Anxiety Biases Audiovisual Processing of Social Signals

Naomi Heffer[a], Anke Karl[b], Crescent Jicol[a], Chris Ashwin[a,*] & Karin Petrini[a,*]

[a]University of Bath, Department of Psychology, United Kingdom; [b]University of Exeter, Mood Disorders Centre, United Kingdom; [*]both authors contributed equally to this manuscript.

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

Supplemental Materials

A. Additional Methodological Details

A.1 Sample Size and Power Calculation

Sample size was selected based on the large effect size for the emotion by anxiety interaction effect reported by Koizumi et al. [18]. An a priori power calculation stipulated a sample size of 34 for 95% power to detect the previously found effect, which was inflated to 55 for this study to account for the 'winner's curse effect', which refers to the fact that the first published study to find an effect normally has an inflated effect size [67].

A.2 Edits Applied to All Stimuli

The average sound amplitude of all clips was normalised to -9.9dBFS using Audacity® v.2.3.2 (Audacity Team, 2019) before the clips were exported into Adobe Premiere Pro 2017 (Adobe, 2017). For all stimuli, the mean length of the audio and visual clips, as well as the mean length for happy and angry clips (shown in Table A1), were compared using independent t-tests, or Mann-Whitney U tests where data was not normally distributed, to check that there were not significant differences in the length of stimuli across different emotions and modalities ($p \geq .060$ for all comparisons). When audio and visual clips were paired together to create the audiovisual displays, the resulting clips were the length of whichever component clip was longer (either audio or video). The clips were exported as H.264 (mp4) files with a resolution of 800 by 600 pixels, a frame rate of 30fps, and Advanced Audio Coding (AAC) audio with a sample rate of 44.1kHz. The stimuli were presented on a 47 x 27cm monitor via a 1920 x 1080 resolution video player, and via Sony WH-CH700N Noise-Cancelling headphones, at a mean sound intensity of 60 dB, although volume was adjusted according to participant preference.

Table A1

Stimulus clip length in different modality and emotion conditions

	Mean (SD) Stimulus Length (ms)			
Stimulus Type	Audio	Visual	Happy	Angry
Bodies-Voices	3068 (180)	3003 (240)	3096 (288)	2976 (44)
Faces-Voices	1571 (385)	1433 (215)	1500 (386)	1503 (236)
Circles-Tones	1400 (0)	1400 (0)	1400 (0)	1400 (0)

A.3 Faces-Voices

Owing to existing evidence that facial cues are judged as being more reliable than vocal cues in previous experiments using similar audiovisual stimuli [18,26], the reliability of the visual face cues was reduced by using Adobe Premiere Pro 2017 (Adobe, 2017) to add a 30% Gaussian blur over the eyes, nose and mouth region of the faces. Blurring was chosen as a means of reducing the reliability of the visual information to emulate real-life conditions when visual information may be less reliable, e.g., looking at a person from a distance. It was important to equalise the reliability of the visual and audio cues as the extent of multisensory integration has been shown to be dependent on the relative reliability of component sensory cues, with greater multisensory facilitation effects occurring when the different senses are similarly reliable (e.g., [60]).

Six actors portrayed the facial expressions, each contributing an angry and a happy expression. These stimuli were validated as showing the target emotion in a British sample by running a small pilot experiment ($N = 12$) where individuals were asked to make a forced choice of happy or angry for each stimulus. Stimuli from eight actors were piloted and the six that were most accurately recognised (*M* = 96%) were selected for inclusion. Faces were presented on a light grey background. The face region of each stimulus (eyes, nose and mouth) subtended a visual angle of 5° and the full image covered a visual angle of 28° degrees wide and 20° degrees in height at a viewing distance of 57cm. Luminosity and contrast levels for the facial stimuli were measured in Gimp 2.8 (The GIMP Team, 1997-2004) and statistically compared to check that there was not significant variation in these low-level visual properties between angry and happy stimuli (p's $\geq .662$ for all comparisons).

Voice clips of the same sentences being spoken with angry and happy prosody were not available for each actor in the EU-Emotion Stimulus set [41,42], so in order to create happy and angry versions of each stimulus, actors speaking the same sentences with different emotional prosody were matched on voice similarity to form pairs. Eight pairs of matched voices by a total of five different actors were piloted and the six pairs that were recognised as their target emotion with the highest accuracy (*M* = 84%) were selected for inclusion. To create the audiovisual face-voice stimuli, the same voice-actors were always matched to the same face-actor.

A.4 Body Motion-Voices

From the set of clips in Piwek et al. [39], six angry and six happy audiovisual displays that portrayed the emotions at medium intensity, and which had previously been recognised with high accuracy (≥75%), were selected, thus ensuring that we avoided ceiling effects while allowing the correct emotion to be identified at a level appropriately above chance. The point-light stimuli were presented on a plain black background and the area of the screen occupied by the interacting figures subtended a visual angle of 7° at a viewing distance of 57cm. Finding the mean recognition accuracy for emotions across the visual, auditory and audiovisual congruent displays used in the Piwek et al. study [39] revealed that identification accuracy was higher in the audiovisual (82%) and auditory-only (78%) groups than the visual group (62%), indicating that the auditory information was a more reliable cue to emotion than the visual information. Hence, reliability of the auditory stimuli had to be decreased to a level that approximated the visual stimuli by high-cut filtering of the audio clips. A high-cut filter attenuating sound above 280Hz was applied to the auditory stimuli to maintain the vocal prosody and intonation while decreasing the auditory reliability to a level more similar to the visual stimuli [23]. High-cut filtering was chosen as a means of reducing the reliability of the auditory information as this modification emulates real-life conditions, such as listening to voices over a fence or in the rain, and therefore maintains ecological validity [68]. The procedures applied to standardise these stimuli in terms of low-level visual properties etc. are described in Piwek et al. [39].

A.5 Circles-Tones

The visual stimuli, the circles, were created using the ellipse drawing tool in Paint 3D (Microsoft, 2016) to draw a light grey circle (300 pixels in diameter) which was transformed to six different levels of brightness, ranging from dark grey to white, using GIMP 2.8 (The GIMP Team, 1997-2004) image editing software. These images were exported from GIMP as PNG files and made into short video clips using Adobe Premiere Pro 2017 (Adobe, 2017). The circles were presented on a plain black background and the visual angle of the circle diameter was 10° at a viewing distance of 57cm. The area of the circles approximated the area subtended by the face region (eyes, mouth and nose) for the face-voice stimuli. The circle stimuli were piloted with a small group ($N = 17$) to validate whether they were consistently recognised as the target emotion. Individuals were asked to make a forced choice of happy or angry for each stimulus. Eight stimuli were piloted and six were selected which were recognised as their target emotion at least 67% of time ($M = 78\%$).

The auditory stimuli were pure (sine wave) tones created using an online tone generator (onlinetonegenerator.com, 2016). The original set of eight tones which was piloted ranged from 50

Hz to 1245 Hz progressing upward in musical fifths (in accordance with the stimuli used in Ward et al. [69]) but excluding the two middle fifths to enhance contrast between ‘low’ and ‘high’ pitch tones, facilitating discrimination of ‘happy’ tones from ‘angry’ tones. Initial piloting revealed that low tones were reliably recognised as angry, but high tones above 367 Hz were also consistently recognised as angry. Feedback from participants suggested that this was because the very high tones were perceived as being ‘louder’ (despite being normalised for sound amplitude) and ‘sometimes painful’, so the decision was made to try and limit the frequency of the pure tones to a smaller range that was more reflective of the range of tones in human speech, thus cutting out the ‘painful’ high tones and making the pure tone stimuli more comparable in terms of frequency to the social speech stimuli. The fundamental frequency of speech for men is typically 85-180 Hz and for women it is typically 165-255 Hz [70, 71], and so the tone stimuli were changed from progressing in fifths from 50 Hz (as in Ward et al. [69]) to progressing in perfect fourths, again with the middle two tones removed to increase contrast in pitch between ‘happy’ and ‘angry’ tones. The highest frequency tone used in this new set was 375 Hz. After these adjustments, all the tones were recognised as their target emotion 68% of the time on average.

A.6 Additional Procedural Details

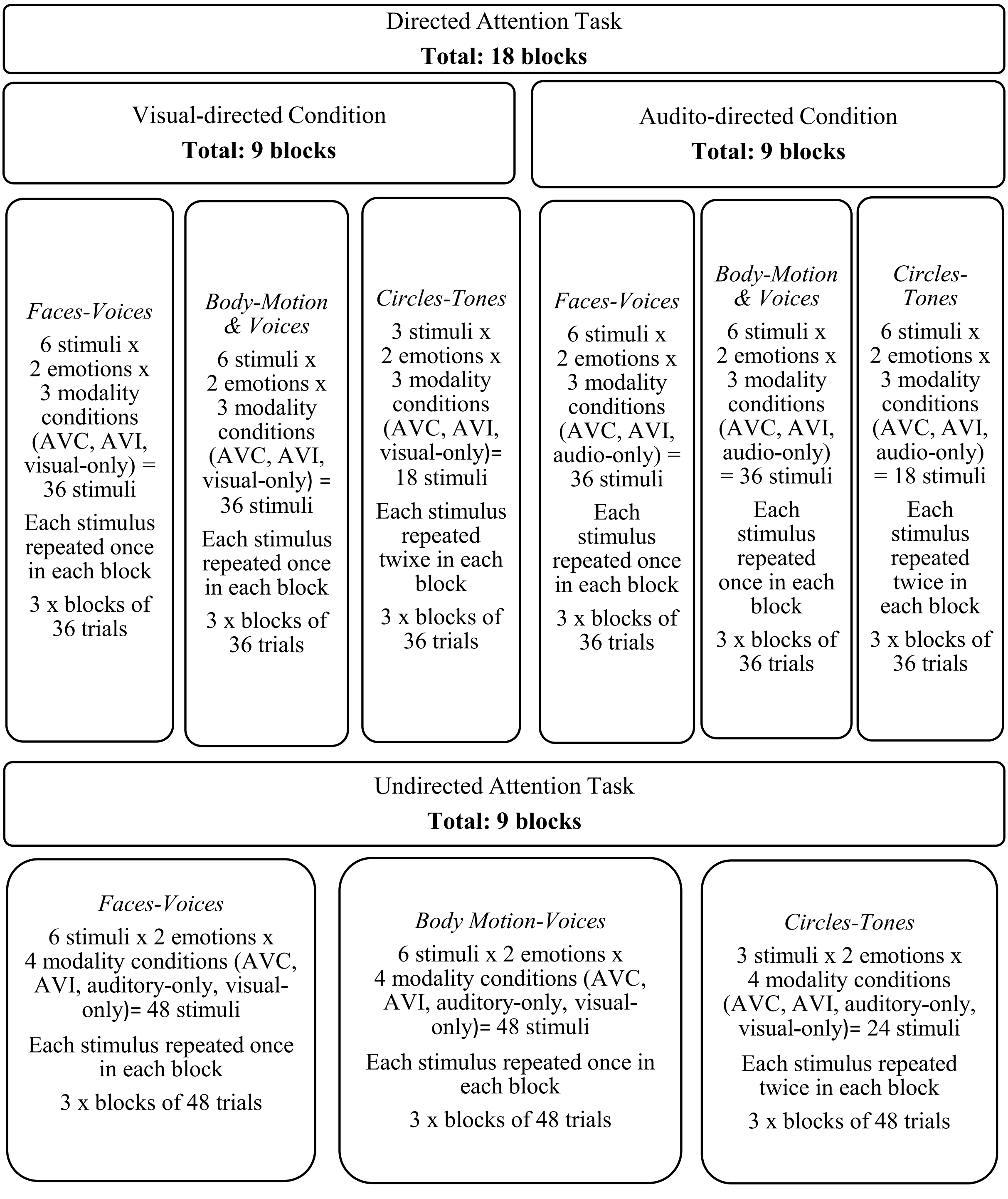

Figure A1. Diagram showing the number of stimuli and trials for each of the tasks and stimulus types. AVC = audiovisual congruent, AVI = audiovisual incongruent.

B. Additional Results

Throughout this section all main effects from all ANCOVA analyses are reported in the relevant tables. Given that there were many potential interaction effects, we report the significant interactions for each ANCOVA, but only summative results for the non-significant interactions to aid readability. Results highlighted in bold are those which have already been reported in the main manuscript.

B.1 Crossmodal Interference

Crossmodal Interference Effects (see section 3.1 of main text) were analysed separately for each stimulus type using a mixed factorial ANCOVA with Anxiety Group (lower or higher) as a between-participant factor, Attended Modality (visual or audio) and Emotion (angry or happy) as within-participant factors, and Autistic Traits (AQ score) as a covariate. Please refer to the main manuscript (Table 2) for the full set of ANCOVA results for Crossmodal Interference Effects calculated based on percentage accuracy data. Here, descriptive statistics are reported for Crossmodal Interference Effects – Accuracy and RTs – and ANCOVA results for RTs.

Table B1

Mean (SD) Crossmodal Interference (Accuracy [%]) in the Directed Attention Task in lower- and higher trait anxiety groups as a function of the target emotion

Stimulus Type	Target Emotion	Anxiety Group	
		Lower	Higher
Faces-Voices	Happy	26.79(12.63)	27.98 (12.47)
	Angry	26.69 (12.82)	28.57 (10.14)

Bodies-Voices	Happy	16.47 (11.11)	19.35 (11.80)
	Angry	14.78 (11.62)	19.05 (13.50)
Circles-Tones	Happy	37.10 (10.86)	33.33 (14.46)
	Angry	34.42 (11.32)	35.02 (13.89)

Note: crossmodal interference calculated by subtracting percentage accuracy in the audiovisual incongruent trials from percentage accuracy in the audiovisual congruent trials for each participant. Larger interference effects indicate that processing of incongruent information in the to-be-ignored modality was more likely to interfere with the perception of emotion from information in the attended modality.

Table B2

Mean (SD) Crossmodal Interference (Reaction Times [ms]) in the Directed Attention Task in lower- and higher trait anxiety groups as a function of the target emotion

Stimulus Type	Target Emotion	Anxiety Group	
		Lower	Higher
Faces-Voices	Happy	-114 (179)	-55 (158)
	Angry	-29 (202)	-61 (167)
Bodies-Voices	Happy	-96 (160)	-135 (223)
	Angry	-41 (265)	-45 (294)
Circles-Tones	Happy	-84 (170)	-51 (58)
	Angry	-61 (110)	-40 (99)

Note: crossmodal interference calculated by subtracting mean reaction time for correct responses in the audiovisual incongruent trials from the mean reaction time in the audiovisual congruent trials for each participant. Larger negative values indicate slower reaction times in the audiovisual incongruent condition relative to the audiovisual congruent condition.

Table B3

ANCOVA results for Crossmodal Interference Effects (Reaction Times)

Stimulus Type	Source	*df*	*F*	*p*	Effect Size (partial η^2)
Faces-Voices	Attended Modality	1, 49	6.65	.013	.12
	Emotion	1, 49	1.57	.216	.03
	AQ Score	1, 49	0.05	.822	<.01
	Anxiety Group	1, 49	0.08	.780	<.01
	Interaction terms	1, 49	≤1.57	≥.216	≤.03
Bodies-Voices	Attended Modality	1, 51	0.16	.693	<.01
	Emotion	1, 51	0.57	.452	.01
	AQ Score	1, 51	0.35	.558	.01
	Anxiety Group	1, 51	0.09	.760	<.01
	Interaction terms	1, 51	≤0.60	≥.443	≤.01
Circles-Tones	Attended Modality	1, 22	0.43	.519	.02
	Emotion	1, 22	0.95	.342	.04
	AQ Score	1, 22	3.02	.096	.12
	Anxiety Group	1, 22	1.85	.187	.08
	Interaction terms	1, 22	≤2.13	≥.159	≤.09

Accuracy and reaction times on unimodal trials in the Directed Attention Task were analysed to determine whether differences in unimodal perception were likely to be driving the observed differences in Crossmodal Interference Effects in this task. The data for each stimulus type was analysed separately using a mixed factorial ANCOVA with Anxiety Group (lower or higher) as a between-participant factor, Modality (visual or audio) and Emotion (angry or happy) as within-participant factors, and Autistic Traits (AQ score) as a covariate.

Table B4

Mean (SD) Accuracy (%) in unimodal conditions of the Directed Attention Task in lower- and higher trait anxiety groups

Stimulus Type	Stimulus Modality & Emotion		Anxiety Group	
			Lower	Higher
Faces-Voices	Audio			
		Happy	77.78 (15.42)	78.57 (16.95)
		Angry	79.17 (16.74)	83.13(11.46)
	Visual			
		Happy	88.09 (9.88)	81.55 (10.59)
		Angry	96.03 (6.21)	95.83 (4.69)
Bodies-Voices	Audio			
		Happy	69.84 (15.45)	71.63 (18.97)
		Angry	62.90 (17.11)	63.69 (14.30)
	Visual			
		Happy	73.02 (16.19)	67.86 (13.30)
		Angry	58.73 (16.45)	62.90 (11.01)
Circles-Tones	Audio			
		Happy	78.57 (17.42)	83.53 (12.33)

	Angry	95.44 (6.06)	89.68 (10.98)
	Visual		
	Happy	98.02 (3.45)	97.42 (4.66)
	Angry	95.63 (5.93)	96.03 (7.54)

Table B5

Mean (SD) Reaction Times (ms) in unimodal conditions of the Directed Attention Task in lower- and higher trait anxiety groups

Stimulus Type	Stimulus Modality & Emotion	Anxiety Group	
		Lower	Higher
Faces-Voices	Audio		
	Happy	1689 (318)	1753 (372)
	Angry	1799 (552)	1750 (361)
	Visual		
	Happy	954 (249)	956 (196)
	Angry	949 (210)	947 (194)
Bodies-Voices	Audio		
	Happy	3192 (466)	3221 (455)
	Angry	3160 (695)	3201 (391)
	Visual		
	Happy	2898 (862)	2960 (671)
	Angry	2789 (948)	2824 (701)

Circles-Tones	Audio			
		Happy	1055 (249)	1068 (276)
		Angry	1028 (248)	1064 (299)
	Visual			
		Happy	636 (140)	694 (187)
		Angry	664 (101)	736 (185)

Table B6

ANCOVA results for Accuracy (%) perceiving emotions from unimodal displays in the Directed Attention Task

Stimulus Type	Source	*df*	*F*	*p*	Effect Size (partial η^2)
Faces-Voices	Modality	1, 53	4.86	.032	.08
	Emotion	1, 53	5.35	.025	.09
	AQ Score	1, 53	1.37	.248	.03
	Anxiety Group	1, 53	<0.01	.958	<.01
	Anxiety Group x Modality	1, 53	4.85	.032	.08
	All other interaction terms	1, 53	≤2.23	≥.141	≤.04
Bodies-Voices	Modality	1, 53	3.35	.073	.06
	Emotion	1, 53	4.03	.050	.07
	AQ Score	1, 53	1.32	.257	.02
	Anxiety Group	1, 53	0.01	.927	<.01
	Interaction terms	1, 53	≤2.76	≥.103	≤.05

Circles-Tones	Modality	1, 53	9.17	.004	.15
	Emotion	1, 53	7.23	.010	.12
	AQ Score	1, 53	0.53	.471	.01
	Anxiety Group	1, 53	<0.01	.973	<.01
	Anxiety Group x Emotion	**1, 53**	**4.33**	**.042**	**.08**
	Anxiety Group x Emotion x Modality	**1, 53**	**6.34**	**.015**	**.11**
	Modality x Emotion	1, 53	4.01	.051	.07
	All other interaction terms	1, 53	≤0.63	≥.431	≤.01

Note: the results in bold are those presented in the main manuscript.

The analyses of accuracy for unimodal trials in the Directed Attention Task (Table B6) revealed that for the body-motion and voice stimuli there were no differences in performance between the two anxiety groups, but that there were differences for the faces-voices and circles-tones. Following up on the significant interaction terms in the ANCOVA for the circles-tones showed that individuals in the higher trait anxiety group perceived the emotion from the angry audio cues (i.e., low tones) less accurately than individuals in the lower trait anxiety group, $p = .031$, 95% CI [0.51, 10.50], $d = 0.60$. Following up on the significant anxiety group by modality interaction for the faces-voices shows that there was no significant difference in performance between anxiety groups for audio stimuli, $p = .242$, 95% CI [-9.12, 2.35], while there was a trend that did not reach significance for the visual stimuli, $p = .063$, 95% CI [-0.18, 6.57], $d = 0.52$.

Table B7

ANCOVA results for mean Reaction Times (ms) for correct responses to unimodal displays in the Directed Attention Task

Stimulus Type	Source	*df*	*F*	*p*	Effect Size (partial η^2)
Faces-Voices	Modality	1, 53	85.54	<.001	.62
	Emotion	1, 53	1.23	.273	.02
	AQ Score	1, 53	0.05	.818	<.01
	Anxiety Group	1, 53	0.01	.912	<.01
	Interaction terms	1, 53	≤1.69	≥.199	≤.03
Bodies-Voices	Modality	1, 53	0.13	.718	<.01
	Emotion	1, 53	0.12	.730	<.01
	AQ Score	1, 53	0.04	.842	<.01
	Anxiety Group	1, 53	0.10	.754	<.01
	Interaction terms	1, 53	≤2.10	≥.153	≤.04
Circles-Tones	Modality	1, 53	19.41	<.001	.27
	Emotion	1, 53	1.22	.274	.02
	AQ Score	1, 53	0.74	.394	.01
	Anxiety Group	1, 53	0.41	.527	.01
	Interaction terms	1, 53	≤2.87	≥.096	≤.05

B.2 Emotional Bias for Incongruent Audiovisual Displays

Response Bias (see section 3.2 of main text) was analysed by mixed factorial ANCOVA with Anxiety Group (lower or higher) as a between-participant factor, Visual Emotion (happy or angry) as a within-participant factor and Autistic Traits (AQ score) as a covariate. Relevant descriptive statistics are summarised in Table B8. The results from the ANCOVA analyses are summarised in Table B9.

Table B8

Mean (SD) Response Bias for emotionally incongruent audiovisual displays in the Undirected Attention Task for lower-and higher trait anxiety groups

Stimulus Type	Anxiety Group	
	Lower	Higher
Faces-Voices	0.1409 (0.1653)	0.2460 (0.2162)
Bodies-Voices	-0.0119 (0.2651)	-0.0218 (0.2953)
Circles-Tones	0.1210 (0.2602)	0.1131 (0.3248)

Note: response bias was calculated for each participant by subtracting the proportion of ‘happy’ judgements from the proportion of ‘angry’ judgements for audiovisual incongruent trials in the Undirected Attention Task. A score of 1 indicates that the participant always responded ‘angry’, while a score of -1 indicates that the participant always responded ‘happy’. A score of 0 indicates that there was no response bias.

Table B9

ANCOVA results for Response Bias on emotionally incongruent trials in the Undirected Attention Task

Stimulus Type	Source	*df*	*F*	*p*	Effect Size (partial η^2)
Faces-Voices	Visual Emotion	1, 53	36.26	<.001	.41
	AQ Score	1, 53	1.82	.183	.03
	AQ Score x Visual Emotion	1, 53	5.23	.026	.09
	Anxiety Group	**1, 53**	**5.55**	**.022**	**.10**
	All other interaction terms	1, 53	1.57	.216	.03
Bodies-Voices	Visual Emotion	1, 53	4.57	.037	.08
	AQ Score	1, 53	0.59	.445	.01
	Anxiety Group	1, 53	0.12	.734	<.01
	Interaction terms	1, 53	≤1.54	≥.220	.03
Circles-Tones	Visual Emotion	1, 53	<0.01	.968	<.01
	AQ Score	1, 53	0.49	.487	.01
	Anxiety Group	1, 53	0.09	.771	<.01
	Interaction terms	1, 53	≤0.17	≥.685	<.01

Note: the results in bold are those presented in the main manuscript.

In addition to the results reported in section 3.2 of the main manuscript, Table B9 shows that there was also a significant interaction effect between Visual Emotion and Autistic Traits on Response Bias for the faces-voices. As this interaction cannot help us to explain the effect of anxiety group on multisensory response bias, it is not directly relevant to our research questions. However, out of interest about what was driving this effect, we followed up on the interaction by running two Pearson correlation tests to examine the relationship between autistic traits and response bias separately for stimuli with happy vs. angry visual cues.

The results revealed that Autistic Traits were negatively correlated with Response Bias when the visual emotion was angry, $r = -.33$, $p = .012$, and that there was a trend towards a positive association between Response Bias and Autistic Traits when the visual emotion was happy, $r = .23$, $p = .085$. These results showed that participants with higher levels of autistic traits were less likely to respond ‘happy’ to emotionally incongruent audiovisual displays when the visual emotion was happy and the audio emotion was angry, and were less likely to respond angry when the visual emotion was angry and the audio emotion was happy. These findings show that individuals with higher levels of autistic traits weighted visual information from the facial expressions less and auditory information from the voices more when making emotion judgements about incongruent face and voice expressions. This is different to the typical pattern of visual dominance which is observed among the general population when judging emotion from combined face and voice expressions [26,72].

Our finding that individuals with higher levels of autistic traits weighted the less reliable voice information more in their judgements, rather than the more informative facial cues, is consistent with previous findings which indicate that people with autism and higher levels of autistic traits show a tendency to rely on less informative social cues during social processing and interactions. Individuals with autism or higher levels of autistic traits show a particular tendency to focus on cues where the information is audio-relevant, such as gazing towards the mouth, the shape of which influences vocal acoustics, rather than the typically more informative eye region during social processing [73–75].

B.3 Multisensory Facilitation

To determine whether there were likely to be differences in the extent of multisensory facilitation between the two anxiety groups, a mixed factorial ANCOVA was used to analyse Multisensory Gain scores (see section 3.3 of main text) for each of the three stimulus types, with Anxiety Group (lower or higher) as a between-participant factor, Emotion (angry or happy) as a within-participant factor and Autistic Traits (AQ Score) as a covariate. The results of the ANCOVA analyses for Multisensory Gain scores based on accuracy and reaction times are reported in Tables B12 and B13, respectively.

Table B10

Mean (SD) Accuracy (%) and Multisensory Gain (%) in the Undirected Attention Task in lower- and higher trait anxiety groups as a function of stimulus modality

Stimulus Type	Modality	Anxiety Group	
		Lower	Higher
Faces-Voices	Audio	79.27(11.63)	79.84(10.06)
	Visual	90.48(6.43)	88.48(4.67)
	AVC[a]	91.96(6.06)	88.39 (7.60)
	AV Gain[b]		
	Happy	-0.99(10.80)	-3.18(10.63)
	Angry	-3.57(7.44)	-0.19(6.68)
Bodies-Voices	Audio	64.29(9.52)	66.27(10.02)
	Visual	64.68(6.36)	65.58(7.61)
	AVC	68.15(7.91)	70.73(8.37)
	AV Gain		
	Happy	-11.31(12.96)	-9.13(12.85)

	Angry	1.39(13.58)	2.78(13.40)
Circles-Tones	Audio	84.72 (9.93)	83.33(10.72)
	Visual	95.14(5.54)	95.04(4.17)
	AVC	92.86(7.15)	92.86(5.88)
	AV Gain		
	Happy	-5.75 (9.62)	-6.35(10.22)
	Angry	-3.57(5.70)	0.59(10.07)

Note: [a]AVC = audiovisual congruent; [b]AV Gain = difference between accuracy in the audiovisual congruent condition and that in the best unimodal condition, i.e., $AVGain = p(AV) - max\{p(A), p(V)\}$, where p is the percentage of correct responses in each condition. Positive scores indicate multisensory gain and negative scores indicate better performance with only one of the senses.

Table B11

Mean (SD) Reaction Times (ms) in the Undirected Attention Task in lower- and higher trait anxiety groups as a function of stimulus modality

Stimulus Type	Modality	Anxiety Group	
		Lower	Higher
Faces-Voices	Audio	1799(341)	1870(328)
	Visual	1209(245)	1321(334)
	AVC[a]	1471(355)	1536(332)
	AV Gain[b]		
	Happy	272(164)	217(227)

	Angry	208(217)	235(201)
Bodies-Voices	Audio	3243(576)	3188(671)
	Visual	3083(610)	3036(778)
	AVC	3175(664)	3123(744)
	AV Gain		
	Happy	200(300)	246(363)
	Angry	114(216)	64(133)
Circles-Tones	Audio	1143(271)	1140(197)
	Visual	931(288)	956(286)
	AVC	936(258)	1012(240)
	AV Gain		
	Happy	104(149)	130(148)
	Angry	19(144)	49(131)

Note: [a]AVC = audiovisual congruent; [b]AV Gain = difference between accuracy in the audiovisual congruent condition and that in the best unimodal condition (i.e., the condition with the shortest mean reaction time). Negative scores indicate multisensory gain and positive scores indicate better performance with only one of the senses.

Table B12

ANCOVA results for Multisensory Gain Scores (Accuracy) in the Undirected Attention Task

Stimulus Type	Source	*df*	*F*	*p*	Effect Size (partial η^2)
Faces-Voices	Emotion	1, 53	0.87	.356	.02
	AQ Score	1, 53	0.12	.729	<.01

	Anxiety Group	1, 53	0.20	.655	<.01
	Anxiety Group x Emotion	**1, 53**	**3.02**	**.088**	**.05**
	All other interaction terms	1, 53	0.95	.335	.02
Bodies-Voices	Emotion	1, 53	5.43	.024	.09
	AQ Score	1, 53	4.97	.030	.09
	Anxiety Group	1, 53	0.04	.835	<.01
	Interaction terms	1, 53	≤0.34	≥.565	≤.01
Circles-Tones	Emotion	1, 53	<0.01	.947	<.01
	AQ Score	1, 53	1.09	.301	.02
	Anxiety Group	1, 53	1.79	.186	.03
	Interaction terms	1, 53	≤1.18	≥.283	≤.02

Note: the results in bold are those presented in the main manuscript.

The results in Table B12 showed a trend towards an interaction between anxiety group and emotion for the face and voice stimuli which is described in more detail in the main manuscript text (see section 3.3). There was also a significant main effect of Autistic Traits for the body-motion and voice stimuli. As this main effect of autistic traits cannot help us to explain any of the effects of anxiety group on multisensory emotional processing, it is not directly relevant to our research questions. However, out of interest about what was driving this effect, we followed up the ANCOVA analysis by running a Pearson correlation test to examine the relationship between autistic traits and overall Multisensory Gain scores, averaged across the two stimulus emotions.

The results revealed that Autistic Traits were positively associated with Multisensory Gain scores for the body-motion and voice stimuli, $r = .31$, $p = .020$. This suggests that individuals with higher levels of autistic traits benefited more from redundant sensory information when perceiving emotion from audiovisual body-motion and voice displays, relative to body-motion and voice emotion cues presented in isolation. This finding contrasts with previous research suggesting that autistic individuals integrate audiovisual information less efficiently than neurotypical individuals [34,35] and suggests that higher levels of autistic traits may not affect multisensory facilitation for biological motion and voice information in the same way that they affect integration of face and voice cues. Consistent with this argument, Feldman et al. [34] showed that the association between autism symptomatology and difficulties with audiovisual integration is most evident in studies using linguistic social stimuli, and to a lesser extent, faces.

The stimuli used in this body-motion and voice task were quite different to those used in previous studies of autism, as participants had to rely on more abstract cues from point-light displays showing body gestures, and voice stimuli which were filtered with high frequency noise so as to obscure the linguistic properties of the statements and force participants to rely only on tone of voice. In contrast to the face and voice stimuli used in previous tasks, for the body-motion and voice stimuli in this study, the actors in the displays were facing each-other, rather than facing the participant, which means that the participant is likely to have a adopted a third-person perspective of the interaction. It may be that individuals with higher levels of autistic traits are better able to incorporate both sources of information, audio and visual cues, when they are able to act as an objective observer of an interaction, rather than feeling that they are directly involved.

Table B13

ANCOVA results for Multisensory Gain Scores (Reaction Times) in the Undirected Attention Task

Stimulus Type	Source	*df*	*F*	*p*	Effect Size (partial η^2)
Faces-Voices	Emotion	1, 53	0.23	.632	<.01

	AQ Score	1, 53	1.05	.311	.02
	Anxiety Group	1, 53	<.01	.985	<.01
	Interaction terms	1, 53	≤2.08	≥.155	≤.04
Bodies-Voices	Emotion	1, 52	2.38	.129	.04
	AQ Score	1, 52	0.25	.622	.01
	Anxiety Group	1, 52	0.03	.866	<.01
	Interaction terms	1, 52	≤1.25	≥.268	≤.02
Circles-Tones	Emotion	1, 53	0.46	.503	.01
	AQ Score	1, 53	0.17	.684	<.01
	Anxiety Group	1, 53	0.92	.341	.02
	Interaction terms	1, 53	≤1.25	≥.268	≤.02

In order to rule out any explanation that anxiety group differences in multisensory performance on the Undirected Attention Task were being driven by differences in unimodal perceptual performance, we also carried out a further mixed factorial ANCOVA for each stimulus type to analyse accuracy and mean reaction times for unimodal trials in this task, with Anxiety Group (lower or higher) as a between-participant factor, Emotion (angry or happy) and Modality (audio or visual) as within-participant factors, and Autistic Traits (AQ Score) as a covariate. These results are reported for accuracy and reaction times in Tables B14 and B15, respectively. The results show that there were no differences between anxiety groups in terms of accuracy or mean reaction times for perception of emotion from unimodal displays on the Undirected Attention Task for any stimulus type.

Table B14

ANCOVA results for Accuracy on unimodal trials in the Undirected Attention Task

Stimulus Type	Source	*df*	*F*	*p*	Effect Size (partial η^2)
Faces-Voices	Modality	1, 52	11.83	.001	.19
	Emotion	1, 52	7.84	.007	.13
	AQ Score	1, 52	0.01	.917	<.01
	Anxiety Group	1, 52	0.13	.725	<.01
	Interaction terms	1, 52	≤2.51	≥.119	≤.05
Bodies-Voices	Modality	1, 53	3.44	.069	.06
	Emotion	1, 53	4.43	.040	.08
	AQ Score	1, 53	0.09	.766	<.01
	Anxiety Group	1, 53	0.92	.341	.02
	AQ Score x Modality	1, 53	3.98	.051	.07
	All other interaction terms	1, 53	≤1.83	≥.182	≤.03
Circles-Tones	Modality	1, 53	10.56	.002	.17
	Emotion	1, 53	6.52	.014	.11
	AQ Score	1, 53	1.85	.180	.03
	Anxiety Group	1, 53	<0.01	.953	<.01
	AQ Score x Modality x Emotion	1, 53	3.75	.058	.07
	All other interaction terms	1, 53	≤2.07	≥.156	≤.04

Table B15

ANCOVA results for Reaction Times on unimodal trials in the Undirected Attention Task

Stimulus Type	Source	*df*	*F*	*p*	Effect Size (partial η^2)
Faces-Voices	Modality	1, 53	79.20	<.001	.60
	Emotion	1, 53	0.07	.794	<.01
	AQ Score	1, 53	1.08	.303	.02
	Anxiety Group	1, 53	0.68	.414	.01
	Interaction terms	1, 53	≤2.29	≥.136	≤.04
Bodies-Voices	Modality	1, 53	0.80	.375	.02
	Emotion	1, 53	0.92	.343	.02
	AQ Score	1, 53	0.14	.710	<.01
	Anxiety Group	1, 53	0.03	.862	<.01
	Interaction terms	1, 53	≤1.29	≥.261	≤.02
Circles-Tones	Modality	1, 53	21.09	<.001	.29
	Emotion	1, 53	0.12	.727	<.01
	AQ Score	1, 53	0.43	.514	.01
	Anxiety Group	1, 53	<0.01	.974	<.01
	Modality x Emotion	1, 53	3.20	.079	.06
	All other interaction terms	1, 53	≤1.99	≥.164	≤.04

C. Data Pre-Processing and Assumption Testing

Normality. For all ANCOVA analyses, tests were run to check whether the data met the required parametric assumptions, and where assumptions were violated suitable corrections were applied. We plotted the distribution of the various dependent variables examined in our analyses for each combination of the groups of our within- and between-participant factors and tested for significant deviations from normality using the Shapiro-Wilk test. Where significant deviations from the normal distribution were evident from the results of the Shapiro-Wilk test, plotting the distribution of the dependent variable as a histogram and comparing to the normal distribution revealed that in the majority of cases, the visualised data distributions still approximated the normal distribution. In addition to this, it has been shown that ANOVA analysis is very robust to type I errors even when there are moderate violations of the assumption of normality [76], and especially when sample sizes are reasonably large and equal [77], so we deemed it was still acceptable to use a parametric ANCOVA approach.

Outliers. We identified univariate outliers for our dependent variables for each combination of the groups of our within- and between-participant factors as datapoints falling outside the median value by more than three times the median absolute deviation (MAD). This has been highlighted as a more robust measure on which to base the identification of outliers compared to using the mean and standard deviation, which are both statistics that are particularly sensitive to distortion by outliers [78]. Univariate outliers were dealt with by winsorizing (they were replaced with the next most extreme value for the relevant anxiety group in the dataset). We also checked for multivariate outliers by calculating the Mahalanobis Distances (MDs) for the combination of independent variables examined in each of the ANCOVA analyses and then comparing the calculated MDs to a chi-square distribution with the same degrees of freedom. Cases where the MD was significantly different from the values predicted by the chi-square distribution (at an α-level of .001) were considered multivariate outliers. However, after winsorizing univariate outliers, very few multivariate outliers were detected in any of the datasets for any of the analyses. A summary of the number of outliers in each of the datasets is given in Table C1.

Table C1

Summary of the number of univariate outliers winsorized in datasets contributing to each of the analyses

Dependent Variable for ANCOVA	Total Number of Observations	Stimulus Type	No. of Outliers
Crossmodal Interference Effects – Accuracy (Directed Attention Task)	Within subject factors: • Emotion [angry, happy] • Attended modality [audio, visual] Between subject factors: • Anxiety group [lower, higher] Total observations: 224	Faces-Voices Bodies-Voices Circles-Tones	Higher Anx: 3 No outliers identified Lower Anx: 3 Higher Anx: 7
Crossmodal Interference Effects – RTs (Directed Attention Task)	Within subject factors: • Emotion [angry, happy] • Attended modality [audio, visual] Between subject factors: • Anxiety group [lower, higher] Total observations: 224	Faces-Voices Bodies-Voices Circles-Tones	Lower Anx: 5 Higher Anx: 6 Lower Anx: 8 Higher Anx: 7 Lower Anx: 8 Higher Anx: 9
Unimodal Accuracy (Directed Attention Task)	Within subject factors: • Emotion [angry, happy]	Faces-Voices	Lower Anx: 3 Higher Anx: 3

	• Modality [audio, visual]	Bodies-Voices	Lower Anx: 1 Higher Anx: 4
	Between subject factors: • Anxiety group [lower, higher]	Circles-Tones	Lower Anx: 2 Higher Anx: 12
	Total observations: 224		
Unimodal RTs (Directed Attention Task)	Within subject factors: • Emotion [angry, happy] • Modality [audio, visual]	Faces-Voices	Lower Anx: 8 Higher Anx: 5
		Bodies-Voices	Lower Anx: 4 Higher Anx: 1
	Between subject factors: • Anxiety group [lower, higher]	Circles-Tones	Lower Anx: 15 Higher Anx: 13
	Total observations: 224		
Response Bias (Undirected Attention Task)	Within subject factors: • Visual emotion [angry, happy]	Faces-Voices	Lower Anx: 4
	Between subject factors: • Anxiety group [lower, higher]	Bodies-Voices	Lower Anx: 1
	Total observations: 112	Circles-Tones	No outliers identified
Multisensory Gain–Accuracy (Undirected Attention Task)	Within subject factors: • Emotion [angry, happy]	Faces-Voices	Lower Anx: 1 Higher Anx: 1
	Between subject factors: • Anxiety group [lower, higher]	Bodies-Voices	No outliers identified
	Total observations: 112	Circles-Tones	Lower Anx: 4 Higher Anx: 5

Multisensory Gain–RTs (Undirected Attention Task)	Within subject factors: • Emotion [angry, happy] Between subject factors: • Anxiety group [lower, higher] Total observations: 112	Faces-Voices	Lower Anx: 2 Higher Anx: 1
		Bodies-Voices	Lower Anx: 1 Higher Anx: 7 1 participant excluded from the lower anxiety group due to being a multivariate outlier
		Circles-Tones	Lower Anx: 6 Higher Anx: 4
Unimodal Accuracy (Undirected Attention Task)	Within subject factors: • Emotion [angry, happy] • Modality [audio, visual] Between subject factors: • Anxiety group [lower, higher] Total observations: 224	Faces-Voices	Higher Anx: 3 1 participant excluded from the higher anxiety group due to being a multivariate outlier
		Bodies-Voices	Lower Anx: 1
		Circles-Tones	Lower Anx: 5 Higher Anx: 11
Unimodal RTs (Undirected Attention Task)	Within subject factors: • Emotion [angry, happy] • Modality [audio, visual] Between subject factors: • Anxiety group [lower, higher] Total observations: 224	Faces-Voices	Lower Anx: 10 Higher Anx: 1
		Bodies-Voices	Lower Anx: 2 Higher Anx: 3
		Circles-Tones	Lower Anx: 8 Higher Anx: 7

Note: RTs = reaction times.

Sphericity. It was not necessary to check for sphericity as none of the within-participant factors that were examined had more than two levels.

Homogeneity of Regression Slopes. In order to check that the b-coefficients for the covariate (Autistic Traits) were equal across anxiety groups for all analyses, we tested for a significant interaction effect between Anxiety Group and the level of Autistic Traits prior to running the full factorial ANCOVA model. For all analyses we found no significant interaction between Anxiety Group and Autistic Traits, suggestive of homogeneity of regression slopes (i.e., regression slopes for the effects of Autistic Traits on the behavioural outcome measures did not differ across Anxiety Groups).

Linearity. We tested to see if the relation between the covariate (Autistic Traits) and the dependent variables was linear by plotting the relationships in a scatter matrix. This revealed that for most analyses there was no visible relationship at all between the level of autistic traits and the behavioural outcomes measured from the tasks, but even where relationships were very weak, or non-existent, these could be roughly approximated as being linear.

Approach to missing data. Only reaction times (RTs) for correct responses were entered into any of the analyses of RTs, and because of this, there were occasionally cases where participants had no RT data for a particular condition because they gave no correct responses. There was only data missing for this reason from the Crossmodal Interference analyses. Participants with missing data were excluded from the analysis (as shown in Table C2). More than half of participants had missing RT data for one of the stimulus emotion and/or modality conditions for the circle-tone stimuli, so this analysis cannot be considered as being very informative.

Table C2

Summary of cases excluded from the Crossmodal Interference Reaction Time analyses due to missing values

Stimulus Type	Cases Excluded
Faces-Voices	4 participants excluded (1 Lower Anx; 3 Higher Anx)
Bodies-Voices	2 participants excluded (1 Lower Anx; 1 Higher Anx)
Circles-Tones	31 participants excluded (15 Lower Anx; 16 Higher Anx)