



Published in final edited form as:

Q J Exp Psychol (Hove). 2020 July ; 73(7): 1066–1081. doi:10.1177/1747021819890289.

Positive social feedback alters emotional ratings and reward valuation of neutral faces

Katherine S Young^{1,2}, Anni M Hasratian³, Christine E Parsons⁴, Richard E Zinbarg^{5,6}, Robin Nusslock⁵, Susan Y Bookheimer⁷, Michelle G Craske²

¹Social, Genetic and Developmental Psychiatry (SGDP) Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

²Anxiety and Depression Research Center (ADRC), Department of Psychology, University of California, Los Angeles (UCLA), Los Angeles, CA, USA

³Anxiety and Depression Research Center (ARC), Department of Psychology, Southern Methodist University, Dallas, TX, USA

⁴Interacting Minds Centre (IMC), Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

⁵Department of Psychology, Northwestern University, Evanston, IL, USA

⁶The Family Institute at Northwestern University, Evanston, IL, USA

⁷Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles (UCLA), Los Angeles, CA, USA

Abstract

Evaluation of facial and vocal emotional cues is vital in social interactions but can be highly influenced by characteristics of the observer, such as sex, age, and symptoms of affective disorders. Our evaluations of others' emotional expressions are likely to change as we get to know them and anticipate how they are likely to behave. However, the role of associative learning in the evaluation of social cues remains poorly understood. In this study, we investigated whether emotional ratings (valence and arousal) and reward valuation ("liking" and "wanting" measures) of neutral facial expressions can be altered through associative learning. We also examined whether emotional ratings and reward valuation varied with symptoms of anxiety and depression, disorders known to impair socio-affective functioning. Participants ($N = 324$) were young adults, ranging in scores across dimensions of depression and anxiety symptoms: "general distress" (common to depression and anxiety), "anhedonia-apprehension" (more specific to depression), and "fears" (more specific to anxiety). They rated neutral faces and completed a probabilistic learning

Article reuse guidelines: sagepub.com/journals-permissions

Corresponding author: Katherine S Young, Social, Genetic and Developmental Psychiatry (SGDP) Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, Memory Ln, Camberwell, London SE5 8AF, UK. katherine.s.young@kcl.ac.uk

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplementary material

Supplementary material is available at: <http://journals.sagepub.com/doi/suppl/10.1177/1747021819890289>.

task that paired images of neutral faces with positive or negative social feedback. Results demonstrated that pairing neutral faces with positive social feedback increased ratings of arousal, valence, and reward valuation (both “liking” and “wanting”). Pairing neutral faces with negative feedback reduced valence ratings and reduced “wanting,” but did not impact arousal ratings or “liking.” Symptoms of general distress were associated with negative bias in valence ratings, symptoms of anhedonia-apprehension were associated with reduced “wanting,” and symptoms of fears were associated with altered accuracy over trials. Notably, the association between general distress and negative bias was reduced following the associative learning task. This suggests that disrupted evaluation of social cues can be improved through brief training.

Keywords

Anxiety; depression; learning; face processing; emotion; reward

Introduction

We are highly sensitive to facial and vocal emotional cues from others. Telling a joke to a friend, for example, we eagerly await their response, anticipating whether they will laugh or not, comparing it to other times we made them laugh and subsequently updating our internal view of their disposition and sense of humour. However, facial and vocal emotional cues are inherently ambiguous, and reactions to social cues can differ by characteristics of the observer, including sex, age, and psychopathology (e.g., Hall & Matsumoto, 2004; Leppänen et al., 2004; Mill et al., 2009; Yoon & Zinbarg, 2008; Young et al., 2017). How we react to these social cues, and how it influences our behaviour in the future, involves both evaluative and associative learning processes, which have been shown to be altered in anxiety disorders and depression.

Information processing biases that favour negative over positive stimuli are theorised to play a central role in maintaining symptoms of affective disorders (Gotlib & Joormann, 2010; Mathews & MacLeod, 2005). To date, studies of biased information processing of social cues have focused on evaluation of stimuli at a single point in time (e.g., rating the valence of face stimuli viewed once). However, evaluation of social cues may also be altered through interactions with the same individual, for instance, by learning about their character through repeated exposure to their emotional responses. This study aimed to investigate whether this type of associative learning could alter emotional ratings and reward valuation of neutral facial expressions, and whether symptoms of anxiety and depression impact these processes. Improved understanding of how emotional responses to social cues can be altered through learning from daily life experiences may offer insight into how perceptual biases arise, and how they might be more effectively treated.

The role of learning in emotional ratings and reward valuation of social cues

The study of socio-affective responses to human faces has primarily taken a stimulus-driven approach, focusing on the physical attributes of faces which are more or less pleasant/ attractive, and how configurations of facial muscles communicate emotions (Adolphs, 2002; Ekman & Cordaro, 2011; Hahn & Perrett, 2014). These studies have demonstrated some

universalities in how individuals from different countries and cultures perceive facial attractiveness and the emotions being communicated by certain expressions (Ekman & Friesen, 1971; Elfenbein & Ambady, 2002; Rhodes et al., 2001). Nevertheless, observer-based processes can also impact evaluation of emotional cues from others. When we interact with a person we know well, we are often able to anticipate their responses, even in a novel situation. This relies on recall of associations encoded in memory that were formed during prior interactions with that individual. Recall of learned information is one “top-down” process that may impact the anticipation and evaluation of socio-emotional cues (Wieser & Brosch, 2012).

Prior experimental work has demonstrated that even simple conditioning procedures can modulate perceived valence and reward value of face stimuli. For example, pairing a face stimulus with positive statements was shown to increase subsequent likeability ratings for that face, whereas pairing with negative statements decreased likeability (Davis et al., 2009). In another study, associating faces with negative biographical information led to ratings of more negative valence in neutral facial expressions (Suess et al., 2015). These studies have typically relied on linguistic-based learning and paradigms with 100% reinforcement rates (i.e., one face is always presented with negative information). However, linguistic-based learning is not fully representative of daily social interactions, where we often learn about others’ emotional tendencies based on non-linguistic facial and vocal expressions of emotion. When we learn about individuals in this way, we are confronted with a more complex set of response contingencies, where individuals sometimes respond positively and sometimes respond negatively. Just because an individual reacts negatively in one context does not mean that person will always react negatively, or will react negatively in a similar context on a different day. We do not know whether learning based on probabilistic non-linguistic feedback can similarly impact emotional ratings and reward valuation of social cues. In addition, prior work was limited to a single measure of emotional or reward responses (rating of likeability or valence alone). As the current dominant model of emotion encompasses both valence and arousal, emotional responses can be more comprehensively assessed by including ratings of both of these dimensions (Posner et al., 2005). Responding to a “reward” stimulus is also a multi-faceted process, with separable constructs relating to its anticipation, experience and learning about the reward (Morris & Cuthbert, 2012; Rømer Thomsen et al., 2015), that may not be adequately captured in a single rating of likeability. Measuring “wanting” through effortful behaviour (in the form of a key-pressing task) is thought to additionally provide a more objective, or implicit, measure of reward valuation (Aharon et al., 2001; Parsons et al., 2011).

Our first goal was to investigate how probabilistic affective social feedback would impact emotional responses (ratings of valence and arousal) and reward valuation (measures of “liking” and “wanting”) of neutral facial stimuli. In a computer-based learning task, participants viewed pairs of neutral facial expressions. They were instructed to find out, through trial and error, which was the “happier” and “sadder” person in each pair. In each trial, they could select one of the faces in each pair which would then turn into a positive expression (happy face plus laugh sound) or a negative expression (sad face plus cry sound). Six different faces were paired with different probabilities of positive and negative feedback to investigate the impact of different contingencies on learning processes and changes in

emotional ratings and reward valuation. We predicted that overall positive social feedback would be associated with increased ratings of valence and greater reward valuation, whereas negative social feedback would be associated with decreased ratings of valence and reduced reward valuation. In addition, we predicted that the extent of changes in emotional ratings and reward valuation would relate to the amount of positive versus negative feedback.

Disrupted socio-affective processing with symptoms of anxiety and depression

Bias in information processing that favours negative over positive information has been proposed as a vulnerability and maintenance factor in depression and anxiety (Bistricky et al., 2011; Everaert et al., 2017; Gotlib & Joormann, 2010; Mathews & MacLeod, 2005; Mobini et al., 2013). These biases have been widely observed in relation to social stimuli with studies showing small but consistent negative interpretation bias among individuals with symptoms of anxiety disorders and depression when rating neutral or ambiguous facial or vocal expressions (Beevers et al., 2009; Bourke et al., 2010; Gebhardt & Mitte, 2014; Gollan et al., 2008; Joormann & Gotlib, 2006; Leppänen et al., 2004; Yoon & Zinbarg, 2008; Young et al., 2017).

In addition to negative bias, responses to socio-affective cues could be disrupted by altered reward processing and affective learning in anxiety and depression. Disrupted reward functioning is related to a cluster of symptoms known as “anhedonia,” the loss of motivation, interest, and pleasure of previously enjoyable experiences (American Psychiatric Association, 2013). Although common among individuals with depression, anhedonia is a transdiagnostic symptom cluster that can also affect individuals with anxiety disorders (Cooper et al., 2018; Rømer Thomsen et al., 2015). In experimental studies, symptoms of anhedonia have been linked to decreased motivation on effortful motor tasks for monetary (Treadway et al., 2009) or social rewards (Fussner et al., 2018).

Disrupted learning from affective stimuli has also been observed among individuals with anxiety and depression. Anhedonia has been linked to reduced learning in the context of rewarding stimuli (Pizzagalli et al., 2008; Whitton et al., 2015). Symptoms of anxiety are associated with heightened reactivity to threat, leading to overgeneralisation of learned fears and a reduced capacity to extinguish such learned associations (Pittig et al., 2018). For social cues, one recent study demonstrated that individuals with higher levels of symptoms of social anxiety disorder were more accurate at learning contingencies between face stimuli and likelihood of positive or negative feedback (Abraham & Hermann, 2015).

The second goal of this study was to replicate previous findings linking anxiety and depression with disruptions in emotional ratings and reward valuation of social stimuli. We had three specific hypotheses: (1) negative bias in emotional responses to neutral facial expressions (valence ratings) would be associated with symptoms common to anxiety and depressive disorders (a symptom dimension referred to as “general distress”), (2) symptoms of anhedonia (based on a symptom dimension we refer to as “anhedonia-apprehension”) would be related to reduced reward value of neutral facial stimuli, and (3) reduced learning performance (as assessed by cumulative accuracy) in response to positive stimuli would be related to anhedonia-apprehension, and heightened learning performance in response to negative stimuli would be related to anxiety-specific symptoms (a symptom dimension

referred to as “fears”). The third goal was to explore whether associative learning through social feedback would impact predicted associations between general distress and negative bias and between anhedonia-apprehension and reward value. We examined this by assessing: (1) whether training on a probabilistic feedback task impacted these associations, and (2) whether any changes observed were specific to stimuli presented during the training phase or generalised to non-trained stimuli.

Methods

Participant recruitment

Participants were recruited for the Brain, Motivation and Personality Development (BrainMAPD) study, a multisite longitudinal project investigating positive and negative valence functioning in late adolescence to early adulthood based at the University of California, Los Angeles and Northwestern University. Participants were recruited based on their scores on self-reported trait Neuroticism (Eysenck Personality Questionnaire-Neuroticism (EPQ-N); Eysenck & Eysenck, 1975; Kelley et al., 2019) and Reward Sensitivity (Behavioural Activation Scale (BAS); Carver & White, 1994) from among a total of 2,461 who completed these screening instruments. Participants were recruited to ensure sampling from high/mid/low ranges (tertiles) on both scales, with oversampling from the two diagonals of the bivariate space defined by the quasi-orthogonal EPQ-N and BAS scales (i.e., high EPQ-N/high BAS, low EPQ-N/low BAS, mid EPQ-N/mid BAS, high EPQ-N/low BAS, and low EPQ-N/high BAS). Other inclusion criteria were right-handed, no magnetic resonance imaging (MRI) contraindications (due to the inclusion of MRI assessments in the BrainMAPD project), fluent in English, not colour blind (requirement for a different experimental task), and aged 18–19 years old at the time of recruitment. A total of 324 participants recruited into the study completed the current task (214 female, mean age = 19.52 years, standard deviation (*SD*) = 0.73, see Table 1 for details and racial/ethnic composition of sample). The sample size was based on power calculations for the larger longitudinal study, aiming to detect small effect sizes with 80% power among 150 participants completing multiple assessments (accounting for attrition over a 3-year period). Participants provided written, informed consent and all procedures were approved by the institutional review board (IRB) at each institution.

Although this study was designed to use a dimensional approach to investigate broad symptom domains, diagnostic interviews were also conducted on the majority of participants ($n = 294$; 90.7%). Participants were assessed using the Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; DSM-5) (First & Williams, 2016), a semi-structured diagnostic interview. The proportions of individuals who met diagnostic criteria for anxiety and depressive disorders are reported in Table 1.

Self-report assessment of anxiety and depression symptoms

Symptoms of anxiety and depression were assessed in a dimensional framework, using factor analytic methods to generate scores across distinct symptom clusters. In previous work, a tri-level model of anxiety and depression was identified based on factor analyses of self-reported symptoms in studies of adolescents and adults (Naragon-Gainey et al., 2016;

Prenoveau et al., 2010). These analyses identified one “broad” and two “intermediate” symptom factors: general distress (common to anxiety and depression), fears (more specific to anxiety disorders), and anhedonia-apprehension (more specific to depression; see statistical analysis section below for details on factor score calculation).

Participants completed 101 questionnaire items selected from self-report measures of symptoms of anxiety and depression. Sixty-seven of these items were those that Prenoveau et al. (2010) used to create their tri-level hierarchical model, originating from five self-report measures: the Fear Survey Schedule-II (FSS; Geer, 1965), the Albany Panic and Phobia Questionnaire (APPQ; Rapee et al., 1994), the Self-Consciousness subscale of the Social Phobia Scale (SPS; Mattick & Clarke, 1998; Zinbarg & Barlow, 1996), the Inventory to Diagnose Depression (IDD; Zimmerman et al., 1986), and the Mood and Anxiety Symptom Questionnaire (MASQ; Watson et al., 1995). The remaining 34 items were the full scales of the Penn State Worry Questionnaire (PSWQ; Meyer et al., 1990) and the Obsessive-Compulsive Inventory Revised (OCIR; Foa et al., 2002), included to better characterise symptoms of generalised anxiety disorder (GAD) and obsessive-compulsive disorder (OCD).

The 50-item FSS (Geer, 1965) examines symptoms representative of specific phobia. The FSS asks participants to identify how much fear they would experience if they encountered a particular situation or stimulus (0 = *none*, 3 = *some fear*, 6 = *terror*). The participants in this study answered all seven of the FSS questions used by Prenoveau et al. (2010).

The APPQ (Rapee et al., 1994) consists of 22 items that examine fear of sensation-producing activities along with agoraphobic scenarios. Like the FSS, the original version of this questionnaire asks participants how much fear they would feel in each of the listed experiences (0 = *no fear*, 5 = *moderate fear*, 8 = *extreme fear*). Those in this study answered 10 questions used by Prenoveau et al. (2010).

The 13-item Self-Consciousness subscale of the SPS (Mattick & Clarke, 1998; Zinbarg & Barlow, 1996) examines sensitivity to social evaluation. This sensitivity is a key component of social phobia. The original version of this questionnaire asks how typical a statement is of the participant (0 = *not at all typical of me*, 2 = *Moderately*, 4 = *extremely typical of me*). Participants in this study answered eight items used by Prenoveau et al. (2010).

The 21-item IDD (Zimmerman et al., 1986) assesses depression symptoms such as anhedonia and hopelessness. Each IDD item contains five statements. The participants decide which of the statements best reflect how they have been feeling in the past week. Individuals in this study answered eight of the original items.

The 90-item MASQ (Watson et al., 1995) measures symptoms of a broad range of anxiety and depressive disorders. The original MASQ asks participants to describe to what extent they have had certain symptoms over the past week (1 = *Not at all*, 3 = *Moderately*, 5 = *Extremely*). Thirty-four of these items identified by Prenoveau et al. (2010) were used in this study.

The PSWQ (Meyer et al., 1990) contains 16 items that assess worry, the key symptom of GAD. The original version of this measure uses a 5-point Likert-type scale where individuals identify how typical a given statement is of their life in general. The scale ranges from 1 = *not at all typical* to 5 = *very typical* and participants completed all items.

The 18-item OCI-R (Foa et al., 2002) self-report measure examines key symptoms of OCD. The original OCI-R uses a 5-point Likert-type scale to assess how prevalent the symptoms of OCD are in a participant's life. The scale ranges from 0 = *not at all* to 4 = *extremely*, and participants completed all items.

Social learning task

This task was an adapted version of a probabilistic social learning task (Parsons, Young, Bhandari, et al., 2014). Probabilistic learning tasks have been used to study learning patterns in response to positive and negative reinforcement (Frank et al., 2004) and in responses to threatening cues in social anxiety (Abraham & Hermann, 2015). In this task, participants learn to associate images of neutral faces with audio-visual positive and negative social feedback (smile/laugh or frown/cry). Faces are presented in pairs, and participants are instructed to find out who is the “happier” and the “sadder” person in each pair. They can select one of the two faces which then immediately changes to an image of a happy face, paired with a laugh sound, or a sad face, paired with a cry sound (see Figure 1). We used multimodal (facial and vocal) feedback rather than unimodal (facial) feedback, as these cues often co-occur in social interactions. A second goal was to maximise task engagement, as the attentional capture and salience of audio-visual stimuli has been shown to be greater than that of unimodal stimuli (Koelewijn et al., 2010). There are three pairs of faces presented with different contingencies of happy and sad feedback (see “training phase” for more details). This allows investigation of learning in response to positive and negative feedback at different levels of difficulty. Before and after the task, participants complete measures of arousal, valence, pleasantness (“liking”), and motivation (“wanting”; see below). This allows investigation of: (1) the extent to which learning changes emotional ratings and reward valuation of neutral faces, and (2) individual differences in these processes. The total task duration ranged from 12–29 min ($M = 16.94$ min; $SD = 2.24$). Each phase of the task is described in detail below.

Ratings of valence, arousal, and pleasantness (“liking”)

Participants completed a series of ratings of neutral female facial expressions (stimuli from NimStim database; Tottenham et al., 2009). A set of 12 neutral faces were rated on three scales: arousal, valence, and pleasantness. Scales were horizontal visual analogue scales (Figure 1) with anchors as follows: “very relaxed” to “very excited” (arousal), “very unhappy” to “very happy” (valence), “very unpleasant” to “very pleasant” (pleasantness, a measure of “liking”). Six of these stimuli were subsequently presented during the training phase of the task, whereas the other six were just rated before and after training. Responses were made by mouse-click on a 0–100 visual analogue scale (with no upper time limit on responding). These ratings were completed before and after the training phase (described below). Stimuli were presented in a randomised order that varied across participants, ratings scales, and time points (before or after training). Note that only female facial expressions

were included as the training task used positive and negative emotional vocalisations from the OxVoc database (Parsons et al., 2014), in which only female negative vocalisations are available.

Motivation to view (“wanting”)

Participants also completed a motivation to view measure in which they could vary the duration of viewing each neutral face stimulus. This measure has been used in prior studies examining the “wanting” component of reward valuation, by assessing effort expended to “consume” or view the reward (Aharon et al., 2001; Parsons et al., 2011). Stimuli appear on screen for a default duration of 6 s, participants can repeatedly press the “up” key on the keyboard to incrementally increase viewing time for each face, or the “down” key to decrease viewing time (maximum duration of 12 s, minimum duration of 2 s, each keypress corresponds to an increment of 250 ms). A vertical bar indicating time remaining is displayed on-screen, with the bar “moving down” to indicate time passing (Figure 1). This measure was also completed before and after training. Although not explicitly instructed, participants might reason that shortening the duration of stimulus viewing would shorten the overall duration of the task. If this was the case, we would not observe differences on our comparison of interest, the change in motivation to view from before to after training (by stimulus type, see Supplementary Materials for further discussion of this issue). We did not observe that participants were simply acting to reduce the task duration.

Training phase: probabilistic learning

Six of the twelve neutral stimuli used in the pre-training ratings were presented during the training phase, whereas the other six stimuli remained unseen during this period. Training consisted of a probabilistic learning task in which participants learn to associate neutral facial images with varying likelihood of receiving positive or negative social feedback. Social feedback was provided in the form of emotional facial expressions (happy or sad) and emotional vocalisations (laughter or cry sounds). On each trial, participants view two faces, are asked to select one face (using “up” and “down” keyboard keys), and then receive feedback (happy face and laugh sound, or sad face and cry sound) from the face they select (see Figure 1). Note that the image of the individual used for the neutral facial expression was the same as that for the positive and negative feedback, so by selecting a neutral face, participants then saw the same individual smile and laugh or frown and cry (this is a different procedure to Abraham & Hermann, 2015, in which different faces were used for feedback).

Participants were instructed that for each pair of faces, “there is one happy and one sad person,” and that “like in real life, the happy person will not always be happy and the sad person will not always be sad.” Individual faces varied in the likelihood of positive and negative feedback, creating three levels of difficulty. In the first pair of faces, one face led to positive feedback on 80% of trials and negative feedback on 20% of trials, while the other face led to negative feedback on 80% of trials and positive feedback on 20% of trials. This was the easiest to learn pair. The other pairs had contingencies of 70% versus 30% and 60% versus 40%. Participants completed two rounds of 60 trials (120 trials total, 40 trials for each pair). In one round, participants were instructed to “find the ‘happier’ person, and continue

to always select this person, even if they sometimes appear to be sad.” In the other round, they were instructed to find the “sadder” person, so all participants viewed all faces under both versions of instructions. This ensured approximately equal levels of positive and negative feedback across the task (mean percentage of positive feedback trials = 49.70%, negative = 50.30%, $SD = 4.57\%$). The order of rounds was randomised across participants. The allocation of face pairs with rates of positive and negative feedback was counterbalanced across participants. This ensured that when assessing changes in valence, arousal, pleasantness and motivation to view before and after training, any changes observed were independent of the specific features of individual faces.

Testing phase

Immediately following the training phase, participants completed a testing phase in which they viewed pairs of faces and were instructed to “choose the person that ‘feels’ the most happy, based on what you have learned during the previous task. If you are not sure which one to pick, just go with your gut instinct.” Responses were made using the “up” and “down” keys and participants were asked to respond as quickly and accurately as possible. Participants received no feedback on their performance, and face pairs were fully mixed so that each face was presented with every other face twice (total 30 trials).

Statistical analysis

Prior to the analysis relating to the experimental task described here, confirmatory factor analysis (CFA) was conducted to test whether the tri-level symptom model (Prenoveau et al., 2010) provided a good fit to the self-reported symptom data in the current sample provided at the time of the behavioural testing session. These analyses are described in full in Kramer et al. (2019). In brief, the CFA was conducted using Mplus Version 8 statistical software treating all items as categorical using robust weighted least squares estimation (WLSMV) using all available information (i.e., accommodating missing data). Model goodness of fit was evaluated using three fit indices including the comparative fit index (CFI; Bentler, 2004), the root mean square error of approximation (RMSEA; Styger, 1989), and the weighted root mean square residual (WRMR; DiStefano et al., 2018; Yu, 2002). The WRMR, like the standardised root mean square residual (SRMR; Bentler, 1995) for continuous data, measures the (weighted) average differences between the sample and estimated population variances and covariances. As WRMR is considered experimental and its developers caution users to not rely heavily on it (e.g., DiStefano et al., 2018), we supplemented it by re-specifying the items as continuous to obtain a SRMR estimate. To conclude good model fit, we adopted the following cutoffs: $CFI \geq .90$, $RMSEA \leq .06$, $WRMR \leq 1.0$, and $SRMR \leq .08$ (DiStefano et al., 2018; Hu & Bentler, 1998; Yu, 2002). Model fit was good: $CFI = .97$, $RMSEA = .021$ (90% confidence interval = [.018, .024]), $WRMR = .94$, and $SRMR = .05$. We saved factor score estimates from this model and used them to represent the tri-level model symptom dimensions of General Distress, Anhedonia-Apprehension, and Fears in our analyses relating these symptom dimensions to performance in the probabilistic social learning task (full item loadings available in Supplementary Materials Table S1). It should also be noted that these three factor scores are quasi-orthogonal. Thus, the correlations of the General Distress factor scores equalled .08 ($p = .17$) and $-.07$ ($p = .24$), respectively, with the Fears and Anhedonia-Apprehension factor scores.

Similarly, the correlation between Fears and Anhedonia-Apprehension factor scores equalled .07 ($p = .22$). Consequently, associations with each dimension's factor score can be considered unique of the others.

In addition, face validity of tri-level factor scores was assessed by performing correlations between factor scores and diagnostic status (presence of anxiety disorder, dummy coded; major depressive disorder, dummy coded). In line with the expected structure of the tri-level model, anxiety disorder diagnostic status was significantly correlated with General Distress and Fears factor scores ($r = .50, p < .001$; $r = .12, p = .035$, respectively), but not Anhedonia ($r = -.08, p = .16$), whereas depression diagnostic status was significantly correlated with General Distress and Anhedonia factors scores ($r = .33, p < .001, r = -.14, p = .015$), but not Fears scores ($r = .04, p = .409$).

Aim 1.—Repeated-measures ANOVAs with orthogonal, polynomial analysis of trends were performed to investigate linear associations between stimulus type (80/70/60/40/30/20% positive feedback) and change in ratings of arousal, valence, and pleasantness, as well as change in viewing times on the motivation to view task. To compare whether predominantly positive feedback had a greater effect on ratings than predominantly negative feedback, mean absolute change scores were calculated for “positive” stimuli (paired with 80%, 70%, 60% positive feedback) and “negative” stimuli (paired with 40%, 30%, 20% positive feedback). Paired samples t -tests were used to assess differences in the extent of change in ratings of arousal, valence, and pleasantness, as well as mean viewing times.

Aim 2.—Linear regression analyses were used to investigate relationships between symptom dimensions of anxiety and depression (general distress, fears and anhedonia-apprehension factors) and measures of arousal, valence, pleasantness, and motivation to view before training and changes in these measures from before to after training (using estimates of linear trends in the relationship between change in ratings and stimulus type). Linear and quadratic terms of cumulative accuracy slopes were calculated as a measure of learning performance across trials during the training phase. Quadratic terms were used in addition to linear terms to capture the asymptotic nature of learning during this task. Regression analyses were also used to assess performance on the testing phase which was quantified as the linear slope of the relationship between performance accuracy and stimulus type.

Aim 3.—Multiple regression analyses were conducted to investigate whether significant relationships between dimensions of the tri-level model and behavioural measures prior to training changed over time (from pre to post-training), including sex as a covariate. These analyses were repeated for faces seen during the training phase and for unseen faces to test whether any changes observed generalised to untrained stimuli. All analyses were conducted using SPSS (v.24), and multiple regression analyses were conducted using the PROCESS macro for SPSS (Hayes, 2017).

Results

All data were examined for outliers and there were no values falling outside the first and third quartiles ± 1.5 times the interquartile range.

Training phase performance

Overall, accuracy on the social feedback learning task was high, with performance varying in line with the feedback contingencies. The highest accuracy during training was for the easiest to learn pair (80–20 pair; $M = 82.30\%$, $SD = 18.93$), followed by the 70–30 pair ($M = 76.90\%$, $SD = 20.98$), and the lowest performance accuracy for the 60–40 pair ($M = 66.37\%$, $SD = 25.43$). There were significant linear and quadratic trends in cumulative performance accuracy across trials, linear $F(1, 323) = 302.64$, $p < .001$, $\eta^2 = .48$, and quadratic $F(1, 323) = 34.18$, $p < .001$, $\eta^2 = .10$; Figure 2a. There was also a significant linear trend in accuracy scores across stimulus pairs, $F(1, 323) = 71.41$, $p < .001$, $\eta^2 = .18$. Performance increased linearly as uncertainty within the face pair was reduced.

Testing phase performance

During the testing phase, participants demonstrated accurate learning of response contingencies, with the 80% happy face being the most frequently selected on the forced-choice task (choosing which face was the “happier” individual) and the 20% happy face being least frequently selected. There was a significant linear trend in the relationship between face stimulus type and task performance, proportion of trials chosen as “more happy,” $F(1, 323) = 967.66$, $p < .001$, $\eta^2 = .75$; Figure 2b.

Aim 1: does social learning affect emotional ratings and reward valuation of social cues

Examining participant ratings of arousal, valence, pleasantness, and responses on the motivation to view task across the six neutral faces presented before and after the training and testing phases, we found the following.

Arousal.—There was a significant linear trend in the relationship between stimulus type and change in arousal rating, $F(1, 323) = 23.59$, $p < .001$, $\eta^2 = .07$. Arousal ratings non-significantly increased for stimuli paired with 20% and 30% positive feedback and significantly increased for stimuli paired with 40% or more positive feedback. Mean change in arousal ratings increased linearly as the percentage positive feedback increased (see Table 2 for full details).

Valence.—There was a significant linear trend in the relationship between stimulus type and change in valence rating, $F(1, 323) = 252.28$, $p < .001$, $\eta^2 = .44$. Thus, the change in valence ratings increased linearly as the percentage positive feedback increased. More specifically, valence ratings significantly decreased for stimuli paired with 20% positive feedback, non-significantly decreased for stimuli paired with 30% positive feedback, non-significantly increased for stimuli paired with 40% positive feedback, and significantly increased for stimuli paired with 60% or more positive feedback (Table 2).

Pleasantness (“liking”).—There was a significant linear trend in the relationship between stimulus type and change in pleasantness rating, $F(1, 323) = 129.14, p < .001, \eta^2 = .29$. Just as was the case for arousal and valence, the change in pleasantness ratings increased linearly as percentage positive feedback increased. More specifically, pleasantness ratings non-significantly decreased for stimuli paired with 20% positive feedback, non-significantly increased for stimuli paired with 30% positive feedback, and significantly increased for stimuli paired with 40% or more positive feedback (Table 2).

Motivation to view (“wanting”).—There was a significant linear trend in the relationship between stimulus type and change in viewing times for face stimuli, $F(1, 323) = 65.77, p < .001, \eta^2 = .17$. Consistent with the other three dependent variables, the change in viewing times increased linearly as percentage positive feedback increased. More specifically, viewing times significantly decreased for stimuli paired with 20% positive feedback, non-significantly decreased for stimuli paired with 30% positive feedback, non-significantly increased for stimuli paired with 40% and 60% positive feedback, and significantly increased for stimuli paired with 70% and 80% positive feedback (Table 2, note that significant differences in pre to post-training viewing times, including significant increases for some stimuli, indicate that participants were not solely motivated to decrease overall task duration, see Supplementary Materials for further discussion).

Comparison of positive versus negative feedback.—Paired samples *t*-tests demonstrated that there were significant differences in the absolute magnitude of change by feedback type (“majority positive” [mean of 60%, 70%, 80% happy] versus “majority negative” [mean of 40%, 30%, 20% happy]) for ratings of arousal, $t(323) = 5.92, p < .001, d = .34$; valence, $t(323) = 6.37, p < .001, d = .37$; and pleasantness, $t(323) = 5.53, p < .001, d = .31$, as well as performance on the motivation to view task, $t(323) = 4.83, p < .001, d = .27$. Changes following majority positive feedback were significantly greater in absolute magnitude than those following majority negative feedback.

We did not collect ratings of emotional responses to the positive and negative feedback stimuli as part of this task, so we conducted a post hoc online experiment (hosted by Prolific; <https://app.prolific.co>) to examine whether there were differences in valence and arousal of positive versus negative feedback stimuli. An independent sample ($N = 30$) rated positive and negative face/voice pairs for valence and arousal. We computed an “absolute difference” valence score (i.e., how “different from neutral” stimuli were) and conducted paired samples *t*-tests to examine differences in ratings of valence and arousal. Results demonstrated no significant “absolute difference” in valence ratings between positive and negative stimuli ($p > .05$). There was a significant difference in arousal ratings, such that arousal was higher for positive compared with negative stimuli ($p = .003$, see Supplementary Materials for full details). This greater arousal of positive stimuli may account for why positive feedback had a larger impact than negative feedback in changing responses to neutral face stimuli.

Sex differences.—As the face and voice stimuli were all female, we examined whether there were any effects of participant sex on performance during the training and testing phases of the task, as well as on the changes in valence, arousal, pleasantness, and

motivation to view measures. All analyses demonstrated no significant effect of sex (all p 's $\geq .05$, see Supplementary Materials Table S2 for details).

Aim 2: do symptom dimensions of anxiety and depression relate to aspects of social functioning?

Emotional responses to neutral facial expressions.—Linear regression analyses demonstrated significant associations between the general distress factor and mean valence ratings of neutral faces prior to training ($p < .001$, Table 3). Individuals with higher scores on the general distress factor rated neutral facial expressions more negatively. There were no other significant relationships between factors of the tri-level model or sex and ratings of arousal and valence prior to training (all p 's $> .05$). There were also no significant relationships between factors of the tri-level model and individual estimates of linear slopes in the change of ratings from pre- to post-training by face type (all p 's $> .05$).

Reward value of neutral facial stimuli.—There was a significant association between the anhedonia-apprehension factor and mean viewing times of neutral faces prior to training (Table 3). Individuals with higher levels of anhedonia-apprehension had lower viewing times for the neutral face stimuli. There were no other significant relationships between factors of the tri-level model or sex and measures of reward value (ratings of pleasantness and motivation to view) prior to training (all p 's $> .05$). There were no significant associations between factors of the tri-level model or sex and estimates of individual slopes of change in responses from pre- to post-training by face type (all p 's $> .05$).

Learning from social feedback

Training.—Regression analyses demonstrated significant associations between the fears factor and linear slopes of cumulative performance accuracy for the 80–20 and 70–30 pairs ($\beta = .19, p = .001$; $\beta = -.18, p = .001$, respectively), and between the general distress factor and linear slopes of cumulative performance accuracy for the 70–30 pair ($\beta = .12, p = .04$). Individuals with higher levels of fears had steeper slopes (more rapid learning) on trials for the easiest to learn pair (80–20), and shallower slopes (slower learning) on trials for the medium-difficulty pair (70–30), whereas individuals with higher levels of general distress had steeper slopes on this pair. Other tri-level model factors and sex were not significant predictors in these models. There were no significant associations between cumulative performance accuracy in the hardest to learn pair (60–40, $p > .05$). There were also no significant associations between factors of the tri-level model and estimated quadratic trends (all p 's $> .05$).

Testing.—There were no significant associations between factors of the tri-level model and performance during the testing phase (p 's $> .05$).

Aim 3: changes in associations with symptoms from before to after training

General distress and valence.—multiple regression analysis demonstrated a significant change in the relationship between general distress and mean valence ratings from pre- to post-training. General distress, time, and the interaction between these variables overall significantly predicted mean valence ratings, overall model fit: $F(4, 319) = 20.77, p < .001$,

$R^2 = .21$. General distress was a significant predictor of valence ratings ($\beta = -.23, p = .001$), as was time (pre- or post-training: $\beta = .43, p < .001$). There was also a significant interaction effect of general distress and time ($\beta = .16, p = .024$). Examination of simple slopes demonstrated that prior to training, there was a significant relationship between general distress and mean valence rating ($\beta = -.22, p < .001$), such that higher levels of general distress were associated with more negative valence ratings. After training this relationship was no longer significant ($\beta = -.002, p = .98$). Thus, the interaction was driven by training reducing the relationship between general distress and mean valence ratings (Figure 3a).

Anhedonia and viewing time.—multiple regression analysis demonstrated no significant change in the relationship between anhedonia-apprehension and viewing times from pre- to post-training. Anhedonia-apprehension, time, and the interaction between these variables overall significantly predicted mean valence ratings, overall model fit: $F(4, 319) = 2.86, p < .024, R^2 = .035$. Anhedonia-apprehension was a significant predictor of viewing time, $\beta = .23, t(644) = -2.86, p = .004$, but time was not ($\beta = .05, p = .33$). The interaction between anhedonia-apprehension and time was also not significant ($\beta = -.09, p = .26$; Figure 3b).

Test of generalisability.—As with the stimuli presented during the task, there was a significant change in the relationship between general distress and mean valence ratings of stimuli not presented during the task from pre- to post-training. General distress, time, and the interaction between these variables overall significantly predicted mean valence ratings, overall model fit: $F(4, 323) = 6.04, p < .001, R^2 = .07$. General distress was a significant predictor of valence ratings ($\beta = -.16, p = .04$), as was time (pre- or post-training; $\beta = .20, p < .001$). There was also a significant interaction effect of general distress and time ($\beta = .20, p = .01$). Examination of simple slopes demonstrated that prior to training, there was a significant relationship between general distress and mean valence rating ($\beta = -.18, p = .001$), and after training, this relationship was no longer significant ($\beta = .009, p = .87$). Thus, the interaction was driven by training reducing the relationship between general distress and mean valence ratings.

Comparison of the association between anhedonia-apprehension, time, and the interaction between these two variables significantly predicted mean viewing times for unseen faces, $F(4, 319) = 5.06, p = .001, R^2 = .06$. Anhedonia-apprehension was a significant predictor of viewing time ($\beta = -.21, p = .007$), and individuals with higher levels of anhedonia-apprehension demonstrated shorter viewing durations. Time was also a significant predictor ($\beta = -.20, p < .001$), and overall viewing durations were shorter for unseen faces after training, compared with before. There was no significant interaction between anhedonia-apprehension and time ($\beta = -.12, p = .13$).

Discussion

In this study, we demonstrated three main sets of significant findings. First, we observed that a brief probabilistic learning task led to more positive emotional ratings and a greater reward value for neutral face stimuli paired with predominantly positive feedback. There was limited evidence for more negative emotional ratings and reduced reward value for neutral

face stimuli paired with predominantly negative feedback. Second, we found the following associations between emotional and reward processing and symptom dimensions of anxiety and depression: (1) participants with higher levels of general distress (a factor common to anxiety and depression) showed more negative bias in ratings of valence of neutral faces; (2) participants with higher levels of anhedonia-apprehension had decreased motivation to view neutral faces, and (3) participants with higher levels of fears and general distress had altered performance accuracy across trials. Third, we observed that the association between general distress and negative bias was significantly reduced following training.

Positive social feedback alters emotional ratings and reward valuation of neutral facial expressions

Comparing performance before and after training, neutral faces paired with predominantly positive feedback (70% or greater) were rated as higher in arousal, valence, and pleasantness (the “liking” component of reward valuation) and resulted in longer viewing durations (a measure of motivation, the “wanting” component of reward valuation). Faces paired with 60% positive/40% negative feedback were rated as higher in arousal, valence, and pleasantness, with no change in viewing durations. Faces paired with 40% positive/60% negative feedback were also rated as higher in arousal and pleasantness, with no change in valence or viewing duration. There was no change in ratings for stimuli paired with 30% positive/70% negative feedback. Stimuli paired with 20% positive/80% negative feedback were rated as lower in valence and had lower viewing durations, with no change in arousal or pleasantness. Overall, greater change was demonstrated among faces paired with higher probabilities of positive feedback. In a direct comparison, we observed greater absolute change in arousal, valence, pleasantness, and viewing time following “majority positive feedback” (collapsing across 80%, 70%, 60% positive feedback conditions) than following “majority negative feedback” (collapsing across 20%, 30%, 40% positive feedback conditions). Notably, these effects were observed after a short period of training (duration $M = 6.87$ min, $SD = 0.71$ min).

This perhaps surprising finding suggests that positive social feedback is more effective than negative social feedback at changing emotion ratings and reward valuation of ambiguous facial expressions. Returning to the example earlier, when telling joke to a friend, a positive response might impact how positively we view that friend, potentially affecting our likelihood of seeking them out or telling them more jokes in the future. Negative feedback to our jokes, however, may not substantially impact our view of that individual. These findings are comparable to prior work using a similar paradigm with infant faces instead of adult faces (Parsons, Young, Bhandari, et al., 2014). Other work has demonstrated that both positive and negative statements can alter the reward value of face stimuli (Davis et al., 2009), whereas valence of face stimuli was altered by negative, but not positive biographical information (Suess et al., 2015). Together, these findings suggest that emotional ratings and reward valuation of neutral faces can be manipulated using other sources of information, but that the specific type of information provided may be important in determining the direction of this effect.

One potential explanation for the limited change in emotional and reward responses following negative feedback is that the sad face/cry sound negative feedback may have been insufficiently aversive or salient to impact responses to neutral facial expressions. We selected smile/laugh and frown/cry feedback as stimuli that are high and low in valence (respectively) but relatively low in arousal. Direct comparisons of the arousal level of happy and sad facial expressions from the NimStim face set demonstrated no significant differences in arousal level (Smith et al., 2013). There were also no significant differences in arousal or motivation to respond to adult laughter and cry vocalisations from the OxVoc sounds set (Parsons, Young, Craske, et al., 2014). However, to further investigate this, we performed a post hoc experiment, examining ratings of the positive and negative feedback stimuli. We observed a difference in ratings of arousal for combined face/voice pairs, such that “happy” feedback (smile plus laughter) was significantly higher in arousal than “sad” feedback (frown plus cry). Future work might aim to use negative multimodal stimuli that are matched in arousal to the positive stimuli used here (perhaps using angry/threatening or disgust cues) to further examine the efficacy of negative social feedback in altering responses to neutral facial stimuli.

Relationships between symptom dimensions of anxiety and depression and social functioning

We replicated previous findings demonstrating a negative bias in the valence of neutral facial expressions, with more negative ratings associated with higher levels of general distress (a symptom factor common to anxiety and depression). This finding is in line with previous work demonstrating negative bias in depression or anxiety disorders. Here, we show that this bias is associated with symptoms that are common to both disorders indicating a potentially shared transdiagnostic process. This complements prior prospective work demonstrating the mediating role of negative bias in the relationship between behavioural inhibition, anxiety, and depression at different stages of development and highlights the transdiagnostic relevance of these behaviours (Connolly et al., 2016; Price et al., 2016; White et al., 2017).

We also found that negative bias in ratings of valence among neutral faces was significantly reduced following social feedback training. On average, individuals demonstrated increased mean positive valence ratings of neutral faces from pre- to post-training, but these effects were larger for individuals with higher symptoms of general distress, effectively reducing negative bias by the end of the task. Interestingly, this change in negative bias also generalised to neutral faces that were not presented during the training. One possible explanation for this effect is that through associative training, attention is directed more towards the stimuli presented, rather than towards internal mood states. A change in attention might then result in more objective rating of other face stimuli, perhaps relying more on information from physical features of faces than the observer’s own biases. Prior work has demonstrated that interventions targeting attention (e.g., attention bias modification) significantly reduce symptoms of anxiety (Hakamata et al., 2010). The role of attention was not tested in this study but could be addressed in future work by altering instructions to differentially direct attention (e.g., to compare responses when instructed to focus on the eyes or the mouth during feedback, or to focus on your own internal reaction to stimuli).

We also demonstrated that higher levels of anhedonia-apprehension symptoms (more specific to depression) were associated with decreased motivation to view neutral face images, prior to training. Anhedonia was not associated with altered ratings of pleasantness, demonstrating a dissociation between the “liking” and “wanting” components of reward valuation. This finding is consistent with previous work demonstrating associations between symptoms of anhedonia and reduced effortful behaviours (Treadway et al., 2012). Previous work has demonstrated this effect in a decision-making task in which participants can win a monetary reward. We found evidence of a similar effect in the absence of an explicit reward, when participants are simply making keypress responses to view images of neutral faces. Although social cues are thought to carry an inherent reward value (Krach et al., 2010), this finding suggests that anhedonia-apprehension may have an impact on effortful behaviours even when the reward value of a stimulus is not explicit. We found no significant change in the relationship between anhedonia-apprehension and motivation to view neutral faces after training. As a hypothesised deficit in reward-related functioning, this effect is perhaps unsurprising, suggesting that simple associative learning was not sufficient to change motivation to view neutral stimuli. A lack of association between anhedonia-apprehension and performance on training trials suggests that the absence of change was not attributable to reduced associative learning. It may be that manipulation of this association is not amenable through probabilistic learning, or perhaps that more extensive training is required.

Finally, we found that higher levels of fear symptoms (specific to anxiety disorders) and general distress (common to anxiety and depression) were associated with altered slopes of performance accuracy during training. This is largely consistent with prior work demonstrating more accurate learning of probabilistic contingencies among individuals with higher levels of anxiety symptoms (Abraham & Hermann, 2015). Our findings were observed within a group of individuals with a range of symptoms across different types of anxiety disorders, suggesting that disrupted learning is related to “fear” symptoms that are a feature of different disorders, rather than being specific to social anxiety. Examining disrupted learning patterns across individual pairs of stimuli, we observed that higher scores on the “fears” symptom dimension were associated with steeper performance slopes (indicating faster learning) on the easiest-to-learn pair, less steep slopes on the medium-difficulty pair, and no effect on the hardest pair. We also observed that symptoms of general distress were associated with a steeper slopes on the medium-difficulty pair, but no effects on the other pairs. Unlike prior work demonstrating associations between symptoms of anhedonia-apprehension and reduced reward learning (e.g., Pizzagalli et al., 2008), we found no association between anhedonia-apprehension and performance accuracy during training.

Although we do not have a clear explanation for these findings, it is possible that different dimensions of symptoms implicated in anxiety and depression may interact to affect socio-emotional learning. One limitation of the current task design, which may be masking a clearer pattern of effects, is that learning was reinforced by both positive and negative feedback. Participants were instructed to approach positive stimuli in one half of the training (“find the happier person”) and approach negative stimuli in the other half (“find the sadder person”). It was implied by the task design and instructions that successful approach of one stimulus simultaneously meant successful avoidance of the other (i.e., selecting the correct “happy” face not only provided positive social feedback but also prevented exposure to

negative social feedback). Prior work has demonstrated disrupted patterns of approach and avoidance tendencies that vary with symptoms of anxiety and depression (e.g., Heuer et al., 2007; Trew, 2011). In the current task, different contingencies of positive and negative reinforcement across stimulus pairs may differentially engage disrupted approach and avoidance tendencies. In future, we could design a task to separate the positive and negative reinforcers (for example, comparing positive feedback with neutral, or no feedback) to examine these effects.

Strengths and limitations

These effects were demonstrated in a large sample of young adults with a wide range of self-reported symptoms of anxiety and depression. Comparison with clinical samples will be of interest in future work to see if the patterns observed here at a dimensional level are replicated, or are potentially more pronounced. It will also be important to establish whether negative bias and disrupted motivation can be affected by social feedback learning in these populations. The findings presented here were observed during a testing session that occurred immediately after the end of the training phase. The temporal stability of these effects would be of interest, particularly whether they persist beyond the end of the experimental session. Although we demonstrate that negative interpretation bias is modifiable through brief training, this approach does not necessarily translate directly into therapeutic intervention. Similar work in the domain of attentional bias training held much promise for novel therapeutic intervention, yet effect sizes remain modest (Mogoa e et al., 2014). What these findings do suggest is that core appraisal processes are modifiable through associative learning and that, at least in the case of social stimuli, positive feedback may be a particularly effective approach. Here, we tested only the effects of happy and sad facial expressions as social feedback, further investigation of whether other types of expression (particularly anger, fear or disgust expressions as negative feedback) might have different effects on modifying valence and arousal responses to neutral facial expressions. It also remains to be seen whether the effects observed here translate to non-social stimuli. Finally, as this task involved learning, differences in cognitive abilities might affect performance, and inclusion of an intelligence quotient (IQ) measure in future would allow investigation of this possibility.

Conclusion

In sum, we demonstrate that emotion ratings and reward valuation of neutral faces are readily altered through associative learning. The extent of the change in emotional responses and reward value was linearly associated with the ratio of positive-to-negative feedback, with positive feedback overall leading to greater change in emotional responses and reward value than negative feedback. Negative bias in valence of neutral faces was associated with the symptom factor “general distress,” a set of symptoms that are common to anxiety disorders and depression. Notably, this effect was reduced following learning based on probabilistic feedback, demonstrating that this bias can be modified through intervention. Disrupted motivation to view neutral faces was associated with anhedonia-apprehension, a cluster of symptoms specific to depression. Altered cumulative performance accuracy during the training phase of the task was associated with symptoms of “fears” and “general

distress.” These results suggest that brief associative learning can impact perceptual and affective processes implicated in anxiety and depression.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Institute of Mental Health (NIMH) of the National Institutes of Health (NIH) under award number R01MH100117, and was awarded to M.G.C., S.Y.B., R.N., and R.E.Z.

References

- Abraham A, & Hermann C (2015). Biases in probabilistic category learning in relation to social anxiety. *Frontiers in Psychology*, 6, Article 1218. 10.3389/fpsyg.2015.01218
- Adolphs R (2002). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews*, 1(1), 21–62. 10.1177/1534582302001001003 [PubMed: 17715585]
- Aharon I, Etcoff N, Ariely D, Chabris CF, O'Connor E, & Breiter HC (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron*, 32(3), 537–551. [PubMed: 11709163]
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).
- Beevers CG, Wells TT, Ellis AJ, & Fischer K (2009). Identification of emotionally ambiguous interpersonal stimuli among dysphoric and nondysphoric individuals. *Cognitive Therapy and Research*, 33(3), 283–290. 10.1007/s10608-008-9198-6 [PubMed: 20046979]
- Bentler PM (1995). *EQS structural equations program manual*. Multivariate Software.
- Bentler PM (2004). *EQS 6.1 structural equations program manual*. Multivariate Software.
- Bistricky SL, Ingram RE, & Atchley RA (2011). Facial affect processing and depression susceptibility: Cognitive biases and cognitive neuroscience. *Psychological Bulletin*, 137(6), 998–1028. 10.1037/a0025348 [PubMed: 21895353]
- Bourke C, Douglas K, & Porter R (2010). Processing of facial emotion expression in major depression: A review. *Australian and New Zealand Journal of Psychiatry*, 44(8), 681–696. 10.3109/00048674.2010.496359
- Carver CS, & White TL (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology*, 67(2), 319–333. 10.1037/0022-3514.67.2.319
- Connolly SL, Abramson LY, & Alloy LB (2016). Information processing biases concurrently and prospectively predict depressive symptoms in adolescents: Evidence from a self-referent encoding task. *Cognition and Emotion*, 30(3), 550–560. 10.1080/02699931.2015.1010488 [PubMed: 25707445]
- Cooper JA, Arulpragasam AR, & Treadway MT (2018). Anhedonia in depression: Biological mechanisms and computational models. *Current Opinion in Behavioral Sciences*, 22, 128–135. 10.1016/j.cobeha.2018.01.024 [PubMed: 29503842]
- Davis FC, Johnstone T, Mazzulla EC, Oler JA, & Whalen PJ (2009). Regional response differences across the human amygdaloid complex during social conditioning. *Cerebral Cortex*, 20(3), 612–621. 10.1093/cercor/bhp126 [PubMed: 19556348]
- DiStefano C, Liu J, Jiang N, & Shi D (2018). Examination of the weighted root mean square residual: Evidence for trustworthiness? *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3), 453–466.

- Ekman P, & Cordaro D (2011). What is meant by calling emotions basic. *Emotion Review*, 3(4), 364–370. 10.1177/1754073911410740
- Ekman P, & Friesen WV (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2), 124–129. [PubMed: 5542557]
- Elfenbein HA, & Ambady N (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128(2), 203–235. [PubMed: 11931516]
- Everaert J, Podina IR, & Koster EH (2017). A comprehensive meta-analysis of interpretation biases in depression. *Clinical Psychology Review*, 58, 33–48. 10.1016/j.cpr.2017.09.005 [PubMed: 28974339]
- Eysenck HJ, & Eysenck SBG (1975). *Manual of the Eysenck Personality Questionnaire (Junior and Adult)*. Hodder & Stoughton.
- First MB, & Williams JB (2016). *SCID-5-CV: Structured Clinical Interview for DSM-5 Disorders: Clinician Version*. American Psychiatric Association.
- Foa EB, Huppert JD, Leiberg S, Langner R, Kichic R, Hajcak G, & Salkovskis PM (2002). The Obsessive-Compulsive Inventory: Development and validation of a short version. *Psychological Assessment*, 14(4), 485–496. [PubMed: 12501574]
- Frank MJ, Seeberger LC, & O'Reilly RC (2004). By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. *Science*, 306(5703), 1940–1943. 10.1126/science.1102941 [PubMed: 15528409]
- Fussner LM, Mancini KJ, & Luebbe AM (2018). Depression and approach motivation: Differential relations to monetary, social, and food reward. *Journal of Psychopathology and Behavioral Assessment*, 40(1), 117–129. 10.1007/s10862-017-9620-z
- Gebhardt C, & Mitte K (2014). Seeing through the eyes of anxious individuals: An investigation of anxiety-related interpretations of emotional expressions. *Cognition and Emotion*, 28(8), 1367–1381. 10.1080/02699931.2014.881328 [PubMed: 24499045]
- Geer JH (1965). The development of a scale to measure fear. *Behaviour Research and Therapy*, 3(1), 45–53. [PubMed: 14340595]
- Gollan JK, Pane HT, McCloskey MS, & Coccaro EF (2008). Identifying differences in biased affective information processing in major depression. *Psychiatry Research*, 159(1), 18–24. 10.1016/j.psychres.2007.06.011 [PubMed: 18342954]
- Gotlib IH, & Joormann J (2010). Cognition and depression: Current status and future directions. *Annual Review of Clinical Psychology*, 6, 285–312. 10.1146/annurev.clinpsy.121208.131305
- Hahn AC, & Perrett DI (2014). Neural and behavioral responses to attractiveness in adult and infant faces. *Neuroscience & Biobehavioral Reviews*, 46, 591–603. 10.1016/j.neubiorev.2014.08.015 [PubMed: 25199981]
- Hakamata Y, Lissek S, Bar-Haim Y, Britton JC, Fox NA, Leibenluft E, Ernst M, & Pine DS (2010). Attention bias modification treatment: A meta-analysis toward the establishment of novel treatment for anxiety. *Biological Psychiatry*, 68(11), 982–990. 10.1016/j.biopsych.2010.07.021 [PubMed: 20887977]
- Hall JA, & Matsumoto D (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, 4(2), 201–206. 10.1037/1528-3542.4.2.201 [PubMed: 15222856]
- Hayes AF (2017). *Introduction to mediation, moderation, and conditional process analysis. A regression-based approach* (2nd ed.). The Guilford Press.
- Heuer K, Rinck M, & Becker ES (2007). Avoidance of emotional facial expressions in social anxiety: The approach–avoidance task. *Behaviour Research and Therapy*, 45(12), 2990–3001. [PubMed: 17889827]
- Hu L-T, & Bentler PM (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424–453. 10.1037/1082-989x.3.4.424
- Joormann J, & Gotlib IH (2006). Is this happiness I see? Biases in the identification of emotional facial expressions in depression and social phobia. *Journal of Abnormal Psychology*, 115(4), 705–714. [PubMed: 17100528]
- Kelley NJ, Kramer AM, Young KS, Echiverri-Cohen AM, Chat IK-Y, Bookheimer SY, Nusslock R, Craske MG, & Zinbarg RE (2019). Evidence for a general factor of behavioral activation system

- sensitivity. *Journal of Research in Personality*, 79, 30–39. 10.1016/j.jrp.2019.01.002 [PubMed: 30853731]
- Koelewijn T, Bronkhorst A, & Theeuwes J (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, 134(3), 372–384. [PubMed: 20427031]
- Krach S, Paulus FM, Bodden M, & Kircher T (2010). The rewarding nature of social interactions. *Frontiers in Behavioral Neuroscience*, 4, Article 22. 10.3389/fnbeh.2010.00022
- Kramer AM, Kelley NJ, Chat IK, Young KS, Nusslock R, Craske MG, & Zinbarg RE (2019, 4 3). Replication of a tri-level model of anxiety and depression in a sample of young adults. *PsyArXiv*. 10.31234/osf.io/8mpd2
- Leppänen JM, Milders M, Bell JS, Terriere E, & Hietanen JK (2004). Depression biases the recognition of emotionally neutral faces. *Psychiatry Research*, 128(2), 123–133. 10.1016/j.psychres.2004.05.020 [PubMed: 15488955]
- Mathews A, & MacLeod C (2005). Cognitive vulnerability to emotional disorders. *Annual Review of Clinical Psychology*, 1, 167–195. 10.1146/annurev.clinpsy.1.102803.143916
- Mattick RP, & Clarke JC (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*, 36(4), 455–470. [PubMed: 9670605]
- Meyer TJ, Miller ML, Metzger RL, & Borkovec TD (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy*, 28(6), 487–495. [PubMed: 2076086]
- Mill A, Allik J, Realo A, & Valk R (2009). Age-related differences in emotion recognition ability: A cross-sectional study. *Emotion*, 9(5), 619–630. 10.1037/a0016562 [PubMed: 19803584]
- Mobini S, Reynolds S, & Mackintosh B (2013). Clinical implications of cognitive bias modification for interpretative biases in social anxiety: An integrative literature review. *Cognitive Therapy and Research*, 37(1), 173–182. 10.1007/s10608-012-9445-8
- Mogoa e C, David D, & Koster EH (2014). Clinical efficacy of attentional bias modification procedures: An updated meta-analysis. *Journal of Clinical Psychology*, 70(12), 1133–1157. 10.1002/jclp.22081 [PubMed: 24652823]
- Morris SE, & Cuthbert BN (2012). Research domain criteria: Cognitive systems, neural circuits, and dimensions of behavior. *Dialogues in Clinical Neuroscience*, 14(1), 29–37. [PubMed: 22577302]
- Naragon-Gainey K, Prenoveau JM, Brown TA, & Zinbarg RE (2016). A comparison and integration of structural models of depression and anxiety in a clinical sample: Support for and validation of the tri-level model. *Journal of Abnormal Psychology*, 125(7), 853–867. [PubMed: 27732022]
- Parsons CE, Young KS, Bhandari R, Ijzendoorn MH, Bakermans-Kranenburg MJ, Stein A, & Kringelbach ML (2014). The bonnie baby: Experimentally manipulated temperament affects perceived cuteness and motivation to view infant faces. *Developmental Science*, 17(2), 257–269. 10.1111/desc.12112 [PubMed: 24330495]
- Parsons CE, Young KS, Craske MG, Stein AL, & Kringelbach ML (2014). Introducing the Oxford Vocal (OxVoc) Sounds database: A validated set of non-acted affective sounds from human infants, adults, and domestic animals. *Frontiers in Psychology*, 5, Article 562. 10.3389/fpsyg.2014.00562
- Parsons CE, Young KS, Kumari N, Stein A, & Kringelbach ML (2011). The motivational salience of infant faces is similar for men and women. *PLOS ONE*, 6(5), Article e20632.
- Pittig A, Treanor M, LeBeau RT, & Craske MG (2018). The role of associative fear and avoidance learning in anxiety disorders: Gaps and directions for future research. *Neuroscience & Biobehavioral Reviews*, 88, 117–140. 10.1016/j.neubiorev.2018.03.015
- Pizzagalli DA, Iosifescu D, Hallett LA, Ratner KG, & Fava M (2008). Reduced hedonic capacity in major depressive disorder: Evidence from a probabilistic reward task. *Journal of Psychiatric Research*, 43(1), 76–87. 10.1016/j.jpsychires.2008.03.001 [PubMed: 18433774]
- Posner J, Russell JA, & Peterson BS (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(3), 715–734. [PubMed: 16262989]

- Prenoveau JM, Zinbarg RE, Craske MG, Mineka S, Griffith JW, & Epstein AM (2010). Testing a hierarchical model of anxiety and depression in adolescents: A tri-level model. *Journal of Anxiety Disorders*, 24(3), 334–344. 10.1016/j.janxdis.2010.01.006 [PubMed: 20171054]
- Price RB, Rosen D, Siegle GJ, Ladouceur CD, Tang K, Allen KB, Ryan ND, Dahl RE, Forbes EE, & Silk JS (2016). From anxious youth to depressed adolescents: Prospective prediction of 2-year depression symptoms via attentional bias measures. *Journal of Abnormal Psychology*, 125(2), 267–278. 10.1037/abn0000127 [PubMed: 26595463]
- Rapee RM, Craske MG, & Barlow DH (1994). Assessment instrument for panic disorder that includes fear of sensation-producing activities: The Albany Panic and Phobia Questionnaire. *Anxiety*, 1(3), 114–122. [PubMed: 9160559]
- Rhodes G, Yoshikawa S, Clark A, Lee K, McKay R, & Akamatsu S (2001). Attractiveness of facial averageness and symmetry in non-Western cultures: In search of biologically based standards of beauty. *Perception*, 30(5), 611–625. 10.1068/p3123 [PubMed: 11430245]
- Rømer Thomsen K, Whybrow PC, & Kringelbach ML (2015). Reconceptualizing anhedonia: Novel perspectives on balancing the pleasure networks in the human brain. *Frontiers in Behavioral Neuroscience*, 9, Article 49. 10.3389/fnbeh.2015.00049
- Smith E, Weinberg A, Moran T, & Hajcak G (2013). Electrocardiac responses to NIMSTIM facial expressions of emotion. *International Journal of Psychophysiology*, 88(1), 17–25. 10.1016/j.ijpsycho.2012.12.004 [PubMed: 23280304]
- Syger JH (1989). EzPATH: A supplementary module for SYSTAT and SYSGRAPH. SYSTAT.
- Suess F, Rabovsky M, & Abdel Rahman R (2015). Perceiving emotions in neutral faces: Expression processing is biased by affective person knowledge. *Social Cognitive and Affective Neuroscience*, 10(4), 531–536. 10.1093/scan/nsu088 [PubMed: 24948155]
- Tottenham N, Tanaka JW, Leon AC, McCarry T, Nurse M, Hare TA, Marcus DJ, Westerlund A, Casey BJ, & Nelson C (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, 168(3), 242–249. 10.1016/j.psychres.2008.05.006 [PubMed: 19564050]
- Treadway MT, Bossaller NA, Shelton RC, & Zald DH (2012). Effort-based decision-making in major depressive disorder: A translational model of motivational anhedonia. *Journal of Abnormal Psychology*, 121(3), 553–558. 10.1037/a0028813 [PubMed: 22775583]
- Treadway MT, Buckholtz JW, Schwartzman AN, Lambert WE, & Zald DH (2009). Worth the ‘EEfRT’? The effort expenditure for rewards task as an objective measure of motivation and anhedonia. *PLOS ONE*, 4(8), Article e6598. 10.1371/journal.pone.0006598
- Trew JL (2011). Exploring the roles of approach and avoidance in depression: An integrative model. *Clinical Psychology Review*, 31(7), 1156–1168. [PubMed: 21855826]
- Watson D, Weber K, Assenheimer JS, Clark LA, Strauss ME, & McCormick RA (1995). Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *Journal of Abnormal Psychology*, 104(1), 3–14. [PubMed: 7897050]
- White LK, Degnan KA, Henderson HA, Pérez-Edgar K, Walker OL, Shechner T, Leibenluft E, Bar-Haim Y, Pine DS, & Fox NA (2017). Developmental relations among behavioral inhibition, anxiety, and attention biases to threat and positive information. *Child Development*, 88(1), 141–155. 10.1111/cdev.12696 [PubMed: 28042902]
- Whitton AE, Treadway MT, & Pizzagalli DA (2015). Reward processing dysfunction in major depression, bipolar disorder and schizophrenia. *Current Opinion in Psychiatry*, 28(1), 7–12. 10.1097/YCO.000000000000122 [PubMed: 25415499]
- Wieser MJ, & Brosch T (2012). Faces in context: A review and systematization of contextual influences on affective face processing. *Frontiers in Psychology*, 3, Article 471. 10.3389/fpsyg.2012.00471
- Yoon KL, & Zinbarg RE (2008). Interpreting neutral faces as threatening is a default mode for socially anxious individuals. *Journal of Abnormal Psychology*, 117(3), 680–685. 10.1037/0021-843x.117.3.680 [PubMed: 18729619]
- Young KS, Parsons CE, LeBeau RT, Tabak BA, Sewart AR, Stein A, Kringelbach ML, & Craske MG (2017). Sensing emotion in voices: Negativity bias and gender differences in a validation study of

the Oxford Vocal (“OxVoc”) sounds database. *Psychological Assessment*, 29(8), 967–977. 10.1037/pas0000382 [PubMed: 27656902]

Yu CY (2002). Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes [Unpublished doctoral dissertation, University of California, Los Angeles].

Zimmerman M, Coryell W, Corenthal C, & Wilson S (1986). A self-report scale to diagnose major depressive disorder. *Archives of General Psychiatry*, 43(11), 1076–1081. [PubMed: 3767597]

Zinbarg RE, & Barlow DH (1996). Structure of anxiety and the anxiety disorders: A hierarchical model. *Journal of Abnormal Psychology*, 105(2), 181–193. [PubMed: 8722999]

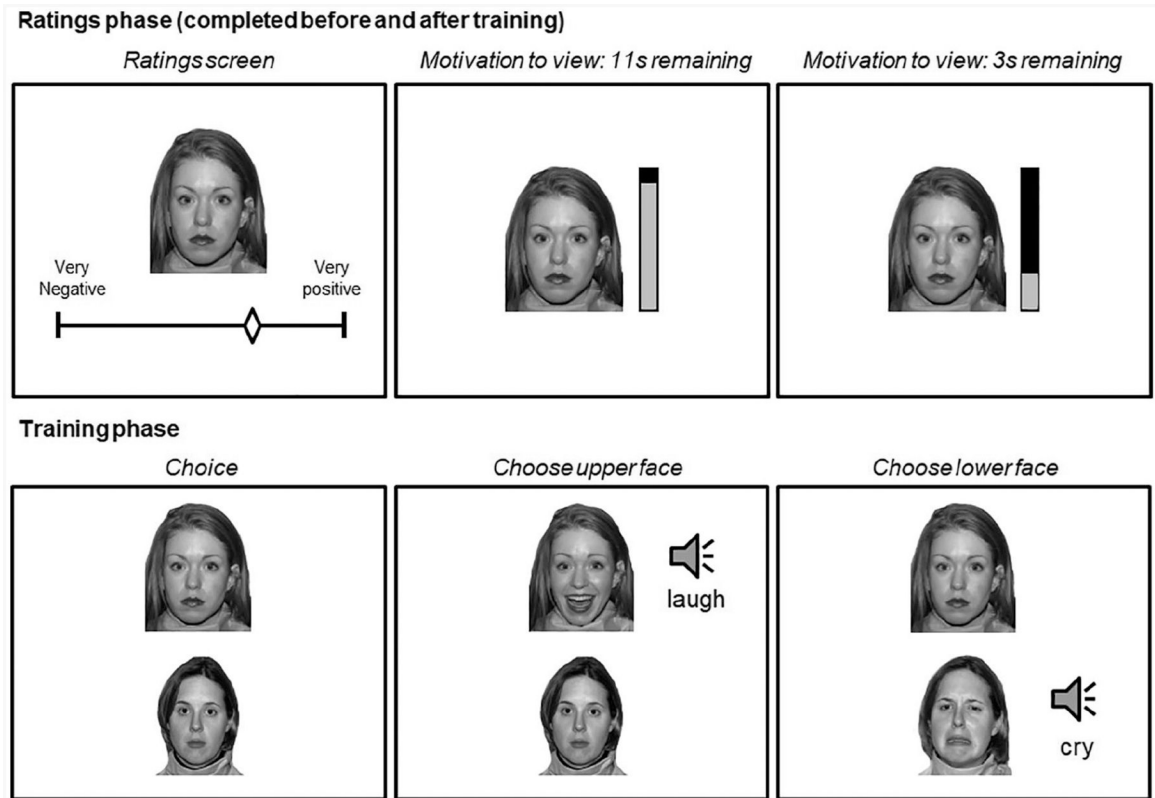


Figure 1. Example screens presented during experiment. Upper: neutral faces were rated on arousal, valence and pleasantness (left), and participants completed a keypress “motivation to view” measure (mid and right) before and after training. Lower: during the training phase, participants learned to associate different faces with different probabilities of positive and negative feedback, by selecting one of two neutral faces (left) and receiving either positive (mid) or negative (right) social feedback.

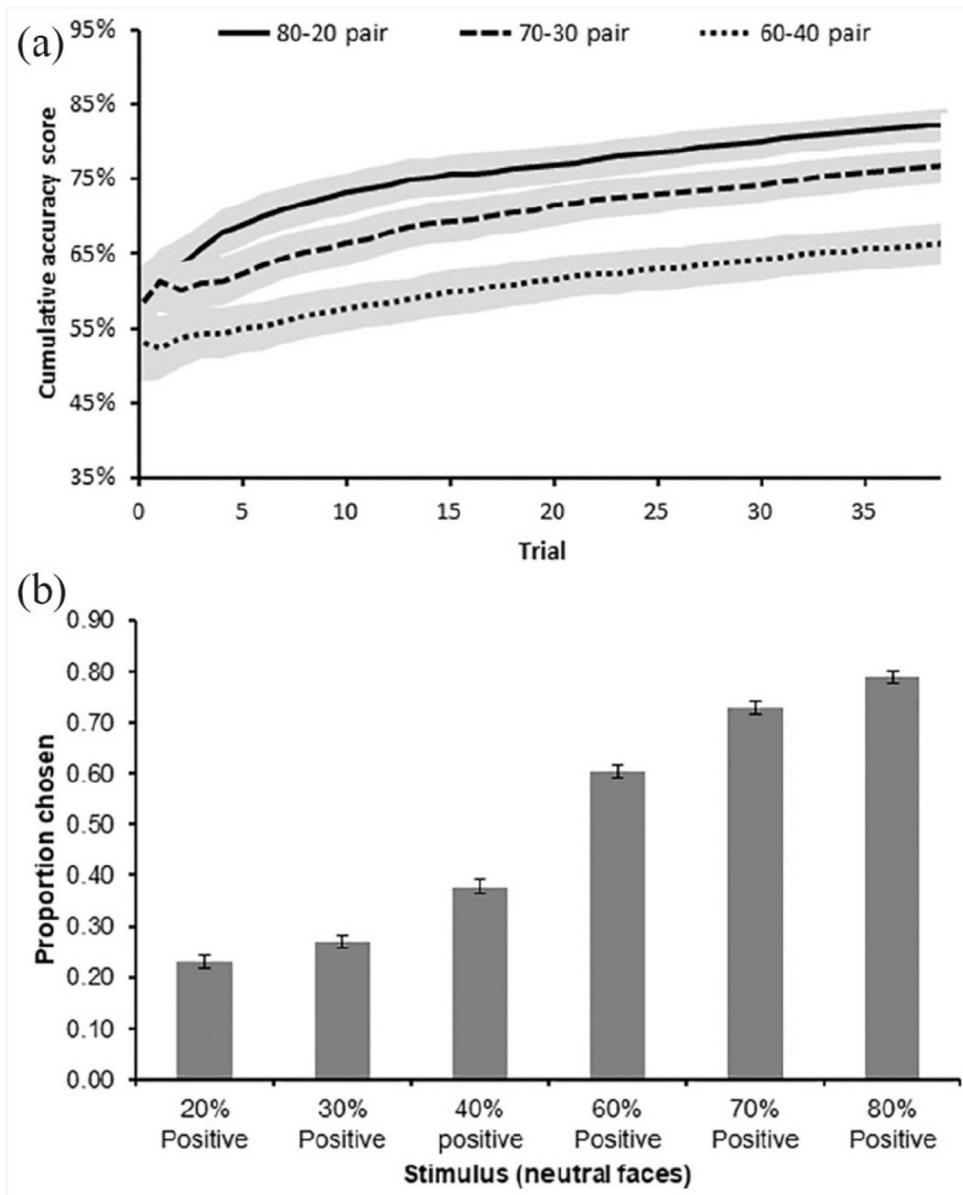


Figure 2. Training and testing phase data. (a) Cumulative accuracy scores across trials are separated by face pair. Performance accuracy was highest for the 80–20 pair, followed by the 70–30 pair, and then the 60–40 pair (shaded areas represent 95% confidence intervals). (b) Proportion of trials selected as “more happy” during the testing phase increased corresponding with stimulus contingencies during training phase (error bars indicate mean ± standard error).

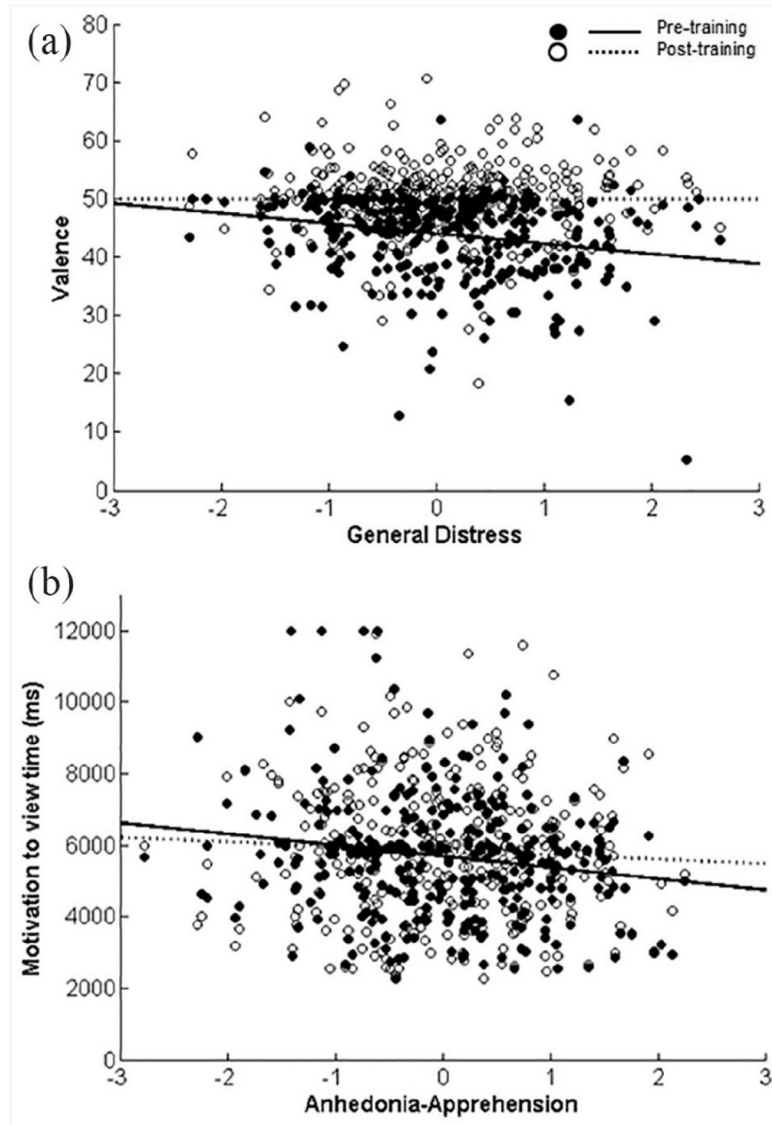


Figure 3. Changes in associations with symptoms from pre to post-training. (a) There was evidence of a negative bias prior to training that was significantly reduced following training. (b) Higher levels of anhedonia-apprehension (lower scores) were associated with shorter viewing times before training, but this effect was not significantly reduced following training.

Table 1.Demographics and diagnostic status of participants included in the study ($N = 324$).

Demographic variable		
Age (M, SD)	19.52	0.73
Sex	211F, 112M, IT ^a	
Race/ethnicity ($N, \%$)		
White	167	51.54
Black or African American	28	8.64
Asian	99	30.56
American Indian or Alaska Native	7	2.16
Multiracial	22	6.79
Hispanic or Latino	82	25.31
Diagnostic status, $n = 294$ ($N, \%$)		
One or more anxiety disorders ^b	99	33.67
Generalised anxiety disorder	33	33.33
Obsessive compulsive disorder	10	10.10
Panic disorder	10	10.10
Post-traumatic stress disorder	10	10.10
Social anxiety disorder	54	54.55
Separation anxiety disorder	1	1.01
Specific phobia	31	31.31
Depressive disorder ^c	27	9.18
Major depressive disorder	21	7.14
Persistent depressive disorder	6	2.04

^aTransgender (this participant was excluded from sex difference analyses),^b62 individuals met criteria for more than one disorder,^c24 of these individuals also met criteria for one or more anxiety disorders.

Table 2.

Mean, standard error, and 95% confidence intervals for change in emotional responses (arousal and valence ratings) and reward value (pleasantness ratings and viewing time) of face stimuli.

Stimulus type (% positive feedback)	Mean	Standard error	95% confidence interval	
			Lower bound	Upper bound
Change in arousal rating				
20% Positive	1.03	1.00	-0.94	2.99
30% Positive	1.87	0.96	-0.02	3.76
40% Positive	3.25 [*]	1.02	1.23	5.27
60% Positive	6.10 [*]	1.08	3.98	8.23
70% Positive	7.03 [*]	1.11	4.85	9.22
80% Positive	7.38 [*]	1.25	4.92	9.84
Change in valence rating				
20% Positive	-2.18 [*]	.94	-4.02	-0.34
30% Positive	-1.51	.87	-3.21	0.19
40% Positive	1.32	.90	-0.45	3.08
60% Positive	10.05 [*]	.85	8.39	11.72
70% Positive	12.76 [*]	.83	11.12	14.40
80% Positive	15.63 [*]	.92	13.81	17.44
Change in pleasantness rating ("liking")				
20% Positive	-0.03	0.97	-1.94	1.88
30% Positive	1.41	0.99	-.54	3.36
40% Positive	3.92 [*]	0.95	2.05	5.79
60% Positive	8.89 [*]	0.95	7.02	10.76
70% Positive	11.30 [*]	0.95	9.42	13.18
80% Positive	13.62 [*]	1.01	11.63	15.60
Change in viewing time (ms, "wanting")				
20% Positive	-425.93 [*]	113.22	-648.67	-203.19
30% Positive	-229.17	119.17	-463.62	5.29
40% Positive	41.67	144.63	-242.87	326.20
60% Positive	162.81	146.76	-125.92	451.54
70% Positive	572.53 [*]	151.95	273.59	871.48
80% Positive	836.42 [*]	158.89	523.84	1149.00

* Significant difference ($p < .05$).

Table 3.

Results of regression analyses investigating the relationship between tri-level factor scores and pre-training measures of emotion (arousal and valence) and reward value (pleasantness and viewing time).

	<i>B</i>	<i>SE b</i>	β
Arousal			
General distress	-.20	.61	-.02
Fears	.30	.65	.03
Anhedonia-apprehension	.65	.61	.06
Sex	-.35	1.17	-.02
Valence			
General distress	-1.78	.43	-.23*
Fears	-.15	.47	-.02
Anhedonia-apprehension	.15	.44	.02
Sex	.53	.83	.04
Pleasantness			
General distress	1.52	.95	.09
Fears	.53	.50	.06
Anhedonia-apprehension	.47	.53	.05
Sex	-.32	.50	-.04
Viewing time			
General distress	130.91	104.74	.07
Fears	30.69	111.94	.02
Anhedonia-apprehension	-318.69	104.32	-.17*
Sex	195.40	200.05	.06

SE: standard error.

* $p < .01$