High Trait Anxiety Enhances Optimal Integration of Auditory and Visual Threat Cues

Naomi Heffer[a], Molly Gradidge[a], Anke Karl[b], Chris Ashwin[a], & Karin Petrini[a]

[a]University of Bath, Department of Psychology;

[b]University of Exeter, Mood Disorders Centre

Corresponding Author:

Miss Naomi Heffer, PhD candidate

Department of Psychology

University of Bath

Claverton Down

Bath, BA2 7AY

United Kingdom

Email: n.r.heffer@bath.ac.uk

Acknowledgements

This work was supported by a PhD Studentship from the Medical Research Council (MRC) under award number 2110628. The MRC did not play a role in the study design or conduct of the research, or the preparation of the article.

Declaration of Interest

No potential conflicts of interest were reported by the authors.

Data Statement

All data underlying the statistical analyses in this manuscript are openly available from the University of Bath Research Data Archive at https://doi.org/10.15125/BATH-01023.

Abstract

Background and Objectives

Emotion perception is essential to human interaction and relies on effective integration of emotional cues across sensory modalities. Despite initial evidence for anxiety-related biases in multisensory processing of emotional information, there is no research to date that directly addresses whether the mechanism of multisensory integration is altered by anxiety. Here, we compared audiovisual integration of emotional cues between individuals with low vs. high trait anxiety.

Methods

Participants were 62 young adults who were assessed on their ability to quickly and accurately identify happy, angry and sad emotions from dynamic visual-only, audio-only and audiovisual face and voice displays.

Results

The results revealed that individuals in the high anxiety group were more likely to integrate angry faces and voices in a statistically optimal fashion, as predicted by the Maximum Likelihood Estimation model, compared to low anxiety individuals. This means that high anxiety individuals achieved higher precision in correctly recognising anger from angry audiovisual stimuli compared to angry face or voice-only stimuli and compared to low anxiety individuals.

Limitations

We tested a higher proportion of females, and although this does reflect the higher prevalence of clinical anxiety among females in the general population, potential sex differences in multisensory mechanisms due to anxiety should be examined in future studies.

Conclusions

Individuals with high trait anxiety have multisensory mechanisms that are especially fine-tuned for processing threat-related emotions. This bias may exhaust capacity for processing of other emotional stimuli and lead to overly negative evaluations of social interactions.

Keywords: Anxiety; Emotion; Multisensory processing; Negative bias; Maximum Likelihood Estimation; Race Model

1. Introduction

Emotions are rarely perceived by way of a single sensory modality in everyday life. During social interactions humans typically receive sensory input relating to the emotions of others from different types of cues, such as facial expression, voice prosody and body motion. Therefore, a full ecological understanding of emotion requires effectively combining cues from different sensory modalities (Klasen et al., 2012; Schirmer & Adolphs, 2017). Research has shown that emotions are recognised more accurately and efficiently when relevant auditory and visual cues are integrated, than when auditory or visual cues are processed in isolation (e.g. Collignon et al., 2008; De Gelder & Vroomen, 2000; Love et al., 2013; Piwek et al., 2015) - an effect known as multisensory facilitation.

The ability to accurately attribute emotional states to other people is considered key to successful social interactions, and biases interpreting the emotions of others have been linked to interpersonal problems in psychiatric populations, including people with heightened anxiety levels. Individuals with anxiety disorders and heightened trait anxiety show attentional biases towards threat-related information and tend to interpret the emotions of others in a more negative manner (e.g. Cisler & Koster, 2010; Richards et al., 2002; Yoon & Zinbarg, 2008). Threat bias models of anxiety propose that these emotional biases play a causal role in the aetiology and/or maintenance of anxiety symptoms (e.g. Beck & Clark, 1997; Eysenck et al., 2007; Mathews & Mackintosh, 1998; Mogg & Bradley, 2016, 2018). However, investigation of these biases has been dominated by studies examining perception of visual cues only. Hence, it is still unclear how anxiety affects emotion perception when information is available from multiple sensory modalities.

One of the few previous studies to investigate multisensory emotion processing with relation to trait anxiety is Koizumi et al., (2011). Koizumi et al. recruited a sample of non-clinical individuals with different levels of anxiety traits and measured their accuracy interpreting emotion from emotionally congruent (e.g. happy facial expression coupled with happy voice prosody) and incongruent (e.g. happy facial expression coupled with angry voice prosody) audiovisual displays. Koizumi et al., (2011) showed that when angry cues were presented alongside happy emotional cues, individuals with higher trait anxiety levels showed a tendency to allocate increased attention to angry cues, resulting in a biased and less accurate interpretation of happy cues. Whilst the findings from this study showed that audiovisual processing of emotional face and voice cues was influenced by trait anxiety levels, they cannot tell us specifically whether audiovisual integration of emotional cues is altered by anxiety. This is an important question to address, given that multisensory integration is one of the key mechanisms by which our sensory systems overcome the perceptual ambiguity that is inherent in our environment to allow us to make accurate judgements about the world around us (e.g. King & Calvert, 2001; Tseng et al., 2015).

From the limited studies that have been conducted in this area, there is some preliminary evidence to suggest that audiovisual integration of emotional cues may be altered by anxiety. Campanella et al., (2010) showed that brain activity associated with perceptual processing was only able to discriminate between participants prone to anxiety and neurotypical participants under multisensory task conditions, and not during tasks requiring unimodal processing only (Campanella et al., 2010, 2012; Delle-Vigne et al., 2015). Furthermore, Yoon & Hong (2020) showed that individuals with a lower predisposition for anxiety detected neutral faces faster when accompanied by neutral voices, but individuals more prone to anxiety did not show the same multisensory facilitation effect. Whilst the findings of Yoon & Hong (2020) suggest that individuals predisposed to higher anxiety levels may show atypical audiovisual integration, an effect of anxiety was detectable only for neutral stimuli and not emotional stimuli, which appears somewhat contradictory to previous studies that have revealed differences in perceptual processing of emotional stimuli related to anxiety levels (e.g. Koizumi et al., 2011; Quadflieg et al., 2007; Richards et al., 2002; Veerapa et al., 2020).

Peschard & Philippot (2017) investigated how social anxiety affected the interpretation of angry and neutral face and voice cues by comparing emotional processing from unimodal and audiovisual expressions between individuals with different social anxiety levels. They found that individuals with lower and higher social anxiety levels benefited to a similar extent in terms of accuracy and reaction times when emotional faces and voices were presented together. However, this study only examined audiovisual integration of angry and neutral emotional cues and used static pictures of facial expressions and voices as stimuli. Given that multisensory integration of sensory signals is largely dependent on temporal synchrony (Stein & Meredith, 1993), it is not clear that participants would have necessarily integrated static face cues with dynamic voice cues. This may have produced limited audiovisual integration in both the lower and higher social anxiety groups, resulting in an apparent lack of differences between groups.

Another recent study (Heffer et al., 2021) examined emotion perception from angry and happy faces and voices and reported that multisensory facilitation effects for angry stimuli are enhanced for individuals with higher trait anxiety levels. However, this study suffered from the same limitation as the Peschard & Philippot (2017) study due to the lack of temporal synchrony between face and voice cues in the audiovisual stimuli. Also, because only happy and angry expressions were tested in Heffer et al. (2021), it is unclear whether the observed multisensory bias in the higher trait anxiety group was indicative of a general negativity bias or was specific to threat-related emotions.

The present study compared individuals with low versus high trait anxiety on audiovisual integration of emotional cues during perception of happy, angry, and sad expressions from dynamic faces and voices. Participants made emotional judgements in three different stimulus modality conditions: visual-only, audio-only and audiovisual. Investigating perception of happy, angry and sad emotions allowed us to determine whether differences in audiovisual integration of emotional cues between individuals with low and high trait anxiety spanned both positive and negative emotions, or whether differences were specific to threat-related emotional cues (e.g. anger) or to negative emotions more generally (e.g. anger and sadness). We used a model-based analysis to elucidate how multisensory integration of emotional information operates in high trait anxiety. This was done by comparing performance in the audiovisual condition for both high and low trait anxiety groups to performance predicted by the Maximum Likelihood Estimation (MLE) model (Ernst & Banks, 2002; Rohde et al., 2016) and Miller's Race Model (Miller, 1982; Ulrich et al., 2007).

Hence, the aim of the present study was to examine whether audiovisual integration of emotional cues operates in a similar way for individuals with low vs. high trait anxiety. Based on previous findings, we hypothesised that both groups would show a significant facilitation effect in terms of accuracy and response speed for the audiovisual condition compared to the unimodal conditions. However, we also hypothesised that group differences would be evident for audiovisual integration of threat-related cues (i.e. specifically for anger) in the high anxiety group, based on theories of enhanced processing of threat-related information in anxiety disorders.

2. Methods and Materials

2.1 Participants

A total of 62 participants took part in the study (49 women, 13 men; mean age = 24.31, *SD* = 9.17). The sample consisted predominantly of undergraduate psychology students at the

University of Bath. Participants were selected from a pool of 109 volunteers who expressed an interest in participation and filled in an online screening questionnaire. Participants were selected for inclusion based on their scores on the trait subscale of the Spielberger State-Trait Anxiety Inventory (STAI-T, Y-Form; Spielberger, 1983). High anxiety individuals were defined as those scoring 48 or more on the STAI-T ($n = 35$) and low anxiety individuals were defined as those scoring 37 or below ($n = 27$). For more information on the selection criteria, please refer to section 2.2.1.

There was no significant difference in age between the low ($M = 24.89$, $SD = 8.47$) and high ($M = 23.86$, $SD = 9.78$) trait anxiety groups, $t(60) = 0.44$, $p = .664$, 95% CI[-3.70, 5.76], and the sex ratio was also not significantly different across the two groups (8 males in the low vs. 5 males in the high trait anxiety group; $p = .209$, Fisher's exact test). The mean trait anxiety score in the high trait anxiety group ($M = 56.31$, $SD = 6.50$) was significantly greater than the mean score for the low trait anxiety group ($M = 30.67$, $SD = 4.72$), $t(60) = -17.27$, $p < .001$, 95% CI[-28.62, -22.68], $d = 4.51$.

Out of the original 109 volunteers who were recruited via word-of-mouth and social media to complete the online screening questionnaire, 75 were eligible to take part in the study. The remaining 34 individuals were not eligible because they had a trait anxiety score in the middle range of possible scores, indicating that they had neither particularly low nor particularly high trait anxiety levels. The study had to be ended early due to government restrictions put in place as a result of the coronavirus pandemic, which meant that 12 individuals who were eligible and willing to take part in the study did not have the opportunity to participate. One additional participant who was eligible to take part based on screening chose to withdraw before completing the study. Overall, of the 75 who were eligible, 62 participated in the full study: 35 in the high anxiety group and 27 in the low anxiety group. The target sample size was 70 (35 in each anxiety group) and was based on the prediction of a medium effect-size, akin to the effect size for the significant between-within interaction effect reported by Yoon & Hong (2020). A summary of participant selection is shown in Figure A1 of the supplemental material.

All participants were at least 16 years of age and self-reported normal or corrected-to-normal vision and hearing. Participants received reimbursement for their time and informed consent was obtained from all participants prior to testing. Further demographic information about participants may be found in section A of the supplemental material. This study received full ethical approval from the Psychology Research Ethics Committee at the University of Bath.

2.2 Materials

2.2.1 Trait Anxiety Measure

The STAI-T was used to measure anxiety levels as part of an online screening process. The trait subscale of the STAI consists of 20 items which assess the frequency of feelings relevant to “anxiety proneness” (1 almost never, 2 sometimes, 3 often, and 4 almost always), such as worry, confidence and security (e.g. “I feel secure”, “I feel nervous and restless”). The STAI-T has been shown to have strong psychometric properties, including high internal consistency, which was also the case in our sample (α = .96).

In the current study, high anxiety individuals were defined as those scoring 48 or more on the STAI-T and scores in this group ranged from 48 to 70. This threshold was selected to reflect previously reported STAI-T norms which suggest that scores for people with diagnosed anxiety disorders, including panic disorder, social anxiety and specific phobia, typically fall in the range 47 to 61 (Bieling et al., 1998; Spielberger, 1983). Therefore, a threshold score of 48 was selected as a way to recruit individuals who likely had clinically meaningful anxiety symptoms, even if they did not have a diagnosed anxiety disorder.

Low anxiety individuals were defined as those scoring 37 or below and scores in this group ranged from 22 to 37. Existing norm data suggests that neurotypical adults have a mean trait anxiety score of approximately 35 (Bieling et al., 1998; Spielberger, 1983). A threshold score of 37 was chosen as a criterion for recruitment into the low anxiety group to ensure that participants in this group had low-to-typical anxiety levels relative to the general population.

2.2.2 Stimuli

Stimuli were 500ms-long videos of female adults vocalising the phrase ‘the door’ in a manner which was angry, happy or sad (Figure 1). The phrase 'the door' was used as a semantically neutral statement so that individuals would have to rely upon emotion cues from facial expression and/or voice prosody in order to judge the speaker’s emotion. This approach reflects similar previous studies where short or single-word semantically neutral phrases were used across the different stimulus emotions (Koizumi et al., 2011; Peschard & Philippot, 2017). Videos were adapted from longer clips taken from the RAVEDESS database (Livingstone & Russo, 2018). Initial piloting revealed that only clips showing females expressing the emotions were reliably recognised for all three emotions, and so only these clips were used in the final study. Clips were edited in Adobe Premiere Pro 2017 (Adobe, 2017) to produce the stimuli in three different conditions: audio-only, visual-only and audiovisual.

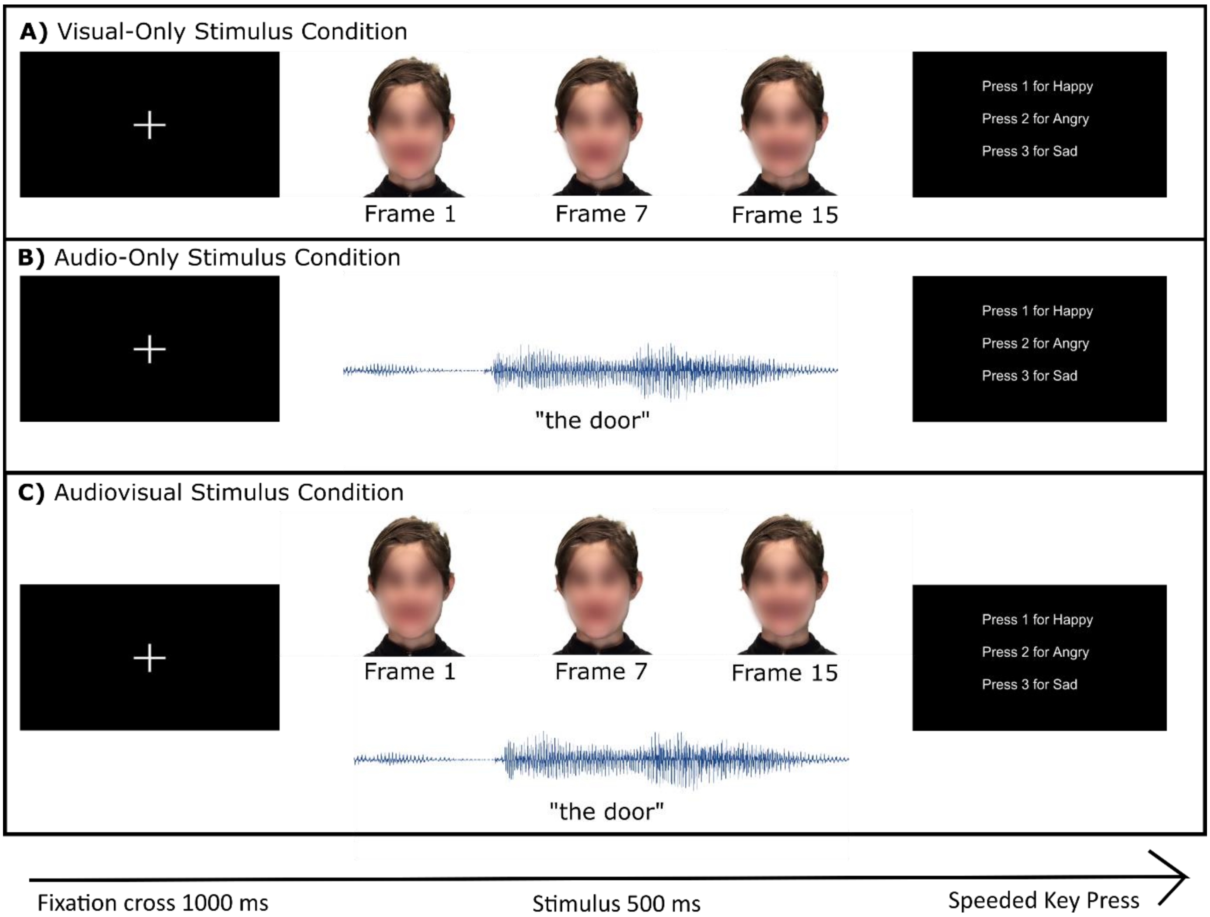

Figure 1. Schematic representation of trials involving the three different stimulus modality conditions, including: (a) visual-only stimuli, (b) audio-only stimuli, and (c) audiovisual stimuli. The dynamic face stimuli were presented with the addition of noise in order to decrease the reliability of visual information; this is explained in more detail in section A of the supplemental material.

To produce the audio-only clips the video was replaced with a black background for each clip, and for the visual-only clips the audio track was muted. In the audiovisual condition audio and visual cues were presented together. A total of 54 clips were used (6 models x 3 emotions [angry, happy and sad] x 3 modality conditions [audio-only, visual-only and audiovisual]). Given existing evidence of visual dominance in audiovisual emotion processing (e.g. Collignon et al., 2008; De Gelder & Vroomen, 2000), the central face region of the videos was degraded by the addition of a 75% Gaussian blur to match the reliability of visual cues more closely to the audio cues. Further details can be found in section A of the supplemental material.

2.3 Procedure

Volunteers expressing an initial interest in participating were emailed a link to complete the STAI-T questionnaire online. Participants were contacted by email to inform them whether or not they were eligible to participate in the main study based on their responses to the screening questionnaire, and eligible participants were invited to come into the lab to complete the audiovisual emotion recognition task.

For the lab portion of the study, participants were individually tested in a dimly lit room. They sat 57cm from a computer monitor and wore noise-cancelling headphones. The task required participants to make judgements about the emotion expressed by the face and voice stimuli (either angry, happy or sad) in different modality conditions (visual-only, audio-only and audiovisual). On each trial, participants were presented with a fixation cross for 1000ms followed by the stimulus. Responses were made by a speeded key press using whichever hand the participant self-reported as being their dominant hand. There were three different response options (i.e. happy, angry and sad) and so three different keys used for responding to stimuli (Figure 1). There were six possible permutations of emotion-key assignments for responding to stimuli, and these were randomised across participants.

Participants were asked to categorize stimuli as quickly as possible, but not to the detriment of accuracy, and were informed prior to starting the task that the stimuli would be very short ("less than a second long") and that the faces would be blurry. Participants completed three practice trials consisting of a visual-only, audio-only and audiovisual stimulus from an additional female actor who did not appear in any of the main trials. After completing the practice trials, participants were given the opportunity to ask questions before starting the task. The task was completed in three five-minute blocks, each consisting of 54 trials. Over the course of the experiment the participants saw each of the 54 stimuli three times. The whole task took approximately 15 minutes to complete.

2.4 Statistical Analysis

We analysed accuracy and reaction time (RT) data by examining how well two established theoretical models of multisensory integration could account for the audiovisual integration performance in each anxiety group. A full description of the data and additional analyses, such as analysis of mean accuracy and RTs by modality condition, are reported in section B of the supplemental material.

2.4.1 Accuracy

We obtained a measure of variance in accuracy for each participant in each condition and examined how well the variance in the audiovisual condition could be predicted by the Maximum Likelihood Estimation (MLE) model. The MLE model is a prominent model used to describe multisensory integration in the general population. This framework suggests that integration of information across multiple senses is achieved via a reliability-weighted linear average of the different sensory signals (e.g. Ernst & Banks, 2002; Rohde et al., 2016). Reliable signals (which produce perceptual estimates with low variance) receive a high weight, while unreliable signals (which produce perceptual estimates with higher variance) receive a low weight. This combination rule is described as 'statistically optimal' because integrating sensory signals in this way should always produce the most reliable representation. This is a useful strategy in everyday life where sensory signals are often noisy and ambiguous. For example, there are situations where emotional information from the face is made less reliable because the person’s face is obstructed or viewed from a distance. In these situations, it would be advantageous to rely more heavily on emotional information from the person's tone of voice, because this information is likely to be a more reliable cue to recognise the expressed emotion when the face of a person is obscured. If individuals' perceptual estimates resemble the predictions of the MLE model, this indicates that they are able to use all the information available in the environment in a more optimal way compared to individuals who tend to rely on one sense more than the other, regardless of the relative reliability of information from the different senses.

The measured variances in accuracy for the visual-only (V) and audio-only (A) conditions for each participant were entered in the MLE model equation ($\sigma_{AV}^2 = \sigma_A^2\sigma_V^2 / (\sigma_A^2 + \sigma_V^2)$) to determine how much lower the variance should be for the audiovisual (AV) condition to resemble optimal multisensory integration.

For each participant, the actual variance in the audiovisual condition was subtracted from the audiovisual prediction made by the MLE model (i.e. *Δ measured* - *predicted*). These 'MLE Difference Scores' represent how much each participant's performance in the audiovisual condition deviated from 'statistically-optimal' performance as predicted by the MLE model and provides a quantified estimation of the perceptual benefit that is gained through multisensory processes. The closer the participant's score is to zero the more closely their performance in the audiovisual condition matched performance predicted by the MLE model, and the greater the attained perceptual benefit (i.e. reduction in variability or uncertainty when recognising an emotion) from the integration of auditory and visual cues. Positive MLE Difference Scores indicate that the reduction in uncertainty when recognising an emotion in the audiovisual condition was greater than the reduction in uncertainty predicted by the model, whereas negative scores indicate that the reduction in uncertainty when recognising an emotion in the audiovisual condition was smaller than the reduction in uncertainty predicted by the model. Therefore, positive scores reflect higher perceptual precision than predicted by the model while negative scores indicate less precise emotion recognition than predicted. Scores that approximate zero reflect that the model well-predicted the level of perceptual precision achieved when participants were able to use both auditory and visual emotional information in their judgements.

2.4.2 Reaction Times

For RT data we investigated whether increases in the speed of response for audiovisual trials reflected a multisensory integrative process by comparing our results to the predictions of Miller's Race Model. Miller (1982) proposed a method for testing whether facilitation of RTs in multisensory conditions is likely to reflect a true multisensory integrative process, compared to the alternative hypothesis that shortening of RTs in multisensory conditions occurs because the multiple component stimuli set up a race for the control of the response and the faster process wins. If RTs obtained in the audiovisual condition of our task were better than the predictions of Miller's Race Model, this would suggest that information from the visual and auditory cues truly interacted to produce the RT facilitation. We tested for violation of Miller's Race Model across five percentiles of the RT distribution in each anxiety group and emotion condition, to determine whether facilitation of RTs resulting from integration of face and voice cues differed between low and high anxiety groups.

Analyses of violation of the Race Model inequality were carried out using the RMITest software which implements the algorithm described in Ulrich et al., (2007) to estimate the cumulative probability distributions of reaction times in the two unimodal conditions and the audiovisual condition, and tests whether RTs in the audiovisual condition are significantly faster than would be predicted by the Race Model (using one-tailed t-tests to test for significant violations of the model across five percentiles of the RT distribution).

Some participants lacked sufficient data in one or more of the stimulus conditions to estimate the cumulative probability distributions of RTs required to run the model. This meant that for a small number of participants some values had to be estimated or cases had to be excluded – a summary can be found in section C of the supplemental material. Only RTs for correct responses were included in the analysis.

3. Results

3.1 Accuracy

To analyse the potential effects of anxiety group on deviation from statistically optimal integration, we entered MLE Difference Scores into a mixed factorial ANOVA with stimulus emotion (angry, happy, sad) as a within-subjects factor and anxiety group (high, low) as a between-subjects factor. The results revealed that there was no significant main effect of anxiety group, F (1, 60) = 0.70, p = .407, η_p^2 = .01, which showed that overall there was no difference in the extent of perceptual benefit through multisensory integration between individuals in the low and high anxiety groups. However, there was a significant two-way interaction effect between emotion and anxiety group, F (2, 120) = 4.43, p =.014, η_p^2 = .07. This showed that there were differences in multisensory integration ability between groups that were contingent on stimulus emotion, as shown in Figure 2.

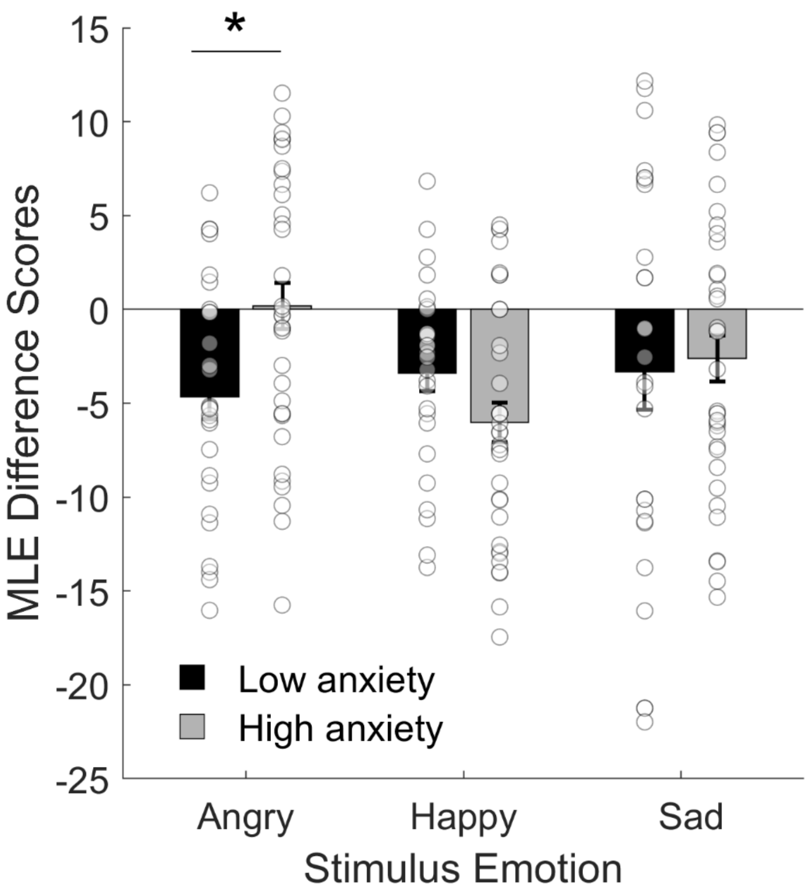

Figure 2. MLE Difference Scores (Δ measured-predicted audiovisual variance) for the high (grey bars) and low (black bars) anxiety groups across the three emotion categories. Error bars show the standard error of the mean MLE Difference Scores in each condition, and circles show the MLE Difference Scores for individual participants. *$p < .05$.

Pairwise comparisons, Bonferroni-corrected, revealed that for the high anxiety group performance was more closely predicted by the MLE model when processing angry stimuli compared to the low anxiety group, $p = .008$, 95% CI [1.29, 8.38], $d = 0.70$. No significant differences were revealed in MLE Difference Scores between anxiety groups for happy or sad stimuli, $p = .749$, 95% CI [-3.78, 5.23], $d = 0.08$, although for happy stimuli the high anxiety group showed a trend towards less optimal integration compared to the low anxiety group, $p = .080$, 95% CI [-5.61, 0.32], $d = 0.46$.

3.2 Reaction Times

As shown in Figure 3, although the audiovisual condition did result in faster RTs overall for the two anxiety groups, we failed to observe significant violation of the Race Model prediction over any percentile of the RT distribution for any stimulus emotion in either the low or high anxiety group, $p \geq .254$, $t \leq 0.67$.

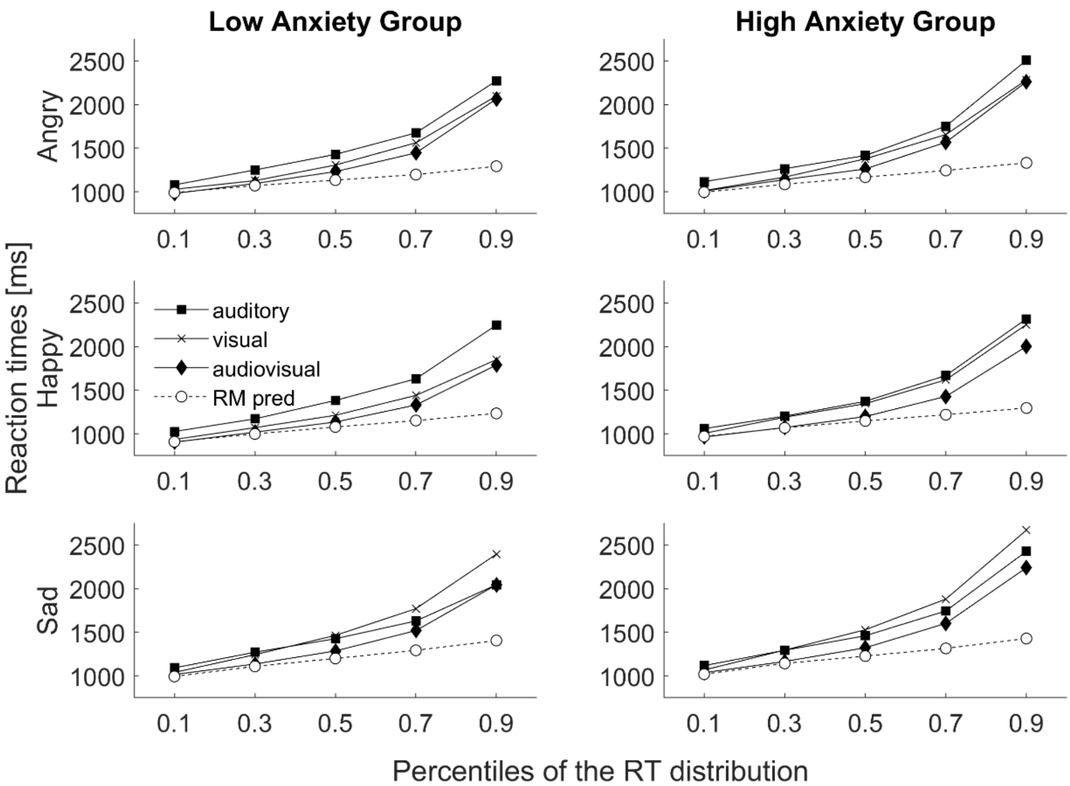

Figure 3. The cumulative probability distributions of the reaction times (RTs) for correct classifications of audiovisual stimuli (diamonds) and their unimodal components (squares for audio-only stimuli, x's for visual-only stimuli), as well as the predicted audiovisual distribution based on the assumptions of the Race Model (open circles, named as 'RM pred'). Distributions are shown separately for each emotion within each of the two anxiety groups.

4. Discussion

In the current study we examined the effect of high and low trait anxiety on audiovisual integration of emotional faces and voices using a model-based analysis. Both anxiety groups exhibited significant multisensory facilitation effects, as evidenced by increases in accuracy and response speed for emotion perception from audiovisual compared to unimodal stimuli (see section B of the supplemental material for these analyses). The results showed that trait anxiety had a significant impact on optimal multisensory integration and that this effect was contingent on stimulus emotion.

High anxiety individuals were more likely to optimally integrate threat-related emotional cues compared to low anxiety participants and compared to perception of other positive or non-threat-related negative emotions. In fact, only audiovisual performance in the high anxiety group for the angry emotion was well predicted by the optimal integration model (Ernst & Banks, 2002; see section B4 of the supplemental material for further supporting results). This means that in contrast to low anxiety participants, individuals in the high anxiety group were able to use all of the information available in an optimal way in order to reduce uncertainty, by weighting the face and voice information in their judgements in a way that was proportional to the relative reliability of the two cues. These findings suggest that individuals with high trait anxiety have multisensory mechanisms that are especially fine-tuned for perception of threat-related emotions.

The finding that individuals with high trait anxiety have a preserved ability to integrate multisensory cues to enhance their emotion perception is consistent with previous research by Peschard & Philippot (2017), who used a similar forced-choice emotion task to compare the response to angry and neutral faces and voices between individuals with lower and higher social anxiety traits. Peschard & Philippot found that both social anxiety groups showed a benefit for the audiovisual condition compared to the unimodal conditions. Our findings not only show that individuals with high trait anxiety retain the capacity to benefit from redundant information in audiovisual displays, but that for angry emotional cues, they do so more optimally than participants with low trait anxiety.

Of the two negative stimulus emotions tested, this effect was specific to audiovisual integration of angry cues. This finding builds on previous studies of multisensory processing in anxiety which have only looked at perception of anger or disgust (Koizumi et al., 2011; Peschard & Philippot, 2017; Yoon & Hong, 2020), which are both emotions that indicate interpersonal and/or physical threat. These studies have been unable to determine whether differences in multisensory processing in anxiety are specific to threat-related emotions or encompass other negative emotions as well. The present findings support theories of anxiety which propose that specific attentional and interpretation biases towards prioritising threat help to both develop and maintain hyperarousal and other anxiety symptoms (Bar-Haim et al., 2007; Beck & Clark, 1997; Cisler & Koster, 2010; Mathews & Mackintosh, 1998; Mogg & Bradley, 2016, 2018; Sussman et al., 2016). Our results suggest that anxiety-related threat biases not only influence how information is attended to and interpreted, but also how information is integrated. The tendency to integrate threat-related emotional cues more optimally may serve to exhaust capacity for processing of other emotional stimuli, leading to negatively biased evaluations of social interactions.

In addition to showing that high trait anxiety individuals were more likely to act as optimal observers when processing angry audiovisual stimuli, our results also revealed a trend towards less optimal integration for happy stimuli for high anxiety individuals. This suggests that individuals with high trait anxiety potentially obtained less perceptual benefit through multisensory integration for happy emotional cues (i.e. less of a reduction in uncertainty when judging emotion from happy audiovisual stimuli compared to happy unimodal stimuli). This trend towards less optimal integration of happy emotional cues in the high trait anxiety group suggests that as well as being more optimal multisensory integrators for threat-related emotions, individuals with high trait anxiety may also be suboptimal integrators when judging positive emotions. This trend is consistent with existing findings suggesting that anxiety is not only associated with a bias towards over-processing threat signals, but also a reduced tendency to orient towards and focus on positive emotional cues (e.g. Chen et al., 2012, 2016; Taylor et al., 2010).

The group differences in optimal multisensory integration for angry stimuli, and the trend towards a difference in optimal integration for happy stimuli, occurred in the absence of any significant differences in accuracy between anxiety groups for stimuli in any single emotion or modality condition (see section B of the supplemental material for these analyses). This suggests that anxiety-related differences were specific to how information from audio and visual cues was combined in audiovisual displays, rather than more general differences in sensory processing.

Although we observed differences between anxiety groups in optimal multisensory integration of emotional cues, we did not observe any group differences in the extent of multisensory benefit for response speed. RTs were shorter in both groups for the audiovisual condition compared to the unimodal conditions (see section B of supplemental material), but RTs in the audiovisual condition did not significantly violate the predictions of the Race Model (Miller, 1982). The lack of violation of Race Model predictions across either group for any stimulus emotion suggests that facilitation of RTs in the audiovisual condition did not necessarily reflect multisensory integration (i.e. neural interaction between visual and auditory cues) but is more likely to be the product of additive multisensory processes.

While the high trait anxiety group showed greater optimal integration of threat cues compared to the low anxiety group, they did not have longer RTs on these trials. This shows that this enhanced perception effect cannot be explained by a speed-accuracy trade-off in the high trait anxiety group, even though participants were told to take as much time as needed to ensure higher accuracy. The absence of anxiety-related differences in multisensory benefit for

RTs, and the lack of group differences in RTs more generally, is consistent with the existing literature showing a lack of differences in response speed for recognising emotions from audiovisual emotional stimuli for individuals with different levels of anxiety and social anxiety traits (Campanella et al., 2010, 2012; Delle-Vigne et al., 2015; Peschard & Philippot, 2017).

4.1 Limitations

While the findings of this study provide novel insight into how audiovisual integration of emotional cues is affected by trait anxiety, there are some limitations which should be noted. The sample size was not equal across the low and high anxiety groups, as only 27 individuals participated in the low anxiety group compared to 35 in the high anxiety group. However, this limitation is unlikely to have drastically affected the results, especially given that individuals in the low and high anxiety groups were recruited on the basis of their extreme scores on the trait anxiety scale, meaning that the two anxiety groups were highly different from one another, thus increasing the statistical power to detect between-groups differences. It has also been suggested that ANOVA F-statistics remain robust to violation of assumptions such as homogeneity of variance so long as the largest to smallest group size ratio is below 1.5, which was the case for our data (Blanca et al., 2017, 2018).

Another limitation relating to the participant sample is that it consisted predominantly of female university students, which raises issues of generalisability to the wider population. Although the high proportion of female participants does to an extent reflect the established finding that clinical anxiety is much more prevalent among females than males (McLean & Anderson, 2009), it also means that the results obtained in this study may not generalize to males.

It should also be noted that this study did not control for the potential influence of individual levels of depression, which given the high comorbidity rate between anxiety and depression may be considered as a limitation (Brown et al., 2001). However, the inclusion of a negative, non-threat-related emotion (i.e., sad) in the stimuli in addition to a threat-related stimulus emotion (i.e., anger) was intended to separate out general negativity bias effects, which are typically associated with depression and low mood (e.g., Bourke et al., 2010), from specific threat bias effects, which are specifically associated with anxiety(e.g., Cisler & Koster, 2010). Future studies could further control for the potential confounding effects of comorbid depression by include a measure of depression symptoms and adding this as a covariate in the analyses.

A further consideration is whether elicitation of emotion in the participants by the face and voice stimuli played a role in subsequent integration and interpretation of the emotional cues. It is known that viewing emotional information can elicit changes in the emotional state of the observer (e.g., Thom et al., 2014), and also that manipulating observer mood can alter subsequent perception of emotional faces (e.g., Niedenthal et al., 2010). In the current study the stimuli may have elicited emotion in participants which then impacted on subsequent emotional judgements. In particular, the angry cues may have elicited a fear response among individuals in the high anxiety group which then contributed to subsequent hypervigilance and enhanced detection of anger from audiovisual displays. It was beyond the scope of this study to measure the wider more complex appraisals or emotional states involved in the task, but future studies might consider asking participants about their felt emotions and how intense they were, and also complement these subjective measures with physiological measures such as heart rate and skin conductance to obtain objective measures of arousal.

4.2 Conclusion

The results of the current study show that while significant multisensory facilitation effects were evident in both anxiety groups, there were important differences in audiovisual integration between individuals with high and low trait anxiety that were contingent on stimulus emotion. High anxiety individuals showed a significantly greater propensity towards optimal integration of angry faces and voices compared to individuals in the low trait anxiety group, indicating that high anxiety individuals were more likely to attain maximal perceptual benefit from audiovisual integration of threat-related emotional cues. Our findings build on existing evidence of multisensory attentional biases towards threat in high trait anxiety by showing that these individuals also display an increased tendency to integrate threat-related multisensory emotional cues. This suggests that anxious individuals have altered multisensory mechanisms that are especially fine-tuned for processing threat-related emotions.

References

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. *Psychological Bulletin*, *133*(1), 1–24. https://doi.org/10.1037/0033-2909.133.1.1

Beck, A. T., & Clark, D. A. (1997). An information processing model of anxiety: Automatic and strategic processes. *Behaviour Research and Therapy*, *35*(1), 49–58. https://doi.org/10.1016/S0005-7967(96)00069-1

Bieling, P. J., Antony, M. M., & Swinson, R. P. (1998). The State–Trait Anxiety Inventory, Trait version: Structure and content re-examined. *Behaviour Research and Therapy*, *36*(7–8), 777–788. https://doi.org/10.1016/S0005-7967(98)00023-0

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, *29*(4), 552-557. https://doi.org/10.7334/psicothema2016.383

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2018). Effect of variance ratio on ANOVA robustness: Might 1.5 be the limit? *Behavior Research Methods*, *50*(3), 937–962. https://doi.org/10.3758/s13428-017-0918-2

Bourke, C., Douglas, K., & Porter, R. (2010). Processing of facial emotion expression in major depression: A review. *Australian & New Zealand Journal of Psychiatry*, *44*(8), 681–696. https://doi.org/10.3109/00048674.2010.496359

Brown, T.A., Campbell, L.A., Lehman, C.L., Grisham, J.R., Mancill, R.B. (2001). Current and lifetime comorbidity of the DSM-IV anxiety and mood disorders in a large clinical sample. *Journal of Abnormal Psychology*, *110*(4), 585-599. https://doi.org/10.1037//0021-843x.110.4.585.

Campanella, S., Bruyer, R., Froidbise, S., Rossignol, M., Joassin, F., Kornreich, C., Noël, X., & Verbanck, P. (2010). Is two better than one? A cross-modal oddball paradigm reveals greater sensitivity of the P300 to emotional face-voice associations. *Clinical Neurophysiology*, *121*(11), 1855–1862. https://doi.org/10.1016/j.clinph.2010.04.004

Campanella, S., Delle-Vigne, D., Kornreich, C., & Verbanck, P. (2012). Greater sensitivity of the P300 component to bimodal stimulation in an event-related potentials oddball task. *Clinical Neurophysiology*, *123*(5), 937–946. https://doi.org/10.1016/j.clinph.2011.10.041

Chen, N. T. M., Clarke, P. J. F., MacLeod, C., & Guastella, A. J. (2012). Biased attentional processing of positive stimuli in social anxiety disorder: An eye movement study. *Cognitive Behaviour Therapy*, *41*(2), 96–107. https://doi.org/10.1080/16506073.2012.666562

Chen, N. T. M., Clarke, P. J. F., MacLeod, C., Hickie, I. B., & Guastella, A. J. (2016). Aberrant gaze patterns in social anxiety disorder: An eye movement assessment during public speaking. *Journal of Experimental Psychopathology*, *7*(1), 1–17. https://doi.org/10.5127/jep.040313

Cisler, J. M., & Koster, E. H. W. (2010). Mechanisms of attentional biases towards threat in the anxiety disorders: An integrative review. *Clinical Psychology Review*, *30*(2), 203. https://doi.org/10.1016/j.cpr.2009.11.003

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain Research*, *1242*, 126–135. https://doi.org/10.1016/j.brainres.2008.04.023

De Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion*, *14*(3), 289–311. https://doi.org/10.1080/026999300378824

Delle-Vigne, D., Kornreich, C., Verbanck, P., & Campanella, S. (2015). The P300 component wave reveals differences in subclinical anxious-depressive states during bimodal oddball tasks: An effect of stimulus congruence. *Clinical Neurophysiology*, *126*(11), 2108–2123. https://doi.org/10.1016/j.clinph.2015.01.012

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429. https://doi.org/10.1038/415429a

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*(2), 336–353. https://doi.org/10.1037/1528-3542.7.2.336

Heffer, N., Karl, A., Jicol, C., Ashwin, C., & Petrini, K. (2021). Anxiety biases audiovisual processing of social signals. Behavioural Brain Research, 410, 113346. https://doi.org/10.1016/j.bbr.2021.113346.

King, A. J., & Calvert, G. A. (2001). Multisensory integration: Perceptual grouping by eye and ear. *Current Biology*, *11*(8), R322–R325. https://doi.org/10.1016/S0960-9822(01)00175-0

Klasen, M., Chen, Y. H., & Mathiak, K. (2012). Multisensory emotions: Perception, combination and underlying neural processes. *Reviews in the Neurosciences*, *23*(4), 381–392. https://doi.org/10.1515/revneuro-2012-0040

Koizumi, A., Tanaka, A., Imai, H., Hiramatsu, S., Hiramoto, E., Sato, T., & de Gelder, B. (2011). The effects of anxiety on the interpretation of emotion in the face–voice pairs. *Experimental Brain Research*, *213*(2), 275–282. https://doi.org/10.1007/s00221-011-2668-1

[dataset] Livingstone, S. R., & Russo, F. A. (2018a). *The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)*. Zenodo. https://doi.org/10.5281/zenodo.1188976

Livingstone, S. R., & Russo, F. A. (2018b). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLOS ONE*, *13*(5), Article e0196391. https://doi.org/10.1371/journal.pone.0196391

Love, S. A., Petrini, K., Cheng, A., & Pollick, F. E. (2013). A psychophysical investigation of differences between synchrony and temporal order judgments. *PLoS ONE*, *8*(1), e54798. https://doi.org/10.1371/journal.pone.0054798

Mathews, A., & Mackintosh, B. (1998). A cognitive model of selective processing in anxiety. *Cognitive Therapy and Research*, *22*(6), 539–560. https://doi.org/10.1023/A:1018738019346

McLean, C. P., & Anderson, E. R. (2009). Brave men and timid women? A review of the gender differences in fear and anxiety. *Clinical Psychology Review*, *29*(6), 496–505. https://doi.org/10.1016/j.cpr.2009.05.003.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*(2), 247–279. https://doi.org/10.1016/0010-0285(82)90010-X

Mogg, K., & Bradley, B. P. (2016). Anxiety and attention to threat: Cognitive mechanisms and treatment with attention bias modification. *Behaviour Research and Therapy*, *87*, 76–108. https://doi.org/10.1016/j.brat.2016.08.001

Mogg, K., & Bradley, B. P. (2018). Anxiety and threat-related attention: Cognitive-motivational framework and treatment. *Trends in Cognitive Sciences*, *22*(3), 225–240. https://doi.org/10.1016/j.tics.2018.01.001

Niedenthal, P. M., Brauer, M., Halberstadt, J. B., & Innes-Ker, Å. H. (2001). When did her smile drop? Facial mimicry and the influences of emotional state on the detection of change in emotional expression. *Cognition & Emotion*, *15*(6), 853-864. https://doi.org/10.1080/02699930143000194

Peschard, V., & Philippot, P. (2017). Overestimation of threat from neutral faces and voices in social anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, *57*, 206–211. https://doi.org/10.1016/j.jbtep.2017.06.003

Piwek, L., Pollick, F., & Petrini, K. (2015). Audiovisual integration of emotional signals from others' social interactions. *Frontiers in Psychology*, *6*, 611. https://doi.org/10.3389/fpsyg.2015.00611

Quadflieg, S., Wendt, B., Mohr, A., Miltner, W. H., & Straube, T. (2007). Recognition and evaluation of emotional prosody in individuals with generalized social phobia: A pilot study. *Behaviour Research and Therapy*, *45*(12), 3096–3103. https://doi.org/10.1016/j.brat.2007.08.003

Richards, A., French, C. C., Calder, A. J., Webb, B., Fox, R., & Young, A. W. (2002). Anxiety-related bias in the classification of emotionally ambiguous facial expressions. *Emotion*, *2*(3), 273. https://doi.org/10.1037/1528-3542.2.3.273

Rohde, M., van Dam, L. C. J., & Ernst, M. (2016). Statistically optimal multisensory cue integration: A practical tutorial. *Multisensory Research*, *29*(4–5), 279–317. https://doi.org/10.1163/22134808-00002510

Schirmer, A., & Adolphs, R. (2017). Emotion perception from face, voice, and touch: Comparisons and convergence. *Trends in Cognitive Sciences*, *21*(3), 216–228. https://doi.org/10.1016/j.tics.2017.01.001

Spielberger, C. D., Gorsuch, R.L., Lushene, R., Vagg, P.R., Jacobs, G.A. (1983). *Manual for the State–Trait Anxiety Inventory, STAI (Form Y): Self-evaluation questionnaire*. Consulting Psychologists Press.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. The MIT Press.

Sussman, T. J., Jin, J., & Mohanty, A. (2016). Top-down and bottom-up factors in threat-related perception and attention in anxiety. *Biological Psychology*, *121*, 160–172. https://doi.org/10.1016/j.biopsycho.2016.08.006

Taylor, C. T., Bomyea, J., & Amir, N. (2010). Attentional bias away from positive social information mediates the link between social anxiety and anxiety vulnerability to a social stressor. *Journal of Anxiety Disorders*, *24*(4), 403–408. https://doi.org/10.1016/j.janxdis.2010.02.004

Thom, N., Knight, J., Dishman, R., Sabatinelli, D., Johnson, D. C., & Clementz, B. (2014). Emotional scenes elicit more pronounced self-reported emotional experience and greater EPN and LPP modulation when compared to emotional faces. *Cognitive, Affective, & Behavioral Neuroscience, 14*(2), 849-860. https://doi.org/10.3758/s13415-013-0225-z

Tseng, H.H., Bossong, M. G., Modinos, G., Chen, K.M., McGuire, P., & Allen, P. (2015). A systematic review of multisensory cognitive–affective integration in schizophrenia. *Neuroscience & Biobehavioral Reviews*, *55*, 444–452. https://doi.org/10.1016/j.neubiorev.2015.04.019

Ulrich, R., Miller, J., & Schröter, H. (2007). Testing the race model inequality: An algorithm and computer programs. *Behavior Research Methods*, *39*(2), 291–302. https://doi.org/10.3758/BF03193160

Veerapa, E., Grandgenevre, P., El Fayoumi, M., Vinnac, B., Haelewyn, O., Szaffarczyk, S., Vaiva, G., & D'Hondt, F. (2020). Attentional bias towards negative stimuli in healthy individuals and the effects of trait anxiety. *Scientific Reports*, *10*(1), 11826. https://doi.org/10.1038/s41598-020-68490-5

Yoon, K. L., & Hong, S. W. (2020). Behavioral inhibition system sensitivity moderates audio-visual neutral information processing. *Journal of Behavior Therapy and Experimental Psychiatry*, 101597. https://doi.org/10.1016/j.jbtep.2020.101597

Yoon, K. L., & Zinbarg, R. E. (2008). Interpreting neutral faces as threatening is a default mode for socially anxious individuals. *Journal of Abnormal Psychology*, *117*(3), 680–685. https://doi.org/10.1037/0021-843X.117.3.680