Let's Face It: Facial Emotion Processing Is NOT Impaired in Bipolar Disorder

Lucy J. Robinson [a*], *BA(Hons), BSc(Hons), PhD, DClinPsy, AFBPsS*
Clinical Psychologist

John M. Gray

Mike Burt

I. Nicol Ferrier

Peter Gallagher, *BSc(Hons), MPhil, PhD, AFBPsS*
Lecturer in Neuropsychology

Affiliations:

[a] School of Psychology, Ridley Building 1, Newcastle University, Queen Victoria Road, Newcastle upon Tyne, NE1 7RU, UK

*Corresponding Author:

Dr Lucy Robinson, address above

Tel: +44 (0) 191 222 8371

Fax: +44 (0) 191 222 7520

e-mail: lucy.robinson2@ncl.ac.uk

Word count – Abstract = 239
Word count – Manuscript = 4,682 (excluding title page, abstract, references, tables, and figures)

ABSTRACT

Objective: Studies of facial emotion processing in bipolar disorder (BD) have reported conflicting findings. Some of the variation can be attributed to differences in the mood state of the samples recruited. This study aimed to investigate facial emotion labelling in euthymic and depressed patients with BD using tasks with static and dynamically-morphed images of different emotions displayed at different intensities.

Method: Study 1 included 38 euthymic patients with BD and 28 controls. Participants completed two tasks: 1) facial emotion labelling of static images of basic emotions (anger, disgust, fear, happy, sad) shown at different expression intensities (20%, 40%, 60%, 80%); 2) the Eyes Test (Baron-Cohen et al reference), which involves recognition of secondary emotions using only the eye region of the face. Study 2 included 53 depressed patients with BD and 47 controls. Participants completed three tasks: 1) facial emotion labelling of dynamically-morphed images of the same five basic emotions shown up to different intensities (30%, 50%, 70%, 100%); 2) the Emotional Hexagon test, which involves labelling facial emotion-blends.

Comment [PeterG1]: REF

Comment [PeterG2]: Perhaps state 2 here if the third test referred to is Benton's faces which was just there as a control measure.

Results: There were no significant differences between patients and controls on any of the measures. This was observed in both the euthymic and depressed groups.

Comment [PeterG3]: Might be worth adding in some details on the new analyses around effect sizes on other measures.

Conclusions: There was no evidence of deficits in facial emotion labelling in euthymic or depressed patients with BD. Methodological variations – ~~especially~~ including~~in~~ the mood state of the participants and the sample sizes recruited – may play a significant role in the variability in findings between studies.

Comment [PeterG4]: I've just suggested this as we might want to make more on the theory that the specific task demands might be a major factor e.g. which emotions are included, how the labels are presented, the time the face is on the screen for etc.

Keywords: Affective disorder, facial emotion labelling, Eyes test, facial expression recognition, emotional hexagon

INTRODUCTION

Bipolar disorder (BD) is a chronic cyclical mood disorder involving periods of elevated mood and periods of depressed mood. Its aetiology is unknown and a large amount of work in recent years has been undertaken to characterise the functional, cognitive and social deficits associated with the illness (Bonnín et al.; Fagiolini et al., 2005; Goetz, Tohen, Reed, Lorenzo, & Vieta, 2007; Green, Cahill, & Malhi, 2007; MacQueen, Young, & Joffe, 2001; Tamsyn E. Van Rheenen & Susan L. Rossell, 2014). BD is considered to lie on a spectrum of mood disturbance, with two primary types most-often studied: BD I in which sufferers experience manic episodes and (typically) depressive episodes, and BD II where sufferers experience less severe elevated mood episodes (hypomania) and depressive episodes (APA, 1994). Emotion processing in BD has received increasing attention in an attempt to understand whether some element of dysfunction in the processing of emotional stimuli plays a part in a disorder where the 'emotional thermostat' seems markedly disturbed (Van Rheenen & Rossell, 2013). Part of that endeavour has involved exploring facial expression recognition in BD to capture emotion-decoding and labelling processes. Given the central importance of emotional expressions in day-to-day communication, deficits or biases in emotion processing could cause marked social impairments that may be of relevance in the experience of mood episodes or in the impaired social functioning seen in BD (Miklowitz, 2011; Sanchez-Moreno et al., 2009).

The findings of studies exploring facial emotion processing in BD are characterised by variability rather than supporting a single deficit or bias in emotion processing (Van Rheenen & Rossell, 2013). This is in large part due to the differences in methods used (e.g. facial image sets, emotion categories used/contrasted with one another, labelling versus discrimination tasks, stimulus display time, response format), the population studied (BD I, BD II, euthymic, symptomatic, pooled samples of BD subtypes/samples) and the sample sizes recruited.

Comment [PeterG5]: We will need to also cite the Kohler meta-analysis and the main findings:

Kohler CG, Hoffman LJ, Eastman LB, Healey K, Moberg PJ (2011): Facial emotion perception in depression and bipolar disorder: A quantitative review. *Psychiatry Res* 188:303-309.

Even in adult samples of euthymic BD patients there is considerable variability in the findings and conclusions of extant studies with some reporting specific deficits in the recognition of particular emotions, e.g. disgust (Harmer, Grayson, & Goodwin, 2002), fear (Vederman et al., 2012; Venn et al., 2004), or sadness (Vederman et al., 2012); others reporting difficulties with emotion discrimination generally (Addington & Addington, 1998; Bozikas, Tonia, Fokas, Karavatos, & Kosmidis, 2006); and others reporting no significant differences in facial expression recognition (Addington & Addington, 1998; Lembke & Ketter, 2002). In symptomatic patients the picture is no clearer with some studies reporting no differences on one or other of: recognition, discrimination or sensitivity (Bellack, Blanchard, & Mueser, 1996; Edwards, Pattison, Jackson, & Wales, 2001; Summers, Papadopoulou, Bruno, Cipolotti, & Ron, 2006; Vaskinn et al., 2007); others reporting differences in recognition (generally without exploring specific emotions (Getz, Shear, & Strakowski, 2003), or worse recognition of fear, but better recognition of disgust (Summers et al., 2006)) or differences in sensitivity (i.e. the 'amount' of any particular emotion that needs to be present for the emotion to be correctly recognised) (Gray et al., 2006; Schaefer, Baumann, Rich, Luckenbaugh, & Zarate, 2010).

Comment [PeterG6]: Harmer actually showed increased recognition of disgust, but then treated this like it was a bad thing – even though they were correctly identify disgust as disgust!

Comment [PeterG7]: Relative to other emotions or vs controls?

To make sense of the disparate and contradictory findings above, further studies are needed to develop our understanding of the extent to which methodological variations in tasks or sample characteristics are affecting results. Studies in relatively large samples of well-characterised patients in clearly-defined mood states administering more than one emotion processing paradigm would go some way to address this gap.

In a recent article in this journal, Van Rheenen & Rossell (T. E. Van Rheenen & S. L. Rossell, 2014) used a series of face-processing paradigms in a pooled sample of patients with BD in different mood states. In the study, three tasks were administered that each employed four basic emotions (happy, sad, anger and fear): emotion labelling of full-intensity dynamically-morphed images; emotion labelling of static images of different emotion intensities (high [100%], medium [75%], and low [50%]); and emotion discrimination of static images using the same three intensity levels. When assessing all 3 tasks simultaneously, patients with BD were significantly less accurate than controls generally, although the effect was not seen for all of the tasks when analysed individually. However significant differences between groups on individual emotions were not evident. This led the authors to conclude that there was evidence of a broad deficit in aspects of emotion processing in BD, with effect sizes in the small to medium range. The comprehensive set of tasks used is undoubtedly a strength of the study and serves to highlight the extent to which methodological variations in task demands may contribute to the varied findings in this field. The patient cohort included a mix of depressed, hypomanic, mixed and euthymic states, which were pooled for the primary analyses. While follow-up analyses indicated no statistical differences were reported between these different mood states, the size of the subgroups and complexity of the analyses in a repeated measures design may have impacted on the statistical power of *post hoc* ~~analyses~~ contrasts to detect differences, which the authors identify as relatively subtle in the group as a whole and which were not detected in all tasks (T. E. Van Rheenen & S. L. Rossell, 2014) ~~and which were not detected in all tasks~~.

In order to further-explore the impact of mood state and task variations on emotion processing deficits in BD, the present investigation outlines a series of two studies designed to maximise the chances of identifying differences between patients and controls. The first study was conducted in a well-characterised sample of prospectively-verified euthymic BD patients and involved emotion labelling of static images of five basic emotions (angry, happy, fearful, sad, disgusted) at different intensity levels and static facial expression recognition of secondary emotions. The second study was conducted in a ~~depressed~~well-characterised sample of depressed bipolar patients, where it was anticipated that any group differences that resulted from emotion processing deficits would be ~~larger~~ maximised as patients were symptomatic (effectively adding state-related effects to the purported trait-related deficit). To maximise ecological validity of the second study, the tasks involved emotion-labelling of dynamic facial expressions (of the same five basic emotions used in the first study) displayed up to 5 different intensity levels as well as a task labelling static images of blends of emotional expressions. It was anticipated that emotion labelling deficits would be observed in euthymic patients compared to controls and that these between group differences would be significantly greater in acutely symptomatic patients, using the more ecologically valid methodology of the second study.

STUDY 1: EUTHYMIA

In order to assess the mood-state independence of basic emotion recognition ability in bipolar disorder, study one focussed on testing patients when euthymic.

METHODS

Participants

Sixty four participants were recruited (n=38 bipolar patients and n=28 controls). Patients were recruited from secondary and tertiary psychiatric services throughout the North East of England. Inclusion criteria comprised: aged between 18-65, a DSM-IV diagnosis of bipolar disorder (confirmed by a psychiatrist using the mood disorders section of the Structured Clinical Interview for DSM-IV (First, Spitzer, Williams, & Gibbon, 1995)) and currently euthymic (scored ≤7 on both the 17-item Hamilton Depression Rating Scale (Hamilton, 1960) and the Young Mania Rating Scale (R. Young, Biggs, Ziegler, & Meyer, 1978) which was prospectively verified for one month before testing (for full details see (Thompson et al., 2005)). Exclusion criteria comprised current alcohol misuse or dependence, history of head injury with loss of consciousness lasting more than 5 minutes, known neurological illness or relevant major medical illness, ECT within the last 6 months, and learning disability or difficulty with fluent use of the English language. Patients were not excluded for use of psychotropic medication or for comorbid anxiety disorders (comorbidities were assessed using the Mini-International Neuropsychiatric Interview (Sheehan et al., 1998)).

Control participants were recruited via local advertisements. They were subject to the same exclusion criteria as the patient sample with the addition of no personal history of psychiatric illness and no family history of bipolar disorder in a first degree relative. The study was reviewed and given ethical approval by Newcastle Research Ethics Committee. All participants gave written informed consent to participate.

Demographic details are provided in table 1.

Measures

Facial Expression Recognition Task – Static Images (FERT-static)

The task used was based on versions used in earlier studies (Harmer et al., 2002, Montagne et al., 2007). Participants were presented with a black and white still facial photograph of a person showing one of 5 facial expressions (angry, disgusted, fearful, happy, or sad) or neutral. The images used were drawn from the Ekman series (Ekman and Friesen, 1976) and were morphed with neutral (Tiddeman, Burt, & Perrett, 2001) to produce expressions which varied in intensity before being masked from the bottom of the chin to the top of the forehead (thereby covering the hair and ears). Four different individuals were used from the Ekman series (2 male, 2 female) each posing the five expressions plus neutral. This meant each of the expressions was shown sixteen times – four times at each of four intensity levels (20%, 40%, 60%, 80%) (5 emotions x 16 presentations = 80 stimuli). The neutral expression was shown four times (once per individual) (84 stimuli in total).

The picture of the face was presented on a black background (333x482 pixels) on the left hand side of the screen for one second (see Figure 1a). After it had displayed, a solid black mask covered the image and the participant was instructed to indicate the expression (see Figure 1b). The words ‘Angry’, ‘Disgusted’, ‘Fearful’, ‘Happy’, ‘Sad’ and ‘Neutral’ were presented on the right hand side of the screen underneath one another listed in alphabetical order (the touch-sensitive area of screen allocated to each response-option was 180x100 pixels). It was not possible for a response to be given when the face was still being displayed.

In order for participants to familiarise themselves with the position and order of the response options, the task began with six practice trials. This involved six presentations of 100% intensity of each of the 5 emotions and one neutral face. The six pictures were of the same individual, who was not used again in the task. The practice trials were presented in the same fixed order to all participants. The 84 experimental trials were presented in a different random order to each participant.

Stimuli were presented using Superlab 4.0 (Cedrus) and responses were recorded using 15” CTX resistive touch screen LCD monitor. Responses were self-paced with the next stimulus appearing only after the participant had responded to the previous stimulus. Void responses were recorded if a participant touched the screen outside of the areas designated for each response option. The outcome measure of interest was the number of correct responses at each intensity level for each emotion. Reaction time was not analysed as participants were not instructed to respond as quickly as possible.

Reading the Mind in the Eyes Test

This task is described in detail by Baron-Cohen, Wheelwright, Hill, Raste, and Plumb (2001). The participant is shown a single picture of the eye region of a face presented on an A4 page. The picture is surrounded by four adjectives describing a mental state (e.g. perplexed, horrified, astonished, intrigued). The participant is instructed to identify which of the words they think best describes what the person in the picture is thinking or feeling and circle their choice on a separate answer sheet. After a single practice item, 36 experimental items are completed one after the other in a self-paced manner. Response time is not recorded. The outcome measure of interest was the number of correct responses.

Procedure

The tests were administered as part of a larger battery of neuropsychological tests. All participants received the FERT-static test before the Eyes Test with unrelated tasks in between. The whole assessment took place over approximately two hours and participants were able to take breaks.

Data Analysis

Data were analysed using SPSS version 17.0. A significance level of $p<0.05$ was adopted. Patient and control groups were compared using independent samples *t*-tests, χ^2-tests or, for tests that involved multiple levels or repetitions, repeated measures ANOVA. For *t*-tests, when Levene's *F*-test identified instances of unequal variance, corrected *p*-values were reported.

RESULTS

FERT-static

The results of the facial expression recognition task are shown in Table 2. The results of a five (emotion: angry, disgusted, fearful, happy, sad) x four (intensity: 20%, 40%, 60%, 80%) x two (group: patient, control) repeated measures ANOVA indicated that there was no significant main effect of group ($F_{1,64}=0.59$, $p=0.45$, $\eta^2=0.01$). There was a significant main effect of emotion ($F_{4,256}=66.44$, $p<0.001$, $\eta^2=0.51$) and intensity ($F_{3,192}=583.77$, $p<0.001$, $\eta^2=0.90$). Follow-up paired t-tests indicated the main effect of emotion reflected that happy expressions were significantly more easily recognised than each of the other emotions (all *p*<0.05) and anger was significantly more poorly recognised than four of the other emotions (all *p*<0.05) but not sadness (*p*=0.097). There was a significant group x intensity interaction ($F_{3,192}$=2.96, *p*=0.034, η^2=0.04) but follow-up independent samples t-tests did not indicate a significant difference between the groups at any intensity level (all *p*>0.084) and so the effect could not be related to particular comparisons. The group x emotion interaction was not significant ($F_{4,256}$=834.59, *p*=0.34, η^2=0.02). The three-way interaction between group, intensity and emotion was not significant ($F_{12,768}$=176.93, *p*=0.91, η^2=0.01). Using an independent samples t-test, there was no significant difference between the two groups for recognition of neutral faces (t_{64}=0.81, *p*=0.42).

Comment [LJR8]: Mike suggested this might be due to ceiling effects at higher intensities, but this doesn't seem to be the case. It seems to be due to slightly worse performance by the patients at the 20% & 40% levels which when aggregated is enough to create a difference, but when considering the intensities separately isn't sufficient for a significant difference

Comment [PeterG9]: Typo?

Comment [PeterG10]: Typo?

Reading the Mind in the Eyes Test

There were no significant differences between patients and controls for this task (patient mean (s.d.) = 26.69 (4.03), control mean (s.d.) = 26.79 (3.5), t_{62}=0.10, *p*=0.93).

SUMMARY OF STUDY 1

In a well-characterised, prospectively verified sample of euthymic patients with BD there were no significant differences in emotion labelling of static images of facial expression of primary or secondary emotions compared to controls. Images were presented at low intensities, which makes the task more difficult and therefore more likely to expose group differences, but despite this no statistically significant differences were observed. The effect sizes indicate small effects (0.2<d<0.5) for recognition of angry, disgusted and fearful expressions at the higher intensity levels indicating poorer recognition by the patient sample. There was a small effect size indicating better recognition of happiness at the lowest intensity for the patient group. Thus there may be subtle differences in processing and labelling emotions that may become ~~more obvious~~clearer when patients are symptomatic or when stimuli are more naturalistic.

Comment [PeterG11]: Might be adding that people weren't at floor either i.e. that it was impossible for either group to do.

Might also be worth noting how this links to the depressed study where the intensities were tweaked slightly in this regard.

STUDY 2: DEPRESSION

In a second study, we aimed to examine emotional expression labelling in bipolar patients who were currently in a depressive episode. We also administered a dynamic version of the facial emotion recognition test, an approach which has been suggested to hold many advantages over typical static displays, including increased ecological validity (for a review see Krumhuber, Kappas, & Manstead, 2013). In addition we administered a standardised, well-validated 'static' facial emotion labelling task from the FEEST battery (Facial expressions of emotion: stimuli and tests).

Comment [PeterG12]: Ref.

Young A, Perret D, Calder A, Sprengelmeyer R, Ekman P (2002): Facial expressions of emotion: stimuli and tests (FEEST). Bury, St. Edmunds: Thames Valley Test Company.

METHODS

Participants

One hundred participants were recruited (n=53 bipolar patients and n=47 matched controls). Recruitment was part of a larger ~~project~~ programme of research examining neuropsychological function as an outcome measure in a pharmacological treatment trial (Gallagher, Gray, & Kessels, 2014; Gallagher, Gray, Watson, Young, & Ferrier, 2014; Watson et al., 2012).

Patients were aged 18 to 65 years with a diagnosis of BD, confirmed using the Structured Clinical Interview for DSM-IV (SCID; First et al., 1995), and were recruited from secondary and tertiary care services in North East of England. All were out-patients and currently in a SCID defined depressive episode ~~(SCID defined)~~. Patients were excluded if they met criteria for any other current Axis I disorder or substance dependence/abuse. All were receiving medication at the time of testing and had remained stable for a minimum of 4 weeks. Healthy control subjects were recruited by general advertisement. All controls were screened prior to testing to exclude anyone with a personal or family history (first-degree) of psychiatric illness, significant medical or neurological illness likely to affect neuropsychological functioning, or history of drug/alcohol abuse. After a complete description of the study, written informed consent was obtained from all participants. The study was approved by the Newcastle and North Tyneside Local Research Ethics Committee.

Demographic details are shown in table 1.

Measures

Facial Expression Recognition Task – Dynamic Images (FERT-dynamic)

Similar to the FERT-static, this version of the task uses faces from the original Ekman and Friesen (1976) set, cropped to isolate the face. Two male and two female faces were used (sets: *jj*, *pe*, *pf*, *mo*). The program rapidly display the images (~50ms per image), which change from neutral (0% intensity) to the full prototypical emotion (100% intensity) in 5% steps, producing a dynamic morphing effect. This 1 second 'stream' can be terminated at any of these steps allowing emotional morphs of 5% increments to be possible. For this study, after a short practice block, 80 trials were randomly administered, divided into 4 blocks, permitting a rest between each block. In total there were 16 trials for each of 5 emotions (happy, sad, anger, disgust, fear). For each of these emotions, 4 intensity levels were used (30%, 50%, 70% and 100%).

Benton Facial Recognition Test (short-form) (Benton, Sivan, Hamsher, Varney, & Spreen, 1983)

The BFRT was administered as a control task to examine general face recognition ability. The short form contains 13 trials with a maximum score of 27. On each item, participants are presented with a target black and white photograph and are asked to choose the target individual from six faces, presented simultaneously with the target photograph.

Facial expressions of emotion: stimuli and tests (FEEST) – Emotional Hexagon

The Emotional Hexagon test from the FEEST was administered according to the standardized instructions (A. Young, Perret, Calder, Sprengelmeyer, & Ekman, 2002). The test utilises a single actor (JJ) from the (Ekman and Friesen (1976)) set displaying 6 emotional expressions (happiness, surprise, fear, sadness, disgust, and anger). Each emotion is blended with the two it is most often confused with, resulting in blends over five continua: happiness–surprise, surprise–fear, fear–sadness, sadness–disgust, disgust–anger; the final blend from anger–happiness completes the circle. The blends are displayed in five different proportions of the two emotions: 90%:10%, 70%:30%, 50%:50%, 30%:70%, 10%:90%. This results in 30 unique stimuli which are displayed randomly 5 times each over the course of the task, giving a total of 150 experimental trials.

RESULTS

Two patients did not complete the emotion recognition tasks so results are presented for the remaining 51 who had full valid data.

FERT-dynamic

The results of the facial expression recognition task using dynamic stimuli in depressed patients are shown in Table 3. The results of a five (emotion: angry, disgusted, fearful, happy, sad) x four (intensity: 30%, 50%, 70%, 100%) x two (group: patient, control) repeated measures ANOVA indicated that there was no significant main effect of group ($F_{1,96}$=2.23, p=0.14, η^2=0.02). There was a significant main effect of emotion ($F_{4,384}$=76.77, p<0.001, η^2=0.44) indicating differences in the accuracy of overall emotional labelling (ranging from happy being the most easily detected; average collapsed across group and intensity = 95.9%, and disgust being the most difficult; 58.4%) and a significant main effect of intensity ($F_{3,288}$=104.30, p<0.001, η^2=0.52), with accuracy increasing with increasing intensity. There was no significant group x emotion interaction ($F_{4,384}$=0.71, p=0.59, η^2=0.01) and no three-way interaction between group, intensity and emotion ($F_{12,1152}$=1.15, p=0.31,

η^2=0.01), although the group x intensity interaction was significant ($F_{3,288}$=2.96, p=0.033, η^2=0.03), with patients being worse at the 30% intensity level compared to controls.

The effect sizes showed a small effect size difference for the recognition of disgust and happiness at the lowest intensity level, indicating poorer recognition by the patients. Small effects were also noted for poorer recognition of fear by the patients at the 30%, 50% and 100% intensity levels. There was a medium effect size ($0.5 \leq d < 0.8$) again showing poorer performance by the patients for the recognition of anger at the lowest intensity level. These are commensurate with the magnitude of effect sizes noted in euthymic patients, not larger as anticipated. As for the euthymic sample, the majority of the calculated effect sizes were $d<0.2$ suggesting minor differences between the groups.

BFRT

BD patients were significantly poorer than controls on the BFRT (t_{98}=-2.41, p=0.02), although this corresponded to only a 1-point difference in performance (BD: mean=22.8, s.d.=2.32; Controls: mean=23.8, s.d.=1.72).

FEEST

Data from the Emotional Hexagon paradigm were available in a sub-set of 51 participants (26 bipolar depressed patients and 25 controls). The results of a six (emotion: angry, disgusted, fearful, happy, sad, surprized) x two (group: patient, control) repeated measures ANOVA indicated that there was no significant main effect of group ($F_{1,49}$=1.56, p=0.22, η^2=0.3) or group x emotion interaction ($F_{5,245}$=0.31, p=0.85, η^2=0.01) (see figure 2). A significant main effect of emotion was observed ($F_{5,245}$=13.66, p<0.0001, η^2=0.22). Pairwise comparisons revealed that overall, while not differing from each other, accuracy for happy and sad faces was significantly higher than for all other emotions. Conversely, while not differing from each other, -accuracy for disgusted, angry and fearful faces was significantly lower than all other emotions (*p*<0.05).

GENERAL DISCUSSION

There were no significant differences between patient and control groups on any of the emotional expression measures used in the present study. Contrary to expectations, differences did not emerge in symptomatic (depressed) groups or as the stimuli became increasingly face-valid (i.e. dynamic expressions) or static facial expressions that were blends of different emotions or secondary emotions). This differs from the recent findings of van Rheenen & Rossell (2014), where a general deficit in emotion recognition and discrimination was observed (T. E. Van Rheenen & S. L. Rossell, 2014). It is worth noting that unlike their study, the present studies did not include measures of emotion discrimination. Nonetheless, van Rheenen & Rossell (2014) noted differences on the emotion recognition measures in their study that were not evident in the present study on similar tasks (emotion recognition of static or dynamic images displayed at different intensity levels). Our sample included only patients in either the euthymic or depressed phase of illness and explored the two groups separately. Combining groups of patients in different symptomatic states and including patients in the manic or hypomanic state could be one reason why the results differ.

Comment [PeterG13]: Can we really say these are more face-valid? I've moved them out of the brackets so they are now just other examples.

The relatively comprehensive set of emotion recognition tests, including paradigms that are generally considered more difficult and therefore more likely to expose a deficit or bias (e.g. static images of low-intensity emotions), combined with larger sample sizes and well-characterised patient groups are strengths of the present study. It cannot be ruled out that low statistical power is a concern, as it remains in many studies in patient samples (Van Rheenen & Rossell, 2013). The present analyses were adequately powered (β≥80%) to identify large effect size differences for main effects of group, however power was lower to detect smaller effect sizes, especially for the interactions. The calculated effect sizes indicated small effects on some measures, although many were also below this threshold (d<0.2). This study adds to others (Addington & Addington, 1998; Bellack et al., 1996; Edwards et al., 2001; Lembke & Ketter, 2002; Vaskinn et al., 2007) that have not reported evidence of significant impairment in facial emotion recognition in euthymic or depressed patients with BD. It is difficult to infer directly from statistical effect size to clinical significance, but it seems this element of emotion processing (specifically the labelling of displayed emotion) may be of limited importance in understanding the presentation of those with this disorder.

~~In contrast,~~It is important to note that the patient samples reported here did show significant neuropsychological deficits with large effect sizes in many domains of ‘cold’ cognition (Gallagher, Gray, Watson, et al., 2014; Robinson, 2010) and therefore the absence of differences is not a consequence of recruiting high-performing patients with BD. To derive a sense of the relative scale of ‘impairment’, it is not uncommon to ascertain the proportion of the patient group falling below the 5th or 10th percentile of the control group (Gallagher, Gray, Watson, et al., 2014; Thompson et al., 2005). In the euthymic sample, the proportion of patients scoring below the 10th percentile for the controls on cognitive measures that were administered alongside the facial expression recognition tests ranged from 2.6%-53.8%. These tests included measures of executive function, verbal declarative memory, working memory and psychomotor speed, and those domains showing the largest proportion of low-scoring patients were executive measures (category fluency, 53.8%) and verbal declarative memory (list-learning total recall, 42.1%). In contrast, the proportion of those scoring below the controls’ 10th percentile on the expression recognition test ranged from 2.6%-15.8% suggesting there is less evidence of potential impairment in the patient group on these measures. Data for the depressed patients showed a similar pattern. The cognitive performance of the depressed sample is detailed elsewhere (Gallagher, Gray, Watson, et al., 2014). In these studies, the depressed BD patients performed significantly worse on 18/26 measures examined, with large effect sizes (d>0.8) on tests of speed of processing, verbal

Comment [MB14]: Key sentence to whole paper. Next one also very important and needs expanded so that the reader understands the difference between the sizes of effects. In some ways the ultimate paper to support the argument which you are making would show that cold cognition was a better predictor of social ability within BD than emotional expression recognition.
But here just showing that the effect sizes are smaller for the facial expression tasks would be interesting.

Comment [LJR15]: I have this data for the euthymic sample. The strongest predictors of social functioning were ‘soft psychological measures’ (e.g. self-esteem, dysfunctional attitudes), then cognitive measures. FER didn’t really feature. We don’t have social functioning measures on the depressed sample. The plan is to publish that analysis separately. This can be a stepping stone towards that.

Comment [PeterG16]: Are we including this data anywhere or have I overlooked it?

learning and specific executive/working memory processes. Almost all tests produced at least one outcome measure on which ~25–50% of the BD sample performed at more than 1 standard deviation (S.D.) below the control mean. ~~—in summary those~~Patients performing below the controls' 10th percentile for measures of accuracy ranged from ~~7.1~~11.3%-47.2%. In contrast for the facial expression recognition task it ranged from 9.8%-23.5%.

Given the extent of these neuropsychological deficits ~~across many domains~~, it ~~would be unsurprising if~~ might be that where individuals with BD have ~~showed~~ shown performance deficits ~~in performance~~ on tasks involving facial expression perception, ~~owing to their more~~this may be because of general difficulties in performing the task, ~~itself~~ rather than deficits ~~specific to~~in facial expression perception *per se*. However, t~~T~~he effect of such general deficits might be expected to be fairly small (since one would hope that the assessments of facial expression perception have a good degree of specificity) and ~~would~~ emerge as significant in a fairly random fashion in some experiments but not others and, within these experiments, in some conditions but not others; ~~. T~~this pattern describes the literature reviewed previously in BD. For example, ~~Within the literature it would be expected that~~where facial expression perception experiments and analyses ~~with a greater~~ overlap with cognitive domains in which individuals with BD have ~~been shown to be deficient~~deficits, it would be more likely to find significant results ~~and results~~ with a greater effect size. It is of interest that in fMRI studies it has been demonstrated that patterns of activation differ according to the demands of the task. Direct matching of emotional facial expressions (i.e. choosing between two faces in which the emotion displayed in one matched that of a target face) has been found to increase amygdala activation while the selection of the label that matches (e.g. 'afraid' or 'angry') results in greater right prefrontal cortex activation (Hariri et al., 2000). Therefore tasks which examine discrimination of emotions compared to direct labelling of emotion may be tapping different aspects of processing.

Formatted: Font: Italic

Comment [PeterG17]: Hariri AR, Bookheimer SY, Mazziotta JC (2000): Modulating emotional responses: effects of a neocortical network on the limbic system. *Neuroreport* 11:43-48.

This These methodological differences may partially account for some of the variability in findings to date. For example, —tasks with a response format that has have a high declarative memory load (e.g.such as remembering the response options or learning the paired-associations between emotions and which response-key to press) or those with complex instructions or time-pressured responses may be more likely to indicate lead to group differences. that actually result from cognitive difficulties rather than a facial expression recognition deficit per se. Future studies should also consider how the specifics of the response format can potentially affect the outcome of studies of this nature. For example, it is important to be mindful that the majority of studies are fixed-choice paradigms (i.e. there is no “don’t know” option). Therefore if stimuli are presented quickly or are ambiguous, participants still have to press one of the options to move to the next trial. Therefore patients (who may simply be slightly slowed in general processing speed or decision making) are more likely to ‘miss’ stimuli and which requires a random response to move on – this is not an emotional processing bias/deficit. It should also be noted that in tasks of this nature, the majority of the responses are in fact ‘negative’, with the ‘happy’ response being the only overt positive emotion. Therefore any form of response bias will lead to a ‘deficit’ in negative emotion perception of some sort.

A further point to consider is how findings in this area are interpreted. For example, results that have demonstrated reduced accuracy of labelling specific expressions have been interpreted as supporting the notion that emotion perception decrements are evident in BD (Vederman et al., 2012). Other studies have interpreted increased correct recognition of specific emotions (e.g. disgust) as possibly being linked to low self-esteem and other cognitive biases in BD (Harmer et al., 2002). It is therefore important to consider the precise nature of the task demands and the social processes being assessed to avoid a situation in which both increased and decreased accuracy is considered as reflecting a 'negative outcome'.

There are a number of limitations of the present study to be considered. Firstly, low statistical power for the interaction analyses has already been mentioned. This difficulty is commonly encountered in this area of investigation and is likely to contribute to the varied findings. More widespread reporting of effect sizes alongside inferential statistics would help clarify whether studies are broadly finding group differences of a similar magnitude or, if not, it may help to identify which methodological variations impact most markedly on group differences. Secondly, not all tests were administered to both patient groups, which raises the possibility that some measures may have shown differences, ~~between groups~~ had ~~all of them~~both groups received the same tasks. However, three of the tasks used the same image set and similar intensities of emotions and all involved a range of difficulty in the stimuli presented, thereby offering the opportunity for even a subtle deficit to become evident. Also, using the two different experimental expression recognition tests suggests the lack of difference is not specific to a methodological feature of one particular task. Furthermore, the depressed sample were administered standardised measures (~~BFRT and~~e.g. the Emotional Hexagon from the FEEST) alongside the other tasks and did show pronounced deficits ~~on~~ in other aspects of ~~the test battery they received~~cognitive function. Thirdly, although we utilised a dynamic emotional expression task to increase the ecological validity of the task, some studies have suggested that ~~the~~ dynamic facial movements ~~actually~~ play only a small role in the ability to identify emotion from facial expressions (Gold et al., 2013). Nonetheless, employing different variants of facial emotion stimuli develops our understanding of the robustness or otherwise of any effect irrespective of an impact on ecological validity.

Based upon our current findings and the mixed findings of the literature we conclude there is little evidence of abnormalities in explicit facial emotion identification in euthymic or depressed patients, within the parameters examined in the present studies. Future studies should address the methodological issues in this area of research, in order to build a more complete picture of emotion processing in BD and how or whether it is of relevance in our understanding of this illness.

References

Addington, J., & Addington, D. (1998). Facial affect recognition and information processing in schizophrenia and bipolar disorder. *Schizophrenia Research, 32*(3), 171-181.

APA. (1994). *Diagnostic and Statistical Manual of Mental Disorders Fourth Edition (DSM-IV)*. Washington DC: American Psychiatric Association.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42*(2), 241-251.

Bellack, A. S., Blanchard, J. J., & Mueser, K. T. (1996). Cue Availability and Affect Perception in Schizophrenia. *Schizophrenia Bulletin, 22*(3), 535-544. doi: 10.1093/schbul/22.3.535

Benton, A., Sivan, A., Hamsher, K., Varney, N., & Spreen, O. (1983). *Contribution to neuropsychological assessment*. NY: Oxford University Press.

Bonnín, C. M., Martínez-Arán, A., Torrent, C., Pacchiarotti, I., Rosa, A. R., Franco, C., . . . Vieta, E. Clinical and neurocognitive predictors of functional outcome in bipolar euthymic patients: A long-term, follow-up study. *Journal of Affective Disorders, In Press, Corrected Proof*.

Bozikas, V. P., Tonia, T., Fokas, K., Karavatos, A., & Kosmidis, M. H. (2006). Impaired emotion processing in remitted patients with bipolar disorder. *Journal of Affective Disorders, 91*(1), 53-56.

Edwards, J., Pattison, P. E., Jackson, H. J., & Wales, R. J. (2001). Facial affect and affective prosody recognition in first-episode schizophrenia. *Schizophrenia Research, 48*(2–3), 235-253. doi: http://dx.doi.org/10.1016/S0920-9964(00)00099-2

Ekman, P., & Friesen, W. V. (1976). *Pictures of Facial Affect*. Palo Alto, CA: Consulting Psychologists Press.

Fagiolini, A., Kupfer, D. J., Masalehdan, A., Scott, J. A., Houck, P. R., & Frank, E. (2005). Functional impairment in the remission phase of bipolar disorder. *Bipolar Disorders, 7*(3), 281-285.

First, M. B., Spitzer, R. L., Williams, J. B. W., & Gibbon, M. (1995). *Structured Clinical Interview for DSM-IV (SCID-IV)*. New York: Biometrics Research Department, New York State Psychiatric Institute.

Gallagher, P., Gray, J. M., & Kessels, R. P. (2014). Fractionation of visuo-spatial memory processes in bipolar depression: a cognitive scaffolding account. *Psychological Medicine*, 1-14. doi: 10.1017/s0033291714001676

Gallagher, P., Gray, J. M., Watson, S., Young, A. H., & Ferrier, I. N. (2014). Neurocognitive functioning in bipolar depression: a component structure analysis. *Psychological Medicine, 44*(5), 961-974. doi: 10.1017/s0033291713001487

Getz, G. E., Shear, P. K., & Strakowski, S. M. (2003). Facial affect recognition deficits in bipolar disorder. *Journal of the International Neuropsychological Society, 9*(4), 623-632.

Goetz, I., Tohen, M., Reed, C., Lorenzo, M., & Vieta, E. (2007). Functional impairment in patients with mania: baseline results of the EMBLEM study. *Bipolar Disorders, 9*(1-2), 45-52. doi: 10.1111/j.1399-5618.2007.00325.x

Gold, J. M., Barker, J. D., Barr, S., Bittner, J. L., Bromfield, W. D., Chu, N., . . . Srinath, A. (2013). The efficiency of dynamic and static facial expression recognition. *J Vis, 13*(5). doi: 10.1167/13.5.23

Gray, J., Venn, H., Montagne, B., Murray, L., Burt, M., Frigerio, E., . . . Young, A. H. (2006). Bipolar patients show mood-congruent biases in sensitivity to facial expressions of emotion when exhibiting depressed symptoms, but not when exhibiting manic symptoms. *Cogn Neuropsychiatry, 11*(6), 505-520. doi: 10.1080/13546800544000028

Green, M. J., Cahill, C. M., & Malhi, G. S. (2007). The cognitive and neurophysiological basis of emotion dysregulation in bipolar disorder. *Journal of Affective Disorders, 103*, 29-42. doi: 10.1016/jjad.2007.01.024

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology Neurosurgery and Psychiatry, 23*, 56-62.

Harmer, C. J., Grayson, L., & Goodwin, G. M. (2002). Enhanced recognition of disgust in bipolar illness. *Biological Psychiatry, 51*(4), 298-304.

Krumhuber, E. G., Kappas, A., & Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: A review. *Emotion Review, 5*(1), 41-46.

Lembke, A., & Ketter, T. A. (2002). Impaired recognition of facial emotion in mania. *American Journal of Psychiatry, 159*(2), 302-304.

MacQueen, G. M., Young, L. T., & Joffe, R. T. (2001). A review of psychosocial outcome in patients with bipolar disorder. *Acta Psychiatrica Scandinavica, 103*(3), 163-170.

Miklowitz, D. J. (2011). Functional impairment, stress, and psychosocial intervention in bipolar disorder. *Current Psychiatry Reports, 13*(6), 504-512. doi: 10.1007/s11920-011-0227-x

Robinson, L. J. (2010). *Neuropsychological Performance, Emotion Processing and Psychosocial Function in Bipolar Disorder.* (PhD), Newcastle University, Newcastle upon Tyne.

Sanchez-Moreno, J., Martinez-Aran, A., Tabares-Seisdedos, R., Torrent, C., Vieta, E., & Ayuso-Mateos, J. L. (2009). Functioning and disability in bipolar disorder: an extensive review. *Psychotherapy & Psychosomatics, 78*(5), 285-297. doi: 000228249 [pii]

10.1159/000228249 [doi]

Schaefer, K. L., Baumann, J., Rich, B. A., Luckenbaugh, D. A., & Zarate, C. A., Jr. (2010). Perception of facial emotion in adults with bipolar or unipolar depression and controls. *Journal of Psychiatric Research, 44*(16), 1229-1235. doi: 10.1016/j.jpsychires.2010.04.024

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*, 22-33; 34-57.

Summers, M., Papadopoulou, K., Bruno, S., Cipolotti, L., & Ron, M. A. (2006). Bipolar I and bipolar II disorder: cognition and emotion processing. *Psychological Medicine, 36*(12), 1799-1809.

Thompson, J. M., Gallagher, P., Hughes, J. H., Watson, S., Gray, J. M., Ferrier, I. N., & Young, A. H. (2005). Neurocognitive impairment in euthymic patients with bipolar affective disorder. *British Journal of Psychiatry, 186*, 32-40.

Tiddeman, B., Burt, M., & Perrett, D. (2001). Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications*, *21*(5), 42-50. doi: 10.1109/38.946630

Van Rheenen, T. E., & Rossell, S. L. (2013). Is the non-verbal behavioural emotion-processing profile of bipolar disorder impaired? A critical review. *Acta Psychiatrica Scandinavica*, *128*(3), 163-178. doi: 10.1111/acps.12125

Van Rheenen, T. E., & Rossell, S. L. (2014). Let's face it: facial emotion processing is impaired in bipolar disorder. *Journal of the International Neuropsychological Society, 20*(2), 200-208. doi: 10.1017/s1355617713001367

Van Rheenen, T. E., & Rossell, S. L. (2014). Objective and subjective psychosocial functioning in bipolar disorder: An investigation of the relative importance of neurocognition, social cognition and emotion regulation. *Journal of Affective Disorders*, *162*(0), 134-141. doi: http://dx.doi.org/10.1016/j.jad.2014.03.043

Vaskinn, A., Sundet, K., Friis, S., Simonsen, C., Birkenaes, A. B., Engh, J. A., . . . Andreassen, O. A. (2007). The effect of gender on emotion perception in schizophrenia and bipolar disorder. *Acta Psychiatrica Scandinavica, 116*(4), 263-270. doi: 10.1111/j.1600-0447.2007.00991.x

Vederman, A. C., Weisenbach, S. L., Rapport, L. J., Leon, H. M., Haase, B. D., Franti, L. M., . . . McInnis, M. G. (2012). Modality-specific alterations in the perception of emotional stimuli in Bipolar Disorder compared to Healthy Controls and Major Depressive Disorder. *Cortex*, *48*(8), 1027-1034. doi: http://dx.doi.org/10.1016/j.cortex.2011.03.017

Venn, H. R., Gray, J. M., Montagne, B., Murray, L. K., Michael Burt, D., Frigerio, E., . . . Young, A. H. (2004). Perception of facial expressions of emotion in bipolar disorder. *Bipolar Disorders, 6*(4), 286-293.

Watson, S., Gallagher, P., Porter, R. J., Smith, M. S., Herron, L. J., Bulmer, S., . . . Ferrier, I. N. (2012). A randomized trial to examine the effect of mifepristone on neuropsychological performance and mood in patients with bipolar depression. *Biological Psychiatry*, *72*(11), 943-949. doi: 10.1016/j.biopsych.2012.05.029

Young, A., Perret, D., Calder, A., Sprengelmeyer, R., & Ekman, P. (2002). *Facial expressions of emotion: stimuli and tests (FEEST)*. Bury, St. Edmunds: Thames Valley Test Company.

Young, R., Biggs, J., Ziegler, V., & Meyer, D. (1978). A rating scale for mania: reliability, validity and sensitivity. *British Journal of Psychiatry, 133*(5), 429-435.

Table 1: Demographic details of the patient samples

	Control		Patient			
	mean	**s.d.**	**mean**	**s.d.**	**t/χ^2**	**p**
Euthymic Group	n=28		n=38			
Demographics						
Age	46.5	10.8	44.8	12.8	0.54	0.592
Male/female						
Male: n (%)	13	(46.4)	17	(44.7)	0.00	0.982
Female: n (%)	15	(53.6)	21	(55.3)		
NART IQ	114.4	8.9	111.2	9.6	1.62	0.110
Years of education	16.8	2.9	15.5	3.8	1.72	0.090
Mood Symptoms						
HDRS-17	-	-	3.8	2.1	-	-
YMRS	-	-	0.7	1.6	-	-
BDI	1.3	1.8	7.2	6.9	-4.76	<0.001
AMRS	2.1	3.3	1.8	3.3	0.44	0.665
Depressed Group	n=47		n=53			
Demographics						
Age	45.0	13.7	47.3	9.6	0.97	0.343
Male/female						
Male: n (%)	28	(59.6)	33	(62.3)	0.08	0.783
Female: n (%)	19	(40.4)	20	(37.7)		
Nart IQ	112.5	11.2	108.9	10.5	1.63	0.107
Years of education	14.4	4.0	14.4	3.2	0.05	0.961
Mood Symptoms						
HDRS-17	-	-	19.7	4.9	-	-
YMRS	-	-	1.5	1.8		
BDI	1.0	1.5	26.0	11.4	10.46	<0.001
AMRS	-	-	-	-	-	-

NART, National Adult Reading Test; HDRS-17, Hamilton Depression Rating Scale 17-item; YMRS, Young Mania Rating Scale; BDI, Beck Depression Inventory; AMRS, Altman Mania Rating Scale

Table 2: Results of the facial expression recognition task in euthymic patients using static stimuli. Means and standard deviations of % correct at each intensity level for each emotion.

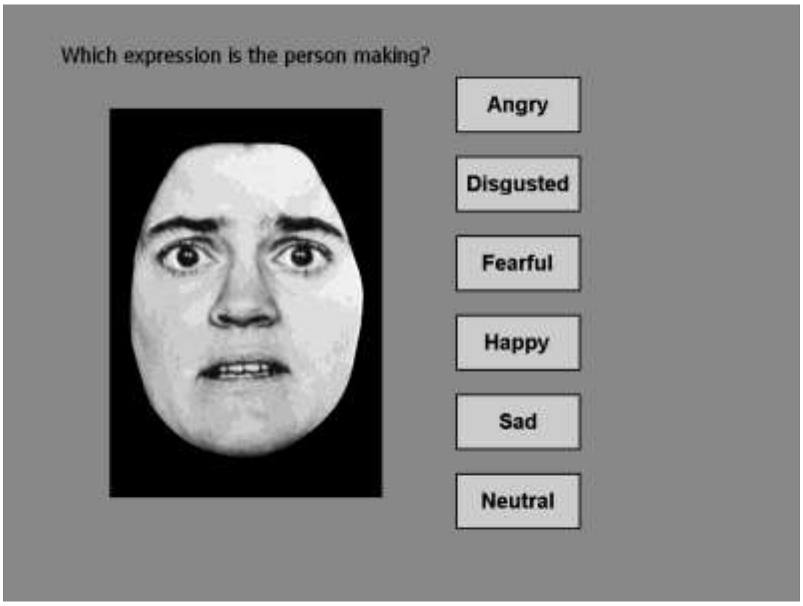
	Control (n=28)		**Bipolar (n=38)**		
	mean	**s.d.**	**mean**	**s.d.**	**d**
Angry					
Correct 20%	9.82	12.43	10.53	14.97	-0.05
Correct 40%	31.25	22.18	32.24	20.06	-0.05
Correct 60%	63.39	31.54	50.66	24.31	0.46
Correct 80%	78.57	24.26	70.39	23.86	0.34
Correct Total %	45.78	16.66	40.95	13.39	0.33
Disgust					
Correct 20%	6.25	14.63	5.26	11.85	0.08
Correct 40%	45.54	24.58	40.13	26.98	0.21
Correct 60%	76.79	22.49	65.79	29.88	0.41
Correct 80%	76.79	25.39	71.71	22.64	0.21
Correct Total %	51.35	14.96	45.73	15.91	0.36
Fear					
Correct 20%	10.71	15.85	16.45	20.37	-0.31
Correct 40%	66.07	26.54	63.82	20.71	0.10
Correct 60%	83.93	20.65	76.97	27.50	0.28
Correct 80%	84.82	21.88	80.92	21.30	0.18
Correct Total %	61.40	15.22	59.54	14.73	0.12
Happy					
Correct 20%	35.71	26.73	46.71	28.58	-0.40
Correct 40%	85.71	18.54	84.87	21.39	0.04
Correct 60%	93.75	12.95	95.39	11.41	-0.14
Correct 80%	95.54	11.89	97.37	7.78	-0.19
Correct Total %	77.70	12.45	81.09	13.36	-0.26
Sad					
Correct 20%	16.96	18.07	18.42	18.09	-0.08
Correct 40%	43.75	26.02	46.05	32.64	-0.08
Correct 60%	59.82	26.65	60.53	25.09	-0.03
Correct 80%	66.07	29.04	65.79	23.55	0.01
Correct Total %	46.67	18.95	47.71	16.86	-0.06
Neutral					
Correct Total %	81.25	23.20	76.32	25.30	0.20
Void	0.21	1.13	0.26	0.76	-0.05

Table 3: Results of the facial expression recognition task in depressed patients using dynamic stimuli. Means and standard deviations of number correct at each intensity level for each emotion.

	Control (n=47)		Bipolar (n=51)		
	mean	**s.d.**	**mean**	**s.d.**	**d**
Angry					
Correct 30%	50.00	26.58	34.80	27.87	0.56
Correct 50%	63.30	28.00	58.82	26.84	0.16
Correct 75%	77.66	23.45	74.51	26.69	0.13
Correct 100%	81.91	24.84	79.90	25.50	0.08
Total%	68.22	17.57	62.01	19.28	0.34
Disgust					
Correct 30%	59.57	24.20	46.08	31.77	0.48
Correct 50%	55.85	29.59	58.82	33.85	-0.09
Correct 75%	62.77	28.96	62.25	28.45	0.02
Correct 100%	61.70	26.50	60.29	31.30	0.05
Total%	59.97	20.96	56.86	25.07	0.13
Fear					
Correct 30%	63.83	22.00	59.80	26.02	0.17
Correct 50%	82.45	20.13	73.53	24.19	0.40
Correct 75%	83.51	19.70	78.92	17.59	0.25
Correct 100%	78.72	22.71	74.51	20.30	0.20
Total%	77.13	16.60	71.69	15.68	0.34
Happy					
Correct 30%	90.96	16.00	86.76	18.27	0.24
Correct 50%	96.81	8.43	96.57	10.02	0.03
Correct 75%	97.87	7.05	98.53	5.94	-0.10
Correct 100%	100.00	0.00	99.51	3.50	0.19
Total%	96.41	5.94	95.34	6.71	0.17
Sad					
Correct 30%	51.60	30.13	52.94	29.00	-0.05
Correct 50%	59.57	29.28	62.75	29.31	-0.11
Correct 75%	62.23	26.52	60.29	31.69	0.07
Correct 100%	70.21	24.80	70.59	24.34	-0.02
Total%	60.90	21.86	61.64	22.40	-0.03

Figure 1: Screen shot of the Facial Expression Recognition Task using static images

a) *Stimulus presentation*



b) *Response phase*

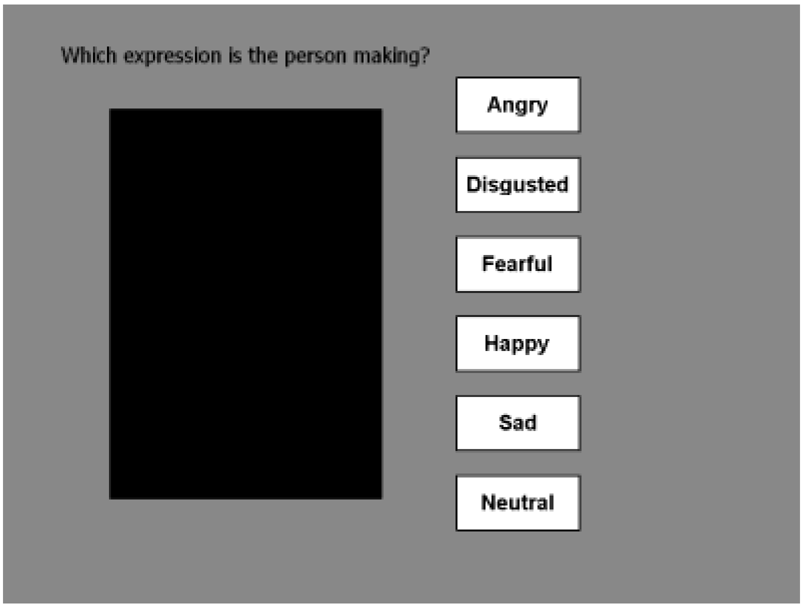


Figure 2: Results of the facial expression recognition task in depressed patients using static blended stimuli (emotional hexagon).