Running Head: Emotional recognition training, response to emotional faces, and mood

Emotional recognition training modifies neural response to emotional faces but does not improve mood in healthy volunteers with high levels of depressive symptoms

Ian S. Penton-Voak [1,8], Sally Adams [2], Katherine S. Button [2], Meg Fluharty [1], Michael Dalili [3], Michael Browning [4], Emily A. Holmes [5], Catherine J Harmer [4], Marcus R. Munafò [1,6,7,8]

1. School of Experimental Psychology, University of Bristol, United Kingdom.

2. Department of Psychology, University of Bath, United Kingdom.

3. School of Social and Community Medicine, University of Bristol, United Kingdom.

4. Department of Psychiatry, University of Oxford, United Kingdom; NIHR Oxford Health Biomedical Research Centre

5. Division of Psychology, Department of Clinical Neuroscience, Karolinska Institute, Sweden.

6. MRC Integrative Epidemiology Unit at the University of Bristol, United Kingdom.

7. UK Centre for Tobacco and Alcohol Studies, University of Bristol, United Kingdom.

8. National Institute for Health Research Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust and University of Bristol.

Word Count: 4700

This work was funded by MRC research grant MR/J011819/1

Corresponding Author: Ian S. Penton-Voak, School of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol BS8 1TU, United Kingdom. T: +44.117.9288667; F: +44.117.9288588; E: i.s.penton-voak@bristol.ac.uk

Abstract

Background

There is demand for new, effective, and scalable treatments for depression, and developing new forms of cognitive bias modification (CBM) of negative emotional processing biases have been suggested as possible interventions to meet this need.

Methods

We report two double blind RCTs, in which analogue samples of volunteers with high levels of depressive symptoms (BDI-ii > 14) completed a brief course of emotion recognition training (a novel form of CBM using faces) or sham training. In study one (N=36), participants completed a post training emotion recognition task whilst undergoing fMRI to investigate neural correlates of CBM. In study two (N=190), measures of mood were assessed post training, and at 2-week and 6-week follow-up.

Results

In both studies, CBM resulted in an initial change in emotion recognition bias, which (in study two) persisted for 6 weeks after the end of the CBM phase. In study one, CBM resulted in increases neural activation to happy faces compared to sad faces, with this effect driven by an increase in neural activity in the mPFC and bilateral amygdala. In study two, CBM did not lead to a reduction in depressive symptoms on the BDI-ii, or on related measures of mood, motivation and persistence, or depressive interpretation bias at either 2 or 6-week follow-ups.

Conclusions

CBM of emotion recognition appears to have effects on neural activity that are similar in some respects to those induced by SSRI administration (study one), but we find no evidence that this had any later effect on self-reported mood in an analogue sample of non-clinical volunteers with low mood (study two).

Background

Mood disorders, dominated by major depression, constitute a substantial burden of disease. NICE guidelines recommend psychotherapy for mild depression, and cognitive-behavioural therapy for moderate depression, but these therapies typically require individual intervention and therefore, while cost-effective, are expensive. Novel approaches are needed to improve treatments for depression, and to prevent relapse.

Understanding emotional signals is critical to successful social functioning but is disrupted in many psychiatric disorders (Cotter, Granger, Backx, Hobbs, Looi & Barnett, 2018). Negative processing biases may play a role in the onset and maintenance of depression. Neurocognitive models suggest that antidepressant medications have early effects on emotional processing biases that result in therapeutic benefit only after sufficient time has elapsed to allow interaction with others, in which these effects lead to more positive social interactions (Warren, Pringle & Harmer, 2015). In support of these models, fMRI studies have demonstrated that SSRIs change responses to emotional expressions, and that such changes are associated with later improvement in mood (Warren, Pringle and Harmer, 2015).

Given the proposed causal role played by emotion processing in depression, biases in this area may provide a potential target for behavioural, rather than pharmacological, intervention (Penton-Voak, Munafo, & Looi, 2017). We have developed a cognitive bias modification (CBM) technique which targets the recognition of facial expression of emotions by initially assessing the threshold for detecting one emotion over another in an ambiguous expression (e.g., a blend of happiness and sadness), and then providing feedback to shift this threshold (e.g., to favour identification of happiness over sadness). Preliminary results from adults recruited from the general population indicate robust and generalizable effects on emotion perception (Griffiths et al, 2015; Dalili et al, 2016; Penton-Voak et al, 2013). An early stage randomised controlled trial (RCT) with participants recruited from the general population on the basis of high levels of depressive symptoms on the Beck Depression Inventory ii (BDI-ii) also indicated that this intervention may have therapeutic benefit on

positive affect which persists for at least two weeks (Penton-Voak et al 2012). This is consistent with recent models of the action of antidepressant medication, which suggest that drug treatment has early effects on emotional processing bias including the ability to detect positive versus negative facial expressions (Harmer, Goodwin et al. 2009, see also Holmes et al 2018) . Here we investigated the neural correlates of our emotional recognition CBM intervention, and the therapeutic potential of this intervention.

Several studies show that SSRIs have robust effects on emotion processing in the amygdala (e.g. Harmer et al, 2006, Godlewska et al, 2012, for review, see Warren, Pringle & Harmer, 2015), which plays a key role in detecting the salience of emotional stimuli in the environment (Sander, Grafman, & Zalla, 2003; Santos, Mier, Kirsch, & Meyer-Lindenberg, 2011). The medial network has substantial amygdaloid and limbic connections (Price & Drevets, 2010), and altered neural activation is seen in the medial prefrontal cortex in individuals suffering from mood disorders, although the pattern of this activation varies widely between studies (Lemogne et al., 2009; Yoshimura et al., 2010, Grimm et al., 2009; Renner et al., 2015). Similarly, mood related changes in activity are found in the dorsolateral prefrontal cortex (dlPFC), a cortical area associated with the control of attention that helps regulate the amygdala through indirect inhibitory input (Davidson, 2000; Drevets, 2001). A meta-analysis of studies measuring the neural response to affective stimuli showed greater response in the amygdala, insula, and dorsal anterior cingulate cortex, and lower response in the dorsal striatum and dlPFC to negative stimuli in depressed individuals relative to healthy controls (Hamilton et al., 2012). Additionally, a review by Disner *et al.* (2011) found that biased processing of emotional stimuli in depression is associated with greater amygdala reactivity, as well as left dlPFC hypoactivity and right dlPFC hyperactivity.

Study 1 aimed to identify changes in the neural correlates of emotion recognition following this novel CBM in an analogue sample of participants with high levels of depressive symptoms. We administered five days of the emotion recognition training intervention (or a sham training procedure) and then scanned participants using fMRI while performing a face perception task that has been previously used to investigate the effects of SSRIs on the

processing of emotion facial expressions (Godlewska et al 2012). We hypothesised that emotional recognition training would reduce amygdala responses to negative facial expressions. We also hypothesised that training would alter activity in the occipital cortex, as it is highly connected to the amygdala and is sensitive to attentional change in response to emotional stimuli, and the prefrontal cortex, which exerts effects on circuitry implicated in pharmacological and psychological treatment for depression. Based on previous findings, we established the following areas as our regions of interest (ROIs) for comparing neural activation in individuals with low mood in our intervention and control conditions: the bilateral amygdala, the mPFC, bilateral dlPFC, and the occipital cortex.

Study 2 was an early phase RCT, again using an analogue sample of participants recruited from the general population on the basis of high levels of depressive symptoms on the Beck Depression Inventory ii (BDI-ii), in a direct replication of earlier work (Penton-Voak et al 2012), using a larger sample with long-term follow-up. The CBM procedure was identical to Study 1 – participants were randomised to receive either five days of the emotion recognition training intervention, or a sham training procedure. Participants completed a series of assessments of mood and anxiety at 2-week and 6-week follow-up after the end of treatment. We hypothesised that participants randomised to the emotion recognition training intervention would reduce lower symptoms of depression on the BDI-ii over the previous two weeks at 6-week follow-up (our primary outcome).

Methods: **Study One**

Participants. We recruited adults from the staff and students at the University of Bristol and from the general population who reported depressive symptoms (defined as a score of 14 or more on the BDI-ii) (Beck, Steer, & Brown, 1996). Participants were recruited via email lists and local advertisements.

Participants provided informed consent and inclusion/exclusion criteria were assessed. Screening consisted of structured clinical interview for DSM-IV: Clinical Interview Schedule; CIS-R (Lewis, Pelosi, Araya, & Dunn, 1992), the Altman Self-Rating Mania Scale;

ASRM (for bipolar disorder) (Altman, Hedeker, Peterson, & Davis, 1997) and medical history. After screening we also collected data on age, sex, ethnicity, alcohol, tobacco and caffeine use, previous history of depression (treated and non-treated), intelligence (National Adult Reading Test, NART) (Nelson, 1982), number of years of education, social network size (SNS), and current and past history of psychiatric treatment. Criteria for exclusion were a diagnosis of primary anxiety disorder, psychosis, bipolar disorder, or substance dependence (other than nicotine and caffeine) as defined by DSM-IV; current use of an illicit drug (except cannabis); being at clinically significant risk for suicidal behaviour; use of psychotropic medication in the last 5 weeks prior to the study; major somatic or neurological disorders and concurrent medication that could alter emotional processing (including active treatment with counseling, cognitive behavioural therapy, or other psychotherapies).

The study was approved by the Faculty of Science Research Ethics Committee at the University of Bristol. On completion of the final study session, participants were reimbursed £60 for their time and expenses.

Study design and intervention. An experimental collaborator at the Bristol Randomised Trials Collaboration used minimization to allocate participants to either a training procedure designed to promote the perception of happiness over sadness in ambiguous emotional expressions, or a control procedure designed to elicit no change in perception of emotional expression, in order to ensure the groups were balanced for baseline BDI-ii symptoms (grouped according to a score of 14-19, or 20+). Testing was double-blind. The CBM intervention consists of three phases. First, in the baseline phase, images from a 15 face morph sequence that runs from happy to sad facial expressions are presented one at a time, with participants asked to judge whether the face is happy or sad. This allows the 'balance point' at which participants shift from a 'happy' judgement to a 'sad' judgement to be calculated in terms of the number of images in the 15-face sequence that a participant, on average, would classify as happy. We take this as a measure of cognitive bias. In the training phase, feedback (correct/incorrect) is used to shift the participant's balance point. In the training condition, the 'correct' classification is shifted towards 'happy';

the two images nearest the balance point that the participant would have previously classified as ‘sad’ at baseline are considered ‘happy’ in terms of providing feedback. Feedback in the control condition is based directly on baseline performance, and has no effect on responses. Sessions last 20 minutes and are fully automated. Methods are described in detail elsewhere (Penton-Voak et al 2012, 2013). Participants completed computerised training or control procedures once a day over five consecutive days (Monday to Friday). fMRI acquisition took place after the completion of training during the last session. The study protocol was registered prior to data collection (ISRCTN 50125738).

Mood Assessment. Mood assessments via questionnaire measures were taken at baseline and at the end-of-treatment. End-of-treatment follow-up included a visual analogue scale rating of how friendly the participant thought the experimenter was, to ensure that there were no differences between treatment conditions that may have affected blinding. The questionnaire measures included the Beck Depression Inventory (BDI-ii) (Beck, 1996), the Beck Anxiety Inventory (BAI) (Beck, 1988), the Hamilton Rating Scale for Depression (HAM-D) (Hamilton, 1960), and the Positive and Negative Affect Schedule (PANAS) (Watson, Clark, & Tellegen, 1988).

Functional MRI Behavioural Task. During fMRI scanning, participants completed a sex discrimination task involving the rapid presentation of sad, happy, and fearful facial expressions. In this task, thirteen 30 sec blocks of a baseline fixation cross were interleaved with twelve 30 sec blocks of the emotional task – four blocks of sad, four blocks of happy, and four blocks of fear. During each emotional block participants viewed 10 emotional faces (5 female) from a standardised image set (Tottenham et al, 2009). Each face was presented for 150 ms and participants were asked to report the sex of the face using a keypad. The experiment lasted 8.5 min.

Our main contrast of interest was happy>sad. We examined happy>fear and happy>sad+fear to explore whether effects generalised to other negative emotions. We also examined the three “emotion” > rest contrasts to explore which emotions underpinned any

observed effects. Where group differences for emotion contrasts were significant, mean percent signal change values were extracted for each participant and compared across conditions to characterise the specific effect. Functional MRI data acquisition, pre-processing and statistical analysis are described in the Supplementary material.

Results

Characteristics of Participants. 36 participants (24 female) aged 18-33 years (*M* = 22, *SD* = 4) were recruited. Due to a randomisation error, there were 19 participants in the intervention condition and 17 participants in the control condition. All participants were right-handed. The characteristics of participants by condition are shown in Table 1. A CONSORT diagram is shown in supplementary material, Figure C1.

Insert Table 1 about here.

Behavioural Results. Participants in the intervention condition showed a shift in balance point compared to participants in the control condition after 5 sessions, adjusting for their session 1 baseline balance point (adjusted mean difference 4.65, 95% CI = 2.95 to 6.36, P < .001). Mean balance points at baseline and test for intervention and control conditions are presented in Figure S1 in supplementary material. A mixed model ANOVA of questionnaire score data with a between-subjects factor of training condition (intervention, control) and within-subjects factor of time (baseline, follow-up) indicated evidence of main effect of time across measures (*F*s [1, 33] = 6.66 to 9.59, Ps ≤ .014), reflecting an improvement of mood from baseline to follow-up, except for the PANAS positive and negative scores (*F*s [1, 33] = 2.08 to 3.06, Ps ≥ .089), where the direction of effect was consistent with other measures but the statistical evidence weaker. We found no evidence of a main effect of training condition in any measures (*F*s [1, 33] = 0.07 to 2.72, Ps ≥ .10), or any evidence of an interaction between time and training condition across measures (*F*s [1, 33] = 0.24 to 2.68, Ps ≥ .11). Due to a programming error, behavioural data from the sex discrimination task were not recorded.

Functional MRI Results (regions of interest). Due to a lost imaging data file, we analysed the fMRI data of 35 participants (19 intervention, 16 control). Our ROI analyses showed evidence of increased activation to the happy>sad contrast in the intervention condition relative to the control condition, but only in the left, and not the right, amygdala (FWE corrected $P < .05$, central coordinates 57, 61, 27; see Figure 1, top panel). There were no group differences on the happy>sad contrast in the other ROIs (occipital cortex, dlPFC, or mPFC).

Insert Figure 1 about here.

Training also increased BOLD activation to happy>fear and happy>sad+fear contrasts in the left amygdala. These group differences were driven by increased BOLD activation to happy faces in the intervention condition compared to the control condition, with higher BOLD activation to the happy>rest contrast in both the left and right amygdala and also in the mPFC (see Figure 1, bottom panel). Percent signal change in activation for happy faces relative to rest for both the intervention and control conditions in the bilateral amygdala and mPFC is shown in Figure 2. There were no group differences for sad>rest, fear>rest, or sad>fear, and no evidence for increased activation in any contrasts for the control condition relative to the intervention condition. There was no evidence for group differences on any contrasts in any other ROIs.

Insert Figure 2 about here.

To further investigate the effect of training in the left and right amygdala between conditions for each of our three “emotion” > rest contrasts, we conducted a post-hoc repeated measures mixed model ANOVA of percent signal change with a between-subjects factor of training condition (intervention or control) and within-subjects factors of hemisphere (left or right) and emotion (happy, sad, or fear). We observed evidence of a main effect of

training condition (*F* [1, 33] = 6.53, P = .015), where participants in the intervention condition showed greater activation across contrasts relative to the control group. We also found a main effect of hemisphere (*F* [1, 33] = 12.10, P = .001), where participants showed greater activation in the right amygdala compared to the left amygdala. We did not find evidence for any interactions between factors (P s > .22). Activation for each condition by contrast and hemisphere is shown in Figure S2. Independent samples t-tests indicated greater activation for the intervention condition relative to the control condition for the happy>rest contrast in both the left (mean difference = 2.65, 95% CI 0.044 to 0.334, P = .012) and right (mean difference = 2.80, 95% CI 0.069 to 0.436, P = .008) amygdala. We also found evidence of greater activation for the intervention condition relative to the control condition for the fear>rest contrast in the right amygdala (mean difference = 2.18, 95% CI 0.010 to 0.286, P = .036).

Conclusions: study 1

Our results suggest that emotion recognition training increases neural activation to happy faces compared to sad faces, driven by an increase in neural activity for happy faces. We see this increase in activation for this contrast at both the whole brain level (see supplementary material) and among our a priori ROIs, specifically the mPFC and bilateral amygdala. Our ROI analyses also indicated increased activation for the intervention condition relative to the control condition in the left amygdala for the happy>sad, happy>fear, happy>sad+fear contrasts. We did not find differences in neural activation between conditions for our other contrasts in either our whole brain analyses or in our other ROIs, the bilateral dlPFC and the occipital cortex. Participants in the intervention condition did not show any clear improvements on measures of depressive symptoms or mood relative to controls at end of treatment following emotion recognition training.

Our finding of clusters of activation for our happy>sad+fear contrast at the whole brain level in both the left amygdala and brainstem may be explained by the amygdaloid projections underpinning the limbic system. The increase in neural activation for happy expressions for the intervention compared to the control condition, resembles changes seen

following antidepressant administration. Although effects of SSRIs on amygdala activity in response to positive emotional faces have been reported and replicated, they are less robust than changes in response to negative facial expressions. This is important mechanistically, as anhedonia is characterized by depressed amygdala responses to happy faces (Keedwell et al, 2005).

Increased neural activation to happy faces has been observed following both acute and prolonged antidepressant administration, both in healthy and depressed individuals (Warren, Pringle & Harmer, 2015). Increased activation to positive emotional information following antidepressant treatment has been observed across a large brain network, including the amygdala, mPFC, parahippocampal gyrus, and extra-striate cortex. While these changes may occur in the absence of any effects on participants' mood, it has been proposed that the early production of a positive bias in emotional processing may be predictive of ultimate symptom improvement in depressed patients (see Warren, Pringle, & Harmer, 2015 for a review). As we did not find any group differences in activation across our contrasts in the bilateral dlPFC and the occipital cortex, we find no evidence that our CBM intervention alters attention to emotional expressions, nor does it modify the way these faces are perceived by the visual system. Our analyses suggest that emotion recognition training may increase the salience of positive emotional expressions indexed by increased neural activation in the amygdala in our intervention v control groups.

While our results indicate that completing a course of emotion recognition training alters neural activation associated with the perception of happy facial expressions, this fMRI study was not powered to detect mood outcomes when comparing participants in intervention and control conditions. Study 2 addresses this question.

Methods: **Study Two**

Participants. We recruited adults who reported depressive symptoms (defined as a score of 14 or more on the BDI-ii) from the same population as Study 1.

Upon arrival, participants provided informed consent and inclusion/exclusion criteria were assessed as in Study One.

The study was approved by the Faculty of Science Research Ethics Committee at the University of Bristol. On completion of the final study session, participants were reimbursed £60 for their time and expenses.

Study design and intervention. As in Study One, participants were allocated to condition using minimisation to balance baseline BDI-ii scores by an experimental collaborator, and testing was double-blind. The CBM intervention and control procedure were the same as in Study One, and participants again completed computerised training or control procedures once a day over five consecutive days (Monday to Friday).

Mood Assessment. Mood assessments via questionnaire measures were taken at baseline and at the end-of-treatment. Questionnaire measures included the BDI-ii, the BAI, HAM-D, and the PANAS.

Other Measures. Social network size was assessed at baseline by asking participants to rate the number of close friends (whom respondents report feeling close to and whom they believe they could confide in)they have on a 5-point scale ranging from 0 (none) to 4 (four or more). Participants repeated this process, rating the number of contacts and acquaintances. A contact or acquaintance was defined as a person known by sight or known to someone, but not intimately.

Behavioural assessments (Emotion Recognition Task, Scrambled Sentences Test, and the Fishing Game) were taken at the end of treatment, and at 2-week and 6-week follow-up. 6-week follow-up also included a visual analogue scale rating of how helpful the participant thought the experimenter was, to ensure that there are no differences between treatment groups.

The emotion recognition task was a 45 trial task that was identical to the baseline block of the training procedure (i.e. no feedback was given). This was administered to determine whether any chance in bias induced by the task persisted to follow-up. The

Fishing Game (Pictet et al, 2011) and Scrambled Sentence Task (Rude et al 2002) are described in supplementary material.

Statistical Analysis. We used linear regression to evaluate the effect of training on mood at 6-week follow-up. Analyses were conducted with adjustment for the minimization factor only, and with additional adjustment for age, sex, ethnicity, previous history of treatment for depression, and baseline mood (for analyses of mood variables only). The primary outcome was depressive symptoms over the last 2 weeks assessed using the BDI-ii at 6-week follow-up. Secondary outcomes included depressive symptoms measured using the HAM-D, and positive and negative affect assessed using the PANAS. Subgroup analyses were conducted stratified by whether participants meet criteria for clinical depression, number of episodes of depression, age at first episode, and whether participants had depression with or without anxiety. We also analysed the impact of social network size on treatment effect.

Our preliminary data indicated an effect size of $d = 0.43$ at 2-week follow-up, corresponding to a difference of 3 points on the Positive and Negative Affect Schedule (PANAS). This suggested that a sample size of 172 would be required to achieve 80% power at an alpha level of 5%. This sample size gave us equivalent power to detect a difference of 5 points on the BDI-ii at 6-week follow-up (our primary outcome), which we considered would be clinically significant. We aimed to recruit 190 participants to accommodate potential attrition. The study protocol was registered prior to data collection (ISRCTN17767674) (Adams et al, 2013).

Results: Study two

Characteristics of Participants. 190 participants (138 female) aged 18 to 39 years ($M = 21$, $SD = 4$) were recruited. Participant characteristics by are shown in Table 1. A CONSORT flow diagram is in supplementary material, figure C2.

Primary Outcome. We found no evidence of a reduction in depressive symptoms on the BDI-ii at 6-week follow-up (our primary outcome) in the intervention condition compared with the control condition in either the unadjusted (mean difference 0.35, 95% CI -2.41 to 3.10, P = 0.80) or adjusted (mean difference 0.10, 95% CI -2.39 to 2.58, P = 0.94) models.

Secondary Outcomes. There was no evidence of a difference between the two conditions on the BDI-ii at any other time points, or on any other mood measures. These results are shown in Table 2. We found no evidence of a difference on the Scrambled Sentences Test (unadjusted mean difference 0.48, 95% CI -0.94 to 1.90, P = 0.51; adjusted mean difference 0.30, 95% CI -1.12 to 1.72, P = 0.68), or the Fishing Game (unadjusted mean difference 0.23, 95% CI -2.24 to 2.70, P = 0.85; adjusted mean difference 0.28, 95% CI -2.24 to 2.79, P = 0.83) at 6-week follow-up. However, we did find clear evidence of a difference on the Emotion Recognition Task at 6-week follow-up (unadjusted mean difference -2.91, 95% CI -3.67 to -2.14, P < 0.001; adjusted mean difference -2.84; 95% CI -3.63 to -2.06, P < 0.001), indicating that the effect of the intervention on this particular cognitive bias persisted beyond the treatment phase.

Insert Table 2 about here.

Planned Sub-Group Analyses. Subgroup analyses, both unadjusted and adjusted, did not indicate any evidence of improved mood in the intervention condition compared to the control condition among participants with a diagnosis of clinical depression, number of previous episodes of depression, age at first episode among those with a previous episode, and among those with high levels of anxiety symptoms. Similarly, social network size had no effects on our results.

Unplanned Exploratory Analyses. Given the lack of an effect of this CBM technique on mood at any time point, we explored whether emotion recognition bias may instead serve as a cognitive biomarker for depressed mood, by calculating the correlation between pre-training balance point at session 1 and self-reported measures of mood at the

same time point. We found evidence of consistent, albeit relatively weak, correlations across most measures (BDI-ii: r = -.18, P = .018; HAM-D: r = -.17, P = .021; BAI: r = -.11, P = .12; PANAS Positive: r = +.23, P = .001; PANAS Negative: r = -.03, P = .67). At 6 week follow-up, these patterns of correlation were still largely present although attenuated (BDI-ii: r = -.08, P = .286; HAM-D: r = -.17, P = .035; BAI: r = -.06, P = .44; PANAS Positive: r = +.16, P = .049; PANAS Negative: r = +.07, P = .37). These results should be treated with caution given the experimental manipulation of balance point.

Conclusions

Our results suggest that a novel form of emotion recognition training induces a change in a cognitive bias (here, training people to classify faces as happy under ambiguity) that persists for 6 weeks after the end of treatment but does not reduce depressive symptoms on the BDI-ii, or on related measures of mood, motivation and persistence, or depressive interpretation bias between end of treatment and at 6 week follow up in an analogue sample of volunteers with low mood. We found no evidence of specific sub-groups that benefited from the intervention. However, we did find evidence that emotion recognition bias may serve as a cognitive biomarker for depressed mood (and in particular low positive affect), and hence may act as a marker of treatment success.

These two studies present evidence that a simple, automated CBM task leads to training effects that increase amygdala response to happy faces at end of treatment (Study 1) and have a behavioural effect that persists for at least six weeks (Study 2). There is no evidence, however, that this form of cognitive bias modification has any downstream effects on either questionnaire measures of mood, or behavioural measures of anhedonia. Given the robust nature of the training effects, these findings provide little support of a causal role for emotion processing biases, as operationalized here (a bias to recognise happy faces) in the onset or maintenance of depression. Other biases have not been assessed and it is unknown how cognitive biases may combine in this context (cf. Hirsch, Mathews and Clark, 2006). A further and clear limitation of the current work is that it employs analogue and not

clinical samples, which may not be appropriate to test mood outcomes. These results highlight the difficulty of translating interventions to mood outcomes, but provide a biomarker model which can be used in future investigations to optimise effects.

One possibility is that the emotional training task does not generalize to other situations in which any therapeutic effects of a modified bias in responding to happy faces may be realized (e.g., social interactions). Although the training effect transfers to other faces in an experimental context (e.g., the face task in Study 1, which employs different faces to the training task, see also Dalili et al, 2017), there is currently little evidence that this bias generalizes to real world encounters with others. A further RCT employing a modified version of the CBM technique reported here aiming to reduce social anxiety in adolescent participants also showed weak, but positive results. Although there was no decrease in social anxiety, participants in the intervention group showed lower depressive symptoms at 2-week follow up (Rawdon et al 2018).

Recent meta-analyses of cognitive bias modification studies (e.g. Cristea et al, 2015) indicate inconsistent effects across a range of paradigms aiming to manipulate bias with therapeutic effect. Grafton et al (2017) note that this meta-analysis does not discriminate between studies that attempt to change a cognitive bias but fail to do so, and those studies that successfully modify bias (which, as predicted, have stronger therapeutic effects). Our studies show excellent target engagement (responses to faces are changed robustly by this CBM procedure) but our mood measures show no change. Additionally, however, Grafton et al suggest that mood measures per se may not be the best outcome measures for CBM studies, which may serve to reduce emotional vulnerability to further challenges. Our mood state outcomes do not investigate this possibility. However, while a recent study of our CBM technique (Peters et al, 2017) with healthy participants showed little evidence of transfer of bias modification to a variety of cognitive tasks thought to be impacted by low mood, there was weak evidence of transfer to a measure of the impact of stressful events in daily life, particularly in those participants with higher baseline anxiety. This is consistent with Grafton et al's reasoning, and may warrant further research.

Alternatively, individual differences in emotion processing may play no causal role in the onset or maintenance of depression, and may be a cognitive biomarker of depression rather than a therapeutic target. However, this conclusion seems premature given the robust behavioural effects on emotion perception and mechanistically interesting neural responses we report here, and the large literature on the potential causal role that emotional perception plays in depression. Therefore, further work is justified to examine the potential of this and related CBM techniques, perhaps as adjunct therapies to pharmacological or other psychological treatments (Holmes et al, 2018).

Acknowledgements

MRM and SA are members of the UK Centre for Tobacco and Alcohol Studies, a UKCRC Public Health Research: Centre of Excellence. Funding from British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged. This work was supported by the Medical Research Council (MR/J011819/1) and supported by researchers at the NIHR Oxford Health Biomedical Research Centre and by the NIHR Biomedical Research Centre at University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health and Social Care. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Conflict of Interest

IPV and MRM are co-directors of Jericoe Ltd. a company that designs and sells software for psychological assessment.

References

ADAMS, S., PENTON-VOAK, I. S., HARMER, C. J., HOLMES, E. A. & MUNAFO, 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.

ALTMAN, E. G., HEDEKER, D., PETERSON, J. L. & DAVIS, J. M. 1997. The Altman Self-Rating Mania Scale. *Biological Psychiatry,* 42**,** 948-955.

BECK, A. T., BROWN, G., EPSTEIN, N. & STEER, R. A. 1988. An inventory for measuring clinical anxiety - psychometric properties. *Journal of Consulting and Clinical Psychology,* 56**,** 893-897.

BECK, A.T., STEER, R.A., & BROWN, G.K. 1996. Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation.

BECKMANN, C. F., JENKINSON, M. & SMITH, S. M. 2003. General multilevel linear modeling for group analysis in FMRI. *Neuroimage,* 20**,** 1052-1063.

BOUBELA, R. N., KALCHER, K., HUF, W., SEIDEL, E. M., DERNTL, B., PEZAWAS, L., NASEL, C. & MOSER, E. 2015. fMRI measurements of amygdala activation are confounded by stimulus correlated signal fluctuation in nearby veins draining distant brain regions. *Scientific Reports,* 5.

COTTER, J., GRANGER, K., BACKX, R., HOBBS, M., LOOI, C. Y. & BARNETT, J. H. 2018. Social cognitive dysfunction as a clinical marker: A systematic review of meta-analyses across 30 clinical conditions. *Neuroscience and Biobehavioral Reviews,* 84**,** 92-99.

CRISTEA, I. A., KOK, R. N. & CUIJPERS, P. 2015. Efficacy of cognitive bias modification interventions in anxiety and depression: meta-analysis. *British Journal of Psychiatry,* 206**,** 7-16.

DALILI, M. N., PENTON-VOAK, I. S., HARMER, C. J. & MUNAFO, M. R. 2015. Meta-analysis of emotion recognition deficits in major depressive disorder. *Psychological Medicine,* 45**,** 1135-1144.

DALILI, M. N., SCHOFIELD-TOLOZA, L., MUNAFO, M. R. & PENTON-VOAK, I. S. 2017. Emotion recognition training using composite faces generalises across identities but not all emotions. *Cognition & Emotion,* 31**,** 858-867.

DAVIDSON, R. J., JACKSON, D. C. & KALIN, N. H. 2000. Emotion, plasticity, context, and regulation: Perspectives from affective neuroscience. *Psychological Bulletin,* 126**,** 890-909.

DISNER, S. G., BEEVERS, C. G., HAIGH, E. A. P. & BECK, A. T. 2011. Neural mechanisms of the cognitive model of depression. *Nature Reviews Neuroscience,* 12**,** 467-477.

DREVETS, W. C. 2001. Neuroimaging and neuropathological studies of depression: implications for the cognitive-emotional features of mood disorders. *Current Opinion in Neurobiology,* 11**,** 240-249.

GODLEWSKA, B. R., NORBURY, R., SELVARAJ, S., COWEN, P. J. & HARMER, C. J. 2012. Short-term SSRI treatment normalises amygdala hyperactivity in depressed patients. *Psychological Medicine,* 42**,** 2609-2617.

GRAFTON, B., MACLEOD, C., RUDAIZKY, D., HOLMES, E. A., SALEMINK, E., FOX, E. & NOTEBAERT, L. 2017. Confusing procedures with process when appraising the impact of cognitive bias modification on emotional vulnerability. *British Journal of Psychiatry,* 211**,** 266-271.

GRIFFITHS, S., JARROLD, C., PENTON-VOAK, I. S. & MUNAFO, M. R. 2015. Feedback training induces a bias for detecting happiness or fear in facial expressions that generalises to a novel task. *Psychiatry Research,* 230**,** 951-957.

GRIMM, S., BOESIGER, P., BECK, J., SCHUEPBACH, D., BERMPOHL, F., WALTER, M., ERNST, J., HELL, D., BOEKER, H. & NORTHOFF, G. 2009. Altered Negative BOLD Responses in the Default-Mode Network during Emotion Processing in Depressed Subjects. *Neuropsychopharmacology,* 34**,** 932-943.

HAMILTON, J. P., ETKIN, A., FURMAN, D. J., LEMUS, M. G., JOHNSON, R. F. & GOTLIB, I. H. 2012. Functional Neuroimaging of Major Depressive Disorder: A Meta-Analysis

and New Integration of Baseline Activation and Neural Response Data. *American Journal of Psychiatry,* 169**,** 693-703.

HAMILTON, M. 1960. A RATING SCALE FOR DEPRESSION. *Journal of Neurology Neurosurgery and Psychiatry,* 23**,** 56-62.

HARMER, C. J., MACKAY, C. E., REID, C. B., COWEN, P. J. & GOODWIN, G. M. 2006. Antidepressant drug treatment modifies the neural processing of nonconscious threat cues. *Biological Psychiatry,* 59**,** 816-820.

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**,** 102-108.

HIRSCH, C. R., CLARK, D. M. & MATHEWS, A. 2006. Imagery and interpretations in social phobia: support for the combined cognitive biases hypothesis. *Behavior Therapy,* 37**,** 223-36.

HOLMES, E. A., GHADERI, A., HARMER, C. J., RAMCHANDANI, P. G., CUIJPERS, P., MORRISON, A. P., ROISER, J. P., BOCKTING, C. L. H., O'CONNOR, R. C., SHAFRAN, R., MOULDS, M. L. & CRASKE, M. G. 2018. The Lancet Psychiatry Commission on psychological treatments research in tomorrow's science. *Lancet Psychiatry,* 5**,** 237-286.

JENKINSON, M., BANNISTER, P., BRADY, M. & SMITH, S. 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage,* 17**,** 825-841.

KEEDWELL, P. A., ANDREW, C., WILLIAMS, S. C. R., BRAMMER, M. J. & PHILLIPS, M. L. 2005. The neural correlates of anhedonia in major depressive disorder. *Biological Psychiatry,* 58**,** 843-853.

LEMOGNE, C., LE BASTARD, G., MAYBERG, H., VOLLE, E., BERGOUIGNAN, L., LEHERICY, S., ALLILAIRE, J. F. & FOSSATI, P. 2009. In search of the depressive self: extended medial prefrontal network during self-referential processing in major depression. *Social Cognitive and Affective Neuroscience,* 4**,** 305-312.

LEWIS, G., PELOSI, A. J., ARAYA, R. & DUNN, G. 1992. Measuring psychiatric disorder in the community – a standardized assessment for use by lay interviewers. *Psychological Medicine,* 22**,** 465-486.

NELSON, H. E. 1982. *The National Adult Reading Test (NART): test manual.* Windsor: NFER-Nelson.

PENTON-VOAK, I. S., BATE, H., LEWIS, G. & MUNAFO, M. R. 2012. Effects of emotion perception training on mood in undergraduate students: randomised controlled trial. *British Journal of Psychiatry,* 201**,** 71-72.

PENTON-VOAK, I. S., MUNAFO, M. R. & LOOI, C. Y. 2017. Biased Facial-Emotion Perception in Mental Health Disorders: A Possible Target for Psychological Intervention? *Current Directions in Psychological Science,* 26**,** 294-301.

PENTON-VOAK, I. S., THOMAS, J., GAGE, S. H., MCMURRAN, M., MCDONALD, S. & MUNAFO, M. R. 2013. Increasing Recognition of Happiness in Ambiguous Facial Expressions Reduces Anger and Aggressive Behavior. *Psychological Science,* 24**,** 688-697.

PETERS, S. E., LUMSDEN, J., PEH, O. H., PENTON-VOAK, I. S., MUNAFO, M. R. & ROBINSON, O. J. 2017. Cognitive bias modification for facial interpretation: a randomized controlled trial of transfer to self-report and cognitive measures in a healthy sample. *Royal Society Open Science,* 4.

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**,** 885-891.

PRICE, J. L. & DREVETS, W. C. 2010. Neurocircuitry of Mood Disorders. *Neuropsychopharmacology,* 35**,** 192-216.

RAWDON, C., MURPHY, D., MOTYER, G., MUNAFO, M. R., PENTON-VOAK, I. & FITZGERALD, A. 2018. An investigation of emotion recognition training to reduce symptoms of social anxiety in adolescence. *Psychiatry Research,* 263**,** 257-267.

RENNER, F., SIEP, N., LOBBESTAEL, J., ARNTZ, A., PEETERS, F. & HUIBERS, M. J. H. 2015. Neural correlates of self-referential processing and implicit self-associations in chronic depression. *Journal of Affective Disorders,* 186**,** 40-47.

RUDE, S. S., WENZLAFF, R. M., GIBBS, B., VANE, J. & WHITNEY, T. 2002. Negative processing biases predict subsequent depressive symptoms. *Cognition & Emotion,* 16**,** 423-440.

SANDER, D., GRAFMAN, J. & ZALLA, T. 2003. The human amygdala: an evolved system for relevance detection. *Reviews in the Neurosciences,* 14**,** 303-316.

SANTOS, A., MIER, D., KIRSCH, P. & MEYER-LINDENBERG, A. 2011. Evidence for a general face salience signal in human amygdala. *Neuroimage,* 54**,** 3111-3116.

STODDARD, J., SHARIF-ASKARY, B., HARKINS, E. A., FRANK, H. R., BROTMAN, M. A., PENTON-VOAK, I. S., MAOZ, K., BAR-HAIM, Y., MUNAFO, M., PINE, D. S. & LEIBENLUFT, E. 2016. An Open Pilot Study of Training Hostile Interpretation Bias to Treat Disruptive Mood Dysregulation Disorder. *Journal of Child and Adolescent Psychopharmacology,* 26**,** 49-57.

TOTTENHAM, N., TANAKA, J. W., LEON, A. C., MCCARRY, T., NURSE, M., HARE, T. A., MARCUS, D. J., WESTERLUND, A., CASEY, B. J. & NELSON, C. 2009. The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research,* 168**,** 242-249.

WARREN, M. B., PRINGLE, A. & HARMER, C. J. 2015. A neurocognitive model for understanding treatment action in depression. *Philosophical Transactions of the Royal Society B-Biological Sciences,* 370**,** 12.

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**,** 1063-1070.

WOOLRICH, M. W., BEHRENS, T. E. J., BECKMANN, C. F., JENKINSON, M. & SMITH, S. M. 2004. Multilevel linear modelling for FMRI group analysis using Bayesian inference. *Neuroimage,* 21**,** 1732-1747.

WOOLRICH, M. W., JBABDI, S., PATENAUDE, B., CHAPPELL, M., MAKNI, S., BEHRENS, T., BECKMANN, C., JENKINSON, M. & SMITH, S. M. 2009. Bayesian analysis of neuroimaging data in FSL. *Neuroimage,* 45**,** S173-S186.

YOSHIMURA, S., OKAMOTO, Y., ONODA, K., MATSUNAGA, M., UEDA, K., SUZUKI, S. & YAMAWAKI, S. 2010. Rostral anterior cingulate cortex activity mediates the relationship between the depressive symptoms and the medial prefrontal cortex activity. *Journal of Affective Disorders,* 122**,** 76-85.

Table 1. Characteristics of Participants (Studies 1 and 2)

	Study 1		Study 2	
	Intervention (*n* = 19)	Control (*n* = 17)	Intervention (*n* = 95)	Control (*n* = 95)
Age	21 (4)	23 (4)	22 (4)	22 (5)
Sex (female)	13 (68%)	11 (65%)	69 (73%)	69 (73%)
Ancestry (European)			69 (73%)	64 (67%)
NART Score	36.47 (6.70)	33.00 (8.73)	38.27 (7.08)	38.26 (6.86)
Years of Education	15.18 (1.63)	16.53 (2.85)	15.57 (2.43)	15.90 (2.44)
CISR Score	16.84 (9.34)	15.06 (8.56)	17.71 (9.79	16.78 (11.14)
ASRM Score	3.42 (2.39)	3.00 (1.66)	2.88 (2.17)	3.00 (2.39)
BDI-ii Screening	25.21 (8.50)	24.12 (6.75)	25.00 (7.48)	24.55 (8.72)
BDI-ii Baseline	19.00 (9.10)	18.18 (6.57)	21.05 (9.95)	20.93 (10.13)
BDI-ii End-of-Treatment	14.37 (5.73)	16.41 (6.96)	17.63 (9.81)	16.98 (10.71)
BDI-ii Follow-Up (2-week)	n/a	n/a	16.15 (9.81)	15.73 (10.99)
BDI-ii Follow-Up (6-week)	n/a	n/a	13.17 (9.62)	14.01 (10.23)
BAI Total Baseline	12.95 (8.20)	14.94 (8.54)	14.60 (8.85)	15.86 (10.39)
BAI Total End-of-Treatment	9.95 (6.70)	12.71 (10.80)	11.30 (7.96)	11.07 (9.05)
BAI Follow-Up (2-week)	n/a	n/a	10.83 (9.83)	10.34 (9.60)
BAI Follow-Up (6-week)	n/a	n/a	10.33 (9.34)	10.15 (9.05)
HAM-D Total Baseline	15.05 (5.34)	15.47 (5.43)	13.25 (5.68)	13.41 (6.38)
HAM-D Total End-of-Treatment	11.74 (5.63)	14.81 (5.12)	9.13 (5.12)	8.90 (5.58)
HAM-D Follow-Up (2-week)	n/a	n/a	9.43 (5.76)	9.53 (6.53)
HAM-D Follow-Up (6-week)	n/a	n/a	8.08 (5.45)	9.06 (6.06)
PANAS Positive Score Baseline	17.26 (6.45)	17.59 (4.35)	16.91 (5.33)	18.06 (7.33)
PANAS Positive Score End-of-Treatment	18.05 (7.15)	19.41 (5.81)	17.81 (6.55)	18.69 (7.36)
PANAS Positive Follow-Up (2-week)	n/a	n/a	18.30 (6.78)	19.69 (7.75)
PANAS Positive Follow-Up (6-week)	n/a	n/a	19.80 (8.36)	19.99 (7.90)
PANAS Negative Score Baseline	15.53 (5.38)	16.94 (5.87)	15.84 (5.55)	15.77 (6.06)
PANAS Negative Score End-of-Treatment	13.53 (3.39)	16.65 (6.08)	15.10 (4.82)	14.54 (5.23)
PANAS Negative Follow-Up (2-week)	n/a	n/a	14.76 (5.14)	15.20 (6.15)
PANAS Negative Follow-Up (6-week)	n/a	n/a	14.34 (4.47)	14.31 (5.17)
Experimenter Friendliness	8.73 (1.59)	8.69 (1.52)	8.29 (1.66)	8.48 (1.65)

Values represent mean (standard deviation) for continuous variables, and number (percentage) for categorical variables.

Table 2. Effects of Emotion Recognition Training on Mood Symptoms in Study 2.

		End of Treatment			Follow-Up (2 weeks)			Follow-Up (6 weeks)		
		Estimate	95% CI	P	Estimate	95% CI	P	Estimate	95% CI	P
BDI-II	Unadjusted	-0.59	-3.33 to 2.15	0.67	-0.26	-3.18 to 2.66	0.86	0.35	-2.41 to 3.10	0.80
	Adjusted	-0.50	-2.50 to 1.50	0.62	-0.75	-3.11 to 1.60	0.53	0.10	-2.39 to 2.58	0.94
BAI	Unadjusted	-0.19	-2.50 to 2.12	0.87	-0.34	-3.08 to 2.39	0.81	-0.15	-2.77 to 2.47	0.91
	Adjusted	-1.14	-2.56 to 0.28	0.12	-1.71	-3.96 to 0.53	0.13	-1.10	-3.36 to 1.16	0.34
HAM-D	Unadjusted	-0.21	-1.72 to 1.30	0.79	0.17	-1.61 to 1.94	0.85	1.02	-0.62 to 2.65	0.22
	Adjusted	-0.36	-1.52 to 0.81	0.55	-0.18	-1.56 to 1.20	0.80	0.80	-0.63 to 2.24	0.27
PANAS Positive	Unadjusted	0.88	-1.14 to 2.89	0.39	1.35	-0.80 to 3.49	0.22	0.17	-2.20 to 2.54	0.89
	Adjusted	0.20	-1.37 to 1.76	0.81	0.49	-1.31 to 2.28	0.59	-0.44	-2.54 to 1.66	0.68
PANAS Negative	Unadjusted	-0.52	-1.96 to 0.91	0.47	0.49	-1.17 to 2.14	0.56	-0.02	-1.43 to 1.38	0.97
	Adjusted	-0.63	-1.70 to 0.45	0.25	0.36	-0.95 to 1.67	0.59	-0.05	-1.35 to 1.25	0.94

BDI-ii: Beck Depression Inventory; HAM-D Hamilton Rating Scale for Depression; PANAS: Positive and Negative Affect Schedule. Adjusted analyses include age, sex, ethnicity, previous history of treatment for depression, and baseline mood score as covariates.