

Title: Shifting threat criterion for morphed facial expressions reduces negative affect

Samantha L. B. O'Brien^{1,2}, Bruce K. Christensen^{1,3}, & Stephanie C. Goodhew^{1,4}

¹*Research School of Psychology, The Australian National University*

²Samantha.O'Brien@anu.edu.au

³Bruce.Christensen@anu.edu.au

⁴Stephanie.Goodhew@anu.edu.au

Corresponding Author: Samantha L. B. O'Brien

Email: Samantha.O'Brien@anu.edu.au

Address: Research School of Psychology (Building 39)

The Australian National University, Canberra, 2601

Declarations of interest: None

Running head: Shifting threat criterion for morphed facial expressions reduces negative affect

Abstract

It is well-established that anxiety and/or depression are associated with a negative bias when interpreting ambiguous information. This study tested the novel hypothesis that the criterion one sets for judging a stimulus as threatening is a core aspect of this bias. A sample of 174 participants were divided into neutral ($n = 87$) and threatening ($n = 87$) training conditions. Participants performed a facial expression detection task, in which criterion was shifted in the liberal (threatening condition) or conservative (neutral condition) direction via differential reward contingencies. Training conditions were successful in inducing large shifts in criterion as intended. There was also a small change in sensitivity in the neutral condition, however, the manipulation is still considered successful given the substantive effect size for change in criterion compared to change in sensitivity. As predicted, conservative criterion-training resulted in significantly lower levels of negative affect post-training. No significant change was found for liberal criterion-training on negative affect. Positive affect also decreased across time regardless of condition. Overall, the reduction in negative affect following conservative criterion-training demonstrates that modifying criterion impacts affect and identifies criterion setting as a potential target in the treatment of mental health disorders with prominent negative affect.

Keywords: Interpretation bias, signal detection theory, criterion, ambiguous, facial expressions

Shifting threat criterion for morphed facial expressions reduces negative affect

Shifting threat criterion for morphed facial expressions reduces negative affect

Shifting threat criterion for morphed facial expressions reduces negative affect

Individuals experiencing symptoms of anxiety and/or depression consistently judge emotionally ambiguous stimuli as more negatively valenced than participants free of psychopathology; this effect is known as *interpretation bias*. Interpretation bias has been demonstrated across judgements of ambiguous scenarios (Chen et al., 2019; Chen et al., 2020; Mathews & MacLeod, 2002; Stopa & Clark, 2000), facial expressions (Chen et al., 2019; Maoz et al., 2016), and homophones and homographs (Mathews et al., 1989, Mogg et al., 2004). Cognitive theories of anxiety and depression postulate that information biases such as these play a significant role in the maintenance of negative affective disorders (Beck, 2008; Clark & Wells, 1995; Matthews & MacLeod, 2005). Accordingly, cognitive bias modification (CBM) methodologies, which reinforces participants for less negative judgements, have been shown to reduce anxiety and depression, even to the point of removing some participants from relevant diagnostic categories (Hirsch et al., 2016; Hirsch et al., 2018; Mathews & Mackintosh, 2000; Mathews et al., 2007; Salemink et al., 2007, 2009).

The most commonly used CBM paradigm for the modification of interpretation bias (CBM-I) requires participants to provide resolutions of ambiguous scenarios (Mathews & Mackintosh, 2000). Valence is controlled by the use of word fragments with one possible meaningful resolution which allows for the manipulation of valence based on condition (see Mathews & Mackintosh, 2000 for details of classic CBM-I methodology). There is general consensus that the manipulation of interpretation bias, using CBM-I procedures, reduces the symptoms of anxiety and depression (Clerkin & Teachman, 2011; Hirsch et al., 2018; Hirsch et al., 2020; Hirsch et al., 2021; Hoppitt et al., 2010; Steinman & Teachman, 2014). However, this effect is not uniform, as some studies fail to demonstrate symptom reduction and/or adequate generalisation to other measures of interpretation bias (Salemink et al., 2010;

Shifting threat criterion for morphed facial expressions reduces negative affect

Standage et al., 2010). To resolve these mixed findings, more research into the underlying mechanisms responsible for interpretation bias is required. This could allow for more targeted training programs, resulting in more consistent findings. Here we posit that signal detection theory (SDT) provides a useful framework for exploring such underlying mechanisms.

Signal detection theory (SDT) is a framework used to understand fixed choice decision-making under uncertainty, such as ambiguity (Green & Swets, 1966). It allows for the separation of two main components in the decision-making process—sensitivity and response bias. Sensitivity is the ability to distinguish target stimuli from a non-target stimulus, whereas response bias is the general propensity of the observer to identify a stimulus as the target (Abdi, 2007; Green & Swets, 1966). Response bias is directly dependent on the criterion set by the observer. In the present context, this corresponds to the amount of threat signal needed to identify a stimulus as threatening. If, during a task requiring the detection of a threatening facial expression, one sets a criterion where less information is needed to decide that a given face was threatening (i.e., increasing one's propensity to endorse the face as threatening), they would be viewed as having set a more *liberal* criterion. Whereas, setting a criterion where the observer required more information (i.e., increasing one's propensity to reject the face as threatening) would suggest a more *conservative* criterion. Importantly, where one sets their criterion is malleable, and implicitly shifts in accordance to experimental design factors such as feedback, task demands, and varying base rates of presentation (Ackermann & Landy, 2015; Coombs, et al., 1970; Estes & Maddox, 1995; Han & Dobbins, 2008, 2009; Healy & Kubovy, 1978; Rhodes & Jacoby, 2007).

The SDT framework has been applied to understanding related phenomena, including attentional bias. Attentional bias occurs when preferential attention is given to a particular

Shifting threat criterion for morphed facial expressions reduces negative affect

category or dimension of information (MacLeod & Mathews, 2012). Attentional bias towards threat, therefore, describes the preferential processing of threatening information, such as angry facial expressions, at the expense of attending to other, non-threatening information. A substantial body of evidence demonstrates that individuals suffering from a variety of clinical anxiety disorders and those with sub-clinical levels of anxiety possess an attentional bias towards threat (for a meta-analysis, see Bar-Haim et al., 2007). Importantly, several studies conclude that biased attention towards threat is driven by group differences in response bias and *not* sensitivity (Becker, & Rinck, 2004; Liu et al., 2014; Manguno-Mire et al., 2005; Windmann & Krüger, 1998). That is, individuals with anxiety are not better at identifying threat, but instead set a lower threshold to identify stimuli as threatening compared to healthy controls.

In a similar vein, we propose that criterion setting is the underlying mechanism responsible for negative interpretation bias – i.e., that individuals with anxiety and depression set a lower threshold to *interpret* something as threatening. To our knowledge, the current study is the third study to utilise SDT analyses in an interpretation bias context. However, previous studies that have utilized this framework do not provide a rationale for its use and as such do not specify criterion as the underlying mechanism responsible (Eysenck et al., 1991; Mathews & Mackintosh, 2000). Instead, these studies use SDT as an analytic tool and do not draw clear conclusions about the relative contributions of criterion and sensitivity. Consequently, the current study was designed to specifically investigate the impact of criterion setting on affect.

Theoretical and empirical advances suggest that anxiety and depression can be integrated within a common underlying construct, namely negative affectivity (NA; Watson & Clark, 1984; Watson et al., 1988). NA is defined as a general dimension of psychological

Shifting threat criterion for morphed facial expressions reduces negative affect

distress that subsumes a broad range of negative mood states, including anxiety and sadness (as well as emotions such as hostility, scorn, fear and disgust; Watson & Clark, 1984; Watson et al., 1988). Moreover, large-scale factor and meta-analytic studies have concluded that the most valid models of mental illness architecture situate anxiety and depressive disorders alongside each other, nested under a superordinate factor that embodies NA (e.g., Krueger, 1999; Krueger & Markon, 2006; Markon, 2010). In contrast, positive affectivity (PA) represents a second, broad dimension of mood that subsumes an array of positive emotions including happiness, enthusiasm, contentment, vigour, and warm heartedness (Watson & Tellegen, 1985). Importantly, NA and PA are dissociable constructs that demonstrate an inverse relationship with one another (Kercher, 1992; Tellegen et al., 1999). As a result, disorders robustly characterised by NA also manifest low levels of PA (Watson et al., 1988). In line with this conceptualisation of anxiety and depression, the current study focusses on high levels of NA and low levels of PA as core indicators of distress-related psychopathology and, in particular, seeks to explore the impact of altering one's criterion for judging the valence of human face photographs on these indicators.

This aim was achieved in the present study by measuring both positive and negative affect (Positive and Negative Affect Schedule; PANAS) and SDT sensitivity and criterion for affective judgements at baseline (Block 1), before participants were assigned to one of two training conditions designed to differentially shift criterion: (a) the threatening training condition that reinforced judgements towards threat; or, (b) the neutral training condition that reinforced judgements towards neutrality (Block 2). The PANAS and the SDT detection task were then re-administered to assess if the training impacted affect and sensitivity or criterion (Block 3).

Shifting threat criterion for morphed facial expressions reduces negative affect

It was predicted that experimentally shifting criterion in this way would result in changes in affect. In particular, it was hypothesized that rewarding participants for identifying ambiguous stimuli as threatening would shift their criterion in a more liberal direction (i.e., more easily accept facial expressions as threatening). In turn, it was hypothesized that shifting criterion to a more liberal setting would increase their NA and decrease their PA. Furthermore, it was hypothesized that rewarding participants for identifying ambiguous stimuli as neutral would shift their criterion in a more conservative direction, which would correspondingly increase their PA and decrease their NA.

Method

Participants

One-hundred and eighty-six individuals (144 females and 42 males, average age = 20.98 years), volunteered to participate in this experiment. Participants were recruited from the Australian National University student population, as well as from the surrounding community via research participation website and social media ads. Participants received either one unit of course credit or \$15 compensation for their time. Data for seven participants is not reported as they did not complete the full experiment due to withdrawal ($n = 4$), computer errors ($n = 2$), and failing the practice ($n = 1$). The data for 12 participants were excluded from further analysis due to scoring ≥ 7 on the Michigan Alcohol Screening Test (MAST)—indicating problematic drinking behaviours (Hedlund & Vieweg, 1984; Selzer, 1971; Selzer, et al., 1975). The final sample consisted of 174 individuals (137 females, 37 males, average age = 21.07). We sought to achieve a large sample size given the likely small effect size for this proof of concept study and thus aimed for a usable sample of 150 participants. Final sample size was larger than all comparably relevant studies reported in popular meta-analyses and systematic reviews in the field (Hallion & Ruscio, 2011; Fodor et al., 2020).

Measures and materials

Questionnaires

Questionnaires were completed using paper format. Participants completed the MAST (Selzer, 1971). The MAST is a 24-item self-report scale measuring alcohol use and consumption in order to assess problem drinking. Respondents were instructed to “read each statement and circle yes or no if that particular statement applies to you”. Weighted total scores range from 0 to 53. The MAST has been well studied and has sound psychometric properties (Laux et al., 2002; 2004). MAST was utilised as a screening tool given higher alcohol consumption among university students, and heavy alcohol use and alcohol use disorders being associated with cognitive and executive functioning deficits, changes in brain structures, and impairments in facial emotion recognition particularly with anger expressions such as those used in the current study (Bora & Zorlu, 2017; Hallett et al., 2012; Rehm et al., 2019).

Participants also completed the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988) on two occasions throughout the experiment. The PANAS consists of two 10-item mood scales assessing current levels of positive and negative affect. Items are mood-related adjectives for PA (e.g., *interested, excited, attentive*) and NA (e.g., *distressed, nervous, scared*). Respondents were instructed to “rate each item on a scale from one to five – 1 ‘very slightly or not at all’, 2 ‘a little’, 3 ‘moderately’, 4 ‘quite a bit’ and 5 ‘extremely’ – descriptive of how you are feeling right now”. As such, possible scores range from 10 to 50 for each of the mood scales. There are many different time periods that can be used when administering the PANAS. The momentary instruction, as is used in the current study, allows for short intervals between administrations which is commonly utilised in research (with similar or shorter time intervals between administrations as those in the current study) to

Shifting threat criterion for morphed facial expressions reduces negative affect

capture pre- and post- intervention manipulation effects on affect (Birk et al., 2011; Holmes et al., 2009; Jones & Shapre, 2014; Kopetz et al., 2017; Meneguzzo et al., 2020; Sedgmond et al., 2020). The PANAS had very high internal consistency in the current sample, as determined by Cronbach's alpha of .92 for the PA scale and .83 for the NA scale. This is consistent with previous research (Crawford & Henry, 2004).

Stimuli

Neutral and threatening facial expressions were selected from the NimStim face stimulus database (Tottenham et al., 2009) and were morphed from neutral to threatening in 10% increments using Abrosoft FantaMorph (Version 4 2002–2007 Edition; see Figure 1 for an example of the successive stages of a morphed face set). A total of 110 images were created from morphing 10 individuals faces (5 female; 5 male) and were used in both the baseline (Block 1) and follow-up detection tasks (Block 3). All of the face photographs subtended approximately 8° wide x 10° high of visual angle and were presented individually in the centre of the screen.

Figure 1

An Example of Successive Morphing from Neutral to Threatening in 10% Morph Increments



Note. 0% threatening (far left) to 100% threatening (far right). 11 images constitute the image set for this individual. There are 10 image sets used for the detection task in Block's 1 and 3.

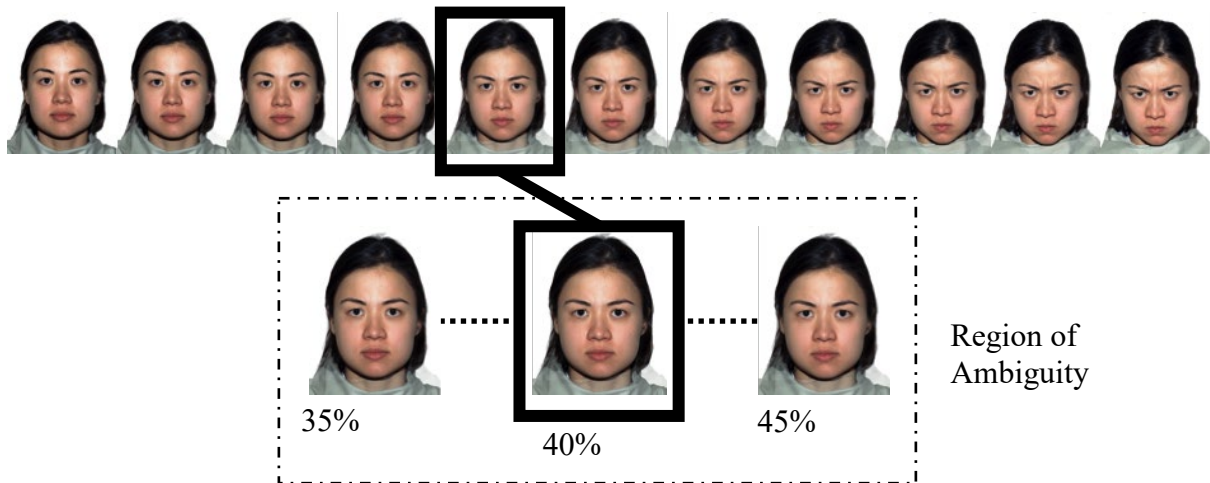
Ambiguous stimuli

Shifting threat criterion for morphed facial expressions reduces negative affect

The criterion-training task utilised ambiguous facial expressions. For each of the image sets used in Blocks 1 and 3, the most ambiguous morph was identified¹. A specified region of ambiguity ($\pm 5\%$ morph increments around the most ambiguous morph for each image set; see Figure 2) for each of the 10 faces was used to create a total of 30 ambiguous facial expression stimuli used in the criterion training block.

Figure 2

Region of Ambiguity for Morphed Facial Expressions



Note. The most ambiguous face for this image set was the 40% threatening morph. In order to create a larger sample of stimuli for Block 2, three ambiguous morphs were used to create a region of ambiguity— the precise morph that was most ambiguous, plus the morphs 5% above and below this value.

Procedure and Design

Participants provided written informed consent prior to participation. Participants then completed the pre-training PANAS questionnaire, followed by the MAST. Next, participants

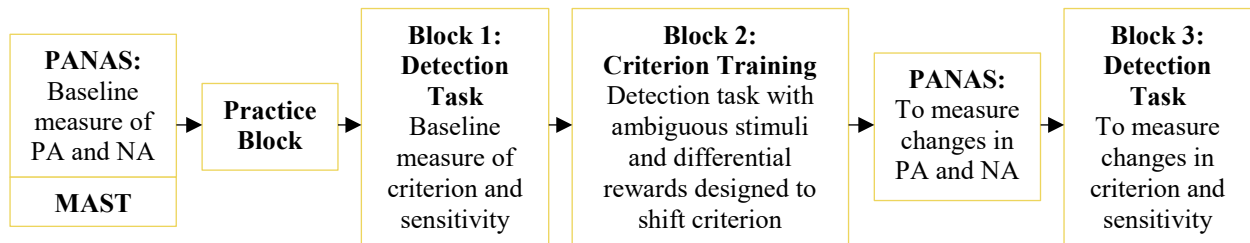
¹ During pilot testing of stimuli, participants were required to identify each face as either neutral or threatening and were also required to rate their confidence of their decision. The most ambiguous morph for each face was identified by average ratings closest to 50%, demonstrating the point of minimum consensus among participants. Please contact corresponding author for more details.

Shifting threat criterion for morphed facial expressions reduces negative affect

completed the practice block, then the experimental block designed to provide baseline measures of criterion and sensitivity, followed by the training block, and then the post-training measurement of PANAS, and finally the block to provide post-training assessment of criterion and sensitivity (see Figure 3 for a schematic representation). The entire experiment took approximately one hour to complete. The experimental task consisted of three blocks—details of each block are described in the following sections. In all experimental blocks, participants were required to detect whether the face presented was threatening (signal) or neutral (noise). The forced choice detection task, used in each block, was created and displayed using the Psychophysics toolbox in MATLAB (Brainard, 1997). Participants' heads were steadied with the use of a chinrest fixed at 44cm viewing distance, which was kept constant throughout the duration of the experiment.

Figure 3

Schematic Representation of Overall Experiment Order



Practice Block

Prior to the commencement of the experimental task, participants had to pass the Practice Block. This consisted of 12 trials designed to familiarise participants with the task. For each response, participants received accuracy feedback in the form of the words ‘correct’ or ‘incorrect’ appearing on the monitor. Accuracy was determined by pilot testing to identify the average point at which participants switched from stating a face was neutral to stating it was threatening. This point was calculated for each individual set of morphs and the average

Shifting threat criterion for morphed facial expressions reduces negative affect

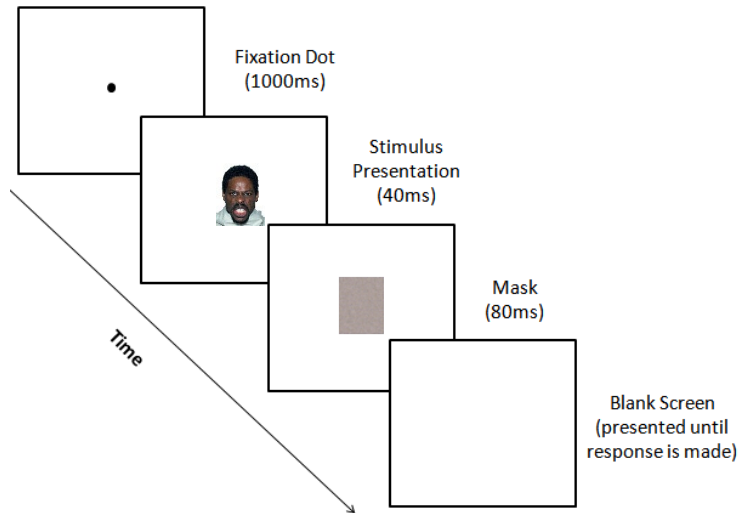
change over points was either at a 30% or 40% morph level for each set. Participants were permitted five attempts of the practice block in order to achieve the minimum accuracy requirement, after which the task was deemed too difficult for them and their participation was terminated as to not exhaust or frustrate participants, as well as to stay within the testing time period that participants consented to. Only one participant was excluded at the practice stage.

Block 1: Signal Detection Task Measuring Baseline Criterion and Sensitivity

This block consisted of a standard signal detection task (see Figure 4) administered to determine baseline criterion and sensitivity. A total of 220 trials were used in this block, with each morphed face presented twice in a random sequence. This number of trials met the norm for SDT studies (e.g., Becker & Rinck, 2004; Frenkel et al., 2009; Manguno-Mire et al., 2005; Winton et al., 1995; Windmann & Kruger, 1998). One rest break was given halfway through this block; the length of which was at the discretion of each participant. On each trial a fixation dot was presented in the centre of the screen, followed by a facial expression presented for 40ms, which was then masked with a grated grey image for 80ms. Participants were instructed to “press the ‘Z’ key for neutral or ‘?’ key for threatening after the face has been presented and the screen has gone blank”. Responses were not timed and no feedback was given in this block.

Figure 4

Signal Detection Task Used in Blocks 1 and 3



Note. Schematic example of the sequence of the signal detection task in Blocks 1 and 3 with a threatening stimulus. The inter-trial period was 507ms.

Block 2: Criterion-Training Detection Task

This block consisted of a signal detection task with differential rewards and feedback according to condition. Participants were informed that that this block was longer and more difficult compared to the first block. They were specifically informed that the faces in this block were morphed closer together so it is more difficult to distinguish between neutral and threatening. Participants were told they would receive feedback due to the increased length and difficulty of this block. They were told they would win 10 points when they were correct, and lose 10 points when they were incorrect. They were also informed that points would accumulate between each break period. The threatening training group received positive feedback (“Correct!” displayed on the screen in green text) and points (+10 points) for indicating that an ambiguous facial expression was threatening. Conversely, when responding that an ambiguous face was neutral, the threatening training group received negative

Shifting threat criterion for morphed facial expressions reduces negative affect

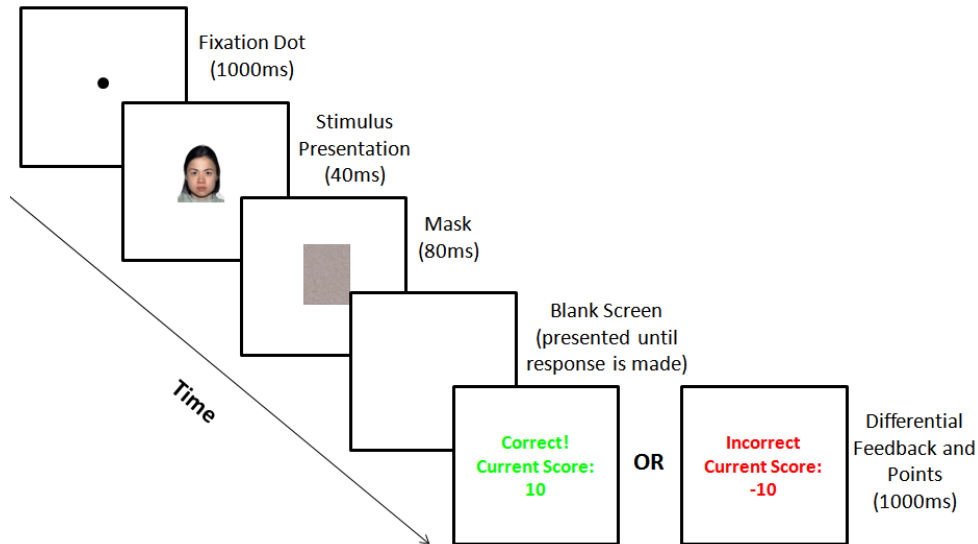
feedback (“Incorrect” displayed on the screen in red text) and were deducted points (-10 points). Points accumulated as trials continued and total points were displayed after each trial (similar to a game). The neutral training group received the same positive feedback and points for indicating that an ambiguous facial expression was neutral, and negative feedback when responding that it was threatening. Please see Figure 5 for a schematic representation of the task for ambiguous trials and a visual depiction of possible feedback options. These conditions were designed to promote liberal criterion setting for threat in the threatening training condition and conservative criterion setting for threat in the neutral training group. The assignment of participants into these two test conditions was counterbalanced.

There was a combined total of 320 trials in this block—300 ambiguous and 20 catch. Stimuli for the ambiguous trials was from the ‘region of ambiguity’ dataset. On these trials, the differential rewards and feedback were given to participants. Ambiguity was essential to allow feedback to be believable (in line with CMB-I literature). Catch trials consisted of 100% threatening and 100% neutral facial expressions and accurate feedback was given on these trials. These were included to reassure participants that pressing one response the entire time would not always give positive feedback.

One designated rest break was given for each quarter of the trials in this block due to its increased length compared to the other blocks. At this time, participants were presented with a summary of their performance in that quarter, indicated by the total amount of points they earned. Tallied points reset after each rest break, to encourage earning a higher score for the next quarter. The length of these designated breaks with score presentation was at the discretion of each participant.

Figure 5

Ambiguous Trial in Signal Detection Task Used in Block 2



Note. Schematic representation of an *ambiguous* trial in Block 2. On all ambiguous trials, the threatening training group received the “Correct” feedback response when responding that the presented face was threatening and received the “Incorrect” feedback response when responding that the presented face was neutral. The neutral training group received opposite feedback. Scores accumulated within each quarter of trials, separated by rest breaks. The inter-trial period was 507ms.

Block 3: Signal Detection Task to Assess Changes in Criterion and Sensitivity

An identical signal detection task to Block 1, re-administered to measure changes in criterion and sensitivity to assess if the experimental criterion manipulations were successful.

Calculations of Criterion and Sensitivity

Consistent with standard SDT procedures (Green & Swets, 1966), criterion was measured by β and sensitivity measured by d' . The β and d' parameters were calculated using Hit and False Alarm rates from the discrimination tasks in Blocks 1 and 3 for each participant. A Hit was coded if a participant correctly identified a threatening facial

Shifting threat criterion for morphed facial expressions reduces negative affect

expression as threatening. A False Alarm (FA) response was coded if a participant incorrectly identified a neutral facial expression as threatening. From these variables, the equations below were used to calculate β and d' , where z = standardized scores. Positive β values indicate a bias toward neutral and negative β values indicate a bias toward negative. Higher d' values reflect a greater ability to distinguish a target stimulus from a non-target stimulus.

$$d' = z(Hits) - z(FA) \qquad \beta = \frac{z(Hit)+z(FA)}{2}$$

Data Analysis

A mixed-Analysis of Variance (ANOVA) served as the principal statistical procedure where training (i.e., threat or neutral) served as the between-subjects independent variable (IV) and Time (i.e., pre- versus post- criterion-training) served as the within-subjects IV. Across analyses the dependent variables (DVs) of interest were: criterion, sensitivity, NA and PA.

Results and Discussion

Data cleaning and screening

Data appeared to be approximately normally distributed for dependent variables (criterion, sensitivity, and PA) grouped by training condition (i.e., Neutral versus Threatening). Data was positively skewed for the dependent variable NA grouped by training condition (i.e., Neutral versus Threatening). This was not troublesome for the analysis as two-way mixed ANOVA is considered robust to violations of normality, there is a fairly large sample size and distributions in both groups were similarly skewed for this variable, altogether providing valid results (Laerd Statistics, 2015). Data screening identified multiple outliers. All analyses were run with and without outliers. There were no significant differences in reported results, as such, analyses are reported with outliers included in

Shifting threat criterion for morphed facial expressions reduces negative affect

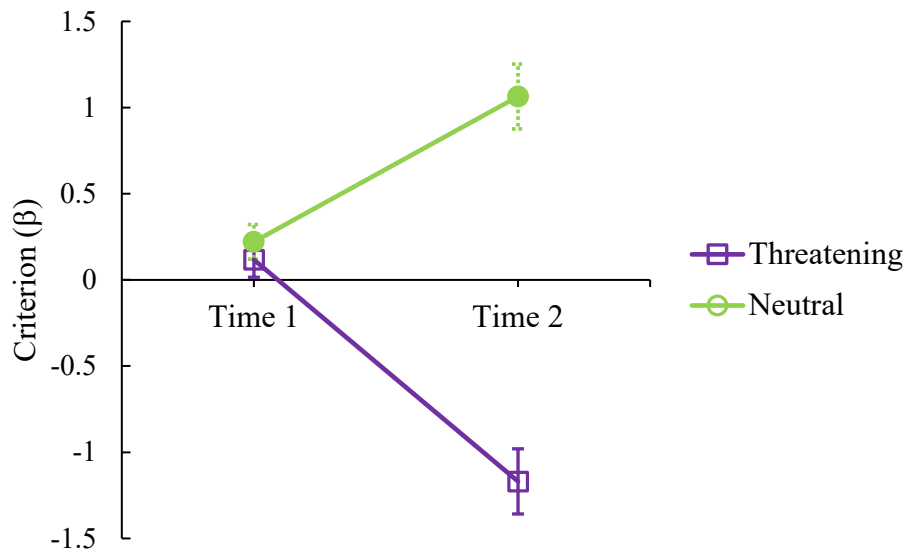
analyses (see supplementary material for reported analyses excluding outliers). Raw data are available via OSF here: <https://osf.io/u2whf/>.

Did the training manipulation impact criterion?

As a necessary manipulation check, we first tested whether the training impacted criterion setting in the threat detection task. A mixed-ANOVA revealed that the Training by Time interaction on criterion was significant, $F(1, 171) = 295.27, p < .001, \eta_p^2 = .633$ (see Figure 6). This interaction reflected the fact that criterion significantly increased in the neutral training condition from Time 1 to Time 2 ($F(1, 86) = 120.97, p < .001, \eta_p^2 = .584, M = .842, SE = .077$) and significantly decreased in the threatening training condition, ($F(1, 85) = 173.77, p < .001, \eta_p^2 = .672, M = 1.286, SE = .098$). These changes were in the expected direction and indicate that training towards threatening facial expressions resulted in a more liberal criterion being adopted following training, whereas training towards neutral facial expressions resulted in a conservative criterion being adopted. The magnitude of this effect was large (i.e., partial eta-squared was greater than 0.25; Tabachnick & Fidell, 2013).

Figure 6

Group Means for Criterion Between Training Conditions



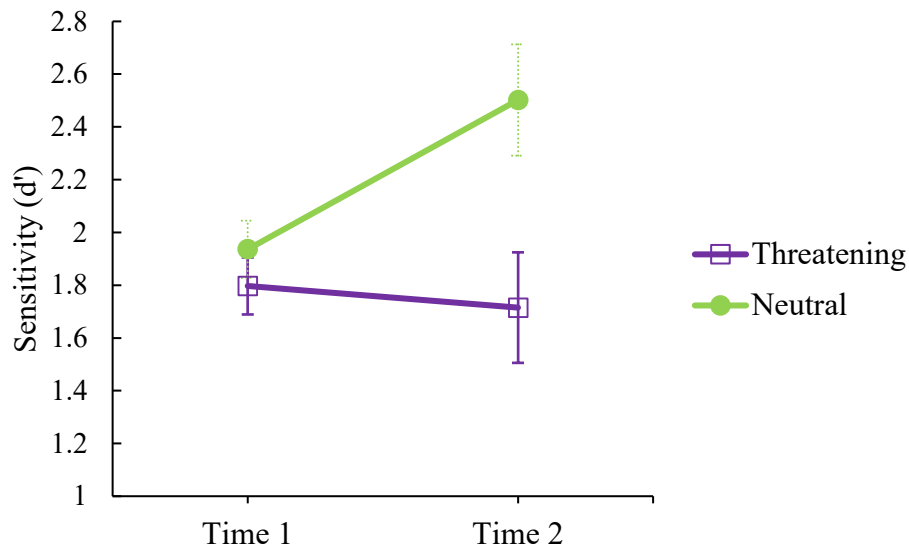
Note. Means for criterion for assessing the target face as threatening, as a function of training condition (threatening and neutral) across time (measured before and after training). $\beta < 0$ indicates a bias toward threat and $\beta > 0$ indicates a bias toward neutral. Error bars represent 95% confidence intervals.

Did criterion training impact sensitivity?

A mixed-ANOVA also revealed a significant Time by Training interaction on sensitivity, $F(1, 171) = 21.45, p < .001, \eta_p^2 = .111$, whereby sensitivity significantly increased in the neutral training group from Time 1 to Time 2 ($F(1, 86) = 32.67, p < .001, \eta_p^2 = .275, M = 0.57, SE = 0.10$), but did not significantly differ in the threatening training condition from Time 1 to Time 2 ($F(1, 85) = 0.69, p = .41, \eta_p^2 = .008$; see Figure 7).

Figure 7

Group Means for Sensitivity Between Training Conditions



Note. Means for sensitivity for assessing the target face as threatening, as a function of training condition (threatening and neutral) across time (measured before and after training). Higher values indicate greater sensitivity. Error bars represent 95% confidence intervals.

Although changes in sensitivity were not predicted (instead it was anticipated that training would impact criterion only), it is not surprising due to the substantial effect size found for the shift in criterion across time and training condition. Although criterion and sensitivity are theoretically and psychometrically distinct, they are functionally related and, from this perspective, a dramatic change in criterion would predict an associated change in sensitivity, although to a lesser degree (Lynn & Feldman Barrett, 2014). Moreover, the size of the change in sensitivity was much smaller (approximately six times smaller) than that for criterion, with η_p^2 's of .111 and .633, respectively (that is, accounting for 11.1% versus 63.3% of variance). In addition, there was a strong positive correlation between change in criterion and change in sensitivity in the neutral training condition, $r(85) = .73, p < .001$

Shifting threat criterion for morphed facial expressions reduces negative affect

(change scores were calculated by subtracting values at Time 1 from Time 2 for both criterion and sensitivity). This lends support to the hypothesis that criterion and sensitivity are associated and that a large shift in criterion would result in a shift in sensitivity.

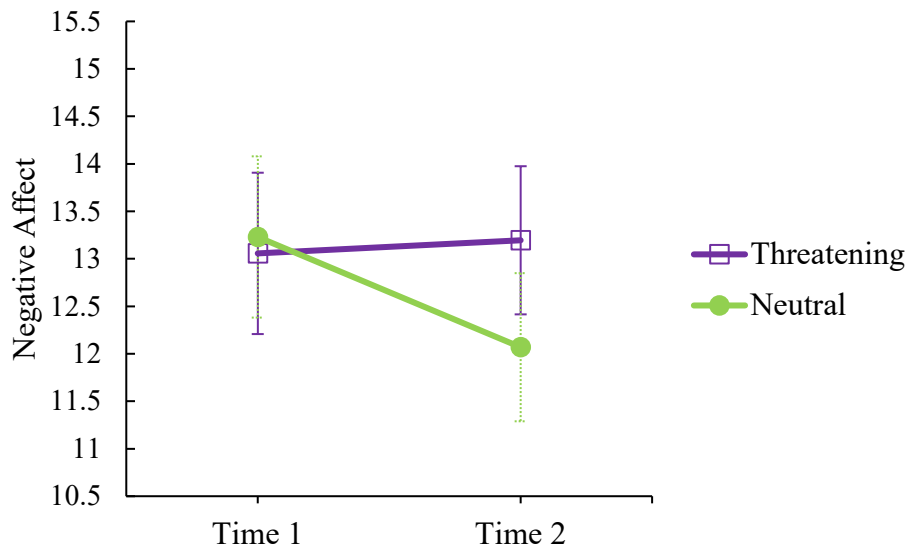
Unfortunately, due to the assumption of linearity being violated in the threatening training condition, correlation between change scores was not able to be assessed. Other studies have also demonstrated a lack of complete independence between criterion and sensitivity (e.g., Wylie et al., 2021). As such, although we were not able to purely isolate a change in criterion, the manipulation is considered successful given its much greater effect on criterion than sensitivity.

Did criterion training have a differential impact on levels of NA?

A mixed-ANOVA revealed a significant Training by Time interaction on level of NA, $F(1, 172) = 7.96, p = .005, \eta_p^2 = .044$ (see Figure 8). This interaction reflected the fact that NA significantly decreased in the neutral training condition from Time 1 to Time 2, $F(1, 86) = 12.87, p = .001, \eta_p^2 = .13$ ($M = 1.16, SE = 0.32$). This supports the hypothesis that shifting criterion to become more conservative in assessing facial expressions as threatening is associated with decreased levels of NA. However, the level of NA did not significantly differ in the threatening training group between time points, $F(1, 86) = 0.18, p = .675, \eta_p^2 = .002$, thereby failing to support the hypothesis that shifting criterion more liberally would increase levels of NA.

Figure 8

Group Means for NA Between Training Conditions



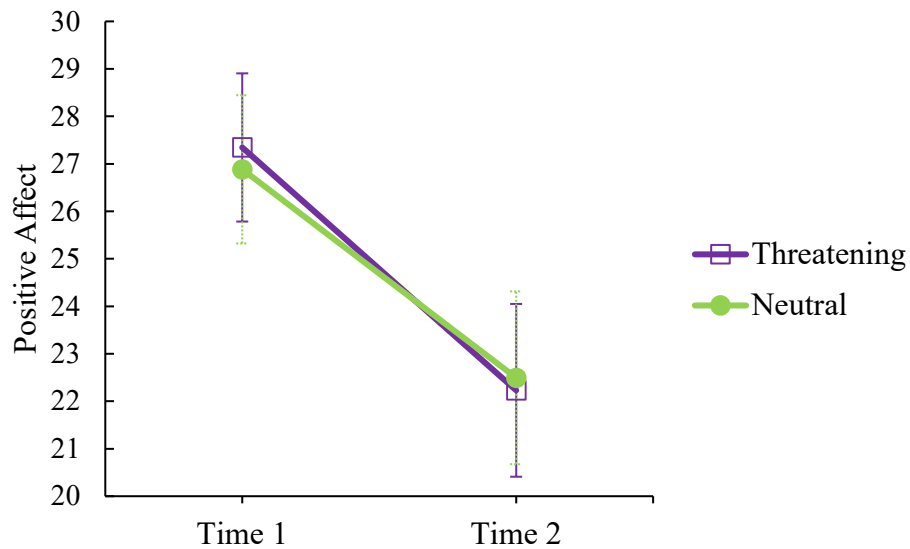
Note. Means for NA, as a function of training condition (threatening and neutral) across time (measured before and after training). Higher values indicate greater negative affect. Error bars represent 95% confidence intervals.

Did criterion training have a differential impact on levels of PA?

It was also hypothesized that shifting criterion to become more liberal in assessing facial expressions as threatening would decrease levels of PA. Conversely, it was hypothesized that when criterion was shifted in the conservative direction there would be an increase in levels of PA. As visually depicted in Figure 9, PA significantly decreased over time, but did not interact with training condition.

Figure 9

Group Means for PA Between Training Conditions



Note. Means for PA, as a function of training condition (threatening and neutral) across time (measured before and after training). Higher values indicate greater positive affect. Error bars represent 95% confidence intervals.

Accordingly, a mixed-ANOVA revealed a significant main effect for Time on PA, $F(1, 172) = 117.75, p < .001, \eta_p^2 = .406$, such that participants' level of PA significantly decreased over Time ($M = 4.7, SE = 0.44$). As per graphical representation, the mixed-ANOVA unsurprisingly revealed no significant Time by Training interaction on PA, $F(1, 172) = 0.683, p = .41, \eta_p^2 = .004$. Altogether this indicates that there was a general decrease in PA over time, which was not significantly moderated by training group, $F(1, 172) = 0.01, p = .931, \eta_p^2 < .001$.

Discussion

The current study sought to explore whether experimentally altering how people make decisions about the magnitude of threat manifest in human facial expressions can change

Shifting threat criterion for morphed facial expressions reduces negative affect

their affect, namely their levels of NA and PA. To our knowledge, this was the first empirical test of this question. To alter participants' decisions, they were exposed to differential reinforcement training for judging ambiguous facial expressions as either threatening or neutral. Results confirmed that training had the hypothesized differential effect – training shifted criterion towards the rewarded judgement. A modest shift in sensitivity was also observed in the neutral training condition, which while not predicted is not surprising, given the empirical association between criterion and sensitivity (Lynn & Feldman Barrett, 2014). Critically, however, the shift observed in criterion was demonstrably greater in magnitude than that observed in sensitivity. The training manipulation was therefore very effective—reflected by both the significant training by time interaction on criterion and also its large effect size. In the following sections, the impact that this training manipulation had on participants' positive and negative affect will be considered.

Impact on NA

Training that induced a bias to identify ambiguous faces as neutral resulted in significantly lower levels of NA post-training. This reduction in NA supports this study's main hypothesis and provides potential treatment targets for future research.

In contrast, training that induced a bias to identify ambiguous faces as threatening did not result in a significant increase in NA. This should not be taken, however, as clear evidence against the potential for criterion shifts to increase NA as there are multiple potential explanations for this lack of significance. Considering the large and consistent decrease in PA across the experiment, with no related significant increase in NA, there may be something systematically driving down reports of increases in NA. One possible cause for this is social desirability, whereby individuals may be unwilling to report increases in NA. That is, participants may be less willing to admit to socially stigmatised emotions such as

Shifting threat criterion for morphed facial expressions reduces negative affect

‘hostile’ and ‘irritable’, especially in contrast to socially desirable positive emotions such as ‘interested’, ‘enthusiastic’, ‘alert’ and ‘attentive’. Furthermore, if social desirability has attenuated reports of NA, PANAS items and the paper format of the questionnaire may have exacerbated this effect. In support of this explanation, the PANAS has been found to be particularly vulnerable to such social desirability responses due to the obviousness of the valence of the test items indicating positive and negative affectivity (Alliger & Dwight 2000; Furr & Bacharach, 2008). Additionally, although no difference has been found in levels of socially desirable responding between paper- and computer- formats (Dodou, & de Winter, 2014), in this instance, the questionnaire was completed with the participant and the researcher in the same room. Participants then handed questionnaires back to the researcher directly after completion, which, may have increased socially desirable responding.

Alternatively, it is possible that words on the NA scale of the PANAS (e.g., jittery, nervous, afraid) could capture anticipatory anxiety when being presented in a laboratory environment, thereby confounding our results. However, this does not appear to be the case in the current experiment as the reduction in NA only occurs in one condition rather than uniformly. In addition, levels of NA at Time 1 do not appear to be elevated to a degree that would be indicative of this kind of anticipatory anxiety.

An asymmetry in the directional malleability of NA could also be a contributing factor to the lack of significant increase in NA in the threatening training condition . That is, it may be that decreasing NA is more feasible than increasing it, especially in a non-clinical sample. This is might be because non-clinically targeted samples, such as the sample used in the current study, may be relatively less familiar with the experience of high states of NA. Therefore, the relationship between criterion training and increasing NA should be reassessed within a clinical sample. From a treatment perspective, however, it is not problematic if

Shifting threat criterion for morphed facial expressions reduces negative affect

future studies support the current findings that criterion-training has a stronger impact on decreasing NA, since this is generally the goal of treatment for disorders characterised by high NA.

Impact on PA

Given its reciprocal relationship with NA, levels of PA were also assessed. Participants' level of PA significantly decreased over time, regardless of which condition they were assigned. This was supported by anecdotal reports from participants that they found the hour-long experiment boring. In this context, uniform diminished PA might be accounted for by decreases on PANAS items that directly enquire about interest and enjoyment. Future research would benefit from investigating the length and nature of the task on PA. Furthermore, the present study directly manipulated criterion toward or away from threat. Although effective in changing NA, the specificity of these reward contingencies may have been inadequate to produce changes in PA. A strategy to overcome this constraint may be to include training conditions designed to shift criterion more directly *toward positive* judgments (rather than merely *toward neutral* and *away from negative*). Although not explicitly targeting or reporting criterion, studies utilising similar methodologies that have trained interpretation towards positive (e.g., happiness) have demonstrated increased PA post-training (Penton-Voak et al., 2012).

Implications

The current training paradigm overcomes a confound of commonly used CBM-I procedures. In the present study, all participants were exposed to the same ambiguous facial expression stimuli regardless of condition; it was simply the feedback that changed as a function of condition. In contrast, in Mathews and Mackintosh's (2000) CBM-I paradigm, the final to-be-completed disambiguating word fragment was conflated with condition. In that

Shifting threat criterion for morphed facial expressions reduces negative affect

study, participants in the threatening training condition received a negative word fragment resolution, whereas participants in the positive training condition received a positive word fragment resolution. Results from the current paradigm are, in contrast, not confounded by stimulus changes across conditions.

Importantly, the current study supports the novel concept that manipulation of criterion setting has a significant impact on one's level of NA. Further exploration and support of this concept has the potential to explain some of the inconsistency in the CBM-I literature. That is, CBM-I training paradigms may inadvertently target criterion and thus result in a reduction in symptoms of anxiety and depression. Alternatively, when the impact of these procedures on criterion is minimal, this may result in minimal impact on symptomatology. If supported through replication, the current conceptualisation could allow for CBM-I procedures to more effectively target criterion. As a result, outcomes may result in larger magnitude in the reduction of symptomatology, which may be achieved more rapidly and for longer.

Limitations and Future Directions

The current study utilised a non-clinical sample made up of a high proportion of university students to investigate the hypothesis that changes in criterion are linked to changes in affect. While this study provided important initial evidence, it would be useful for future research to test this in a clinical sample. In particular, one issue stemming from using a non-clinical sample was the relatively low levels of NA, including floor (i.e., minimum possible value) scores. A floor score at time one means that a reduction in negative affect at time two is impossible to observe. This means that when a participant exhibited consistent floor NA scores at both time points, it unclear whether this was a true representation of no change in affect or if this change was not captured due to the bounded nature of the scale and already being at

Shifting threat criterion for morphed facial expressions reduces negative affect

floor. While we were able to detect a selective reduction in NA in the conservative criterion training condition despite this issue, it would be useful to explore in a clinical sample with higher levels of NA to overcome such limitations of the current study.

Another limitation was that we were unable to isolate a pure change in criterion without concomitant changes in sensitivity. The effect size for changes in criterion was far greater than that for changes in sensitivity, and thus we are confident that the observed changes in affect are primarily driven by criterion. Further, sensitivity and criterion are not completely orthogonal (e.g., Lynn & Barrett, 2014); indeed, it is unlikely that large changes in criterion can occur without some change in sensitivity. However, this needs to be acknowledged as a limitation and consideration for future research.

Although it did not appear to be an issue in the current study, future research should consider anticipatory anxiety as a confounding factor for uniform reductions in NA, particularly when using the PANAS at the onset of an experiment. This may also be an interesting avenue to explore in future research when comparing lab-based to home-based applications. Future research would also be well advised to include measures of social desirability, as well as physiological measures of PA and NA to overcome the potential for participants to consciously alter their responses (Lang et al., 1993). The inclusion of demand questions in future research is also recommended.

Conclusion

This study provides evidence that altering one's criterion has an impact on affect. In particular, when participants adopted a more conservative criterion regarding whether threat was present in an ambiguous stimulus, NA was reduced. It is unlikely that this effect can be accounted for by changes in sensitivity given the small magnitude of this change. Reducing NA is typically the treatment goal for those living with anxiety and/or depressive disorders.

Shifting threat criterion for morphed facial expressions reduces negative affect

Therefore, these findings provide preliminary, but promising, results that underscore the potential for criterion as a target of psychological treatments (including CBM-I methods) for these disorders.

Acknowledgements: We would like to acknowledge Tamara Gradden for her help in familiarising us with morphing software.

Funding: This research was supported by an Australian Research Council (ARC) Future Fellowship (FT170100021) awarded to S.C.G..

References

- Abdi, H. (2007). Signal detection theory. In N. Salkind (Ed.), *Encyclopedia of measurement and statistics*. (pp. 887-890). Thousand Oaks, CA: SAGE Publications, Inc.
- Ackermann, J. F., & Landy, M. S. (2015). Suboptimal decision criteria are predicted by subjectively weighted probabilities and rewards. *Attention, Perception, & Psychophysics*, *77*(2), 638-658.
- Alliger, G. M., & Dwight, S. A. (2000). A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching. *Educational and Psychological Measurement*, (1).
- Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & Van Ijzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study.
- Beck, A. T. (2008). The evolution of the cognitive model of depression and its neurobiological correlates. *American Journal of Psychiatry*, *165*, 969–977.
- Becker, E., & Rinck, M. (2004). Sensitivity and response bias in fear of spiders. *Cognition and Emotion*, *18*(7), 961-976.
- Birk, J. L., Dennis, T. A., Shin, L. M., & Urry, H. L. (2011). Threat facilitates subsequent executive control during anxious mood. *Emotion*, *11*(6), 1291.
- Bora, E., & Zorlu, N. (2017). Social cognition in alcohol use disorder: a meta-analysis. *Addiction*, *112*(1), 40-48.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433-436.

- Chen, J., Milne, K., Dayman, J., & Kemps, E. (2019). Interpretation bias and social anxiety: does interpretation bias mediate the relationship between trait social anxiety and state anxiety responses?. *Cognition and Emotion*, *33*(4), 630-645.
- Chen, J., Short, M., & Kemps, E. (2020). Interpretation bias in social anxiety: A systematic review and meta-analysis. *Journal of Affective Disorders*.
- Clark, D. M., & Wells, A. (1995). *A cognitive model of social phobia*. In R. G. Heimberg, M. R. Liebowitz, D. A. Hope, & F. R. Schneier (Eds.), *Social phobia: Diagnosis, assessment, and treatment* (p. 69–93). The Guilford Press.
- Clerkin, E. M., & Teachman, B. A. (2011). Training interpretation bias among individuals with symptoms of obsessive compulsive disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, *42*, 33-343.
- Coombs, C.H., Dawes, R.M., Tversky, A. (1970). *Mathematical psychology: An elementary introduction*. New York: Prentice Hall.
- Crawford, J., & Henry, J. (2004). The Positive and Negative Affect Schedule (PANAS): construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, *43*(Part 3), 245-265.
- Dodou, D., & de Winter, J. C. (2014). Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *Computers in Human Behavior*, *36*, 487-495.
- Estes, W. K., & Maddox, W. T. (1995). Interactions of stimulus attributes, base rates, and feedback in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1075.

Shifting threat criterion for morphed facial expressions reduces negative affect

- Eysenck, M. W., Mogg, K., May, J., Richards, A., & Mathews, A. (1991). Bias in interpretation of ambiguous sentences related to threat in anxiety. *Journal of abnormal psychology, 100*(2), 144.
- Fodor, L. A., Georgescu, R., Cuijpers, P., Szamoskozi, S., David, D., Furukawa, T. A., & Cristea, I. A. (2020). Efficacy of cognitive bias modification interventions in anxiety and depressive disorders: a systematic review and network meta-analysis. *The Lancet Psychiatry, 7*(6), 506-514.
- Frenkel, T. I., Lamy, D., Algom, D., & Bar-Haim, Y. (2009). Individual differences in perceptual sensitivity and response bias in anxiety: Evidence from emotional faces. *Cognition and Emotion, 23*(4), 688-700.
- Furr, R. M., & Bacharach, V. R. (2008). *Psychometrics: An introduction*. Los Angeles : Sage Publications, c2008.
- Green, D. M., & Swets, J. A. (1966). *Signal detectability and psychophysics*. New York.
- Hallett, J., Howat, P. M., Maycock, B. R., McManus, A., Kypri, K., & Dhaliwal, S. S. (2012). Undergraduate student drinking and related harms at an Australian university: web-based survey of a large random sample. *BMC Public Health, 12*(1), 1-8.
- Hallion, L. S., & Ruscio, A. M. (2011). A meta-analysis of the effect of cognitive bias modification on anxiety and depression. *Psychological bulletin, 137*(6), 940.
- Han, S., & Dobbins, I. G. (2008). Examining recognition criterion rigidity during testing using a biased-feedback technique: Evidence for adaptive criterion learning. *Memory & Cognition, 36*(4), 703-715.
- Han, S., & Dobbins, I. G. (2009). Regulating recognition decisions through incremental reinforcement learning. *Psychonomic Bulletin & Review, 16*(3), 469-474.

Shifting threat criterion for morphed facial expressions reduces negative affect

- Healy, A. F., & Kubovy, M. (1978). The effects of payoffs and prior probabilities on indices of performance and cutoff location in recognition memory. *Memory & Cognition*, 6(5), 544-553.
- Hedlund, J. L., & Vieweg, B. W. (1984). The Michigan Alcoholism Screening Test (MAST): A comprehensive review. *Journal of Operational Psychiatry*.
- Hirsch, C. R., Krahé, C., Whyte, J., Loizou, S., Bridge, L., Norton, S., & Mathews, A. (2018). Interpretation training to target repetitive negative thinking in generalized anxiety disorder and depression. *Journal of Consulting and Clinical Psychology*, 86(12), 1017.
- Hirsch, C. R., Krahé, C., Whyte, J., Bridge, L., Loizou, S., Norton, S., & Mathews, A. (2020). Effects of modifying interpretation bias on transdiagnostic repetitive negative thinking. *Journal of consulting and clinical psychology*, 88(3), 226.
- Hirsch, C. R., Krahé, C., Whyte, J., Krzyzanowski, H., Meeten, F., Norton, S., & Mathews, A. (2021). Internet-delivered interpretation training reduces worry and anxiety in individuals with generalized anxiety disorder: A randomized controlled experiment. *Journal of consulting and clinical psychology*, 89(7), 575.
- Hirsch, C. R., Meeten, F., Krahé, C., & Reeder, C. (2016). Resolving ambiguity in emotional disorders: the nature and role of interpretation biases. *Annual review of clinical psychology*, 12, 281-305.
- Holmes, E. A., Lang, T. J., & Shah, D. M. (2009). Developing interpretation bias modification as a "cognitive vaccine" for depressed mood: imagining positive events makes you feel better than thinking about them verbally. *Journal of abnormal psychology*, 118(1), 76.

Shifting threat criterion for morphed facial expressions reduces negative affect

- Hoppitt, L., Mathews, A., Yiend, J., & Mackintosh, B. (2010). Cognitive mechanisms underlying the emotional effects of bias modification. *Applied Cognitive Psychology*, 24, 312-325.
- Jones, E. B., & Sharpe, L. (2014). The effect of cognitive bias modification for interpretation on avoidance of pain during an acute experimental pain task. *PAIN®*, 155(8), 1569-1576.
- Kercher, K. (1992). Assessing subjective well-being in the old-old: The PANAS as a measure of orthogonal dimensions of positive and negative affect. *Research on Aging*, 14(2), 131-168.
- Kopetz, C., MacPherson, L., Mitchell, A. D., Houston-Ludlam, A. N., & Wiers, R. W. (2017). A novel training approach to activate alternative behaviors for smoking in depressed smokers. *Experimental and Clinical Psychopharmacology*, 25(1), 50.
- Krueger, R. F. (1999). The structure of common mental disorders. *Archives of General Psychiatry*, 56, 921-926.
- Krueger, R. F., & Markon, K. E. (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology*, 2, 111-133.
- Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30, 261-261.
- Laux, J. M., Newman, I., & Brown, R. (2002). The Michigan Alcoholism Screening Test (MAST): A Psychometric Investigation.

Shifting threat criterion for morphed facial expressions reduces negative affect

- Laux, J. M., Newman, I., & Brown, R. (2004). The Michigan alcoholism screening test (MAST): A statistical validation analysis. *Measurement and Evaluation in Counseling and Development*, 36(4), 209-225.
- Liu, G., Xin, Z., & Lin, C. (2014). Lax decision criteria lead to negativity bias: evidence from the emotional stroop task. *Psychological reports*, 114(3), 896-912.
- Lynn, S. K., & Feldman Barrett, L. (2014). "Utilizing" Signal Detection Theory. *Psychological Science* (Sage Publications Inc.), 25(9), 1663-1673.
- MacLeod, C., & Mathews, A. (2012). Cognitive bias modification approaches to anxiety. *Annual Review of Clinical Psychology*, 8, 189-217.
- Manguno-Mire, G. M., Constans, J. I., & Geer, J. H. (2005). Anxiety-related differences in affective categorizations of lexical stimuli. *Behaviour Research and Therapy*, 43, 197-213.
- Maoz, K., Eldar, S., Stoddard, J., Pine, D. S., Leibenluft, E., & Bar-Haim, Y. (2016). Angry-happy interpretations of ambiguous faces in social anxiety disorder. *Psychiatry research*, 241, 122-127.
- Markon, K. E. (2010). Modeling psychopathology structure: A symptom-level analysis of Axis I and II disorders. *Psychological medicine*, 40(2), 273.
- Mathews, A., & Mackintosh, B. (2000). Induced emotional interpretation bias and anxiety. *Journal of Abnormal Psychology*, 109, 602-615.
- Mathews, A., & MacLeod, C. (2002). Induced processing biases have causal effects on anxiety. *Cognition & Emotion*, 16(3), 331-354.

Shifting threat criterion for morphed facial expressions reduces negative affect

- Mathews, A., & MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. *Annual Review of Clinical Psychology*, 1, 167–195.
- Mathews, A., Richards, A., & Eysenck, M. (1989). Interpretation of homophones related to threat in anxiety states. *Journal of abnormal psychology*, 98(1), 31.
- Mathews, A., Ridgeway, V., Cook, E., & Yiend, J. (2007). Inducing a benign interpretation bias reduces trait anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, 38, 225-236.
- Meneguzzo, P., Collantoni, E., Bonello, E., Busetto, P., Tenconi, E., & Favaro, A. (2020). The predictive value of the early maladaptive schemas in social situations in anorexia nervosa. *European Eating Disorders Review*, 28(3), 318-331.
- Mogg, K., Baldwin, D. S., Brodrick, P., & Bradley, B. P. (2004). Effect of short-term SSRI treatment on cognitive bias in generalised anxiety disorder. *Psychopharmacology*, 176(3-4), 466-470.
- Penton-Voak, I. S., Bate, H., Lewis, G., & Munafo, M. R. (2012). Effects of emotion perception training on mood in undergraduate students: randomised controlled trial. *The British Journal of Psychiatry*, 201(1), 71-72.
- Rehm, J., Hasan, O. S., Black, S. E., Shield, K. D., & Schwarzingler, M. (2019). Alcohol use and dementia: a systematic scoping review. *Alzheimer's research & therapy*, 11(1), 1-11.
- Rhodes, M. G., & Jacoby, L. L. (2007). On the dynamic nature of response criterion in recognition memory: Effects of base rate, awareness, and feedback. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(2), 305.

Shifting threat criterion for morphed facial expressions reduces negative affect

- Salemink, E., van den Hout, M., & Kindt, M. (2007). Trained interpretive bias: validity and effects on anxiety. *Journal of behavior therapy and experimental psychiatry*, 38(2), 212-224.
- Salemink, E., van den Hout, M., & Kindt, M. (2009). Effects of positive interpretive bias modification in highly anxious individuals. *Journal of Anxiety Disorders*, 23 (5), 676-683.
- Salemink, E., van den Hout, M., & Kindt, M. (2010). Generalisation of modified interpretive bias across tasks and domains. *Cognitive and Emotion*, 24, 453-464.
- Sedgmond, J., Chambers, C. D., Lawrence, N. S., & Adams, R. C. (2020). No evidence that prefrontal HD-tDCS influences cue-induced food craving. *Behavioral Neuroscience*, 134(5), 369.
- Selzer, M. L. (1971). The Michigan Alcoholism Screening Test: The quest for a new diagnostic instrument. *American Journal of Psychiatry*, 127(12), 1653-1658.
- Selzer, M. L., Vinokur, A., & van Rooijen, L. (1975). A self-administered short Michigan alcoholism screening test (SMAST). *Journal of studies on alcohol*, 36(1), 117-126.
- Standage, H., Ashwin, C., & Fox, E. (2010). Is manipulation of mood a critical component of cognitive bias modification procedures? *Behaviour Research and Therapy*, 48, 4-10.
- Steinman, S. A., & Teachman, B. A. (2014). Reaching new heights: comparing interpretation bias modification to exposure therapy for extreme fear of heights. *Journal of Consulting and Clinical Psychology*, 82, 404-417.
- Stopa, L., & Clark, D. M. (2000). Social phobia and interpretation of social events. *Behaviour Research and Therapy*, 38, 273-283.

Shifting threat criterion for morphed facial expressions reduces negative affect

- Tabachnick, B.G., & Fidell, L.S. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson Education.
- Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological science, 10*(4), 297-303.
- Tottenham, N., Tanaka, J., Leon, A.C., McCarry, T., Nurse, M., Hare, T.A., Marcus, D.J., Westerlund, A., Casey, B.J., Nelson, C.A. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Research, 168*(3):242-9.
- Watson, D., & Clark, L. A. (1984). Negative Affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin, 96*, 465-490.
- Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological bulletin, 98*(2), 219.
- Watson, D., Clark, L. A., & Carey, G. (1988). Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of abnormal psychology, 97*(3), 346.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063.
- Windmann, S., & Krüger, T. (1998). Subconscious detection of threat as reflected by an enhanced response bias. *Consciousness and Cognition, 7*(4), 603-633.
- Winton, E. C., Clark, D. M., & Edelman, R. J. (1995). Social anxiety, fear of negative evaluation and the detection of negative emotion in others. *Behaviour research and therapy, 33*(2), 193-196.

Shifting threat criterion for morphed facial expressions reduces negative affect

Wylie, G. R., Yao, B., Sandry, J., & DeLuca, J. (2021). Using signal detection theory to better understand cognitive fatigue. *Frontiers in psychology*, 3881.