A multimodal approach to emotion recognition ability in autism spectrum disorders

Catherine R.G. Jones, Andrew Pickles, Milena Falcaro, Anita J.S. Marsden, Francesca Happé, Sophie K. Scott, Disa Sauter, Jenifer Tregay, Rebecca J. Phillips, Gillian Baird, Emily Simonoff, and Tony Charman

Methods: We tested 99 adolescents (mean age 2010 The Authors. Journal of Child Psychology and Psychiatry We found no evidence of a fun-

Keywords: Autism spectrum disorder, emotion recognition, emotion processing, social communication, structural equation modelling.

Conflict of interest statement: No conflicts declared.

Autism spectrum disorders (ASDs), the common clinical term for the pervasive developmental disorders (American Psychiatric Association, 2000; World Health Organisation, 1993), are defined by social and communication difficulties, and deficits related to emotional processing are seen as a hallmark symptom. Basic emotion recognition is a fundamental ‘building block’ of more sophisticated emotional and social understanding and establishing the degree of deficit in ASD is important for ascertaining at what level social-emotional understanding begins to break down for these individuals.

The most widely used emotion recognition task requires identification of the emotional state of faces in a forced-choice paradigm, with the participant being tested on a selection from the six ‘basic’ emotions of happiness, sadness, fear, anger, surprise and disgust (Ekman & Friesen, 1976). Some studies report significantly poorer performance in ASD for either total score or individual emotions (e.g., Ashwin, Chapman, Colle, & Baron-Cohen, 2006; Baron-Cohen, Spitz, & Cross, 1993; Boraston, Blakemore, Chilvers, & Skuse, 2007; Corden, Chilvers, & Skuse, 2008; Pelphrey et al., 2002; Philip et al., in press; Wallace, Coleman, & Bailey, 2008; Wright et al., 2008) but others have found no difference between the ASD and comparison groups (e.g., Castelli, 2005; Grossman, Klin, Carter, & Volkmar, 2000). For those studies that do find a difference, the specific emotions that are problematic vary, with each of the core emotions except for happiness being identified in at least one study but with sadness and fear being the most commonly cited (Ashwin et al., 2006; Boraston et al., 2007; Corden et al., 2008; Pelphrey et al., 2002; Philip et al., in press; Wallace et al., 2008). One explanation for the inconsistent findings is likely to be sample size, with the studies cited here including between 11 and 39 participants with ASD.

In addition, small sample sizes limit exploration of the effect of IQ on performance and many previous studies include exclusively high or low IQ participants.

Research into emotion recognition ability in ASD has also been limited by over-focus on the visual modality, specifically the recognition of emotion in faces. Understanding emotional states in real life involves reading a variety of cues that include tone of voice, non-verbal vocalisations, vocal content, ges-

© 2010 The Authors. Journal of Child Psychology and Psychiatry
© 2010 Association for Child and Adolescent Mental Health.
Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main St, Malden, MA 02148, USA
diagnoses were made using information from the ADI-R (Lord, Rutter, & Le Couteur, 1994) and ADOS-G (Lord et al., 2000) as well as IQ, language and adaptive behaviour measures (see Baird et al., 2006; for details). The 26 participants assigned to the non-ASD group were adolescents who did not reach clinical criteria for an ASD (Baird et al., 2006). Rather, they had a range of primary ICD-10 diagnoses (16 mild mental retardation; 3 moderate mental retardation; 3 specific reading/ spelling disorder; 2 AD/HD; 1 expressive/receptive language disorder; 1 no diagnosis). The remaining non-ASD participants (n = 31) were recruited from local mainstream schools. Parent and teacher report confirmed that all were typically developing; none had a psychiatric or developmental diagnosis, a statement of special educational needs or were receiving medication. The Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003) was collected from parents of 25 of the 31 adolescents; no individual scored 15 or above, which is the cut-off for ASD. Measures of IQ were obtained using the Wechsler Abbreviated Scale of Intelligence (WASIUK; Wechsler, 1999) with the full-scale IQ of the total cohort ranging from 50 to 133. To explore the effects of IQ on performance, the participants were split into low IQ (full-scale IQ < 80) and high IQ (full scale IQ ≥ 80) subgroups. There was a significant difference in IQ between the high IQ subgroups with and without ASD, with a mean full-scale IQ of 104.0 (SD = 11.8) in the non-ASD group and 96.9 (SD = 10.0) in the ASD group (t(91) = 3.07; p < .01). Achieving balance through pairwise and group-wise matching would have lost participants and thus power. Instead, we matched the distributions by weighting the ASD subjects (formally by the ratio of the non-ASD to ASD kernel density estimates of the within-group IQ distributions) and undertaking a weighted analysis in Mplus (Muthén & Muthén, 1998–2007). While it is not possible to use standard likelihood ratio tests with weighted data, testing using appropriate Wald tests remains straightforward. All estimates, confidence intervals and test statistics reported took account of this weighting. Following the weighting procedure, there were no group differences between the ASD and non-ASD participants for age or IQ (t-test, all p > .10; see Table 1).

The study was approved by the South East Research Ethics Committee (05/MRE01/67) and informed consent was obtained from all participants.

**Tasks**

All tasks were programmed in Matlab v6.5 (Mathworks Inc., Sherborn, MA) using Cogent 2000 (Wellcome Department of Imaging Neuroscience, UCL Institute of Neurology, London, UK; http://vislab.ucl.ac.uk/Cogent/) and presented on a Hewlett-Packard laptop computer with a 15” LCD display screen. For the verbal tasks, stimuli were delivered binaurally through headphones (Sennheiser HD 280 pro).

Emotion recognition from facial cues. Facial expressions of emotion task (FE). This task used faces from the Ekman–Friesen test of affect recognition (Ekman & Friesen, 1976). The stimuli (.jpg files) were black and white halftone photographs of male and female faces expressing one of six ‘basic’ emotions.
Verbal vocal expressions of emotion task (V-VE) (i) 2009; Sauter & Scott, 2007). Press) as well as used in previous studies (Garrido et al., have been validated in typical adults (Sauter et al., in press). The stimuli used stimuli designed by SS and DS (Sauter, 2006; Sauter, Calder, Eisner, & Scott, in press). The stimuli (.wav files) were vocal recordings of four actors (2 male; 2 female) expressing one of the six basic emotions verbally. The verbal content was three-digit numbers (e.g., ‘one-hundred and thirty-seven’, ‘five-hundred and forty-five’). Stimulus length was on average 1.84 s (SD .46 s). A total of 30 stimuli were presented, 5 of each emotion, in one of six pseudo-randomised orders.

Non-verbal vocal expressions of emotion task (NV-VE). The stimuli (.wav files) were vocal recordings of four actors (2 male; 2 female) expressing one of the six basic emotions non-verbally. Examples of the non-verbal stimuli include crying vocalisations for sadness and laughing vocalisations for happiness. Stimulus length was on average 1.01 s (SD .52 s). A total of 30 stimuli were presented, 5 of each emotion, in one of six pseudo-randomised orders.

For all tasks, the inter-response interval was 1000 ms.

Design and procedure

Each participant completed the tasks in a random order over two days of testing (interspersed with other tasks). The lag between the two testing sessions averaged at 29 days (SD: 36 days). The laptop was placed directly in front of the participant and the laptop. A laminated A4 response sheet was placed on the desk between the participant and the laptop. The sheet was divided into a 3 x 2 grid, with each grid square containing one of the response options (‘Happiness’, ‘Sadness’, ‘Anger’, ‘Fear’, ‘Surprise’, ‘Disgust’). Before the task began the participant was asked to read the six emotions aloud. If a participant struggled to read the emotions then cartoon drawings of the emotions were added to the response sheet and the examiner reiterated the six response options on each trial as necessary. For each task, participants were told that they were going to see some faces/hear some voices. They were instructed to ‘decide how the person is feeling’ and to choose the word from the response sheet that ‘best describes how that person is feeling’. Participants were allowed to give their answer verbally or point to the word, and the examiner would then input the response using one of six labelled keys on the keyboard. The FE task also included six practice trials prior to the task, to ensure that the task was understood.

Emotion recognition from vocal cues. These tasks used stimuli designed by SS and DS (Sauter, 2006; Sauter, Calder, Eisner, & Scott, in press). The stimuli have been validated in typical adults (Sauter et al., in press) as well as used in previous studies (Garrido et al., 2009; Sauter & Scott, 2007).

(i) Verbal vocal expressions of emotion task (V-VE). The stimuli (.wav files) were vocal recordings of four actors (2 male; 2 female) expressing one of the six basic emotions verbally. The verbal content was three-digit numbers (e.g., ‘one-hundred and thirty-seven’, ‘five-hundred and forty-five’). Stimulus length was on average 1.84 s (SD .46 s). A total of 30 stimuli were presented, 5 of each emotion, in one of six pseudo-randomised orders.

(ii) Non-verbal vocal expressions of emotion task (NV-VE). The stimuli (.wav files) were vocal recordings of four actors (2 male; 2 female) expressing one of the six basic emotions non-verbally. Examples of the non-verbal stimuli include crying vocalisations for sadness and laughing vocalisations for happiness. Stimulus length was on average 1.01 s (SD .52 s). A total of 30 stimuli were presented, 5 of each emotion, in one of six pseudo-randomised orders.

Data analysis

The tasks involved six emotions and three modalities we made use of structural equation models that modelled our theory and design and postulated a more parsimonious set of latent traits that allowed for correlated measurement error. We considered that the tasks measured six correlated emotion-specific recognition abilities (latent traits) for each participant. The participant characteristics of diagnosis and IQ group were allowed to influence item responses only through mean differences in these six traits (the reference category of participants being set to mean zero on each trait with all item thresholds freely estimated and thus able to vary in difficulty). We also considered a model with just a single general emotion recognition factor (i.e., encompassing all emotion conditions); this offers a more powerful test of group differences in gross recognition ability in circumstances where emotion-specific traits are highly correlated.

Since the tasks were of distinct types, notably distinguished by stimulus modality, we also expected correlation in ability by modality. We therefore considered two models. The first, a correlated trait-correlated uniqueness model (CTCU; see Marsh, 1989), modelled the six emotion-specific recognition abilities (traits), as outlined above, and allowed 15 free correlations among the responses within each of the three modalities (see Figure 3). The second, a multitrait-multimethod structure model (MTMM; see, e.g., Campbell & Fiske, 1959, and Loehlin, 2004), was more restrictive in postulating that a participant’s expected performance on a task was the sum of an emotion-specific ability (trait) and a stimulus modality-specific (method) ability. Both trait and stimulus modality factors can be linked to

Table 1 Mean age, verbal IQ, performance IQ and full-scale IQ (SD in brackets) for the non-ASD and ASD groups. The recalculated weighted mean IQ for the ASD groups (ASD-WT) is also shown. Data for all cases, low IQ (FSIQ < 80) and high IQ (FSIQ ≥ 80) shown separately

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>M:F</th>
<th>Age</th>
<th>VIQ</th>
<th>PIQ</th>
<th>FSIQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low IQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-ASD</td>
<td>23</td>
<td>20:3</td>
<td>15:5 (4.1)</td>
<td>65.8 (8.6)</td>
<td>68.5 (9.5)</td>
<td>64.5 (8.4)</td>
</tr>
<tr>
<td>ASD</td>
<td>40</td>
<td>35:5</td>
<td>15:6 (5.0)</td>
<td>64.7 (9.9)</td>
<td>73.1 (13.5)</td>
<td>66.4 (9.5)</td>
</tr>
<tr>
<td>ASD-WT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High IQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-ASD</td>
<td>34</td>
<td>34:0</td>
<td>15:7 (6.9)</td>
<td>100.1 (12.5)</td>
<td>107.1 (11.0)</td>
<td>104.0 (11.8)</td>
</tr>
<tr>
<td>ASD</td>
<td>59</td>
<td>55:4</td>
<td>15:6 (6.0)</td>
<td>92.2 (12.9)</td>
<td>102.4 (10.4)</td>
<td>96.9 (10.0)</td>
</tr>
<tr>
<td>ASD-WT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(happiness, sadness, anger, fear, surprise, disgust). Each stimulus was displayed on the screen until the participant’s response had been input by the examiner. A total of 60 faces were presented, 10 of each emotion, in the same order as the original Ekman–Friesen test.

Structural equation modelling. Data were analysed using structural equation modelling in MPlus 5.2. With 18 different measures a method for dealing with the problems of multiple testing was essential. Since the tasks involved six emotions and three modalities we made use of structural equation models that recognised our theory and design and postulated a more parsimonious set of latent traits that allowed for correlated measurement error. We considered that the tasks measured six (correlated) emotion-specific recognition abilities (latent traits) for each participant. The participant characteristics of diagnosis and IQ group were allowed to influence item responses only through mean differences in these six traits (the reference category of participants being set to mean zero on each trait with all item thresholds freely estimated and thus able to vary in difficulty). We also considered a model with just a single general emotion recognition factor (i.e., encompassing all emotion conditions); this offers a more powerful test of group differences in gross recognition ability in circumstances where emotion-specific traits are highly correlated.

Since the tasks were of distinct types, notably distinguished by stimulus modality, we also expected correlation in ability by modality. We therefore considered two models. The first, a correlated trait-correlated uniqueness model (CTCU; see Marsh, 1989), modelled the six emotion-specific recognition abilities (traits), as outlined above, and allowed 15 free correlations among the responses within each of the three modalities (see Figure 3). The second, a multitrait-multimethod structure model (MTMM; see, e.g., Campbell & Fiske, 1959, and Loehlin, 2004), was more restrictive in postulating that a participant’s expected performance on a task was the sum of an emotion-specific ability (trait) and a stimulus modality-specific (method) ability. Both trait and stimulus modality factors can be linked to
explanatory variables. The MTMM model with positive method factor loadings assumes consistency of differences across modalities, such that if one emotion is easier under one modality then all emotions should be. This offered a framework within which systematic differences across groups by modality (e.g., the ASD group being better with an auditory stimulus) could be tested. However, empirical applications of the MTMM model often fail to converge, or converge on improper solutions (Marsh, 1989).

Since the scores on each task were ordinal and on some tasks some participants performed at or close to ceiling, the analysis used WLSMV (weighted least squares mean and variance adjusted) in Mplus. Reasonable fitting models give Comparative Fit Index (CFI) values of .95 or larger and Root Mean Square Errors (RMSE) of .08 or less. The WLSMV estimator precludes the use of the standard likelihood ratio comparison of models. Instead, models are compared by means of adjusted Wald tests.

SEM-based findings were supported by single and multivariable regression analyses of total and subtotal scores, the latter estimated using a generalized estimating equation (GEE) with an unstructured covariance matrix estimated in Stata 11 (StataCorp, 2009).

Results

**Emotion recognition abilities in ASD vs. non-ASD participants**

The mean emotion recognition scores for the three tasks are shown in Figure 1, plotted as a diagnostic group comparison, and in Figure 2, plotted as an IQ group comparison. For technical reasons or time constraints, 4 participants were not administered the FE task, 4 were not administered the V-VE task, and 3 were not administered the NV-VE task. All participants completed at least one task, so no participants have been excluded from the illustrative tables and figures or the SEM models.

Data from the tasks were ordinal and censored; no simple transformation yielded single task scores suitable for continuous variable methods.

**MTMM model.** Regardless of parameterisation (Marsh, 1989) this model failed to give a positive definitive residual covariance matrix and thus did not provide a model suitable as a basis for inferring group differences. Although more complex method effects are possible, simple additive method components of variance would be expected to result in positively correlated residual errors. Contrary to this expectation, the free correlations estimated among the 45 within-method measurement errors from the CTCU models (described in the next section) were generally small, ranging from -.37 (FE happy with disgust) to .45 (V-VE anger with disgust) with averages of .019 for FE, .113 for V-VE and .053 for NV-VE. Multivariable regression analysis (using GEE) of the three total scores, obtained from summing across the six emotions within each method, gave no evidence for differences in performance either for diagnosis by IQ interaction (Adjusted Wald $\chi^2(3) = 4.05, p = .26$) or for diagnostic group (Adjusted Wald $\chi^2(3) = 2.85, p = .42$) from a model without the diagnosis by IQ interaction. We therefore focused upon the traits (see below), treating the correlations of shared method effects as a nuisance.

Figure 1 Mean scores for each emotion on the three tasks (± SE) compared by diagnostic group: (a) Facial expressions (FE) (b) Verbal vocal expressions (V-VE) (c) Non-verbal vocal expressions (NV-VE) H = Happy; S = Sad; F = Fear; A = Anger; Su = Surprise; D = Disgust

© 2010 The Authors
Figure 3 shows parameter estimates from the CTCU model. As is often the case (see Brown, 2006), although the $\chi^2$ test of model fit indicated imperfect fit ($\chi^2(37) = 65.17$, $p = .003$), both the CFI (.96) and the RMSE (.07) criteria suggested the fit to be satisfactory. As expected and shown on the figure, the standardized loadings for each latent emotion recognition trait factor were all positive. The trait factors explained on average half of the task score variance (ranging between 18% for non-verbal anger and 74% for verbal surprise).

Extending the model to regress the emotion recognition trait factors on the participant groups (IQ, diagnosis and their interaction) showed no significant effect for the interaction (diagnosis by IQ) either individually (1 df Wald test $p$-values: happy .30, sad .16, fear .48, anger .53, surprise .75 and disgust .61) or altogether (adjusted Wald $\chi^2(6) = 2.75$, $p = .84$) (Model $\chi^2(46) = 62.05$, $p = .06$; CFI = .96; RMSE = .047). The conventional statistical approach is to find the most parsimonious model, so the model was refitted without the interaction. This main effects only model, (Model $\chi^2(49) = 74.65$, $p = .01$; CFI = .95; RMSE = .058), gave the coefficient estimates shown in Table 2. Consistently significant and large effects were evident for IQ group but not for diagnosis (the group difference for surprise was significant at the nominal critical $p$-value of .05 but not the Bonferroni corrected value of .008). A combined Wald test for the six diagnosis group differences was not significant ($\chi^2(6) = 9.93$, $p = .13$). The pattern of findings was similar when IQ was modelled as a continuous variable.

The estimated correlations among the traits are shown in Table 3. The smallest of these correlations is .62, suggesting that emotion recognition ability does not have a marked specificity in this sample. We therefore also tested a model in which the six emotion factors were replaced by a single factor for all 18 tasks. As in the six-factor model, correlations among the errors for measures using the same modality were allowed. This model, (Model $\chi^2(41) = 88.34$, $p < .001$; CFI = .94; RMSE: .086) fitted marginally less well that the six-factor solution. Consistent with the six-factor solution model, there was no evidence for differences for diagnosis by IQ interaction ($p > .2$). Further, in the absence of the interaction (Model $\chi^2(56) = 99.45$, $p = .0003$; CFI = .91; RMSE: .071) there was no significant effect of diagnosis (standardized difference = -.079, SE = .063, $p = .208$) but a significant effect of IQ (standardized IQ group difference = .726; SE = .107; $p < .001$). Thus, a one-factor solution, although providing a more powerful test for gross differences by diagnosis, did not alter the pattern of findings.

We used the non-central chi-square method to estimate the power to detect a medium-size group difference of .7 SD on the means of the six latent emotion variables. This gave 67% for happy, 84% for sad, 88% for anger, 87% for fear, 83% for surprise and 77% for disgust. For the single common factor model the power was 96% to detect a group difference of .7 SD and 76% for .5 SD.

CTCU model. Figure 3 shows parameter estimates from the CTCU model. As is often the case (see Brown, 2006), although the $\chi^2$ test of model fit indicated imperfect fit ($\chi^2(37) = 65.17$, $p = .003$), both the CFI (.96) and the RMSE (.07) criteria suggested the fit to be satisfactory. As expected and shown on the figure, the standardized loadings for each latent emotion recognition trait factor were all positive. The trait factors explained on average half of the task score variance (ranging between 18% for non-verbal anger and 74% for verbal surprise).
fusion matrices in Table 4 show the error pattern for the ASD and non-ASD participants. The two groups showed a remarkably similar pattern of errors across the emotions, with the emotion that is most consistently confused with the target being the same for both groups. The only emotion for which this pattern was not identical was for the non-verbal recognition of surprise, where the non-ASD group confused surprise most with disgust and the ASD group confused it most with happiness.

**Discussion**

In an exploration of emotion recognition abilities in ASD, we investigated 99 adolescents with ASD across the IQ spectrum and examined both visual (facial) and auditory (verbal and non-verbal voices) emotion recognition. Using a structural equation modelling (SEM) approach that enabled us to treat emotion recognition ability as a multimodal construct, we found no evidence of a fundamental impairment in emotion recognition ability in adolescents with ASD.

**Emotion recognition ability in ASD**

The SEM approach allowed emotion recognition ability to be modelled as a multimodal construct. By taking a composite approach, we argue that our model better encapsulates the ability to recognise emotional states than the more familiar unimodal approach taken in previous studies in the ASD field. Although we were unable to estimate explicit modality factors as part of an MTMM model, no differences by diagnosis in modality-specific subtotals were found. The weakness of the method effect is congruent with the hypothesis of a general ‘emotion processor’ (e.g., Borod et al., 2000) and reflects data from patients with subcortical lesions who show deficits in recognising specific emotions in the face and voice (Calder, Keane, Manes, Antoun, & Young, 2000).
suggestive of a common neural focus for emotion recognition regardless of modality.

This study did not take a developmental approach, so we cannot discount that an early-years difficulty with recognizing emotions, with associated developmental ramifications, is later compensated for. However, we can conclude that the broader social communication difficulties in our adolescent sample with ASD do not stem from a specific perceptual emotion recognition deficit. This finding was replicated in a one-factor model, which represented emotion recognition ability as a singular multimodal and multi-emotion construct. However, this model does not best reflect neuropsychological constructs of emotion recognition ability that theorize and demonstrate emotion-specific impairments (e.g., Boraston et al., 2007; Calder et al., 2000, 2004; Corden et al., 2008) and, as such, we favour the six-factor solution. The results appear at odds with the

Table 4 a–c Confusion matrix for (a) Facial expressions, (b) Verbal vocal expressions, (c) Non-verbal vocal expressions scores for non-ASD and ASD groups. Columns show each of the six target emotions and the rows show the percentage of responses that were given for the correct answer (in bold) and the five alternative response options. The final column shows the total percentage of answers for each response option.

<table>
<thead>
<tr>
<th></th>
<th>Happy</th>
<th>Sad</th>
<th>Fear</th>
<th>Anger</th>
<th>Surprise</th>
<th>Disgust</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(a) FE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-ASD</td>
<td>97.5</td>
<td>0.7</td>
<td>0.9</td>
<td>0.2</td>
<td>1.1</td>
<td>0.4</td>
</tr>
<tr>
<td>Sad</td>
<td>0.2</td>
<td>68.9</td>
<td>2.9</td>
<td>2.9</td>
<td>0.5</td>
<td>2.7</td>
</tr>
<tr>
<td>Fear</td>
<td>0.2</td>
<td>13.8</td>
<td>54.9</td>
<td>4.3</td>
<td>8.4</td>
<td>3.6</td>
</tr>
<tr>
<td>Anger</td>
<td>0.2</td>
<td>2.7</td>
<td>8.4</td>
<td>74.4</td>
<td>0.5</td>
<td>38.7</td>
</tr>
<tr>
<td>Surprise</td>
<td>2.0</td>
<td>4.7</td>
<td>27.1</td>
<td>4.4</td>
<td>87.6</td>
<td>0.9</td>
</tr>
<tr>
<td>Disgust</td>
<td>0.0</td>
<td>9.1</td>
<td>5.8</td>
<td>14.0</td>
<td>1.8</td>
<td>53.6</td>
</tr>
<tr>
<td>ASD</td>
<td>97.8</td>
<td>1.2</td>
<td>3.0</td>
<td>1.3</td>
<td>1.9</td>
<td>0.6</td>
</tr>
<tr>
<td>Happy</td>
<td>0.2</td>
<td>70.9</td>
<td>3.4</td>
<td>6.1</td>
<td>1.6</td>
<td>4.5</td>
</tr>
<tr>
<td>Sad</td>
<td>0.2</td>
<td>13.3</td>
<td>56.0</td>
<td>4.5</td>
<td>12.1</td>
<td>2.8</td>
</tr>
<tr>
<td>Fear</td>
<td>0.2</td>
<td>1.1</td>
<td>4.4</td>
<td>7.4</td>
<td>68.4</td>
<td>0.6</td>
</tr>
<tr>
<td>Anger</td>
<td>1.3</td>
<td>3.6</td>
<td>23.8</td>
<td>4.8</td>
<td>81.6</td>
<td>1.4</td>
</tr>
<tr>
<td>Surprise</td>
<td>0.3</td>
<td>8.5</td>
<td>6.4</td>
<td>14.8</td>
<td>2.1</td>
<td>44.8</td>
</tr>
<tr>
<td><strong>(b) V-VE:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-ASD</td>
<td>76.1</td>
<td>1.4</td>
<td>2.1</td>
<td>4.6</td>
<td>10.7</td>
<td>6.4</td>
</tr>
<tr>
<td>Sad</td>
<td>1.4</td>
<td>82.5</td>
<td>23.6</td>
<td>0.4</td>
<td>3.6</td>
<td>12.5</td>
</tr>
<tr>
<td>Fear</td>
<td>1.8</td>
<td>7.5</td>
<td>56.4</td>
<td>3.9</td>
<td>2.9</td>
<td>5.4</td>
</tr>
<tr>
<td>Anger</td>
<td>5.7</td>
<td>0.7</td>
<td>3.2</td>
<td>80.4</td>
<td>2.1</td>
<td>8.6</td>
</tr>
<tr>
<td>Surprise</td>
<td>10.0</td>
<td>2.9</td>
<td>10.0</td>
<td>3.9</td>
<td>75.4</td>
<td>12.1</td>
</tr>
<tr>
<td>Disgust</td>
<td>5.0</td>
<td>5.0</td>
<td>4.6</td>
<td>6.8</td>
<td>5.4</td>
<td>55.0</td>
</tr>
<tr>
<td>ASD</td>
<td>71.3</td>
<td>2.3</td>
<td>6.0</td>
<td>7.9</td>
<td>19.4</td>
<td>10.4</td>
</tr>
<tr>
<td>Happy</td>
<td>2.3</td>
<td>84.6</td>
<td>27.9</td>
<td>1.7</td>
<td>1.0</td>
<td>15.0</td>
</tr>
<tr>
<td>Sad</td>
<td>1.9</td>
<td>7.1</td>
<td>54.6</td>
<td>0.4</td>
<td>4.6</td>
<td>6.3</td>
</tr>
<tr>
<td>Fear</td>
<td>5.4</td>
<td>0.2</td>
<td>1.9</td>
<td>79.2</td>
<td>1.5</td>
<td>6.9</td>
</tr>
<tr>
<td>Anger</td>
<td>12.3</td>
<td>1.9</td>
<td>7.7</td>
<td>2.9</td>
<td>67.5</td>
<td>12.3</td>
</tr>
<tr>
<td>Surprise</td>
<td>6.7</td>
<td>4.0</td>
<td>1.9</td>
<td>7.9</td>
<td>5.8</td>
<td>49.2</td>
</tr>
<tr>
<td><strong>(c) NV-VE:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-ASD</td>
<td>90.5</td>
<td>1.1</td>
<td>1.5</td>
<td>0.0</td>
<td>4.0</td>
<td>0.7</td>
</tr>
<tr>
<td>Sad</td>
<td>3.3</td>
<td>86.9</td>
<td>6.9</td>
<td>1.5</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>Fear</td>
<td>0.7</td>
<td>2.5</td>
<td>77.5</td>
<td>6.5</td>
<td>4.4</td>
<td>4.7</td>
</tr>
<tr>
<td>Anger</td>
<td>0.4</td>
<td>2.9</td>
<td>3.6</td>
<td>88.0</td>
<td>0.7</td>
<td>2.9</td>
</tr>
<tr>
<td>Surprise</td>
<td>4.7</td>
<td>2.9</td>
<td>8.0</td>
<td>1.1</td>
<td>85.1</td>
<td>2.2</td>
</tr>
<tr>
<td>Disgust</td>
<td>1.0</td>
<td>3.6</td>
<td>2.2</td>
<td>2.9</td>
<td>4.7</td>
<td>88.7</td>
</tr>
<tr>
<td>ASD</td>
<td>86.6</td>
<td>1.0</td>
<td>2.5</td>
<td>0.4</td>
<td>7.8</td>
<td>0.0</td>
</tr>
<tr>
<td>Happy</td>
<td>3.1</td>
<td>88.0</td>
<td>3.7</td>
<td>1.4</td>
<td>0.8</td>
<td>1.6</td>
</tr>
<tr>
<td>Sad</td>
<td>0.6</td>
<td>3.7</td>
<td>79.0</td>
<td>8.2</td>
<td>7.6</td>
<td>2.9</td>
</tr>
<tr>
<td>Fear</td>
<td>0.0</td>
<td>0.8</td>
<td>2.3</td>
<td>84.5</td>
<td>0.4</td>
<td>3.3</td>
</tr>
<tr>
<td>Anger</td>
<td>8.2</td>
<td>2.3</td>
<td>10.5</td>
<td>2.3</td>
<td>77.1</td>
<td>1.2</td>
</tr>
<tr>
<td>Surprise</td>
<td>1.4</td>
<td>4.1</td>
<td>2.1</td>
<td>2.9</td>
<td>6.0</td>
<td>90.9</td>
</tr>
</tbody>
</table>

2000; Calder, Keane, Lawrence, & Manes, 2004), suggestive of a common neural focus for emotion recognition regardless of modality.

This study did not take a developmental approach, so we cannot discount that an early-years difficulty with recognising emotions, with associated developmental ramifications, is later compensated for. However, we can conclude that the broader social communication difficulties in our adolescent sample with ASD do not stem from a specific perceptual emotion recognition deficit. This finding was replicated in a one-factor model, which represented emotion recognition ability as a singular multimodal and multi-emotion construct. However, this model does not best reflect neuropsychological constructs of emotion recognition ability that theorize and demonstrate emotion-specific impairments (e.g., Boraston et al., 2007; Calder et al., 2000, 2004; Corden et al., 2008) and, as such, we favour the six-factor solution. The results appear at odds with the
majority of face emotion recognition studies, which have found evidence of a deficit in ASD [e.g., Ashwin et al., 2006; Boraston et al., 2007; Corden et al., 2008; Wallace et al., 2008; Wright et al., 2008] (the data on vocal emotion recognition are too sparse to provide a true precedent). However, our study has used a methodologically sound approach that incorporates the largest sample tested to date, the full range of testable IQ and a narrow age-range. It should also be noted that emotion-specific deficits in just one or two emotions have often been reported in the absence of a global (across emotion) deficit, which does not indicate a fundamental emotion recognition dysfunction [e.g., Boraston et al., 2007; Corden et al., 2008; Wright et al., 2008] and the specific emotions that cause difficulty have been variable across studies. Further, studies using different methodologies have also indicated that emotion recognition may be intact in ASD [e.g., Buitelaar, van der Wees, Swaab-Barneveld, & van der Gaag, 1999; Loveland et al., 1997; Ozonoff, Pennington, & Rogers, 1990].

By dividing our groups into those of low (FSIQ < 80) and high (FSIQ ≥ 80) IQ we were also able to investigate the effects of IQ on performance. Participants with a low IQ had significantly worse recognition ability for all emotions, regardless of whether they had an ASD or not, which complements earlier studies demonstrating the importance of IQ on emotion recognition ability [Buitelaar et al., 1999; Loveland et al., 1997]. A more recent study reported that individuals with and without ASD and with a low verbal mental age performed similarly on a measure of facial emotion discrimination [Loveland, Bachevalier, Pearson, & Lane, 2008]. However, for those with a higher verbal mental age, the ASD group performed significantly more poorly than those without a diagnosis. We found no evidence of a Group × IQ interaction in our own dataset. Comparing our data to a developmental study of performance on the Ekman-Friesen faces [Campbell et al., 2006], both low IQ groups scored at an equivalent level to typically developing 6–7-year-olds, whilst the high IQ groups performed similarly to typically developing 15–17-year-olds. In demonstrating that IQ rather than diagnosis is a discriminator of emotion recognition ability, our findings align with Wright et al. (2008) who found that age, sex and IQ accounted for 55% of the variance on a facial emotion recognition task (total score) amongst a group of children with and without ASD, whilst diagnosis only accounted for an additional 1%.

When discussing these results it should be acknowledged that the recognition of emotional cues can be measured in a variety of ways, and this necessarily limits the generalisability of the results. The current design collates performance from independently measured modes and creates a single index of multi-channel processing proficiency. This is distinct from a cross-modal approach, where the ability to simultaneously integrate information from different modalities produces a unitary response. Therefore, our results cannot speak to the ability of individuals with ASD to simultaneously integrate emotional information across modalities and it is possible that deficits might be apparent at the level of complex multi-sensory integration (although see Loveland et al., 1997 for a null result). Further, the stimuli in the current study represent salient exemplars of the chosen emotions. We cannot infer how the adolescents with ASD would have responded to more ambiguous stimuli, which exhibit less intense expressions of emotion; behavioural results have been mixed [e.g., Castelli, 2005; Humphreys, Minshew, Leonard, & Behrmann, 2007], whilst a recent imaging study suggests a lack of neural modulation to changes in the intensity of facial fear in ASD [Ashwin, Baron-Cohen, Wheelwright, O’Riordan, & Bullmore, 2007]. Specific to visual stimuli, there are additional literatures on dynamic facial stimuli [e.g., Gepner, Deruelle, & Grynfeltt, 2001], as well as emotional gesture and bodily expression [e.g., Atkinson, 2009; Hobson, 1986]. However, our null result in such a large sample using multimodal measuring of a very specific and fundamental type of cognitive emotion recognition provides an appropriate platform from which to explore and compare variation in performance across the nuances of emotion recognition.

When discussing the limitations of the study it is also important to recognise that the sample size, although large within the ASD literature, provides good power only for effects of medium size or larger. Although our fit indices were good and our confidence intervals acceptable, the impact of this study would be bolstered by replication in other large samples.

**A circumscribed difficulty in recognising surprise**

We found modest evidence of a circumscribed difficulty with the recognition of surprise. Baron-Cohen et al. (1993) have argued that, unlike emotions such as happiness or sadness that are typically understood at the ‘simple’ (produced by a situation) level, surprise is invariably a ‘cognitive’ emotion (produced by a belief), i.e., understanding the person’s belief is necessary to understand the emotion. The expression of surprise is further complex as, unlike the other core emotions, it can reflect both positive and negative valance or even be affectively neutral. This varied valence has led to discussion that surprise is not an emotion but is a cognitive state that ‘focuses on aspects of knowledge and belief rather than on affect per se’ [Ortony & Turner, 1990, p. 317]. Concurring with our results, Baron-Cohen et al. (1993) reported that low-functioning children with ASD have difficulty in recognising the facial expression of surprise but not happiness or sadness. The majority
of emotion recognition studies have not reported a specific deficit in surprise recognition, although our multimodal approach is better suited to detecting a deficit that is driven by cognitive factors that are not modality-specific. From a different perspective, the social motivation hypothesis (see Dawson, Webb, & McPartland, 2005) asserts that lack of interest in social cues in ASD reduces the attention paid to them, ultimately leading to individuals with a degraded sensitivity for social and emotional nuances. This has echoes of the cognitive avoidance of emotional situations seen in individuals with generalised social anxiety disorder, behaviour that is known to impact upon emotion recognition capabilities (see Montagne et al., 2006 for a discussion). It could be suggested that avoidance of, or inattentiveness to, social situations exponentially impacts the development of recognition of emotions that are (a) less common in everyday life and (b) dependent on complex social scenarios, including beliefs. We argue that, using these criteria, surprise is the emotion that would be most vulnerable. It is also worth acknowledging that many parents and care-givers learn that their child with ASD responds well to routine and finds surprising or unexpected events difficult to manage. This could lead to care-givers moderating their reactions of surprise and also avoiding exposing their child to events that evoke surprise, which speculatively may contribute to under-exposure of this particular emotional expression.

**Emotion recognition processing style is similar in ASD and non-ASD**

For all three tasks, analysis of the confusions individuals made when recognising emotions showed a remarkably similar pattern of errors across the groups with and without ASD. This is compatible with the hypothesis that similar styles of deductive reasoning are underpinning comparison of similar visual or auditory information in both groups. For example, both groups mistook facial expressions of fear for surprise around 25% of the time and mistook facial expressions of surprise for fear around 10% of the time; this is arguably due to orienting to and misinterpreting the wide eyes seen in both emotional states. Previous research into the processing style used during emotion recognition in ASD has been focused in the visual domain. Studies of the pattern of eye gaze during face scanning suggest atypical looking patterns in ASD (Hernandez et al., 2009; Pelphrey et al., 2002; although see van der Geest, Kemner, Verbaten, & van Engeland, 2002 for a null effect). However, this seems to be driven by the increased amount of time spent looking at non-core facial features or outside of the picture. When individuals with ASD are looking at core facial features they show the same basic pattern of looking preferences as controls, spending more time looking at the eyes and starting gaze search with the eyes (Hernandez et al., 2009; Pelphrey et al., 2002). This seems to suggest that individuals with ASD collect the same pertinent perceptual information as those without ASD, but that the quality of this information is degraded due to reduced looking time, which is perhaps driven by a lack of interest in faces, avoidance of faces or difficulties with attention. Of note, there is suggestion that the modification of gaze focus as a function of emotional expression is the same for those with and without ASD (Hernandez et al., 2009; van der Geest et al., 2002). Although we did not collect eye-tracking data and cannot comment directly on the looking patterns of our own participants, the distribution of confusion errors and the comparable level of performance across groups are congruent with the hypothesis that the visual scanning of core facial features in adolescents with ASD is similar to those without ASD.

**Summary**

In recognising ASD as a complex behavioural, perceptual and cognitive disorder it is important to isolate potential contributing factors to the expressed symptoms, which are invariably multifaceted. By targeting emotion recognition ability we were able to explore a relatively basic and fundamental contributor to social and communication competence. Further, the application of SEM enabled us to model emotion recognition ability more realistically as a composite of multi-channel processing proficiency. Our results suggest that basic emotion recognition ability should not be considered in isolation as the source of the social and communication difficulties observed in ASD.

**Acknowledgements**

This study was supported by the Medical Research Council (G0400065) and research at the Centre for Research in Autism and Education is supported by the Clothworkers’ Foundation and the Pears Foundation. We are grateful to the participants and their parents for generously giving their time to take part in this study.

**Correspondence to**

Tony Charman, Centre for Research in Autism and Education, Department of Psychology and Human Development, Institute of Education, 25 Woburn Square, London WC1H 0AA, UK; Email: t.charman@ioe.ac.uk
Key points

- Emotion recognition ability in ASD has primarily focused on emotion recognition of faces and results have been varied.
- Using a large sample \(n = 99\) of adolescents with ASD and matched controls, we modelled emotion recognition as a composite of visual (face) and vocal (verbal and non-verbal) emotion recognition ability.
- There was no evidence of a fundamental difficulty with emotion recognition in the adolescents with ASD, although a circumscribed difficulty with surprise was noted.
- IQ had a large and significant effect on performance, with higher IQ adolescents outperforming lower IQ adolescents.
- Clinically, this suggests that basic emotion recognition should not be considered a source of social and communication difficulties in adolescents with ASD.

References


Humphreys, K., Minshew, N., Leonard, G.L., & Behrmann, M. (2007). A fine-grained analysis of facial expression...
processing in high-functioning adults with autism. Neuropsychologia, 45, 685–695.

Manuscript accepted 2 July 2010