistently to improvements corresponding to a small effect size, however, patients and health professionals might be overoptimistic about its potential, which could lead to frustration and further depression.

Recommendations

Identification of the most effective CBM-I paradigm, and the most meaningful, sensitive measures is important in the development of finding an effective CBM-I treatment for depression. An important goal to pursue is clinically significant change rather than small effect sizes. Further investigation into what constitutes minimal clinically important differences on outcome measures used in CBM-I research would be beneficial. An additional consideration is its potential to be an accessible adjunct to other interventions, whether or not it brings about clinically significant change on its own. Studies should ensure rigor in design, implementation, and reporting; the CONSORT checklist (Schulz et al., 2010) is a helpful tool to follow. When designing interventions and comparators, ethical implications should be considered. Analysis of the underlying mechanism of the intervention, using mediation analysis, is important for inclusion in future studies. Some studies included

measures of acceptability, which should be routinely used to ensure any treatment developed is acceptable. Only one study was implemented in a country where English is not the first language, translating the intervention into Persian (8). The verbal nature of three CBM-I paradigms reduces accessibility for many people, worldwide. Only one study tested the face-morph paradigm (6), which could be adapted to be non-verbal. Further investigation of the effectiveness of non-verbal CBM-I tasks is a recommendation for developing more accessible interventions.

Conclusion

This systematic review has examined ten studies that have used CBM-I interventions with people with depression and dysphoria to investigate its effect on depressive symptoms. The findings were mixed, suggesting the intervention is not currently at a stage to be recommended as a treatment possibility for depression. Identification of the most meaningful outcome measures and effective CBM-I paradigm, perhaps a non-verbal design to maximise accessibility, paired with further rigorous studies are recommendations for further research in this field.

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of *mental disorders.* (5th ed.). Washington, DC: Author.

Andrews, G., Cuijpers, P., Craske, M. G., McEvoy, P., & Titov, N. (2010). Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: A meta-analysis. *PLoS ONE, 10*(5), 1-6.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford Press.

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

Blackwell, S. E., Browning, M., Mathews, A., Pictet, A., Welch, J., Davies, J.,…Holmes, E. A. (2014). Positive imagery-based cognitive bias modification as a web-based treatment tool for depressed adults: A randomised controlled trial. *Clinical Psychological Science* (in press).

Blackwell, S. E., & Holmes, E. A. (2010). Modifying interpretation and imagination in clinical depression: A single case series using cognitive bias modification. *Applied Cognitive Psychology, 24,* 338-350.

Chabrol, H., Teissedre, F., Armitage, J., Danel, M., & Walburg, V. (2004) Acceptability of psychotherapy and antidepressants for postnatal depression among newly delivered mothers. *Journal of Reproductive and Infant Psychology, 22*(1), 5-12.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1996). *Structured clinical interview for DSM-IV Axis 1 disorders: Patient edition (SCID-I/P,*

Version 2.0). New York: Biometrics Research Dept, New York State Psychiatric Institute.

Hallion, L. S., & Ruscio, A. M. (2011). A meta-analysis of the effect of cognitive bias modification on anxiety and depression. *Psychological Bulletin, 137*(6), 940-958.

Harmer, C. J., Goodwin, G. M., & Cowen, P. J. (2009). Why do antidepressants take so long to work? A cognitive neuropsychological model of antidepressant drug action. *The British Journal of Psychiatry, 195,* 102-108.

Hollon, S. D., Stewart, M. O., & Strunk, D. (2006). Enduring effects for cognitive behaviour in the treatment of depression and anxiety. *Annual Review of Psychology, 57,* 285-315.

Karampampa, K., Borgström, F., & Jönsson, B. (2011). Economic burden of depression on society. *Medicographia, 33*, 163-168.

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

Landreville, P., Landry, J., Baillargeon, L., Guerette, A., & Matteau, E. (2001). Older adults' acceptance of psychological and pharmacological treatments for depression. *Journal of Gerontology, Series B, Psychological Sciences and Social sciences, 56*(5), 285-291.

Lang, T. J., Blackwell, S. E., Harmer, C., Davison, P., & Holmes, E. A. (2012). Cognitive bias modification using mental imagery for depression:

developing a novel computerized intervention to change negative thinking styles. *European Journal of Personality, 26,* 145-157.

Lépine, J-P., & Riley, M. (2011). The increasing burden of depression. *Neuropsychiatric Disease and Treatment, 7*, 3-7.

Fox, E., Mackintosh, B., & Holmes, E. (2014). Travellers' Tales in Cognitive Bias Modification Research: A Commentary on the Special Issue. *Cognitive Therapy and Research, 38*(2), 239-247.

Mathews A., & Mackintosh, B. (2000). Induced emotional interpretation bias and anxiety. *Journal of Abnormal Psychology, 109*(4), 602-615.

Mathews, A., & MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. *Annual Review of Clinical Psychology, 1*, 167-195.

Menne-Lothmann, C., Viechtbauer, W., Hohn, P., Kasanova, Z., Haller, S. P., Drukker, M.,…Lau, J. Y. F. (2014). How to Boost Positive Interpretations? A meta-analysis of the effectiveness of cognitive bias modification for interpretation. *PLoS ONE, 9*(6), 1-26.

Micco, J. A., Henin, A., & Hirshfeld-Becker, D. R. (2014). Efficacy of Interpretation Bias Modification in Depressed Adolescents and Young Adults. *Cognitive Therapy and Research, 38*(2), 89-102.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement, *PLoS Med. 6*(6), e1000097.

Newby, J. M., Lang, T., Werner-Seidler, A., Holmes, E., & Moulds, M. L. (2014). Alleviating distressing intrusive memories in depression: A comparison

between computerised cognitive bias modification and cognitive behavioural education. *Behaviour Research and Therapy, 56*(100), 60-67.

Nolen-Hoeksema, S., & Morrow, J. (1991). A prospective study of depression and posttraumatic stress symptoms after a natural disaster: The 1989 Loma Prieta Earthquake. *Journal of Personality and Social Psychology, 61,* 115-121.

Penton-Voak, I. S., Bate, H., Lewis, G., & Munafò, M. R. (2012). Effects of emotion perception training on mood in undergraduate students: Randomised control trial. *The British Journal of Psychiatry, 200,* 1-3.

Pictet, A., Coughtrey, A. E., Mathews, A., & Holmes, E. A. (2011). Fishing for happiness: The effects of generating positive imagery on mood and behaviour. *Behaviour Research and Therapy, 49*(12), 885-891.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *The British Medical Journal, 340*, c332.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E.,…Dunbar, G.C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry, 59*(Suppl 20), 22-33.

Titov, N., Andrews, G., Davies, M., McIntyre, K., Robinson, E., & Solley, K. (2010). Internet Treatment for Depression: A Randomized Controlled Trial Comparing Clinician vs. Technician Assistance. *PLoS ONE, 5*(6), e10939.

Torkan, H., Blackwell, S. E., Holmes, E., Kalantari, M., Neshat-Doost, H. T., Maroufi, M., & Talebi, H. (2014). Positive Imagery Cognitive Bias Modification in Treatment-Seeking Patients with Major Depression in Iran: A Pilot Study. *Cognitive Therapy and Research, 38*(2), 132-145.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070.

Williams, A. D., Blackwell, S. E., Mackenzie, A., Holmes, E. A., & Andrews, G. (2013). Combining imagination and reason in the treatment of depression: A randomized controlled trial of Internet-based cognitive-bias modification and Internet-CBT for depression. *Journal of Consulting and Clinical Psychology, 81*(5), 793-799.

World Health Organization (2013). *Mental health action plan: 2013-2020.* Geneva, Switzerland: Author.

Yiend, J., Lee, J-S., Tekes, S., Atkins, L., Mathews, A., Vrinten, M.,…Shergill, S. (2014). Modifying interpretation in a clinically depressed sample using 'cognitive bias modification-errors': A double blind randomised controlled trial. *Cognitive Therapy and Research, 38*(2), 146-159.

Appendix A: The British Journal of Psychiatry: Instructions for Authors

The British Journal of Psychiatry (BJPsych) is published monthly by the Royal College of Psychiatrists. The BJPsych publishes original work in all fields of psychiatry. Manuscripts for publication should be submitted online via http://submit-bjp.rcpsych.org.

All published articles are peer reviewed. Contributions are accepted for publication on the condition that their substance has not been published or submitted for publication elsewhere, and this includes web-based documents. Authors submitting papers to the BJPsych (serially or otherwise) with a common theme or using data derived from the same sample (or a subset thereof) must send details of all relevant previous publications and simultaneous submissions.

Manuscripts accepted for publication are copy-edited to improve readability and to ensure conformity with house style.

The title should be brief and relevant. Subtitles should not be used unless they are essential. Titles should not announce the results of articles and, except in editorials, they should not be phrased as questions.

The BJPsych supports and expects all authors to follow the World Medical Association's Declaration of Helsinki regarding the ethics of research involving human participants. Assessors for the journal are required to consider the ethics of each submitted work. To facilitate this, every research article submitted to the BJPsych must include a statement that the investigators obtained ethical approval for the study (or an explanation of why ethical approval was not needed), including the name of the ethics committee(s) or review board(s), any allocated reference number of the approval(s) and a statement concerning the participants' informed consent.

A structured abstract of up to 150 words should be given at the beginning of the article, incorporating the following headings: Background; Aims; Method; Results; Conclusions; Declaration of interest. The abstract is a crucial part of the paper and authors are urged to devote some care to ensuring that all the important findings are included within the word limit.

Introductions should normally be no more than one paragraph; longer ones may be allowed for new and unusual subjects. This should be followed by Method, Results and Discussion sections. The Discussion should always include limitations of the paper to ensure balance. Use of subheadings is encouraged, particularly in Discussion sections. A separate Conclusions section is not required.

The article should normally be between 3000 and 5000 words in length (excluding references, tables and figure legends) and normally would not include more than 25 essential references beyond those describing statistical procedures, psychometric instruments and diagnostic guidelines used in the study. All large tables (exceeding half a journal page) will be published only in the online version of the BJPsych (see Online data supplements, below). Authors are encouraged to present key data within smaller tables for print publication. This applies also to review articles and short reports.

Review articles should be structured in the same way as regular papers, but the length of these may vary considerably, as will the number of references. Systematic reviews are preferred and narrative reviews will be published only under exceptional circumstances. Reviews done for the Cochrane Collaboration, the National Institute for Health and Clinical Excellence and other groups likely to be published, or already published, elsewhere, should be

submitted with the latest version of the parent review and its status so that an informed decision can be made about the added value of the submitted paper.

Authors are responsible for checking all references for accuracy and relevance in advance of submission. Reference lists not in the correct style will be returned to the author for correction. References should be numbered in the order that they appear in the text and listed at the end of the article using the Vancouver style (see below), in which the names and initials of all authors are given after the appropriate reference number. If there are more than six authors, the first six should be named, followed by 'et al'.

The authors' names are followed by the full title of the article; the journal title abbreviated (in italics) according to the style of Index Medicus; the year of publication; the volume number (in bold type); and the first and last page numbers. References to book or book chapters should give the titles of the book (and the chapter if selected), names of any authors, name of publisher, names of any editors, and year.

Personal communications need written authorisation (email is acceptable); they should not be included in the reference list. Unpublished doctoral theses may be cited (please state department or faculty, university and degree). No other citation of unpublished work, including unpublished conference presentations, is permissible.

Tables should be numbered and have an appropriate heading. The tables should be mentioned in the text but must not duplicate information. The heading of the table, together with any footnotes or comments, should be self-explanatory. The desired position of the table in the manuscript should be

indicated. Do not tabulate lists, which should be incorporated into the text, where, if necessary, they may be displayed.

Figures should be clearly numbered and include an explanatory legend. Avoid cluttering figures with explanatory text, which is better incorporated succinctly in the legend. 3-D effects should generally be avoided. Lettering should be parallel to the axes. Units must be clearly indicated and should be presented in the form quantity (unit) (note: `litre' should be spelled out in full unless modified to ml, dl, etc.). All figures should be mentioned in the text and the desired position of the figure in the manuscript should be indicated.

Authors must obtain permission from the original publisher if they intend to use figures from other sources, and due acknowledgement should be made in the legend.

Colour figures may be reproduced if authors are able to cover the costs.

Methods of statistical analysis should be described in language that is comprehensible to the numerate psychiatrist as well as the medical statistician. Particular attention should be paid to clear description of study designs and objectives, and evidence that the statistical procedures used were both appropriate for the hypotheses tested and correctly interpreted. The statistical analyses should be planned before data are collected and full explanations given for any post hoc analyses carried out. The value of test statistics used (e.g. t, F-ratio) should be given as well as their significance levels so that their derivation can be understood. Standard deviations and errors should not be reported as ± but should be specified and referred to in parentheses.

Trends should not be reported unless they have been supported by appropriate statistical analyses for trends.

The use of percentages to report results from small samples is discouraged, other than where this facilitates comparisons. The number of decimal places to which numbers are given should reflect the accuracy of the determination, and estimates of error should be given for statistics.

A brief and useful introduction to the place of confidence intervals is given by Gardner & Altman (1990, Br J Psychiatry, 156, 472-4). Use of these is encouraged but not mandatory.

Authors are encouraged to include estimates of statistical power where appropriate. To report a difference as being statistically significant is generally insufficient, and comment should be made about the magnitude and direction of change.

The BJPsych requires authors to submit a completed checklist and flowchart in accordance with the appropriate CONSORT guidelines. The registration details of the trial and a flow chart illustrating the progress of participants through the trial (CONSORT diagram) must be included in the submitted manuscript.

Reports of systematic reviews or meta-analyses of evaluations studies, including randomised controlled trials, must follow the PRISMA guidelines; meta-analyses of observational studies must adhere to the MOOSE guidelines.

For further guidance, authors may refer to the Royal College of Psychiatrists' house style guide.

If the study includes original data, at least one author must confirm that he or she had full access to all the data in the study, and takes responsibility for

the integrity of the data and the accuracy of the data analysis. We strongly encourage authors to make their source data publicly available.

Material related to a paper but unsuitable for publication in the printed journal (e.g. large tables) may be published as a data supplement in the online BJPsych at the Editor's discretion. For very large volumes of material, charges may apply.

All abbreviations must be spelt out on first usage and only widely recognised abbreviations will be permitted.

The use of notes separate to the text should generally be avoided, whether they be footnotes or a separate section at the end of a paper. A footnote to the first page may, however, be included to give some general information concerning the paper.

There is no submission or publication fee for papers published in the BJPsych in the usual way. All papers published in the BJPsych become freely available online 12 months after publication. In a new initiative to maximise access to original research, authors now have the option to make their papers freely available from the time of publication, on payment of an open access charge. This charge is currently £2500 (or US$4500) per article plus VAT where applicable. If you wish to take up this option, contact the BJP Editorial Assistant once your paper has been accepted for publication. For such papers, the requirement for a 12-month delay before release of the manuscript in a public archive is waived, and the final published version may be deposited.

Authors may be asked to provide the raw data at any time up to 5 years after publication of research in the BJPsych.

MAJOR RESEARCH PROJECT COVER SHEET

Effects on Depressive Symptoms of a Web-Based Cognitive Bias Modification-Interpretation (CBM-I) Program for Emotion Recognition: A Randomised Controlled Trial

TRAINEE NAME: Victoria Stephens

SUPERVISORS

PRIMARY: Professors Marcus Munafò and Ian Penton-Voak, School of Psychology, University of Bristol.

SECONDARY: Dr Anna Adlam, College of Life and Environmental Science, University of Exeter.

TARGET JOURNAL: British Journal of Psychiatry

WORD COUNT: 7709 + corrections (excluding abstract, tables, titles, references and appendices)

DECLARATION:

“I certify that all material in this literature review which is not my own work has been identified and properly attributed. I have conducted the work in line with the BPS DCP Professional Practice Guidelines.”

SUBMITTED IN PARTIAL FULFILMENT OF REQUIREMENTS FOR THE DOCTORATE DEGREE IN CLINICAL PSYCHOLOGY

UNIVERSITY OF EXETER

Abstract

Depression is a global problem, causing disability and economic burden. Many people currently do not obtain treatment. Development of more accessible, cost-effective treatments is essential. An identified mechanism by which depression treatments work is through modifying underlying negative cognitive biases, which mediate changes in mood. A specific negative information-processing bias in depression is a tendency to interpret ambiguous facial expressions as sad rather than happy. The emotion recognition task is a treatment paradigm developed as a cognitive bias modification intervention to target this emotion recognition bias. Previous studies showed promising signs that this novel intervention could modify biases in people with low mood outside of laboratory conditions and potential to increase positive affect within laboratory conditions. The current study built on these developments, aiming to investigate, using a randomised controlled trial with follow-up at 2 and 6 weeks, whether a web-based version of the emotion recognition task could reduce depressive symptoms in addition to modifying emotion recognition biases. An analogue sample of 124 participants with low mood was recruited. Evidence was found that the intervention modified participants' biases, compared to the control group but there was no evidence of improvement in mood. Study limitations included a high rate of attrition and non-adherence to the intervention. Future recommendations include modifying the intervention to increase acceptability, investigating generalizability of increased positive bias to different stimuli, and identifying consistent reductions in symptoms of depression before examining its efficacy with a clinical population.

Introduction

Depression is a major cause of disability (Rodgers et al., 2012) characterised by persistent low mood and loss of interest in activities, accompanied by changes in psychomotor activity, sleep, and appetite, lack of energy, feelings of worthlessness, concentration difficulties, and suicidal ideation (American Psychiatric Association, 2000). Evidence-based treatments, such as cognitive-behavioural therapy (CBT), are often delivered face-to-face. This makes them inaccessible for many people, particularly when depression is becoming more prevalent (Ustun, Ayuso-Mateos, Chatterji, Mathers, & Murray, 2004) and there is increasing pressure on healthcare budgets. Innovative, accessible, automated treatments could allow more people to receive help. This study investigated the potential of a web-based programme aimed at modifying emotion recognition biases to reduce depressive symptoms. This section will outline the rationale for this study, first describing cognitive theories of depression with emphasis on negative biases, then cognitive treatments with a focus on cognitive bias modification (CBM).

Cognitive Models of Depression

Cognitive perspectives of how psychological problems develop generally emphasise the role of interpreting ambiguous situations with a negative bias (Hughes, Panzarella, Alloy, & Abramson, 2007). Interpretation of stimuli and other cognitive processes are hypothesised to be underpinned by schemata, composed of core beliefs: how we see ourselves, others, the world, and future (Beck, 1976). These schemata assist us in processing complex information, biasing how we attend to, interpret, and remember information. Low mood is hypothesised to be maintained by negative internal representations of the world and associated negative information-processing biases: overgeneralising and

assuming personal responsibility for negative events, and disqualifying or ignoring positive information. A subtly different theory, depressive realism, suggests people with depression might perceive situations more accurately than people without depression who have optimistic biases (Alloy & Abramson, 1988); both theories emphasise people with depression have a more negative cognitive style than those without depression. It should be noted that other cognitive theories of depression, not discussed here, emphasise processes other than negative biases, including rumination (e.g., Nolen-Hoeksema, 1991) and executive control (e.g., Dalgleish et al., 2007).

Beck's model of depression, which focuses on negative schema and negative biases, led to the development of CBT, a recommended evidence-based treatment for depression (National Institute for Health and Clinical Excellence, 2009). The cognitive component of CBT targets modification of negative automatic thoughts regarding self, others, the world and future (Beck, Rush, Shaw, & Emery, 1979). The inaccessibility of face-to-face therapy has been partly addressed through the development of computerised CBT (cCBT; e.g., Andersson et al., 2005).

CBM Research for Depression

Unlike cCBT, which was designed to replicate traditional CBT, CBM treatments have been developed from experimental research and informed by cognitive science, fitting with an 'experimental medicine' framework (Lang, Blackwell, Harmer, Davison, & Holmes, 2012). Evidence was found showing people with depression have negative information-processing biases, including interpreting (Rude, Wenzlaff, Gibbs, Vane, & Whitney, 2002), attending to (Joorman & Gotlib, 2007), and remembering information (Mathews & MacLeod, 2005). CBM studies developed from theory-testing paradigms to investigating

the effectiveness of CBM as an intervention, targeting negative biases occurring at different levels of cognitive processing, interpretation (Blackwell & Holmes, 2010), attention (Baert, De Raedt, Schacht, & Koster, 2010) and memory (Joormann, Hertel, LeMoult, & Gotlib, 2009), with the aim of mediating changes in mood.

Antidepressant Treatment and CBM

A neurocognitive model of the action of antidepressant medication shows similarities to the underlying mechanism proposed for CBT and CBM. At a neuropsychological level, emotional and social stimuli have been found to be processed in a more positive manner quickly after drug administration (Harmer, Goodwin, & Cowen, 2009). Change in bias is hypothesised to lead to gradual changes in social interactions, followed by improved mood; establishing a self-sustaining virtuous cycle. Compared with people without depression, people with depression generally interpret ambiguous facial expressions as less happy (Bourke, Douglas, & Porter, 2010). This negative bias was found to decrease after a single dose of antidepressants (Harmer et al., 2009) and before subjective changes in mood occur (Tranter et al., 2009). This evidence supports the view that cognitive models of depression, which emphasise the importance of negative biases, fit with the proposed action of antidepressants (Harmer et al., 2009). Therefore, antidepressants, CBT, and CBM might employ similar mechanisms in treating depression, i.e., reducing negative biases to mediate improvements in mood. Regarding antidepressant treatment, many people prefer to avoid taking medication; this underlines the importance of pursuing innovative, non-invasive, cognitive interventions that modify cognitive biases and treat depression.

CBM for Emotion Recognition

A CBM paradigm was developed to directly target emotion recognition biases (Penton-Voak, Bate, Lewis, & Munafò, 2012), similar to biases found to be modulated by antidepressants, described above. This CBM task, described as the emotion recognition task (ERT) in this report, aims to reduce negative biases and promote positive biases when interpreting ambiguous facial expressions. For example, the program can reinforce the identification of ambiguous faces as showing a happy, rather than sad, emotion. The rationale for using facial expressions as a target for interpretation modification to treat depression stems from findings that people with depression, when compared with healthy controls, have a negative bias towards interpreting ambiguous or neutral faces as sad or less happy (Bourke et al., 2010). A negative bias towards interpreting ambiguous faces can maintain depressive schemas about others and the world because people with depression interpret their environments as having more sadness and negativity than people without depression. Promotion of a more positive bias to emotion recognition could reduce negative interpretations that are maintaining depressive schemas about the world, which could mediate a reduction in depressive symptoms.

There have been promising preliminary results in modifying emotion recognition biases, which could provide therapeutic benefit for people with depressive symptoms (Penton-Voak et al., 2012). The hypothesis for the action of the ERT matches the model of the action of antidepressants (Harmer et al., 2009): early changes occur in decreasing negative biases; this is predicted to gradually have a positive impact on social interactions, eventually causing positive changes in subjective mood. Preliminary findings showed the ERT, which promotes the identification of positive over negative emotion, might have

beneficial effects on mood, which persist for at least two weeks (Penton-Voak et al., 2012). The results were generated in a laboratory setting with adults recruited from the general population on the basis of having at least a mild level of depression, as indicated by a score of 14 or more on the Beck Depression Inventory-II (BDI-II; Beck, Steer, & Brown, 1996). Studies investigating the generalizability of the ERT have shown good potential for reducing problematic behaviours in young offenders (Penton-Voak et al., 2013) and effectiveness in modifying emotion recognition biases of people with low mood via portable devices (e.g., smartphones and tablets) outside of laboratory conditions (Brazil, Munafò, & Penton-Voak, 2012).

Penton-Voak and colleagues (2012) recruited 80 participants with low mood for their laboratory-based study. Participants were divided randomly into two groups: one received the ERT; the other, sham-ERT. There was limited attrition: two participants discontinued the intervention and one did not complete the follow-up stage. Issues of attrition are important to consider for measuring the acceptability of the intervention, particularly if it is to be developed as an accessible treatment. Following training, the threshold for identifying ambiguous faces as happy rather than sad was significantly different between groups: the training group had a more positive bias than the control group. Positive affect, measured by the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988), showed a significant increase for the training group compared with the control group. Reductions in low mood, measured by the BDI-II, which was the primary outcome measure, and negative affect, measured by the PANAS, were statistically non-significant between the groups. Penton-Voak and colleagues (2012) noted study limitations including lack of statistical power to detect what they described as “modest (but clinically valuable)

improvements in mood" (p. 72). The study's sample size (*n* = 77) provided 80 percent power to detect, at an alpha level of 0.05, a difference of six points on the BDI-II (assuming *SD* = 10) and five points on the PANAS (*SD* = 7); the minimal clinically important difference was not defined. Additionally, follow-up was 2 weeks post-training and the researchers suggested that assessing mood change over a longer period would be valuable. The current study aimed to take into account these limitations, recruiting a larger sample and assessing mood change both 2 and 6 weeks after training ended.

Brazil and colleagues (2012) demonstrated that the ERT could be effective in modifying cognitive biases outside laboratory conditions when accessed using portable devices such as smartphones. They recruited 88 participants with low mood, defined by a score of 14 or more on the BDI-II. Attrition was 20%: four participants dropped out before the end of the study; technical problems caused 14 participants' data not to be collected. The ERT was installed onto participants' smartphones and tablet computers for independent use over 6 weeks: on average participants chose to use the program on five occasions. The magnitude of the shift in bias for identifying ambiguous faces as happy rather than sad increased with the number of training sessions completed.

Aim of Study and Hypotheses

The aim of the present study was to extend previous research (Brazil et al., 2012; Penton-Voak et al., 2012) by investigating whether the ERT showed signs of therapeutic benefit on low mood when accessed through the internet, outside of laboratory conditions. The main objective was to investigate whether participants experienced a therapeutic benefit on mood measured by the BDI-II,

in addition to a shift in bias for identifying ambiguous faces as happy rather than sad, compared with a control group assigned to a sham-ERT.

The primary hypothesis predicted that (a) participants who accessed the ERT would have fewer symptoms of depression as measured by the BDI-II, at 6 weeks follow-up, than those randomised to the control group. Secondary hypotheses were (b) participants in the training condition would have fewer symptoms of depression at two other time-points, post-intervention and 2 weeks follow-up, as measured by the BDI-II, than the control group. Also, (c) participants in the training group would have less negative mood, and more positive mood as measured by the PANAS, at three time-points, post-intervention, 2 and 6 weeks follow-up, than the control group. It was further hypothesised that (d) participants in the training group would have a more positive bias for interpreting ambiguous facial expressions as happy than sad, as measured by the thresholds obtained at time-points after baseline, compared to the control group. This was a manipulation check to ensure the intervention was having an effect; furthermore, this analysis was used to check that previous work was replicable by this study.

Exploratory analyses were carried out to find whether (a) participants randomised to the training group had fewer symptoms of anxiety, measured by the Beck Anxiety Inventory (BAI; Beck & Steer, 1993), than the control group. The majority of previous CBM intervention studies have investigated effectiveness in reducing anxiety symptoms (Torkan et al., 2014). Further exploratory analyses examined the moderating effects of (b) baseline depression and (c) baseline anxiety on how use of the ERT affected BDI-II outcomes. The intervention might vary in its effectiveness for people with different baseline levels of anxiety and depression. Indeed, Blackwell and

colleagues (2014) found bigger improvements in BDI-II outcomes, after participation in a CBM intervention, for people with fewer previous episodes of depression. Additional exploratory analyses examined similarities between the effect of the ERT and Harmer and colleagues' (2009) model of the mechanism of antidepressant drugs. This was investigated by finding whether early changes in cognitive bias, as measured by the difference between participants' first and last observed ERT thresholds, predicted later mood, as measured by the BDI-II and PANAS. Finally, exploratory analyses were undertaken to investigate the acceptability of the training program; all participants were asked to complete an acceptability measure, post-training. Adherence to the protocol and attrition were also analysed.

Method

Design

A randomised controlled study design was used. Participants were randomised to receive the ERT or a sham-ERT, which did not aim to modify participants' interpretations of facial expressions. Participants were blind to whether they were randomised to the control or treatment group. The researcher was blind to the assignment of participants to groups until email reminders to complete further sessions of training were required to be sent out.

Participants

Adults from the general population contacted the researcher between August 2013 and February 2014 to register their interest in the study (N = 627). Due to time constraints no more participants were registered after this date. Opportunistic sampling was employed by advertising in a variety of places including social media, mental health forums, universities, newsletters for NHS

and local council staff, libraries, churches, cafes, pubs, and shops. The screening survey, run through Bristol Online Surveys (BOS) was completed by 428 people. Of these, 124 (83.2% female) met the inclusion criteria and consented to take part in the study. To be eligible, participants had to score 14 or more on the BDI-II and indicate they were over 18, had normal or corrected vision, were not using illicit drugs (other than cannabis), were not receiving psychiatric medication or psychotherapy, did not have a diagnosed psychiatric disorder (other than depression), major illness, head injury or intellectual disability. These eligibility criteria were chosen to match previous low mood research using the ERT (Penton-Voak et al., 2012; Brazil et al., 2012), and other planned trials (Adams, Penton-Voak, Harmer, Holmes, & Munafò, 2013), enabling helpful comparisons between studies. All contact between the researcher and participants was carried out via the internet. Despite an international recruitment strategy, via social media and mental health forums, ninety-six percent of the randomised participants (*n* = 119) stated they lived in the UK; one participant lived in Greece, one in Ireland, another in the United Arab Emirates, and two in the USA. All participants declared they had a good command of the English language. Participants were aged 18 to 71 years (*Mdn* = 41.00). Additional socio-demographic information, e.g., ethnicity, level of education, and social economic status were not collected. Participants were offered a chance to participate in a raffle to win gift vouchers between the value of £10 and £50, as compensation for taking part. The study received ethics approval from the University of Exeter and was conducted according to Good Clinical Practice guidelines.

Materials

CBM emotion recognition task (ERT).

The ERT involved participants being shown sequences of faces, which morphed from happiness to sadness. Variations of this design have been used in similar studies (Brazil et al., 2012; Penton-Voak et al., 2012). Prototypical "happy" and "sad" composite images were generated from 20 individual male faces showing a sad facial expression and the same 20 individuals showing a happy expression using established techniques (Tiddeman, Burt, & Perrett, 2001). Original images came from the Karolinska Directed Emotional Faces (Lundqvist, Flykt, & Öhman., 1998). These prototypical images were used as endpoints to generate a linear morph sequence, consisting of images that changed incrementally from unambiguously "sad" to unambiguously "happy" with emotionally ambiguous images in the middle. Figure 1 shows an example morph sequence with unambiguous sad and happy endpoints and intermediate stimuli.

Initially, the threshold for detecting one emotion over another in an ambiguous expression (i.e., a blend of happiness and sadness) was assessed; individuals then received feedback to modify this threshold (i.e., to favour the identification of happiness over sadness). On each of the five occasions that participants were asked to use the web-based programme for training it consisted of three phases: (1) baseline, (2) training, and (3) test.

The baseline and test phase each consisted of 45 trials in which each stimulus from the morph sequence was presented to the participant three times. Exemplar faces were from a 15-frame continuum, morphing from happy to sad. Participants' judgments as to whether presented faces were happy or sad were

obtained using a two-alternative forced choice procedure. Images were presented one at a time, in random order for 150 ms. Stimuli were preceded by a fixation cross which was presented for a random period ranging from 1500 to 2500 ms. Subsequent to presentation, and to prevent processing of afterimages, a backward mask of noise was presented for 250 ms, followed by a prompt asking the participant to judge whether the face presented had been happy or sad. This remained onscreen until the participant made a response.

Each trial in the training phase was similar to baseline and test phase trials with respect to inter-trial interval and stimulus presentation, but with the addition of feedback subsequent to the participant's response. In the control condition, the sham-ERT was programmed to give feedback based on participants' baseline thresholds. Original thresholds could vary from 1, indicating all faces were being interpreted as sad, to 15, indicating all faces were being interpreted as happy; the mid-point threshold was 8. The average baseline ERT threshold in a laboratory setting has been reported as 6 for people with low mood and 7 for a healthy control group, indicating those with low mood had significantly more negative bias than those without dysphoria (Penton-Voak et al., 2012). When the ERT was presented via smartphones, the average baseline threshold of people with low mood was 8, i.e., the mid-point (Brazil et al., 2012). The sham-ERT classified responses as "correct" if the participant identified images below the original threshold as sad and above it as happy; any deviation from baseline thresholds was classified as "incorrect". Feedback was a message saying "Correct/Incorrect! That face was sad/happy." In the active intervention condition, the ERT also gave feedback based on participants' baseline thresholds; however, the ERT was programmed to shift the "correct" classification two steps towards the sad end of the continuum (also

illustrated in Figure 1). Hence, the two images nearest the baseline threshold, which participants would previously have classified as sad at baseline, were classified as happy when providing feedback. Six training blocks, consisting of randomised sequences of 31 faces, were given to each participant, resulting in 186 training trials in total. In each training block, participants were shown more images close to their baseline thresholds rather than the extreme ends of the continuum.

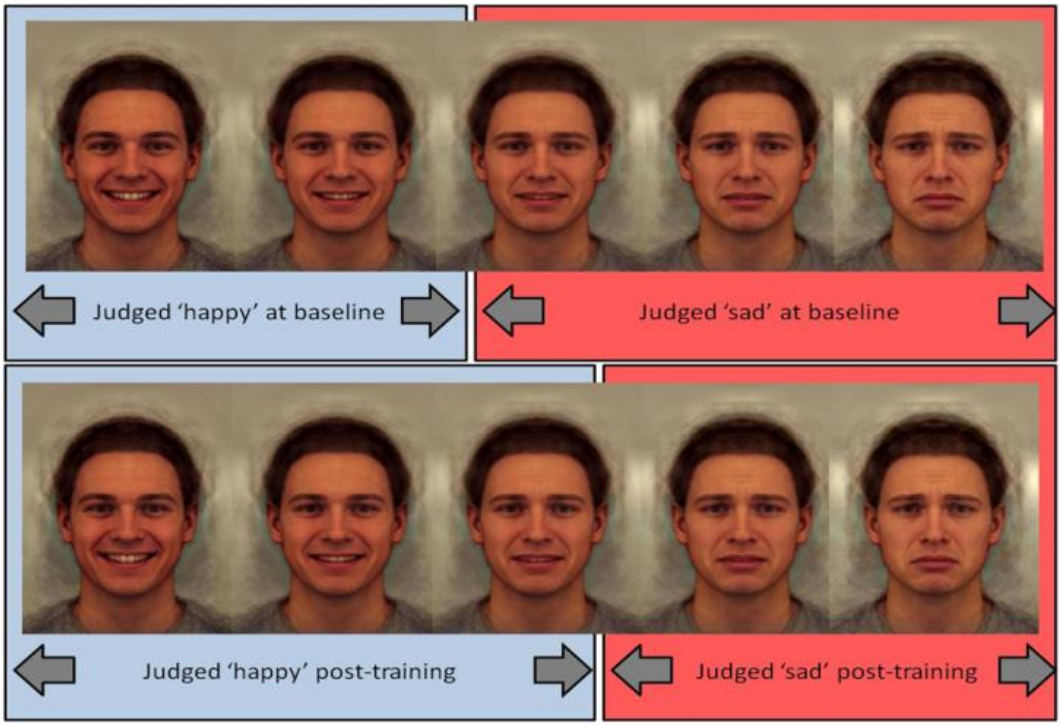

Figure 1. Sample of a face morph sequence: Unambiguous sad and happy endpoints and intermediate stimuli are highlighted; the shift in classification of ambiguous faces as happy rather than sad is also illustrated. From Penton-Voak et al., (2012).

Mood assessments.

Three measures of mood were used: the BDI-II, mood rated over the past week; PANAS, mood rated over the past day; and BAI, mood rated over

the past month. These assessments were used at baseline, and three follow-up times: post-training, 2 weeks, and 6 weeks later. These mood assessments have been used in other studies (Adams et al., 2013; Penton-Voak et al., 2012), which have investigated the ERT with people with dysphoria, enabling results to be more easily compared across studies. The BAI, BDI-II, and PANAS have been found to retain their psychometric properties when used via the internet (Carlbring et al., 2007; Holländare, Andersson, & Engström, 2010; Howell, Rodzon, Kurai, & Sanchez, 2010).

The BDI-II was used to obtain data relating to the severity of participants' depression symptoms. It is a widely used self-report scale (Holländare et al., 2010), composed of 21 questions. Each item yields a score between 0 and 3; total scores range between 0 and 63. Higher scores are associated with higher levels of depressive symptoms. According to the BDI-II manual (Beck et al., 1996), scores of 0 to 13 indicate minimal depression, 14 to 19 mild depression, 20 to 28 moderate depression, and 29 to 63 severe depression. Internal consistency for baseline BDI-II scores of participants in the present study was good ($\alpha = .78$). One week test-retest reliability has been reported to be excellent ($r = .93$; Beck et al., 1996). Content validity has been described as excellent and construct validity is well supported (Dozois & Covin, 2004).

The PANAS obtained data relating to positive and negative affect. Positive affect is the extent to which people feel enthusiastic, alert, active, and appreciative. Negative affect is an orthogonal concept, relating to subjective distress, discontentment, and aversive mood states including anger, guilt, contempt, and fear. Low levels of positive affect indicate sadness and lethargy; whereas, low levels of negative affect indicate serenity and calmness (Watson et al., 1988). The positive subscale is negatively correlated with measures of

depression; the negative subscale positively correlated. Both subscales have explained a significant proportion unique to depression, positive affect significantly more (Crawford & Henry, 2004). The PANAS is a 20-item self-report measure, ten relate to positive mood and ten to negative mood. Participants rated, on a scale from one to five, the extent to which they had experienced certain emotional states over the past day. Total scores for the subscales range from 10 to 50; higher scores are associated with a higher level of affect. Internal consistency for baseline PANAS positive and negative subscale scores of participants in the present study was good (α = .88 for both subscales). Eight-week test-retest reliability, relating to responses for the past week, has revealed a coefficient of .47 for both subscales; good levels of construct validity have also been reported (Watson et al., 1988).

The BAI is a 21-item self-report scale, used to measure severity of anxiety. It was developed to have discriminant validity from depression measures (Beck, Epstein, Brown, & Steer, 1988). Respondents rate on a scale from zero to three the extent to which they have experienced symptoms of anxiety over a given period. Total scores range from 0 to 63; higher scores are associated with higher levels of anxiety. Internal consistency for baseline BAI scores of participants in the present study was good (α = .87). A high level of test-retest reliability has been described (r = .75; Beck & Steer, 1993).

Procedures

Adults were recruited from the general population via advertisements, which directed them to email the researcher. They were given unique usernames to access a screening survey, which was created using BOS. Participants responded to questions relating to consent, inclusion criteria, and demographic data. Baseline mood was assessed at this stage using the BDI-II,

PANAS, and BAI. The researcher checked participants' responses and sent debriefing information to those who did not meet the inclusion criteria. Those who met the criteria were emailed instructions for using the ERT. Each eligible participant was registered onto the program's system; this generated an automatic email, which requested that the program was activated within 7 days. On activation, participants were randomly assigned by the program to the treatment or control group; thus treatment allocation was concealed at this stage. Emails were sent by the researcher to remind participants to activate the program; participants were re-registered if they had not activated the program within the allocated time. Participants were sent a reminder email if they appeared not to be completing the training on 5 days close together, specifically, if there was a gap of 5 days between sessions. Participants were asked to access a post-intervention survey, to complete the BDI-II, PANAS, BAI, and an acceptability measure used by Adams and colleagues (2013; Appendix A), following participants' fifth use of the ERT, or on a day close to when they were predicted to have accessed the fifth session if training was not completed. Emails were sent 2 and 6 weeks later, asking participants to access follow-up surveys consisting of BDI-II, PANAS, and BAI questionnaires. Debriefing information was emailed to each participant following data collection.

Analyses

Sample size and power calculations.

A sample size of $N = 164$ was required to detect, with 80% power at an alpha level of 5%, a decrease in the primary outcome measure, BDI-II, of the magnitude found by Penton-Voak and colleagues (2012). Their study showed the mean of participants' baseline BDI-II scores was 23.05 (s.d. 8.9); following ERT the mean score for the treatment group was 19.3 (s.d. 10.2); on this basis

the effect size was *d* = 0.39. Penton-Voak and colleagues (2012) randomised 80 participants who met similar criteria to the current study after screening 193 people. Based on this information, the current study aimed to screen at least 396 people; 435 people were screened but this only led to 124 participants being randomised. High rates of attrition led to 87 participants' data being analysed at 6 weeks follow-up. This sample size (*N* = 87) provided 80% power, at an alpha level of 5%, to detect a difference of 5.37 points on the BDI-II (assuming s.d. 10) and 3.76 points on the PANAS (assuming s.d. 7).

Linear regression analyses.

Data were analysed using SPSS for Windows, Version 21. Intention-to-treat (ITT) analyses were used, whereby all data from each randomised participant were used. A limitation of the ITT analyses was that only a small proportion of randomised participants completed all five days of the ERT. If participants who do not receive treatment are included in the analysis as participants who did receive treatment, the effect of the treatment is diluted and the analysis is a more conservative estimate, leading to more Type I errors (Gupta, 2011). Therefore, exploratory per-protocol analyses were also carried out. For these linear regressions, only participants who had completed the ERT on at least two occasions (Training group: *N* = 33; Control group: *N* = 34) were included.

A hierarchical regression model was used to examine the primary hypothesis (a) and secondary hypotheses (b) and (c). Methods of data cleaning are described in Appendix B. Appendix C summarises the outcome measures, predictors, and order of entry for each hierarchical regression model. For example, to analyse the primary hypothesis (a), the outcome measure was BDI-II at 6 weeks, the predictors were entered in the following order: baseline BDI-II,

followed by the categorical predictor, i.e., training group (1), control group (0). Exploratory analysis of whether baseline depression was a moderator of BDI-II at 6 weeks followed the above description with an additional predictor, i.e., the product of baseline BDI-II score and group. This third predictor was entered last. Subsequent predictors were included even if zero-order correlations were not significant.

Exploratory analyses were carried out using hierarchical linear regression. These analyses included testing whether participants randomised to the training group had fewer symptoms of anxiety, measured by the BAI, than the control group, and examining the moderating effects of baseline depression and baseline anxiety on how use of the ERT affected BDI-II outcomes. Similarities between the effect of the ERT and Harmer and colleagues' (2009) model of the mechanism of antidepressant drugs were examined by carrying out further exploratory analyses. These investigated whether early changes in cognitive bias, as measured by the difference between participants' first and last observed ERT thresholds, predicted later mood.

ERT threshold manipulation check.

To check hypothesis (d), a graph was used to plot the mean ERT thresholds and 95% confidence intervals for the training and control group at each time-point. As this showed non-overlapping confidence intervals at all time-points other than the first, carrying out linear regression analyses was unnecessary.

Acceptability questionnaire analysis.

The findings from the acceptability questionnaire were summarised in Table 6. Comments made by participants were examined using content analysis

(Berg, 2001). Familiarisation with the data was ensured; this was followed by fitting data to a coding framework as to whether or not participants suggested the task had been or become (a) too long (b) unenjoyable, (c) unengaging, (d) associated with challenges (e) associated with a positive change in mood, and (f) associated with a negative change in mood. The number of participants in each group who had made comments relating to each of these categories was summed. Chi-squared comparisons were made to analyse whether there were differences between responses for the training and control groups.

Categorical data, as to whether participants suggested they would access the program again, would recommend it to a friend, and rated the task instructions as easy to understand, were examined using chi-squared comparisons to compare differences between the control and training group.

Themes relating to specific challenges participants had commented on and ideas for improving the task's accessibility were identified using a thematic analysis, based on a method described by Braun and Clarke (2006). Data familiarisation was ensured, followed by generation of initial codes. Finally, themes relating to challenges and ideas for task improvement were searched for, reviewed, and named.

Results

Demographic Characteristics

Table 1 shows demographic data regarding age, gender, and country of residence of participants who were randomised into the main study. The flow of participants in the study is shown in the CONSORT diagram (Figure 2). Table 2 summarises mood scores and cognitive bias, measured by ERT threshold, at

each time-point for the control and training group; correlations between outcome measures are included in Appendix D.

Table 1

Demographic data of randomised participants

	Training group (N = 66)	*Control group (N = 58)*
Age	*19 to 71 (median = 42.50)*	*18 to 69 (median = 39.00)*
Female	*79% (n = 52)*	*88% (n = 51)*
Non-UK residents	*3% (Greece, n = 1; United Arab Emirates, n = 1)*	*5% (Ireland, n = 1; United States of America, n = 2)*

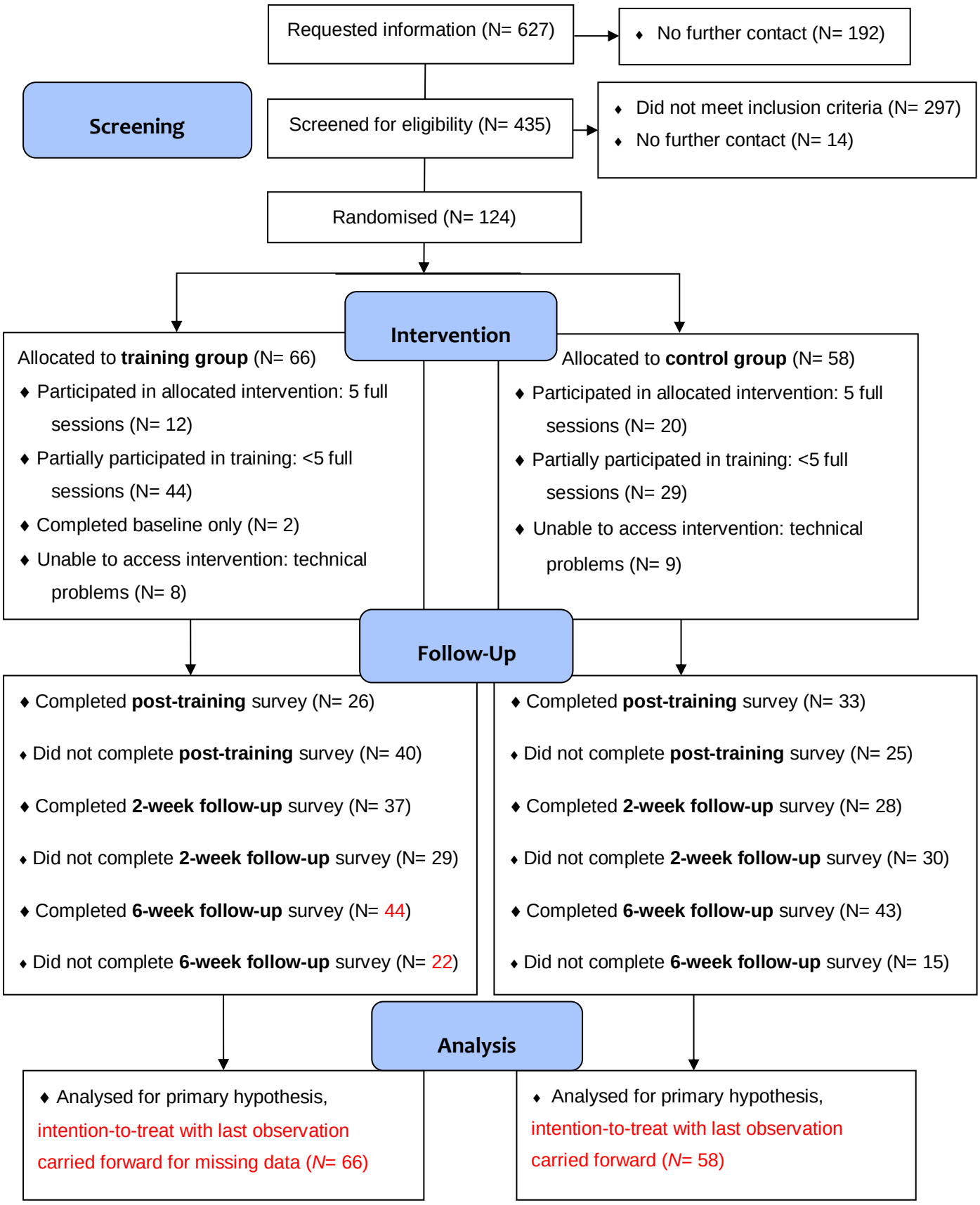

Figure 2. CONSORT Flow Diagram.

Table 2

Outcome measures at baseline and follow-up points

	Pre-intervention M (SD)	Post-intervention M (SD)	2-week follow-up M (SD)	6-week follow-up M (SD)
BDI-II				
Training	21.74 (6.21) *N* = 66	18.50 (8.19) *N* = 26	21.16 (10.10) *N* = 37	18.70 (8.83) *N* = 44
Control	23.21 (8.22) *N* = 58	21.18 (10.30) *N* = 33	19.11 (11.09) *N* = 28	18.33 (9.84) *N* = 43
PANAS: Positive				
Training	19.06 (7.03)	22.54 (6.72)	19.78 (8.80)	21.00 (8.47)
Control	19.78 (6.18)	20.00 (7.27)	20.36 (7.27)	18.42 (7.57)
PANAS: Negative				
Training	18.50 (7.27)	18.38 (6.45)	19.95 (8.74)	17.43 (7.85)
Control	19.00 (7.29)	17.15 (6.46)	16.82 (7.51)	17.26 (6.70)
BAI:				
Training	18.23 (9.05)	15.85 (11.63)	18.32 (11.75)	16.09 (12.49)
Control	16.72 (10.70)	14.03 (12.19)	12.54 (11.16)	14.56 (10.56)
ERT: Threshold				
Training	8.48 (1.25) *N* = 58	11.50 (2.78) *N* = 12		
Control	8.24 (1.45) *N* = 49	8.63 (2.06) *N* = 19		

Results of Hypothesis Testing: Effect of Training on Mood

Table 3 summarises whether the group to which participants were assigned (i.e., training or control group) predicted mood outcomes. There were missing data at each time-point, this was not statistically different between groups. Post-intervention there were missing data for 60.6% of the intervention group and 43.1% of the control group (Chi-square with Yates' correction = 3.12, p = .077); at 2-weeks follow-up 43.9% of the intervention group and 51.7% of the control group (Chi-square with Yates' correction = 0.47, p = .493); at 6-weeks follow-up 33.3% for the intervention group and 25.9% of the control group (Chi-square with Yates' correction = 0.51, p = .477). Further exploratory analyses were carried out to investigate whether data were missing systematically, i.e., differences between characteristics of participants who did and did not complete the questionnaires (Appendix E). One significant difference was found for the training group: participants who completed the questionnaires at 6-weeks follow-up were significantly older than those who did not complete the measures. Significant differences were found for the control group: participants who completed the questionnaires post-intervention and at 2-weeks follow-up had significantly lower baseline ERT thresholds indicating a greater negative bias, than those who did not complete the training program, than those who did not complete the outcome measures.

There is no consensus for how to treat missing data in ITT analyses and different definitions of ITT exist (Alshurafa et al., 2012): (a) ITT is violated if any missing data occur; (b) missing data should be handled in a specified manner, e.g., last observation carried forward; (c) ITT refers to analysis of all data from randomised participants irrespective of how missing data are handled. Two methods of ITT analysis were utilised, which match definitions (b) last

observation carried forward, i.e., all 124 randomised participants were analysed, and (c) no imputation of data, i.e., 87 participants were analysed at 6 weeks follow-up (i.e., all participants who completed the questionnaire at that time-point), 65 at 2 weeks follow-up, and 59 post-intervention. Data were not imputed for the per-protocol analyses.

Similar results for both ITT analyses and per-protocol analyses were found. There was no statistical evidence to support the primary hypothesis (a) that BDI-II scores at 6 weeks follow-up were significantly lower for participants randomised to the intervention group than the control group. Furthermore, there was no statistical evidence supporting the secondary hypotheses (b) that BDI-II scores at 2 weeks follow-up and post-intervention were significantly lower for participants randomised to the intervention group than the control group. Finally, there was no evidence supporting the secondary hypotheses (c) that at 2 and 6 weeks follow-up, and post-intervention, scores on the PANAS negative subscale were significantly lower but significantly higher on the positive subscale for participants randomised to the intervention group than the control group. Contrary to hypothesis (c), there was evidence that scores on the PANAS negative subscale were significantly higher at 2 weeks follow-up for the intervention group than the control group.

Table 3

Linear regression analysis of whether training group vs.control group predicted mood outcome.

		Intention to treat analysis: LOCF (N=124)			*Analysis of all participants: no imputation for missing data*[1]			*Per protocol analysis*[2]		
		B	*95% CI*	*p*	*B*	*95% CI*	*p*	*B*	*95% CI*	*p*
BDI-II	*6-week follow-up*	*1.33*	*-1.14 to 3.81*	*.288*	*1.95*	*-1.38 to 5.28*	*.247*	*2.49*	-1.53 to 6.50	*.220*
	2-week follow-up	*2.02*	*-0.37 to 4.40*	*.098*	*4.15*	*-0.07 to 8.37*	*.054*	*4.09*	-1.05 to 9.23	*.116*
	Post-intervention	*0.32*	*-1.40 to 2.04*	*.713*	*-0.48*	*-3.97 to 3.01*	*.783*	*-1.20*	-5.07 to 2.67	*.537*
PANAS	*6-week follow-up*	*1.86*	*-0.59 to 4.31*	*.135*	*2.32*	*-0.85 to 5.48*	*.149*	*2.75*	-1.36 to 6.86	*.186*
positive	*2-week follow-up*	*-0.24*	*-2.32 to 1.84*	*.818*	*-0.40*	*-4.08 to 3.28*	*.828*	*-1.03*	-5.30 to 3.24	*.630*
	Post-intervention	*0.56*	*-1.19 to 2.30*	*.528*	*2.38*	*-1.02 to 5.77*	*.166*	*1.49*	-2.27 to 5.25	*.429*
PANAS	*6-week follow-up*	*0.55*	*-1.24 to 2.33*	*.547*	*0.48*	*-1.83 to 2.73*	*.695*	*0.13*	-2.56 to 2.81	*.925*
negative	***2-week follow-up***	***2.89***	***1.08 to 4.70***	***.002***	***3.81***	***0.65 to 6.98***	***.019***	***4.15***	**0.40 to 7.89**	***.031***
	Post-intervention	*1.04*	*-0.50 to 2.57*	*.183*	*1.64*	*-1.29 to 4.57*	*.267*	*0.88*	-2.46 to 4.21	*.598*
BAI	*6-week follow-up*	*-0.54*	*-3.23 to 2.14*	*.689*	*-0.37*	*-4.03 to 3.29*	*.840*	*-0.49*	-4.81 to 3.83	*.822*
	2-week follow-up	*0.67*	*-1.80 to 3.14*	*.594*	*2.17*	*-2.54 to 6.87*	*.361*	*1.37*	-4.10 to 6.85	*.616*
	Post-intervention	*.56*	*-1.20 to 2.32*	*.531*	*0.42*	*-3.31 to 4.14*	*.824*	*-0.17*	-4.40 to 4.06	*.936*

Note. B = adjusted mean difference. A positive coefficient indicates a higher outcome score for the training group than the control group. LOCF = last observation carried forward. Bold font indicates a significant effect at alpha level .05. [1] = (*N* = 87) at 6 weeks; (*N* = 65) at 2 weeks; (*N* = 59) post-intervention. [2] = (*N* =) at 6 weeks; (*N* =) at 2 weeks; (*N* =) post-intervention.

Secondary Hypothesis (d): Manipulation Check

Figure 3 illustrates the mean thresholds before and after training sessions on each of the five days, for the training and control groups; the non-overlapping 95 percent confidence intervals demonstrate significant differences between groups (data presented are based on pairwise analysis; a listwise analysis showed similar results. More detailed information is included in Appendix E). The training group had significantly higher ERT thresholds than the control group, indicating a more positive bias, at each time-point, except baseline. The mean threshold at baseline was 8.48 (s.d. 1.25; *n* = 58) for the training group and 8.24 (s.d. 1.45; *n* = 49) for the control group, indicating a relatively positive bias compared with participants in the study by Penton-Voak et al., (2012) and a similar baseline threshold to participants in the study by Brazil and colleagues (2012). Figure 3 shows the control group's mean threshold remained stable across time whereas the training group's increased over two sessions (indicating increased positive bias) then plateaued. The mean threshold after the fifth session was 11.50 (s.d. 2.78; *n* = 12) for the training group and 8.63 (s.d. 2.06; *n* = 19) for the control group. (21 participants from the control group completed the fifth day of training but two of their final thresholds were not recorded due to technical issues).

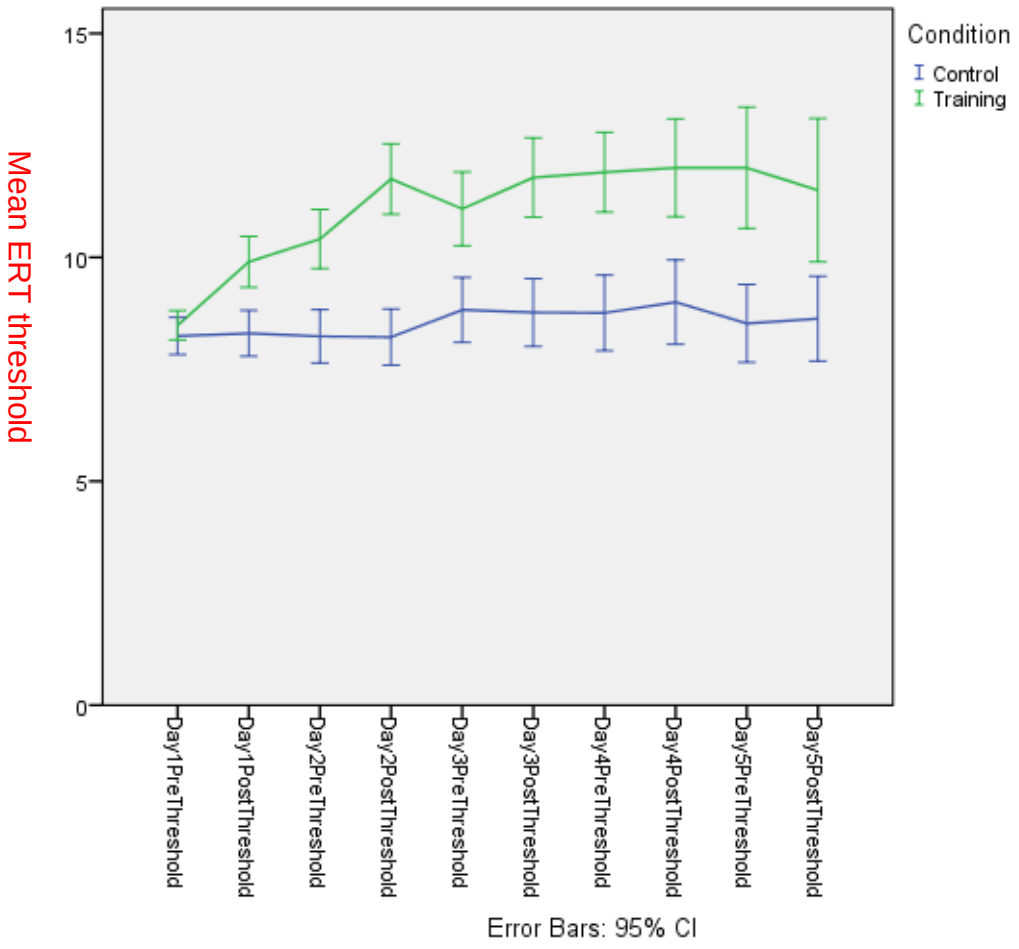

Figure 3. Mean thresholds before and after training sessions on each of the five days, for the training and control groups.

Use of the ERT

Table 4 shows participants from the control group completed significantly more sessions than the training group. Significantly more participants from the control group completed five sessions, as per-protocol. There was no significant difference between how many participants from each group were initially able to access the ERT program.

Table 4

Use of the ERT

	Training group (N = 66)	Control group (N = 58)	
Baseline ERT data obtained	88% (N = 58)^	84% (N = 49)^	Yates' Chi-square = 0.08, *p* = .775
Completed 5 sessions: Per protocol	**N = 12**	**N = 21**	**Yates' Chi-square = 5.12, *p* = .024**
Number of sessions accessed	**Median = 2**	**Median = 4**	**Two-tailed independent-samples Mann Whitney U test, *p* = .048**

Note. ^ = Attrition due to technical problems with ERT program. Bold font indicates significant effect at alpha level .05.

Exploratory analyses were carried out to investigate whether there were differences between participants who completed five sessions of the training program compared with participants who did not complete the training (Table 5; non-parametric and parametric tests produced the same results as far as statistical significance was concerned, therefore, only parametric test statistics have been reported). One significant difference was found for the training group: participants who completed the training program had significantly higher baseline scores on the PANAS positive subscale than those who did not complete the training program. This might reflect that to engage with the ERT a certain baseline level of motivation and enthusiasm is required. One significant difference was found for the control group: participants who completed the training program had significantly lower baseline ERT thresholds indicating a greater negative bias, than those who did not complete the training program. When interpreting these levels of statistical significance it should be taken into account that a number of similar exploratory analyses were undertaken: the

chance of obtaining familywise errors was increased and Type 1 errors could have occurred.

Acceptability Questionnaire

Thirty-nine percent (*n* = 26) of the training group and fifty-seven percent (*n* = 33) of the control group completed the acceptability questionnaire (Table 6). Three of these participants indicated they had not been able to access the program (one training group and two control group participants).

Categorical data.

Most respondents agreed the task instructions were easy to follow. Around half of each group indicated they would use the task again and recommend it to a friend.

Content and thematic analysis.

Approximately seventy percent suggested the task was too long. Some participants suggested the length caused difficulty in concentration, and exacerbation of physical pain including repetitive strain injury, eyestrain, and headache. More than half the respondents from each group indicated they had not found the task enjoyable or engaging or had experienced a decrease in enjoyment and engagement through the task. Some participants suggested enjoyment and engagement could be increased by using expressions from many different people's faces as stimuli. More than half the respondents from each group also reported experiencing challenging issues in completing the task. This included maintaining concentration, categorising ambiguous expressions, processing the images quickly, remembering and finding time to do the task, managing environmental distractions, and coping with feedback suggesting their judgment of emotions was incorrect. One fifth of respondents

from the training group noticed a negative change in mood, including anger, irritability, sadness, boredom, and guilt; whereas one fifth noticed a positive change, including feeling more relaxed, sociable, happy, and less grumpy. A smaller proportion of respondents from the control group noticed a change in mood. There was no significant difference between the responses of each group (when compared using a Yates' chi-square analysis).

Moderation of the Effects of Training on Depressive Symptoms

Linear regression analyses of moderating effects of baseline measures are summarised in Table 7, without imputing data, and Table 8, imputing the last observation carried forward for missing data. When data were not imputed, baseline anxiety was a significant moderator of the effect of training on BDI-II score at 6 weeks follow-up; participants in the training group with lower baseline levels of anxiety experienced an improved effect from training on lowering BDI-II score, relative to the control group. This effect was not significant in the more conservative analysis when data were imputed. At the 5% level of probability, no other analyses showed evidence of a significant moderator.

Table 5

A comparison of characteristics of participants who did and did not complete the training program

Baseline measure		Completed 5 sessions of training (Training: *N* = 12; Control: *N* = 21)	Did not complete training (Training: *N* = 46; Control *N* = 28)		
		Mean (SD)	Mean (SD)	*F*	*p*
PANAS: Positive	**Training**	**24.08 (7.70)**	**18.15 (6.65)**	**4.26**	**.018**
	Control	18.05 (4.33)	20.21 (6.67)	1.78	.179
PANAS: Negative	Training	19.08 (6.79)	18.04 (7.42)	0.35	.703
	Control	18.00 (6.84)	19.54 (8.12)	0.30	.739
BDI-II	Training	20.58 (9.18)	21.83 (5.32)	0.37	.692
	Control	23.67 (7.00)	23.36 (9.19)	0.19	.827
BAI	Training	18.00 (7.63)	18.57 (9.36)	0.16	.855
	Control	17.19 (11.02)	15.50 (11.07)	0.49	.618
ERT Threshold	Training	8.75 (1.22)	8.41 (1.26)	0.69	.409
	Control	**7.76 (0.94)**	**8.61 (1.66)**	**4.36**	**.042**
Age	Training	42.42 (14.99)	41.33 (13.14)	0.08	.922
	Control	41.14 (14.16)	39.64 (12.01)	0.43	.655

Note. Bold font indicates significant effect at alpha level .05

Table 6

Acceptability of the task

	Training *N*=25	Control *N*=31	Yates' Chi-square
Found instructions easy	24 (96%)	29 (94%)	0.16, *p* = .685
Would use the task again	15 (60%)	15 (48%)	0.36, *p* = .551
Would recommend to a friend	10 (40%)	15 (48%)	0.13, *p* = .721
Task too long	17 (68%)	22 (71%)	0.06, *p* = .810
Task not engaging/ became less engaging	13 (52%)	18 (58%)	0.03, *p* = .854
Challenging issues	13 (52%)	18 (58%)	0.03, *p* = .854
Task not enjoyable/ became less enjoyable	16 (64%)	17 (55%)	0.18, *p* = .675
Positive mood change	5 (20%)	5 (16%)	<0.01, *p* = .980
Negative mood change	5 (20%)	4 (13%)	0.13, *p* = .724

Table 7

Linear regression of moderating effects of baseline measures

		Moderation by baseline measure											
		Depression			PANAS positive			PANAS negative			BAI		
BDI-II		B	95% CI	p	B	95% CI	p	B	95% CI	p	B	95% CI	p
	6-week follow-up	-0.10	-0.54 to 0.35	.671	-0.20	-0.73 to 0.33	.462	-0.20	-0.73 to 0.33	.462	**0.33**	**0.01 to 0.66**	**.046**
	2-week follow-up	-0.34	-0.92 to 0.24	.247	-0.04	-0.72 to 0.64	.905	-0.01	-0.57 to 0.58	.980	0.23	-0.29 to 0.74	.389
	Post-intervention	-0.13	-0.61 to 0.35	.581	-0.43	-1.00 to 0.15	.141	-0.11	-0.61 to 0.39	.652	0.12	-0.25 to 0.50	.514

Note. Bold font indicates significant effect at alpha level .05. (*N* = 87) at 6 weeks, (*N* = 65) at 2 weeks; (*N* = 59) post-intervention.

Table 8

Linear regression of moderating effects of baseline measures: LOCF

		Moderation by baseline measure (*N* = 124)											
		Depression			PANAS positive			PANAS negative			BAI		
BDI-II		B	95% CI	p	B	95% CI	p	B	95% CI	p	B	95% CI	p
	6-week follow-up	-0.04	-0.39 to 0.31	.818	-0.10	-0.49 to 0.29	.611	0.20	-0.15 to 0.54	.258	.20	-0.05 to 0.45	.121
	2-week follow-up	-0.17	-0.50 to 0.17	.336	-0.02	-0.40 to 0.35	.909	-0.04	-0.37 to 0.29	.815	0.11	-0.14 to 0.35	.384
	Post-intervention	-0.03	-0.27 to 0.22	.822	-0.22	-0.49 to 0.05	.109	-0.09	-0.33 to 0.15	.447	0.04	-0.13 to 0.22	.632

Note. Bold font indicates significant effect at alpha level .05

Effects of Cognitive Bias Change on Mood

Linear regression analyses aimed at finding out whether changes in cognitive bias mediated changes in mood are summarised in Table 9; analysis using last observation carried forward for missing data was not appropriate. Change in threshold from baseline to last observation was a positive predictor of PANAS positive subscale scores both at 6 weeks follow-up ($p = .013$) and post-training ($p < .01$) at the 5% level of probability, and at 2 weeks follow-up at trend level ($p = .055$). No other analyses showed evidence of participants' final observed cognitive bias threshold predicting any mood outcome measure

Table 9

Linear regression analysis of whether early changes in cognitive bias predicted mood

		B	95% CI	p
BDI-II				
	6-week follow-up	-0.50	-1.24 to 0.24	.179
	2-week follow-up	-0.32	-1.22 to 0.59	.486
	Post-intervention	-0.46	-1.18 to 0.26	.202
PANAS positive				
	6-week follow-up	**0.95**	**0.21 to 1.70**	**.013**
	2-week follow-up	0.75	-0.02 to 1.52	.055
	Post-intervention	**1.29**	**0.64 to 1.93**	**<.001**
PANAS negative				
	6-week follow-up	-0.06	-0.58 to 0.45	.803
	2-week follow-up	0.27	-0.40 to 0.95	.425
	Post-intervention	0.27	-0.38 to 0.92	.404

Note. Bold font indicates significant effect at alpha level .05

Discussion

The present study is the first to evaluate, using an RCT, the effectiveness of a web-based version of the ERT in modifying negative biases for interpreting ambiguous facial expressions and reducing depressive symptoms in a sample of people with low mood. The study aimed to investigate whether this intervention showed signs of therapeutic benefit on participants with low mood as measured by the BDI-II, in addition to a shift in bias for identifying ambiguous faces as happy rather than sad, compared with a control group assigned to a sham-ERT.

Effects of ERT on Mood

There was no statistical evidence to support the primary hypothesis (a) that participants who accessed the ERT had fewer symptoms of depression as measured by the BDI-II, at 6 weeks follow-up, than those randomised to the control group. Furthermore, there was no evidence to support the secondary hypothesis (b) that participants in the training condition had fewer symptoms of depression at two other time-points, post-intervention and 2 weeks follow-up, as measured by the BDI-II, than the control group. There was also no evidence to support hypothesis (c) that participants in the training group had less negative mood, and more positive mood as measured by the PANAS, at three time-points, post-intervention, 2 and 6 weeks follow-up, than the control group. Contrary to hypothesis (c) there was evidence that at 2 weeks follow-up the training group had significantly more negative mood as measured by the PANAS, than the control group.

There are many possible explanations for why hypotheses (a) to (c) were not supported, i.e., no reduction in symptoms of depressive mood was found. The most parsimonious explanation is that the hypotheses were incorrect; the

ERT might not have a beneficial therapeutic effect on reducing depression. Another simple explanation is that the study design was unable to capture the effects; this is expanded upon in the section regarding study limitations.

The null findings partially fit with results from the study by Penton-Voak and colleagues (2012) who found no evidence for the ERT reducing depressive symptoms, measured by the BDI-II and PANAS negative subscale. They did find a significant improvement in positive affect measured by the PANAS, however, which was not replicated in the present study. The explanation that non-significant results in the current study relate to incorrect hypotheses fits with findings from a meta-analysis in which the authors concluded there was no evidence of CBM interventions modifying depression, based on a non-significant, small, unreliable effect (Hallion & Ruscio, 2011). The explanation that the hypotheses were incorrect does not fit with a more recent meta-analysis that found evidence that CBM-I paradigms decreased negative mood (Menne-Lothmann et al., 2014). The meta-analysis found bigger effects when CBM-I paradigms encouraged participants to use mental imagery, which was not a feature of the present study. The hypothesis that a CBM paradigm, such as the ERT, might have therapeutic benefit for people with depression comes from theoretical models, linking negative biases with the aetiology and maintenance of the condition (e.g., Beck et al., 1979), and evidence that modifying negative biases might underpin mechanisms of therapeutic interventions such as CBT and antidepressants (Harmer et al., 2009). This could be described as a compelling rationale for the development of CBM interventions to treat depression. Indeed a CBM-I paradigm might prove effective in reducing symptoms of depression; however an effective target might be a different bias than interpretation of ambiguous facial expressions as targeted by the ERT.

CBM-I studies that have shown evidence of a reduction in symptoms of depression have used different paradigms, pairing ambiguous situations with a positive bias, and encouraging participants to imagine the situations as if they were directly involved, i.e., from a field perspective (e.g., Blackwell & Holmes, 2010; Lang et al., 2012). The focus on processing situations from a field perspective fits with models of depression specifically emphasising a central role for negative self-referent schemata and information processing biases (Hughes et al., 2007), rather than generalised negative cognitive biases.

The only significant finding regarding an effect of the ERT on mood was evidence that the training group had significantly lower mood at 2 weeks follow-up. This could reflect a real finding or represent a Type II error, particularly as many analyses were carried out increasing this possibility. It is important to consider this might be a real effect indicating some participants who accessed the ERT did become lower in mood, highlighting a potential risk of psychological harm. Indeed, by supporting people with depression to have a more positive interpretation of other people's ambiguous expressions, the ERT might actually reinforce negative self-referent interpretations, contributing to the vicious spiral of depression, e.g., everyone seems happy except me.

Exploratory analyses showed that despite no significant improvement in positive affect being found in the present study, increases in positive affect were significantly mediated by increased positive bias, measured by change in ERT threshold. Moderation analyses suggested people with lower levels of baseline anxiety experienced an improved effect from training on lowering BDI-II score at 6-week follow-up. Subjective reports of changes in mood captured by the acceptability measure included negative affect, anger, irritability, sadness,

boredom, and guilt, and positive affect, feeling more relaxed, sociable, happy, and less grumpy.

Effects of ERT on Interpretation Bias

The manipulation check showed significant evidence that the intervention was successful in modifying participants' responses to the ERT, giving support for hypothesis (d) that participants in the training group showed a more positive bias for interpreting ambiguous facial expressions as happy than sad, as measured by the thresholds obtained at time-points after baseline, compared to the control group. This finding was expected in the context of previous studies that had informed the current study design (Brazil et al., 2012; Penton-Voak et al., 2012). Both previous studies investigated use of the ERT with people with dysphoria and found significant increases in positive bias as measured by change in ERT threshold, in laboratory conditions and when delivered via smartphones.

Various explanations could explain the underlying mechanism by which the ERT has consistently been found to modify responses of people with low mood. These findings could indicate an increase in positive interpretation bias that had generalised to a range of ambiguous situations, or the findings might relate to a change in bias specific to processing ambiguous facial expressions. Indeed the stimuli used as the outcome measure were identical to the stimuli used in the training task; therefore, the measured change in positive bias might not transfer to different faces. Furthermore, the change in threshold might indicate a learned response, i.e., participants learned the "right" answer, rather than a change in appraisal of whether ambiguous expressions were happy or sad.

Acceptability of the ERT

The overall impression based on adherence to the intervention and comments obtained from the acceptability questionnaire was that many participants found the ERT had a low level of acceptability. Adherence to the intervention was significantly lower for the training group than the control group. The only difference between training and control conditions was the feedback that encouraged participants in the training group to modify their responses. Arguably, this has some similarities to a CBT therapist challenging a client's maladaptive thoughts, which if done in a persistent, non-empathic manner can be experienced as invalidating, perpetuating depressive schemas of feeling worthless (Katzow & Safran, 2007). Comments obtained on the acceptability questionnaire included many possible reasons why participants found difficulties in engaging with the ERT for five sessions. Issues contributing to low acceptability included the length of the task, which was programmed to be a maximum of 30 minutes and deemed to be too long, causing concentration difficulties and exacerbation of physical pain. The task was viewed as unengaging and unenjoyable particularly as stimuli were all very similar, one man's face showing 15 subtly different expressions. Difficulties categorising the ambiguous expressions, having to process the images quickly, remembering and finding time to do the task, managing environmental distractions, and coping with feedback suggesting their judgment of emotions was incorrect were also described as challenges to acceptability of the intervention.

Study Limitations

Limitations of the study were related to the sample, intervention, outcome measures, and design.

Sample.

There are many limitations related to the study sample. A high rate of attrition and small sample size caused a low level of power to detect small effects on symptoms of depression. Generalizability of the findings is restricted due to the sample being an analogue rather than clinical sample, most participants being female and from the UK, and strict inclusion criteria meaning no participant was receiving professional treatment for their dysphoria.

The high rate of attrition was surprising when compared to low rates in previous studies using the ERT (Brazil et al., 2012; Penton-Voak et al., 2012). Factors influencing attrition could have been lack of motivation, a symptom of depression; people who lacked motivation would have been screened out by the previous studies where there was a requirement to attend the laboratory at the recruitment stage. Indeed the current study found participants who completed the ERT per-protocol had higher baseline levels of positive affect, associated with motivation, than those who did not adhere to the intervention. A further explanation for the attrition is that participants had an unrealistic expectation of what the research would entail, particularly as recruitment was opportunistic and not focused on people who had contact with a university. A lack of face-to-face contact and not enough email reminders is likely to have contributed to study dropout. Indeed, at the 6 week follow-up the researcher had time to send multiple reminders leading to more people completing questionnaires than at other follow-up time-points.

There are many differences between the samples of participants recruited by the current web-based study and the laboratory-based intervention implemented by Penton-Voak and colleagues (2012), which could have

contributed to differences in the studies' results: socio-demographic differences, face-to-face factors, and level of symptoms.

Intervention.

Limitations of the intervention related to being unable to check participants' fidelity to the instructions; they might have asked someone else to do the task or not properly concentrated. Participants were only required to undertake five training sessions; however most did not complete even these. A ceiling effect occurred in shifting participants' thresholds and there was a relatively high baseline threshold compared to the study by Penton-Voak and colleagues (2012).

There were limited stimuli: only one face, which was male, was used. It is unclear what effect the exclusive use of white, male stimuli might have had on a sample of predominantly female participants and people from different ethnic backgrounds. Indeed results from an antidepressant study investigating male participants in Brazil found different interpretation responses to male and female stimuli (Alves-Neto, Guapo, Graeff, Deakin, & Del-Ben, 2010). Use of both male and female stimuli is therefore important for future studies.

Presentation of stimuli in the current study was brief, 150ms exposure, which may be more beneficial for targeting unconscious information-processing biases which occur in anxiety than for targeting the more conscious biases which occur in depression (Teachman, Joormann, Steinman, Gotlib, 2012).

Outcomes.

Changes to specific symptoms related to low mood might not have been detected by the measures of mood, BDI-II and PANAS. The BDI-II might not capture symptom change relating to modifications of emotion recognition bias, a

measure for capturing changes in interpersonal relationships might be more appropriate. Indeed Adams and colleagues (2013) have proposed to measure change in number of close friends of participants in their future trial investigating the effect of ERT on participants with low mood. The BDI-II and PANAS might not have been sensitive in detecting small reductions of dysphoria. No independent measure of interpretation bias was used to check that the ERT had an effect that could be generalized beyond the training stimuli. Participants' ERT thresholds were not measured at 2 and 6 week follow-up time-points to check durability of the effect.

Study design.

Participants' group allocation was revealed to the researcher during the study, leading to a risk that groups could have received differing interaction with the researcher via email. Participants could have guessed the hypothesis which could have influenced their responses.

Study Strengths

The current study had many strengths. It was an RCT that was carefully designed and conducted, and included two follow-up assessments. The study has been reported with the rigour outlined in the CONSORT checklist (Schulz, Altman, & Moher, 2010), ensuring transparency and replicability of the study. The control condition was very similar to the intervention, reducing the influence of confounding factors. The researcher carried out the study with a high regard for ethical issues, ensuring that participants were treated with respect and dignity. No face-to-face contact occurred throughout the study, enabling *in vivo* assessment of a treatment with potential to be delivered without face-to-face contact. The sample was an analogue group of people with low mood from a wide age range, more similar to a clinical population of people with depression

than a sample of young, healthy students. The design was based on a strong theoretical premise and laboratory evidence. A gold-standard self-report depression measure was used to assess change in mood, the BDI-II. Acceptability of the intervention was measured to inform future developments. Mediation analysis was carried out to investigate the proposed underlying mechanism, that changes in bias affect changes in mood.

Clinical, Research, and Theoretical Implications

This study contributes to the growing field of CBM, specifically the development of the ERT as a potential intervention for depression. No evidence was found for the ERT having a beneficial effect on mood. The mechanism by which CBM interventions are proposed to treat depression is in reducing negative bias to mediate changes in mood. Mediation analysis in the current study gave some support for this model showing change in bias was positively related to change in affect, measured by the PANAS positive subscale; caution should be taken in interpreting these analyses because they were exploratory and do not necessarily indicate causation. Harmer and colleagues (2009) suggested a subtly different model where more positive emotion recognition biases, lead to better social interactions, followed by improvements in mood, creating a virtuous cycle; this model is being investigated in a current ERT study (Adams et al., 2013).

Future studies using web-based designs with strict participant exclusion criteria should be aware of the large amount of time and resources required by researchers for the recruitment and screening of participants and prevention of attrition. Automation of further aspects of this study, such as daily emailed reminders to complete the ERT, could have encouraged greater participant

adherence to the intervention and enabled participant allocation to groups to have remained concealed until after all data had been collected.

Exploratory analyses in the present study suggested two subgroups might have benefited more than other participants did from the ERT. Those with higher baseline levels of positive affect were more able to adhere to the intervention; those with lower baseline anxiety were more likely to obtain lower BDI-II scores at 6 weeks follow-up. Future CBM research, including studies investigating the ERT, should consider which subgroups of people might benefit more from interventions.

Problems with acceptability of the ERT were highlighted in the current study. These issues could be helpfully considered when developing and modifying further interventions. There is a need to ensure tasks are engaging, do not cause harm, deterioration in mood or excessive discomfort. The most meaningful outcome measure could be reconsidered, perhaps the ERT might affect specific symptoms of depression. For example, Blackwell and colleagues (2014) did not find a CBM-I intervention improved symptoms of depression measured by the BDI-II but did find an improvement in symptoms of anhedonia measured by a specific selection of BDI-II items.

The ERT intervention should be modified according to the feedback gathered by the acceptability measure and the study design refined before carrying out further studies. Future ERT research should investigate the generalizability of change in bias from specific ambiguous faces to other stimuli and tasks measuring interpretation bias. Evidence that change in bias related to the ERT is generalizable in addition to a replicable finding of reduction in

negative mood or improvement of positive affect should be found before investigating the use of the ERT on a clinical population.

Conclusion

The current study used a randomised controlled trial to investigate the effect of a web-based version of the ERT, a CBM-I intervention, on reducing symptoms of depression in people with low mood. Evidence of a more positive bias was found but no improvement in mood, compared to a control group. Future developments of the ERT should modify the intervention to increase acceptability, ensure generalizability of increased positive bias to different stimuli, and find evidence for consistent reductions in symptoms of depression before investigating its efficacy with a clinical population.

References

Adams, S., Penton-Voak, I. S., Harmer, C. J., Holmes, E. A., & Munafò, M. R. (2013). Effects of emotion recognition training on mood among individuals with high levels of depressive symptoms: Study protocol for a randomised controlled trial. *Trials, 14,* 161.

Alloy, L. B., & Abramson, L. Y. (1988). Depressive realism: Four theoretical perspectives. In L. B. Alloy (Ed.), *Cognitive processes in depression.* New York: Guilford.

Alshurafa, M., Briel, M., Akl, E. A., Haines, T., Moayyedi, P., Gentles, S. J.,…Guyatt, G. H. (2012). Inconsistent definitions for intention-to-treat in relation to missing outcome data: Systematic review of the methods literature. *PLoS ONE, 7*(11), e49163.

Alves-Neto, W. C., Guapo, V. G., Graeff, F. G., Deakin, J. F. W., & Del-Ben, C. M. (2010). Effect of escitalopram on the processing of emotional faces, *Brazilian Journal of Medical and Biological Research, 43*(3), 285-289.

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

Andersson, G., Bergström, J., Holländare, F., Carlbring, P., Kaldo, V., & Ekselius, L. (2005). Internet-based self-help for depression: Randomised controlled trial. *British Journal of Psychiatry, 187,* 456-461.

Baert, S., De Raedt, R., Schacht, R., & Koster, E. H. (2010). Attentional bias training in depression: therapeutic effects depend on depression severity. *Journal of Behavioural Therapy and Experimental Psychiatry, 41*(3), 265-274.

Beck, A. T. (1976). *Cognitive therapy and the emotional disorders.* New York: Meridian.

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology, 56*(6), 893-897.

Beck, A.T., Rush, A.J., Shaw, B.F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford Press.

Beck, A. T. & Steer, R. A. (1993). *Beck Anxiety Inventory Manual.* San Antonio, TX: Harcourt Brace and Company.

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

Berg, B.L. (2001). *Qualitative research methods for the social sciences.* Boston: Allyn & Bacon.

Blackwell, S. E., Browning, M., Mathews, A., Pictet, A., Welch, J., Davies, J.,…Holmes, E. A. (2014). Positive imagery-based cognitive bias modification as a web-based treatment tool for depressed adults: A randomised controlled trial. Clinical Psychological Science (in press).

Blackwell, S. E., & Holmes, E. A. (2010). Modifying interpretation and imagination in clinical depression: A single case series using cognitive bias modification. *Applied Cognitive Psychology, 24,* 338-350.

Bourke, C., Douglas, K., & Porter, R. (2010). Processing of facial emotion expression in major depression: A review. *Australian and New Zealand Journal of Psychiatry, 44*(8), 681-696.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101.

Brazil, M., Munafo, M. R., & Penton-Voak, I. S. (2012). *Effectiveness of smartphone-delivered emotion recognition training on emotion perception.* Unpublished manuscript, School of Experimental Psychology, University of Bristol, UK.

Carlbring, P, Brunt, S., Bohman, S., Austin, D., Richards, J., Ost, L-G., & Andersson, G. (2007). Internet vs. paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. *Computers in Human Behavior, 23(3),* 1421–1434.

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 43,* 245-265.

Dalgleish, T., Williams, J. M., Golden, A. M., Perkins, N., Barrett, L. F., Barnard, P. J.,…Watkins, E. (2007). Reduced specificity of autobiographical memory and depression: The role of executive control. *Journal of Experimental Psychology: Gen*eral. *136*(1), 23-42.

Dozois, D. J. A., & Covin, R. (2004). The Beck Depression Inventory-II (BDI-II), Beck Hopelessness Scale (BHS), and Beck Scale for Suicide Ideation (BSS). In M. J. Hilsenroth, & D. L. Segal. (Eds.), *Comprehensive handbook of psychological assessment, Vol. 2: Personality assessment.* (pp. 50-69). Hoboken, NJ, US: John Wiley & Sons, Inc.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th Ed.). London: SAGE Publications, Ltd.

Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research, 2*(3), 109-112.

Hallion, L. S., & Ruscio, A. M. (2011). A meta-analysis of the effect of cognitive bias modification on anxiety and depression. Psychological Bulletin, 137(6), 940-958.

Harmer, C.J., Goodwin, G. M., & Cowen, P. J. (2009). Why do antidepressants take so long to work? A cognitive neuropsychological model of antidepressant drug action. *British Journal of Psychiatry, 195*(2), 102-108.

Holländare, F., Andersson, G., & Engström, I. (2010). A comparison of psychometric properties between Internet and paper versions of two depression instruments (BDI-II and MADRS-S) administered to clinic patients. *Journal of Medical Internet Research, 12*(5), e49.

Howell, R. T., Rodzon, K. S., Kurai, M., & Sanchez, A. H. (2010). A validation of well-being and happiness surveys for administration via the Internet. *Behavior Research Methods 42*(3), 775-84.

Hughes, M.E., Panzarella, C., Alloy, L.G., & Abramson, L.Y. (2007). A cognitive perspective on mental illness and mental health. In F.T. Durso, R.S. Nickerson, S. Dumais, D.S. Lewandowsky, & T. Perfect (Eds.) Handbook of Applied Cognition (2nd ed., pp. 629-658). New York: John Wiley & Sons Ltd.

Joormann, J. & Gotlib, I.H. (2007). Selective attention to emotional faces following recovery from depression. *Journal of Abnormal Psychology, 116,* 80-85.

Joormann, J., Hertel, P.T., LeMoult, J., & Gotlib, I.H. (2009). Training forgetting of negative material in depression. *Journal of Abnormal Psychology, 118*(1), 34-43.

Katzow, A., & Safran, J. (2007). Recognizing and resolving ruptures in the therapeutic alliance. In P. Gilbert, & R. Leahy (Eds.) *The therapeutic relationship in the cognitive behavioral psychotherapies.* London: Routledge.

Lang, T. J., Blackwell, S. E.**,** Harmer, C., Davison, P., & Holmes, E. A. (2012). Cognitive bias modification using mental imagery for depression: developing a novel computerized intervention to change negative thinking styles. *European Journal of Personality, 26,* 145-157.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces – KDEF.* Stockholm: Karolinska Institutet.

Mathews, A. & MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. *Annual Review of Clinical Psychology, 1,* 167-195.

Menne-Lothmann, C., Viechtbauer, W., Hohn, P., Kasanova, Z., Haller, S. P., Drukker, M.,…Lau, J. Y. F. (2014). How to Boost Positive Interpretations? A meta-analysis of the effectiveness of cognitive bias modification for interpretation. PLoS ONE, 9(6), 1-26.

National Institute for Health and Clinical Excellence (2009). *Depression: The treatment and management of depression in adults.* London: National Institute for Health and Clinical Excellence.

Nolen-Hoeksema, S. (1991). Responses to depression and their effects on the duration of depressive episodes. *Journal of Abnormal Psychology, 100*(4), 569-582.

Penton-Voak, I. S., Bate, H., Lewis, G., & Munafò, M. R. (2012). Effects of emotion perception training on mood in undergraduate students: Randomised control trial. *The British Journal of Psychiatry, 200,* 1-3.

Penton-Voak, I. S., Thomas, J., Gage, S. H., McMurran, M., McDonald S., & Munafò, M. R. (2013). Increasing recognition of happiness in ambiguous facial expressions reduces anger and aggressive behaviour. *Psychological Science, 24*(5), 688-697.

Rodgers, M., Asaria, M., Walker, S., McMillan, D., Lucock, M., Harden, M., Palmer, S., Eastwood, A. (2012). The clinical effectiveness and cost-effectiveness of low-intensity psychological interventions for the secondary prevention of relapse after depression: A systematic review. *Health Technology Assessment, 16*(28), *1-130.*

Rude S. S.; Wenzlaff, R. M.; Gibbs, B.; Vane, J.; Whitney, T. (2002). Negative processing biases predict subsequent depressive symptoms. *Cognition and Emotion, 16(3),* 423-440.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. The British Medical Journal, 340, c332.

Tabachnick, B. G., & Fiddell, L. S. (2006). *Using multivariate statistics* (5th ed.). Needham Heights, MA: Allyn & Bacon, Inc.

Teachman, B. A., Joormann, J., Steinman, S. A., & Gotlib, I. H. (2012). Automaticity in anxiety disorders and major depressive disorder. *Clinical Psychology Review, 32*, 575-603.

Tiddeman, B. P., Perrett, D. I., & Burt, D. M. (2001). Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications, 21*(5), 42-50.

Tranter, R., Bell, D., Gutting, P., Harmer, C., Healy, D., & Anderson, I. M. (2009). The effect of serotonergic and noradrenergic antidepressants on face emotion processing in depressed patients. *Journal of Affective Disorders, 118*, 87-93.

Ustun, T. B., Ayuso-Mateos, J. L., Chatterji, S., Mathers, C. & Murray, C. J. L. (2004). Global burden of depressive disorders in the year 2000. *The British Journal of Psychiatry, 184,* 386-392.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070.

Appendix A: Acceptability Questionnaire for the ERT

Please write as little or as much as you would like, in the spaces provided, in answer to questions 1 to 5.

1. What did you think about the length of the emotion recognition training task?

2. How enjoyable did you find the task?

3. How engaging did you find the task?

4. Did completing the task have an effect on your mood?

5. Did you find the task challenging?

6. Would you do the task again? Yes/No
7. Were the task instructions easy to follow? Yes/No
8. Would you recommend the task to a friend? Yes/No

Appendix B: Methods of Data Cleaning

Data were initially checked for linearity and unusual cases using histograms, boxplots, and scatterplots, and by investigating the individual variables when transformed into z-scores. Scatterplots of outcome scores against baseline scores showed the linearity assumption to be reasonable, as expected. To reduce the impact of extreme outliers on the analysis, such scores were modified so that they were equal to 3.29*SD away from the pooled mean for that variable (as recommended by Tabachnick & Fidell, 2006). Reductions were made to the following cases: the BDI-II post-intervention score of one control participant; the PANAS negative subscale score of one training participant at 6 weeks follow-up, one training participant at two-weeks follow-up, and one training participant post-intervention; the baseline BAI score for one control participant and one training participant, the BAI score of one training participant at 6 weeks follow-up, the post-training BAI score of one control participant, and the baseline threshold of one training participant.

The assumptions of linearity, homoscedasticity, normality, and independence of the residuals were also checked by plotting graphs with standardised scores of the predictor against standardised scores of the residuals, investigating histograms of the residuals, and using a Durbin-Watson analysis to check that scores were between 1 and 3. Checks were also performed to identify any excessive influence of multivariate outliers by checking that values of the Mahalanobis distance were less than 15 (recommended by Field, 2013, for samples of less than 100) and that Cook's distances were less than 1. Assumptions of normality and homoscedasticity were not met in one linear regression analysis relating to whether training

condition predicted PANAS negative subscale data at 2 weeks follow-up. This analysis also showed univariate and multivariate outliers. To increase the robustness of this analysis bootstrapped samples were created. The confidence intervals and standard errors generated from this bootstrapped regression analysis were reported.

Appendix C. Predictors Entered into Hierarchical Linear Regression Models

Table A1

Hierarchical linear regression models indicating how predictors were entered into each analysis.

Analysis	Outcome measure	Initial predictor	Second predictor	Third predictor	Fourth predictor
Primary hypothesis (a)	BDI-II 6-wk follow-up	Baseline BDI-II	Categorical predictor: Training group (1), control group (0).	-	-
Secondary hypotheses (b)	BDI-II 2-wk follow-up	"	"	-	-
	BDI-II post-intervention	"	"	-	-
Secondary hypotheses (c)	PANAS positive 6-wk follow-up	Baseline PANAS positive	"	-	-
	PANAS positive 2-wk follow-up	"	"	-	-

	PANAS positive post-intervention	“	“	-	-
	PANAS negative 6-wk follow-up	Baseline PANAS negative	“	-	-
	PANAS negative 2-wk follow-up	“	“	-	-
	PANAS negative post-intervention	“	“	-	-
Exploratory analysis: Does group predict anxiety score?	BAI 6-wk follow-up	Baseline BAI	“	-	-
	BAI 2-wk follow-up	“	“	-	-
	BAI post-intervention	“	“	-	-
Exploratory analysis: Does change in	BDI-II 6-wk follow-up	Baseline BDI-II	Baseline ERT threshold	Final observed ERT threshold	-
	BDI-II 2-wk follow-up	“	“	“	-

cognitive bias	BDI-II post-intervention	“	“	“	-
predict mood?	PANAS positive 6-wk follow-up	Baseline PANAS positive	“	“	-
	PANAS positive 2-wk follow-up	“	“	“	-
	PANAS positive post-intervention	“	“	“	-
	PANAS negative 6-wk follow-up	Baseline PANAS negative	“	“	-
	PANAS negative 2-wk follow-up	“	“	“	-
	PANAS negative post-intervention	“	“	“	-
Exploratory analysis: Moderation by baseline	BDI-II 6-wk follow-up	Baseline BDI-II	Categorical predictor: Training group (1), control group (0).	Interaction of baseline BDI-II score by group^	-
depression	BDI-II 2-wk follow-up	“	“	“	-
	BDI-II post-intervention	“	“	“	-

Exploratory analysis: Moderation by baseline anxiety	BDI-II 6-wk follow-up	“	“	Baseline BAI	Interaction of baseline BAI score by group^
	BDI-II 2-wk follow-up	“	“	“	“
	BDI-II post-intervention	“	“	“	“

Note. ^ = The interaction predictor was created by generating centred predictors of the baseline score and condition (i.e., group) and multiplying them together. Centred predictors were created to meet assumptions of multi-collinearity by subtracting the pooled mean of the predictors.

Appendix D: Correlations between Outcome Measures

Table A2

Correlations between measures for the training group

	BDI-II Pre	***BDI-II Post***	***BDI-II 2-wk***	***BDI-II 6-wk***	***PANAS +ve Pre***	***PANAS +ve Post***	***PANAS +ve 2-wk***	***PANAS +ve 6-wk***	***PANAS -ve Pre***	***PANAS -ve Post***	***PANAS -ve 2-wk***	***PANAS -ve 6-wk***	***BAI Pre***	***BAI Post***	***BAI 2-wk***	***BAI 6-wk***	***Threshold Pre***	***Threshold Final obs***	***Age***
BDI-II Pre	***1.00***																		
BDI-II Post	***.66****	***1.00***																	
BDI-II 2-wk	***.47****	***.69****	***1.00***																
BDI-II 6-wk	***.48****	***.71****	***.80****	***1.00***															
PANAS +ve Pre	*-.15*	*-.12*	*-.11*	*-.14*	***1.00***														
PANAS +ve Post	*.14*	*-.22*	*-.41*	*-.34*	*-.26*	***1.00***													
PANAS +ve 2-wk	*-.09*	*-.24*	***-.55****	***-.50****	***.50****	***.85****	***1.00***												
PANAS +ve 6-wk	*-.20*	***-.42****	***-.42****	***-.46****	***-.46****	***.64****	***.74****	***1.00***											
PANAS -ve Pre	***.40****	***.45****	*.30*	***.33****	*.12*	*.10*	*.21*	*.23*	***1.00***										
PANAS -ve Post	***.45****	***.48****	***.49****	*.39*	*-.07*	*.07*	*.15*	*.15*	***.59****	***1.00***									
PANAS -ve 2-wk	*.32*	***.72****	***.59****	***.59****	*.04*	*-.14*	*-.12*	*-.09*	***.59****	***.84****	***1.00***								
PANAS -ve 6-wk	*.27*	***.57****	***.50****	***.59****	*-.07*	*.07*	*-.11*	*-.02*	***.68****	***.77****	***.73****	***1.00***							
BAI Pre	***.41****	*.38*	***.41****	***.41****	*.04*	*.02*	*-.21*	*.03*	***.51****	*.35*	***.38****	***.59****	***1.00***						
BAI Post	***.46****	***.76****	***.63****	***.65****	*-.04*	*-.06*	*-.03*	*-.08*	***.54****	***.55****	***.81****	***.62****	***.69****	***1.00***					

BAI 2-wk	***.35****	***.74****	***.61****	***.59****	*-.31*	*-.21*	*-.32*	***-.40****	*.26*	***.58****	***.58****	***.51****	***.57****	***.92****	***1.00***			
BAI 6-wk	***.38****	***.76****	***.56****	***.62****	*-.10*	*-.15*	*-.26*	*-.13*	***.43****	***.56****	***.64****	***.65****	***.61****	***.86****	***.86****	***1.00***		
Threshold Pre	*-.17*	***-.43****	*-.09*	*-.14*	***.30****	*.04*	*.17*	*.09*	*-.03*	*-.15*	*-.05*	*-.18*	*-.10*	*-.26*	*-.20*	***-.33****	***1.00***	
Threshold Final obs	*.20*	*.02*	*-.16*	*-.17*	*.22*	***.64****	***.52****	***.38****	*.15*	*.17*	*.14*	*.04*	*.02*	*.24*	*-.09*	*-.04*	*.23*	***1.00***
Age	*.03*	*-.21*	*-.08*	*-.16*	***.32****	*.12*	*.19*	*.26*	*.08*	*-.11*	*-.21*	*-.10*	*-.01*	*-.40**	*-.30*	*-.18*	*.26*	*.13*

Note. Pairwise analysis. Bold * = p < .05; Bold ** = p < .01.

Table A3

Correlations between measures for the control group

	BDI-II Pre	***BDI-II Post***	***BDI-II 2-wk***	***BDI-II 6-wk***	***PANAS +ve Pre***	***PANAS +ve Post***	***PANAS +ve 2-wk***	***PANAS +ve 6-wk***	***PANAS -ve Pre***	***PANAS -ve Post***	***PANAS -ve 2-wk***	***PANAS -ve 6-wk***	***BAI Pre***	***BAI Post***	***BAI 2-wk***	***BAI 6-wk***	***Threshold Pre***	***Threshold Final obs***	***Age***
BDI-II Pre	***1.00***																		
BDI-II Post	***.74****	***1.00***																	
BDI-II 2-wk	***.78****	***.81****	***1.00***																
BDI-II 6-wk	***.64****	***.66****	***.91****	***1.00***															
PANAS +ve Pre	***-.40****	*-.05*	*-.23*	*-.14*	***1.00***														
PANAS +ve Post	*-.26*	***-.40****	*-.39*	*-.27*	***.57****	***1.00***													
PANAS +ve 2-wk	*-.14*	*-.22*	***-.39****	***-.38****	*.37*	***.71****	***1.00***												
PANAS +ve 6-wk	*-.12*	*-.20*	*-.35*	***-.37****	***.33****	***.57****	***.72****	***1.00***											

	BDI-II Pre	BDI-II Post	BDI-II 2-wk	BDI-II 6-wk	PANAS +ve Pre	PANAS +ve Post	PANAS +ve 2-wk	PANAS +ve 6-wk	PANAS -ve Pre	PANAS -ve Post	PANAS -ve 2-wk	PANAS -ve 6-wk	BAI Pre	BAI Post	BAI 2-wk	BAI 6-wk	Threshold Pre	Threshold Final obs	Age
PANAS -ve Pre	**.56****	**.51****	**.47***	**.36***	-.04	-.01	.16	.15	**1.00**										
PANAS -ve Post	.15	**.41***	.26	.28	.19	.10	.19	.14	**.47****	**1.00**									
PANAS -ve 2-wk	.37	**.46***	**.50****	.34	.23	.03	.00	.01	**.71****	**.71****	**1.00**								
PANAS -ve 6-wk	**.35***	.30	**.46***	**.49****	.17	.25	.11	.08	**.58****	**.56****	**.60****	**1.00**							
BAI Pre	**.53****	.30	.28	.19	-.14	.01	.00	.05	**.41****	.28	.18	.28	**1.00**						
BAI Post	**.40***	**.39***	.01	.09	-.04	-.05	.18	.06	.34	**.47****	.22	.25	**.88****	**1.00**					
BAI 2-wk	**.40***	.27	**.40***	.25	.18	.19	-.10	-.16	.36	**.54****	**.54****	.26	**.69****	**.76****	**1.00**				
BAI 6-wk	**.52****	.27	**.42***	**.40****	.01	.15	.07	-.16	**.35***	**.46****	**.47***	**.50****	**.74****	**.67****	**.76****	**1.00**			
Threshold Pre	-.11	-.02	-.26	-.24	**.37****	.26	.23	.08	.13	.09	.12	.31	.17	.15	.26	.20	**1.00**		
Threshold Final obs	-.17	-.16	-.29	-.29	**.42****	.32	.09	.11	-.12	.00	-.10	-.04	.09	.06	.12	.11	**.69****	**1.00**	
Age	-.23	-.05	-.04	.05	**.36****	.23	.04	-.06	-.04	-.20	-.22	-.07	**-.28***	**-.42***	-.20	-.14	-.03	.11	**1.00**

Note. Pairwise analysis. Bold * = p < .05; Bold ** = p < .01.

Appendix E: Missing Data Comparisons

Table A4

A comparison of characteristics of participants from whom questionnaire responses were and were not obtained post-intervention

Baseline measure		Post-intervention responders (Training: *N* = 26; Control: *N* = 33)	Missing data (Training: *N* = 40; Control *N* = 25)		
		Mean (SD)	Mean (SD)	*F*	*p*
PANAS: Positive	Training	19.50 (7.13)	18.78 (7.05)	0.17	.686
	Control	19.15 (5.51)	20.60 (7.01)	0.78	.382
PANAS: Negative	Training	18.15 (7.03)	18.73 (7.50)	0.10	.758
	Control	19.03 (7.32)	18.96 (7.39)	<0.01	.971
BDI-II	Training	21.73 (6.94)	21.75 (5.78)	<0.01	.990
	Control	24.21 (8.39)	21.88 (7.96)	1.15	.288
BAI	Training	17.46 (8.65)	18.73 (9.37)	0.30	.583
	Control	16.06 (11.46)	17.60 (9.77)	0.29	.592
ERT Threshold(24 v 34)	Training	8.67 (1.13)	8.35 (1.32)	0.89	.349
31v18	**Control**	**7.84 (1.32)**	**8.94 (1.43)**	**7.51**	**.009**

Age	Training	42.42 (13.15)	41.30 (13.28)	0.11	.737
	Control	40.45 (13.43)	41.52 (13.16)	0.09	.764

Note. Bold font indicates significant effect at alpha level .05

Table A5

A comparison of characteristics of participants from whom questionnaire responses were and were not obtained at 2 weeks follow-up

Baseline measure		2 week follow-up responders (Training: *N* = 37; Control: *N* = 28)	Missing data (Training: *N* = 29; Control *N* = 30)		
		Mean (SD)	Mean (SD)	*F*	*p*
PANAS: Positive	Training	19.65 (7.42)	18.31 (6.56)	0.59	.447
	Control	19.96 (5.90)	19.60 (6.54)	0.05	.825
PANAS: Negative	Training	18.70 (7.28)	18.24 (7.38)	0.07	.800
	Control	19.86 (7.87)	18.20 (6.73)	0.75	.391
BDI-II	Training	21.86 (6.51)	21.59 (5.91)	0.03	.858
	Control	24.18 (8.07)	22.30 (8.38)	0.75	.389
BAI	Training	19.78 (7.29)	16.24 (10.69)	2.55	.115
	Control	15.61 (9.28)	17.77 (11.93)	0.59	.447
ERT Threshold	Training	8.51 (1.07)	8.43 (1.50)	0.06	.814

35v23

26v23	**Control**	**7.77 (1.24)**	**8.78 (1.51)**	**6.65**	**.013**
Age	Training	40.79 (12.53)	42.49 (13.73)	0.27	.607
	Control	41.86 (13.51)	40.03 (13.09)	0.27	.604

Note. Bold font indicates significant effect at alpha level .05

Table A6

A comparison of characteristics of participants from whom questionnaire responses were and were not obtained at 6 weeks follow-up

Baseline measure		6 week follow-up responders (Training: *N* = 44; Control: *N* = 43)	Missing data (Training: *N* = 22; Control *N* = 15)		
		Mean (SD)	Mean (SD)	*F*	*p*
PANAS: Positive	Training	20.20 (7.56)	16.77 (5.27)	3.63	.061
	Control	19.65 (5.71)	20.13 (7.60)	0.07	.797
PANAS: Negative	Training	19.00 (7.73)	17.50 (6.32)	0.62	.434
	Control	19.51 (7.67)	17.53 (6.03)	0.82	.370
BDI-II	Training	21.80 (6.68)	21.64 (5.28)	<0.01	.923
	Control	24.07 (8.66)	20.73 (6.40)	1.86	.178
BAI	Training	19.11 (9.54)	16.45 (7.88)	1.27	.264

	Control	16.72 (10.96)	16.73 (10.27)	<0.01	.997
ERT Threshold	Training	8.67 (1.03)	8.00 (1.63)	3.46	.068
39v10	Control	8.15 (1.55)	8.60 (0.97)	0.75	.391
Age	**Training**	**44.00 (13.14)**	**37.23 (12.20)**	**4.08**	**.048**
	Control	40.98 (13.25)	40.73 (13.55)	<0.01	.952

Note. Bold font indicates significant effect at alpha level .05

Appendix F: ERT Adherence and Thresholds

Table A7

Adherence to the ERT and thresholds of participants over five sessions

Participants from whom post-training thresholds were obtained		Threshold for distinguishing happy vs sad faces			
Treatment group (*N* = 66)	Control group (*N* = 58)	Training group		Control Group	
N (%)	*N* (%)	Pre *M (SD (N))*	Post *M (SD (N))*	Pre *M (SD (N))*	Post *M (SD (N))*
Day 1	40 (60.61)	43 (74.14)	8.48 (1.25 (58))	9.90 (1.80 (40))	8.24 (1.45 (49))	8.30 (1.68 (43))
Day 2	28 (42.42)	32 (55.17)	10.41 (2.06 (39))	11.75 (2.08 (28))	8.22 (1.86 (37))	8.22 (1.77 (32))
Day 3	23 (34.85)	26 (44.83)	11.08 (2.02 (24))	11.78 (2.13 (23))	8.83 (1.95 (29))	8.77 (1.93 (26))
Day 4	19 (28.79)	22 (37.93)	11.90 (2.00 (20))	12.00 (2.38 (19))	8.76 (2.11 (25))	9.00 (2.20 (22))
Day 5	12 (18.18)	19 (32.76)	12.00 (2.54 (14))	11.50 (2.78 (12))	8.52 (1.99 (21))	8.63 (2.06 (19))

Appendix G: Ethics Letter of Approval

From: apache@exeter.ac.uk <apache@exeter.ac.uk> on behalf of Ethics Approval System <D.M.Salway@exeter.ac.uk>

Sent: 19 June 2013 13:09

To: Stephens, Victoria

Subject: Your application for ethical approval (2013/377) has been accepted

Ethical Approval system

Your application (2013/377) entitled Effects on depressive symptoms following use of an internet-based programme aimed at modifying emotion recognition has been accepted

Appendix H: Study Information Sheet

Emotion Recognition Training Screening Survey

Study Information

Emotion Recognition Training Using an Internet-Based Programme

Researcher: Vick Stephens, University of Exeter, Trainee Clinical Psychologist. **Supervisors:** Professor Marcus Munafò, Professor Ian Penton-Voak, School of Psychology, University of Bristol

You are being invited to take part in a research study and it is important for you to understand why the research is being done and what it will involve before you decide whether to take part. Please take your time to read the following information carefully and discuss it with your friends, relatives or GP if you wish. Please contact Vick Stephens (lead researcher) if there is anything that is unclear or if you would like more information: vs9494@bristol.ac.uk

Please take your time in deciding whether or not you would like to take part. Thank you for your time.

What is the purpose of this research?
This research project will investigate a newly developed internet-based programme for emotion recognition training and its effects on individuals showing signs of low mood.

Why have I been invited to take part and how is this research useful?
We need to find people from the general public to test the internet-based emotion recognition training programme to investigate whether it shows potential benefits for helping people who have difficulties relating to mood. It is important to test new products on the general population before a clinical population.

Do I have to take part?
No, it is up to you whether you take part. You are free to change your mind at any time during the study without giving a reason, until the data analysis begins in 2014.

What are the requirements to take part?
In order to take part you should be aged 18 or over and have English as your first language or have an equivalent level of fluency.
You would not be able to take part in the study if you are currently using illicit drugs [other than cannabis] or are dependent on drugs [other than nicotine and caffeine]. You should not take part if you have significant current or past psychiatric illness [except depression]. In addition, you should not take part if you are currently using psychiatric medication or have used psychiatric medication in the last 5 weeks. If you are unsure whether you fit these requirements, please contact Vick Stephens (lead researcher) vs9494@bristol.ac.uk.

What's involved?
If you decide to take part you will be directed to a website where you will be asked if you would like to consent to taking part in the study. Participation in this study is entirely voluntary. It is up to you to decide whether or not to do this. If you decide not to take part, or to withdraw, you do not have to give a reason - nobody would be upset. You will also be asked to answer some questions to check your eligibility for the study. If you are found to be eligible, the research will involve you using an internet-based emotion recognition training programme, on 5 occasions, preferably on 5 consecutive days (each training session will take approximately 15 minutes). You will also be asked to complete online questionnaires, which relate to your mood, before and after the 5 training sessions, and again two weeks and six weeks later.
You will be offered an opportunity to take part in a raffle to win vouchers to the value of £10, £15, £20, £30 or £50; you will be able to enter the raffle even if you decide to withdraw from the study.

What are the possible disadvantages and risks of participating in this research?
There is no identified risk in taking part in this study.

Your data
All data will be kept entirely confidential throughout the study and anonymised (i.e. identifying personal details would be removed) on completion of the study. It would not be possible to identify you from any aspect of reporting for this research study. No individual data will be reported.

What will happen to the results of the research study?
When the study has been completed, the data will be analysed and the findings reported in accordance with the guidelines for submitting a major research project for the University of Exeter Doctorate in Clinical Psychology. The findings will also be reported in an appropriate scientific journal or presented at a scientific meeting. You would not be identified in any way. If you would like a copy of the final paper, you may request this.

Who has reviewed the study?
This study has been reviewed and approved of by the University of Exeter Psychology Research Ethics Committee.

Who can I contact for further information?
For further queries, please contact Vick Stephens (lead researcher): vs9494@bristol.ac.uk

If you have any concerns related to your participation in this study please direct them to the Research Ethics Committee via Dr Cris Burgess, c.n.w.burgess@ex.ac.uk

Appendix I: Consent Form

1. Please answer the following questions.

a. Have you been given information which explains the study? Yes/No

b. Have you had an opportunity to ask questions and discuss the study? Yes/No

c. Have you received enough information about the study for you to make a decision about your participation? Yes/No

d. Do you understand that you are free to withdraw from the study, without having to give a reason, at any time? Yes/No

Please read the following statements and indicate below whether you agree or disagree.

2. I hereby fully and freely consent to my participation in this study. I understand the nature and purpose of the procedures involved in this study. I understand and acknowledge that the investigation is designed to promote scientific knowledge and that the the data I provide will be used for no purpose other than research. I understand that the data I provide will be anonymised. I agree to the University of Bristol keeping and processing the data I provide during the course of this study. I understand that these data will be used only for the purpose set out in the information sheet, and my consent is conditional upon the University complying with its duties and obligations under the Data Protection Act.

I agree I disagree

Appendix J: Debriefing Information

Study Debriefing Information

Researcher Vick Stephens, Trainee Clinical Psychologist, University of Exeter:

vs9494@bristol.ac.uk

Title of Research. Effects on Depressive Symptoms Following Use of an Internet-Based Programme Aimed at Modifying Emotion Recognition.

Background. Research shows that people who have depression generally have negative views (biases) about themselves, other people, the world, and the future. It can be helpful to change these negative biases to treat the depression and stop it from returning. Therapies like cognitive-behaviour therapy (CBT) seem to work by modifying negative biases, some research suggests that antidepressants also work in this manner.

Further research has found that people with depression have a negative bias in recognising facial expressions of emotions, i.e., a tendency to judge ambiguous facial expressions as sad rather than happy. Computerised training programs (similar to the one used in this study) can be used to modify this bias, i.e., training people to judge ambiguous faces as happy rather than sad. Furthermore, modification of these biases in emotion recognition could potentially reduce levels of depression. Some research has shown that people who have used the computer program in a laboratory setting have received beneficial effects. The aim of this research project is to investigate whether beneficial results can also be found in a more real-world setting, where people have accessed the training program more independently via the internet.

I plan to present the findings of this research at a conference in the summer of June 2014 and also in a scientific journal.

Depression is a major cause of disability and a quarter of the population may experience at least one depression episode in their lifetime. Computer technologies have great potential for providing inexpensive interventions for mental health conditions, and can contribute to the treatments that are

available for depression to ensure more people can receive help which can be accessed easily.

Please talk to your GP or a counsellor if your own mood is particularly low for any reason.

There are many helplines that you can contact: e.g., the **Support Line Telephone Helpline, 01708 765200.** http://www.supportline.org.uk/

or **The Samaritans, 08457 90 90 90,** jo@samaritans.org http://www.samaritans.org/

Useful self-help and coping tips for dealing with low mood and depression can be found on: www.helpguide.org/mental/depression_tips.htm

Thank you for taking part in this study.

Appendix K: Dissemination Statement

The target journal for this research is the British Journal of Psychiatry. The paper will be adapted to the relevant style and sent for peer review. A summary of the findings will be sent to all people who expressed an interest, at the recruitment stage, in being informed of the results. A summary of the research will also be placed on social media sites that were used for recruitment. The thesis will also be made universally accessible through Open Research Exeter (ORE), the online institutional repository.

Appendix L: Acknowledgements

I would like to give special thanks to Professor Marcus Munafò, Professor Ian Penton-Voak, Dr Anna Adlam, Dr Anke Karl, Dr Sarah Halligan, Dr Tony Wainwright, Dr Heather O'Mahen, Dr Nick Moberly, Dr Janet Smithson, Dr Andy Skinner, Daniel Schien, Dr Sally Adams, Dr Katy Donnelly, Dr Edward Crane, my family and friends, all researchers that supported me with the systematic review, all organisations that supported me in my recruitment, and everyone who contacted me with an interest in participating in the study.