Emotion recognition from the eye region in children with and without Autism Spectrum Disorder in Arab and Scandinavian countries

Sanna Kuusikk-Gauffin1, Sherin Elsheikh1,2, Sven Bölte3,4, Manal Omar5, Geylan Riad6,7, Hanna Ebeling1,2, Arja Rautio8, Irma Moilanen1,7

1PEDEGO Research Unit, Child Psychiatry, Medical Faculty, University of Oulu, Oulu, Finland
2Child Psychiatry Unit, Abbassia Mental Hospital, Cairo, Egypt
3Center of Neurodevelopmental Disorders (KIND), Pediatric Neuropsychiatry Unit, Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden
4Child and Adolescent Psychiatry, Center of Psychiatry Research, Stockholm County Council, Stockholm, Sweden
5Institute of Postgraduate Childhood Studies, Ain Shams University, Cairo, Egypt
6Faculty of Arts, Psychology Department, Helwan University, Cairo, Egypt
7Clinic of Child Psychiatry, Oulu University Hospital, Oulu, Finland
8Arctic Health, Faculty of Medicine and Thule Institute, University of Oulu, Finland

*Corresponding author: sanna.kuusikkogauffin@gmail.com

Abstract

Background: Difficulties in facial emotion recognition (ER) skills are linked to autism spectrum disorder (ASD) in studies performed in Western and Eastern Asian countries. However, there is a paucity of research examining ER skills in Arab countries, where face-covering veils are more common than in Western countries.

Objective: Our aim was to examine basic ER and ER error patterns in Egyptian and Finnish children with and without ASD.

Method: We employed the eye-submodule of the Frankfurt Test and Training of Facial Affect Recognition (FEFA) and the Autism Spectrum Screening Questionnaire (ASSQ).

Results: Arab children with ASD (n = 34, M age = 8.6 years, FSIQ = 96.7) recognized correctly fewer emotions than did Scandinavian children with ASD (n = 32, M age = 12.5 years, FSIQ = 102.8) and Arab typically developing (TD) children (n = 34, M age = 10.3 years, FSIQ = 123.4) in general and specifically on surprise, disgust and neutral scales as well as on a blended emotion scale. Scandinavian children with ASD demonstrated a lower ability to recognize emotions in general and specifically happiness than did Scandinavian TD children. There were no differences between Arab and Scandinavian (n = 28, M age = 13.9 years) TD children in ER accuracy. We found country specific differences in ER error patterns in happiness, sadness and anger: Arab children interpreted these emotions more often as another emotion (happiness = sadness, sadness = anger, anger = sadness and surprise), whereas Scandinavian children interpreted happiness and sadness as neutral expression and anger as disgust. Arab children with ASD labeled sadness and anger in their ER error patterns more negatively than did Arab TD children, but there were no differences between Scandinavian children with ASD and TD in ER error patterns.

Conclusions: The differences between the Arab and Scandinavian children may reflect cultural differences in ER and ER error patterns.

Keywords: Autism; ASSQ; culture; emotion; eye region; face

Introduction

Emotion recognition (ER) skills, such as interpreting another person's facial or bodily expressions, tone of voice, or “reading between lines,” are important tools for navigating in everyday social situations. Previous meta-analyses of ER show that ER is more accurate within the same cultural group, compared to cross-cultural ER (1, 2). However, this in-group advantage appears to diminish in typically developing (TD) populations, when ER is made to rely primarily on the eye region (3, 4).

Fear and disgust recognition appear to be universal and primitive reactions toward external stimuli and not to be dependent on cognitive processing or
learned in the cultural environment (5-8), whereas cultural environment can affect the recognition of other basic emotions (happiness, neutral, surprise and sadness) more strongly (9, 10).

ASD is a lifelong developmental disorder characterized by impairments in social and communication skills as well as by a restricted, repetitive, stereotyped pattern of behavior (11). Difficulties in facial ER skills are linked to ASD in meta-analyses focused mainly in Western countries (12, 13) as well as in one previous cross-cultural study, including Western, Asian, and Middle-Eastern countries (14). Contrary findings also exist, indicating that children with ASD form a heterogeneous group and that full-facial ER deficits are not systematic (15, 16).

Only a few studies, however, have examined basic ER skills from the eye region in children and adolescents with ASD (3, 17, 18), and none of them have examined basic ER skills in children with ASD deriving from Arab countries. In Arab countries, wearing a face-covering veil, such as a niqab or burqa, is more common than in Western countries. Although the custom of wearing a face-covering scarf is becoming less frequent in Egypt, where a part of our study sample derives from, in the year 2013, 10% of Egyptians still thought that women should cover their faces in public (19). Face-covering clothing often limits the recognition of facial emotions to the eye region, which may be critical in children with ASD, who tend to minimize the time spent looking into another person’s eyes (20-25). It appears that detection of fear is impaired in the meta-analyses of ASD (13), and that the recognition of fear requires attention to the eye region (26-28), whereas happiness, disgust, and neutral, on the other hand, can be judged more easily from the mouth region (29-31).

The severity of social deficits in ASD was found to be associated with poor ER skills in some studies (32, 33), and some studies indicated that facial ER might be more dependent on the intellectual ability of children with ASD than in the case of TD children (34). ER differences between ASD and TD may increase in severity with age, suggesting that ER skills do not improve in a similar manner in children with ASD as in their TD peers (12, 18). In TD, facial ER skills improve throughout childhood in social interaction, and the recognition ability of some emotions, such as anger, still improves in adolescence (35).

There are only a limited number of cross-cultural studies examining the ER skills of individuals with ASD (14). Children with ASD in different cultures do differ in their perceptual style (36), neuropsychological abilities (37), behavioral problems, (38) and the severity of comorbidity symptoms (39), and thus they may also differ in various ways in their ER skills. However, it may also be noted that cross-cultural differences in ER could be small between ASD groups, as Fridenson-Hayo et al. (14) recently reported insignificant differences in ER skills between Swedish, British, and Israeli 5 to 9 year old children with ASD, measured with voice recordings, videos of facial and bodily expressions, and with emotional video scenarios.

Another way to study ER is to evaluate ER error patterns. Personal emotional experiences may prime an individual to perceive certain emotions whether the particular emotion is present or not (40). For example, aggressive boys may have a bias toward seeing anger and a threat in another person’s emotional expressions (41-43). There are also differences between collectivist (e.g., Asian, Arab) and individualistic (e.g., Western) cultures in ER error patterns: members of collectivist cultures may avoid seeing negative emotions for the sake of preserving social order, whereas in individualistic cultures negative emotions are more acceptable (44). However, it may be that this tendency is exhibited more clearly in collectivist TD populations than in persons with ASD, due to their deficits in social perception (22, 45). Children’s ER biases may, however, be more relevant in predicting social behavior than in their recognition accuracy (46).

There are only a few studies which have examined the misinterpretation of emotional expression from faces in ASD samples (47-49). In a study by Eack et al. (47), adults with ASD confused happy, sad, and angry faces more often with neutral expressions, compared to TD individuals. Jones et al. (48) reported that adolescents with ASD regularly mislabeled disgust as anger and fear as surprise. Wallace et al. (49) found that adolescents with ASD made similar errors to those made by TD adolescents by labeling sad faces as angry or fear and disgust as angry. In contrast to the study of Eack et al. (47), Jones (48) and Wallace (49) did not include neutral facial expression in their studies. There are no studies which have examined cultural differences in ER error patterns in ASD samples.

The aim of our study was to examine cross-cultural differences (Arab vs Scandinavian) in children with and without ASD in (1) basic ER from the eye region; (2) ER error pattern. On the basis of the literature, we hypothesized that (1) both Arab and Scandinavian children with ASD would perform lower in ER than their TD peers; (2) there are no cross-cultural differences between ASD and TD children in ER accuracy; (3) Arab TD children, representing mainly a collectivist culture, would tend to see emotions more positively in their ER error patterns than Scandinavian TD children, representing an individualistic culture, but there would be no
significant difference between Arab and Scandinavia children with ASD in the ER error patterns.

**Methods**

**Procedure and Participants**

**Arab**

A total of 34 Egyptian children with ASD (M age = 8.6, SD = 2.3, ranging 6.1 to 15.0 years; M Full Scale IQ (FSIQ) measured with Wechsler Intelligence Scale for Children, 3rd edition (WISC-III) = 96.7, SD = 19.3, ranging 71 to 140) (50) were recruited from the Child Psychiatric Outpatient Clinic of the Abbassia Mental Hospital, a tertiary referral public hospital, and from a private child psychiatric clinic in Cairo (37). In total, 34 Egyptian TD children (M age = 10.3, SD = 2.4, ranging 6.2 to 14.9 years; M FSIQ with WISC-III = 123.4, SD = 12.7, ranging 85 to 143) (50) were recruited from mainstream schools in Cairo, Egypt (37).

**Scandinavia**

A total of 32 Finnish children with ASD (M age 12.5, SD = 0.8, ranging 10.6 to 13.9 years; M FSIQ with WISC-III = 102.8, SD = 18.3, ranging 78 to 155) (50) were recruited from the community (51) and from the Child Psychiatric Outpatient Clinic of Oulu University Hospital (52). In total, 28 Finnish TD children (M age = 13.9, SD = 1.3, ranging 10.5 to 15.7 years) following the mainstream educational plan were recruited from mainstream schools in Oulu (53). We did not measure FSIQ of the Finnish TD children, but assumed that it would be within the normal range, as in Finland all children attending mainstream education most probably have FSIQ within the normal range: children with lower than average FSIQ are referred to special education.

A clinical best estimate according to the ICD-10 (11) was used to determine ASD diagnosis in both countries. The Autism Diagnostic Interview (54) and Autism Diagnostic Observation Schedule (55) were used to obtain standardized diagnostic information in these studies to corroborate diagnostic status. Participants’ background information, such as age, IQ, and autistic symptoms are reported in Table 1. Test environments were comparable in both countries: all children were tested at the clinic by health care professionals (psychologist or psychiatrist).

The study was approved by the Rights of Human Subjects in Scientific Research in the General Secretariat of Mental Health in Cairo, Egypt, and the Ethics Committee of the Northern Ostrobothnia Hospital District, Finland. The procedure was explained to all participants and their parents, and written consents were obtained from their parents.

**Measures**

**Parent-reported autistic symptoms**

ASSQ (56) is a 27-item inventory, measuring symptoms in the three main areas of ASD (i.e., social interaction; communication; and restricted and repetitive behavior) as well as motor discoordination (e.g., clumsiness) and other associated symptoms, such as motor and vocal tics. Items were rated on a three-point Likert-type scale (i.e., 0 = normal to 2 = definite abnormality) with scores ranging from 0 to 54 and a higher score reflecting the severity of autistic traits. The psychometric properties of ASSQ are excellent (57). In this study, we were specifically interested in the core ASD symptoms; and therefore, we formed a revised ASSQ (ASSQ-R) total score, excluding five items containing symptoms of motor discoordination and motor and vocal tics (excluded

---

**TABLE 1. Background information of the participants**

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Egyptian ASD</th>
<th>Finnish ASD</th>
<th>Egyptian TD</th>
<th>Finnish TD</th>
<th>ANOVAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys/girls</td>
<td>34/31</td>
<td>32/22</td>
<td>34/22</td>
<td>28/13</td>
<td>F(1.64)</td>
</tr>
<tr>
<td>M(SD)</td>
<td>Age</td>
<td>8.6 (2.3)</td>
<td>12.5 (8.6)</td>
<td>10.3 (2.4)</td>
<td>13.9 (1.3)</td>
</tr>
<tr>
<td>FSIQ</td>
<td>96.7 (19.3)</td>
<td>120.8 (18.3)</td>
<td>123.4 (12.7)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>VIQ</td>
<td>101.4 (20.4)</td>
<td>104.0 (24.0)</td>
<td>133.8 (12.0)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>PIQ</td>
<td>93.1 (18.5)</td>
<td>103.0 (17.3)</td>
<td>107.8 (13.2)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>ASSQ-R</td>
<td>19.6 (6.2)</td>
<td>19.2 (8.2)</td>
<td>8.2 (6.8)</td>
<td>0.8 (1.3)</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Note. ASD, autism spectrum disorder; ASSQ-R, The High-functioning Autism Spectrum Screening Questionnaire -revised; FSIQ, full scale IQ; n/a, not applicable; PIQ, performance IQ; TD, typical development; VIQ, verbal IQ

*p < 0.05; **p < 0.01; ***p < 0.001, two-tailed
items are 9, 10, 20, 21, and 27). The newly formed ASSQ-R includes 22 items, the total score ranging from 0 to 44. The internal consistency for ASSQ-R total score for different groups was: Egyptian children with ASD: $\alpha = .78$; Egyptian TD children: $\alpha = .85$; Finnish children with ASD: $\alpha = .86$ and Finnish TD children: $\alpha = .45$.

**ER test**

We evaluated participants’ ER skills with the eye-submodule of the FEFA (58), which contains 40 black and white photos of various eye-regions, belonging to people from different cultures. Each FEFA eye-submodule photo is shown separately on a computer screen with all six different lexically labeled response options. In addition to lexical choices, researchers in both countries also read response options to participants. The FEFA eye-submodule contains six photos of happiness, eight photos of neutral expressions, four of surprise, three of disgust, four of sadness, six of anger, and nine of blended emotions, which are blended so that they can be labeled correctly with two different emotions. Of the nine blended emotions, five are blended similarly between fear and surprise, and the other four as follows: happiness-surprise, happiness-neutral, sadness-neutral, and anger-disgust. Each answer is scored 0 = incorrect and 1 = correct (in blended emotions, there are two possible correct answers) (58). We reported the sum score for the FEFA total scale (i.e., sum of correct answers of all 40 photos), sum scores for each basic emotion scale separately (i.e., happiness, neutral, surprise, disgust, sadness, and anger), sum score for the blended emotions scale, sum score for the fear-surprise blended emotion scale, error pattern (i.e., frequency of different incorrect responses) of participants’ responses to different basic emotion scales, and at last, we reported how participants labeled five fear-surprise blended emotions, because we wanted to see which was the easier emotion to interpret, fear or surprise. Examples of FEFA eye-region photos are shown in Figure 1.

The FEFA was translated into Finnish (18) and for this study, also into Arabic. FEFA has good psychometric properties, such as re-test stability in TD samples (e.g., 58). The internal consistency for the FEFA total scale in our whole study sample was fair ($\alpha = 0.77$), and varied from poor to fair in individual emotion scales: happiness ($\alpha = 0.66$), neutral ($\alpha = 0.72$), surprise ($\alpha = 0.67$), disgust ($\alpha = 0.57$), sadness ($\alpha = 0.17$), and anger ($\alpha = 0.52$).
**Statistical analyses**

All statistical analyses were performed using the SPSS 23.0 statistical software program for Macintosh. Variables (i.e., age, FSIQ, FEFA scales) were not normally distributed; therefore, the data were log transformed in order to conduct parametric tests. We also found statistically significant age and FSIQ differences between the groups (Table 1) and therefore performed multivariate analysis of covariance (MANCOVA) on FEFA scales between ASD children (Egyptian and Finnish), controlling for significant age difference, and in the case of Egyptian children (ASD vs TD) for significant FSIQ and age difference (Table 1). As we did not collect FSIQ from Finnish TD children, we performed MANCOVAs separately for Finnish children (ASD vs TD) and for TD children (Egyptian vs Finnish), controlling only for age. 

Note: There were two outliers in our data based on FSIQ: One on Scandinavian ASD group with FSIQ of 155 and another in Arab TD group with FSIQ of 85. We run all the analyses with and without these two participants and all the significant differences in between groups comparisons remained the same except one: the significant PIQ difference between Arab and Scandinavian ASD children became non-significant after excluding these two participants. Further, these two participants did not stand out as outliers by their age or by their scores on ASSO-R or FEFA scales. Therefore, we decided to keep these two participants in our data analyses in order to increase statistical power.

We tested error patterns on basic emotion scales with Chi-square ($\chi^2$) and association between parent-reports in ASSO-R and FEFA scales with Partial Correlations ($r$), controlling for age. The effect size was tested using the partialEta-squared ($\eta^2$) statistic and internal consistency of parent-report measures using Cronbach’s Alpha ($\alpha$). All tests of statistical significance are reported in two-tailed form. To control multiple testing on FEFA scales in between-group comparisons, we used Bonferroni correction (0.05/8), resulting in a $p$-value < 0.006 as statistically significant. To control multiple testing on error pattern analyses, we used Bonferroni correction (0.05/6), resulting in a $p$-value < 0.008 to be considered as statistically significant.

**Results**

**Emotion recognition**

MANCOVA (FEFA total scale as dependent variable; country and diagnostic status as fixed factors; age as a covariate) revealed a significant main effect of both diagnostic status ($F(1,127) = 19.1, p < 0.001, \eta^2 = 0.13$), and country ($F(1,127) = 6.2, p = 0.014, \eta^2 = 0.05$) on the FEFA total scale. The interaction between country and diagnostic status was also significant ($F(1,127) = 13.2, p < 0.001, \eta^2 = 0.10$) on the FEFA total scale. The main effect of age ($F(1,127) = 0.8, p = ns, \eta^2 = 0.01$) was not significant on the FEFA total scale.

MANCOVA (FEFA basic emotion scales as dependent variables; country and diagnostic status as fixed factors; age as a covariate) revealed a significant main effect of diagnostic status (Wilk’s Lambda = 0.79, $F(6,117) = 5.2, p < 0.001, \eta^2 = 0.21$) on FEFA basic emotion scales. The interaction between country and diagnostic status was also significant (Wilk’s Lambda = 0.89, $F(6,117) = 2.5, p = 0.025, \eta^2 = 0.11$) on FEFA basic emotion scales. The main effects of country (Wilk’s Lambda = 0.90, $F(6,117) = 2.1, p = 0.060, \eta^2 = 0.10$) and age were not significant (Wilk’s Lambda = 0.93, $F(6,117) = 1.6, p = 0.158, \eta^2 = 0.08$) on FEFA basic emotion scales.

MANCOVA (FEFA blended scale as dependent variable; country and diagnostic status as fixed factors; age as a covariate) revealed a significant main effect of diagnostic status on the FEFA blended scale ($F(1,127) = 16.6, p < 0.001, \eta^2 = 0.12$). The interaction between country and diagnostic status was also significant ($F(1,127) = 9.6, p = 0.047, \eta^2 = 0.03$). The main effects of country ($F(1,127) = 7.4, p = 0.081, \eta^2 = 0.03$) and age ($F(1,127) = 3.1, p = 0.082, \eta^2 = 0.02$) were not significant.

**TABLE 2.** Group comparisons between country on FEFA Scales, corrected with age

<table>
<thead>
<tr>
<th></th>
<th>Egypt</th>
<th>Austria</th>
<th>Finland</th>
<th>Germany</th>
<th>Hungary</th>
<th>Spain</th>
<th>Italy</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEFA (Max score)</td>
<td>M(SE)/Adjusted mean</td>
<td>F(2,127)</td>
<td>M(SE)/Adjusted mean</td>
<td>F(2,127)</td>
<td>M(SE)/Adjusted mean</td>
<td>F(2,127)</td>
<td>M(SE)/Adjusted mean</td>
<td>F(2,127)</td>
</tr>
<tr>
<td>Total (40)</td>
<td>24.4(1.1)/3.2</td>
<td>27.0(1.1)/3.3</td>
<td>18.0</td>
<td>0.014</td>
<td>31.4(0.8)/3.5</td>
<td>30.7(0.9)/3.4</td>
<td>0.04</td>
<td>0.316</td>
</tr>
<tr>
<td>Happiness (6)</td>
<td>4.9(0.4)/1.7</td>
<td>3.6(0.4)/1.5</td>
<td>0.03</td>
<td>0.20</td>
<td>5.3(0.2)/1.8</td>
<td>5.0(0.2)/1.8</td>
<td>0.01</td>
<td>0.680</td>
</tr>
<tr>
<td>Neutral (8)</td>
<td>4.1(0.4)/1.5</td>
<td>5.6(0.4)/1.8</td>
<td>0.15</td>
<td>0.005</td>
<td>5.5(0.4)/1.8</td>
<td>6.2(0.3)/1.9</td>
<td>0.04</td>
<td>0.281</td>
</tr>
<tr>
<td>Surprise (4)</td>
<td>1.6(0.2)/0.8</td>
<td>2.1(0.2)/1.1</td>
<td>0.21</td>
<td>0.005</td>
<td>2.7(0.1)/1.3</td>
<td>2.7(0.3)/1.3</td>
<td>0.01</td>
<td>0.724</td>
</tr>
<tr>
<td>Disgust (3)</td>
<td>1.7(0.2)/0.9</td>
<td>2.2(0.2)/1.1</td>
<td>0.37</td>
<td>0.005</td>
<td>2.4(0.2)/1.2</td>
<td>2.4(0.2)/1.2</td>
<td>0.05</td>
<td>0.214</td>
</tr>
<tr>
<td>Sadness (4)</td>
<td>3.0(0.2)/1.4</td>
<td>3.2(0.2)/1.4</td>
<td>0.01</td>
<td>0.725</td>
<td>3.2(0.2)/1.4</td>
<td>2.9(0.2)/1.3</td>
<td>0.03</td>
<td>0.438</td>
</tr>
<tr>
<td>Anger (6)</td>
<td>4.2(0.3)/1.6</td>
<td>3.6(0.3)/1.5</td>
<td>0.02</td>
<td>0.513</td>
<td>4.7(0.3)/1.7</td>
<td>4.0(0.3)/1.6</td>
<td>0.03</td>
<td>0.412</td>
</tr>
<tr>
<td>Blended (9)</td>
<td>5.7(0.4)/1.8</td>
<td>7.5(1.3)/2.1</td>
<td>0.19</td>
<td>0.001</td>
<td>7.7(1.1)/2.2</td>
<td>8.3(0.8)/2.2</td>
<td>0.13</td>
<td>0.018</td>
</tr>
<tr>
<td>Fear-Surprise (5)</td>
<td>3.2(0.3)/1.3</td>
<td>4.4(0.3)/1.7</td>
<td>0.23</td>
<td>0.000</td>
<td>4.7(0.2)/1.7</td>
<td>4.9(0.2)/1.8</td>
<td>0.06</td>
<td>0.167</td>
</tr>
</tbody>
</table>

Note: ASD, autism spectrum disorder; TD, typical development. $\eta^2$ and p-values derived from adjusted means derived from the log transformation calculation

**p < 0.008; ***p < 0.001, two-tailed
Follow-up between-groups analyses

**ASD**: Egyptian children with ASD recognized correctly fewer emotions than did Finnish children with ASD on the FEFA total scale, FEFA basic emotion scales of surprise, disgust, and neutral as well as on the blended emotion scale and fear-surprise blended emotion scale, when controlling for age (Table 2).

**TD**: There were no statistically significant differences in ER accuracy between Egyptian and Finnish TD children on FEFA scales, when controlling for age (Table 2).

**Egypt**: Egyptian children with ASD recognized correctly fewer emotions than did Egyptian TD children on the FEFA total scale, FEFA basic emotion scales of surprise, disgust, and neutral as well as on the blended emotion scale and fear-surprise blended emotion scale, when controlling for age and FSIQ (Table 3).

**Finland**: Finnish children with ASD recognized correctly fewer emotions than did Finnish TD children on the FEFA total scale and the FEFA basic emotion scale of happiness, when controlling for age (Table 3).

### Error patterns

Egyptian and Finnish children with and without ASD differed statistically significantly in their ER error patterns in happiness, sadness, and anger (Tables 4 and 5). Egyptian children with ASD and TD differed statistically significantly in their ER error patterns in sadness and anger. Finnish children with ASD and TD did not differ statistically significantly in their ER error patterns. Below, we report the error pattern findings in more detail.

**Happiness**: Both Egyptian children with ASD and TD children labeled happiness in their incorrect answers most often as sadness (46.8% vs 52%), whereas both Finnish children with ASD and TD children labeled happiness in their incorrect answers most often as neutral (42% vs 59.3%) (Egyptian ASD vs Finnish ASD, $\chi^2 = 71.3$, df = 5, $p < 0.001$; Egyptian TD vs Finnish TD, $\chi^2 = 65.8$, df = 5, $p < 0.001$).

**Neutral**: All children labeled neutral most often as sadness in their incorrect answers (Egyptian ASD = 64.8%, Finnish ASD = 41.8%, Egyptian TD = 70.9%, Finnish TD = 62%).

**Surprise**: All children labeled surprise most often as fear in their incorrect answers (Egyptian ASD = 40.6%, Finnish ASD = 62.3%, Egyptian TD = 45%, Finnish TD = 58.1%).

**Disgust**: All children labeled disgust most often as anger in their incorrect answers (Egyptian ASD = 50.8%, Finnish ASD = 53.3%, Egyptian TD = 71.8%, Finnish TD = 78.6%).

**Sadness**: Egyptian children with ASD labeled sadness in their incorrect answers most often as anger (50%), whereas Finnish children with ASD labeled sadness most often as a neutral expression (47.8%) ($\chi^2 = 60.6$, df = 5, $p < 0.001$). Egyptian TD children confused sadness most often with surprise (38.1%), whereas Finnish TD children confused sadness most often with neutral (43.8%) ($\chi^2 = 28.7$, df = 5, $p < 0.001$). Egyptian children with ASD labeled sadness more negatively compared to Egyptian TD children ($\chi^2 = 54.7$, df = 5, $p < 0.001$). Finnish children did not differ in the error pattern of sadness ($\chi^2 = 4.5$, df = 3, $p = 0.212$).

**Anger**: Egyptian children with ASD labeled anger in their incorrect answers most often as sadness (44.9%), whereas Finnish children with ASD labeled anger in their incorrect answers most often as disgust (48.7%) ($\chi^2 = 79.2$, df = 5, $p < 0.001$). Egyptian TD

### TABLE 3. Group comparisons within country on FEFA scales

<table>
<thead>
<tr>
<th>FEFA (Max score)</th>
<th>ASD</th>
<th>TD</th>
<th>$\eta^2$</th>
<th>$\rho$</th>
<th>ASD</th>
<th>TD</th>
<th>$\eta^2$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total (40)</td>
<td>24.0(0.9)/3.3</td>
<td>30.4(0.9)/3.3</td>
<td>0.56</td>
<td>0.000***</td>
<td>27.8(3.3)/3.4</td>
<td>31.1(0.8)/3.4</td>
<td>0.18</td>
<td>0.004**</td>
</tr>
<tr>
<td>Happiness (6)</td>
<td>4.7(0.3)/1.8</td>
<td>5.2(0.3)/1.7</td>
<td>0.16</td>
<td>0.014</td>
<td>3.8(1.4)/1.5</td>
<td>4.9(0.3)/1.8</td>
<td>0.17</td>
<td>0.006**</td>
</tr>
<tr>
<td>Neutral (8)</td>
<td>4.0(0.4)/1.7</td>
<td>5.5(0.4)/1.5</td>
<td>0.26</td>
<td>0.000***</td>
<td>5.7(1.7)/1.9</td>
<td>6.2(0.3)/1.9</td>
<td>0.03</td>
<td>0.409</td>
</tr>
<tr>
<td>Surprise (4)</td>
<td>1.5(0.2)/0.9</td>
<td>2.6(0.2)/1.2</td>
<td>0.34</td>
<td>0.000***</td>
<td>2.3(0.7)/1.2</td>
<td>2.7(0.1)/1.3</td>
<td>0.11</td>
<td>0.041</td>
</tr>
<tr>
<td>Disgust (3)</td>
<td>1.5(0.2)/1.0</td>
<td>2.2(0.2)/1.0</td>
<td>0.35</td>
<td>0.000***</td>
<td>2.5(0.6)/1.3</td>
<td>2.5(0.1)/1.2</td>
<td>0.03</td>
<td>0.424</td>
</tr>
<tr>
<td>Sadness (4)</td>
<td>3.1(0.2)/1.4</td>
<td>3.0(0.2)/1.4</td>
<td>0.00</td>
<td>0.973</td>
<td>3.1(0.2)/1.4</td>
<td>3.0(0.2)/1.4</td>
<td>0.03</td>
<td>0.402</td>
</tr>
<tr>
<td>Anger (6)</td>
<td>4.2(0.3)/1.6</td>
<td>4.7(0.3)/1.6</td>
<td>0.05</td>
<td>0.390</td>
<td>3.6(1.2)/1.5</td>
<td>4.0(0.2)/1.6</td>
<td>0.04</td>
<td>0.339</td>
</tr>
<tr>
<td>Blends (9)</td>
<td>5.8(2.4)/1.8</td>
<td>7.7(1.3)/2.2</td>
<td>0.36</td>
<td>0.001***</td>
<td>7.5(3.0)/2.1</td>
<td>8.3(0.8)/2.2</td>
<td>0.14</td>
<td>0.015</td>
</tr>
<tr>
<td>Fear-Surprise (5)</td>
<td>3.1(0.2)/1.4</td>
<td>4.7(0.2)/1.6</td>
<td>0.38</td>
<td>0.000***</td>
<td>4.6(0.8)/1.7</td>
<td>4.9(0.2)/1.8</td>
<td>0.09</td>
<td>0.063</td>
</tr>
</tbody>
</table>

*Note. ASD, autism spectrum disorder; TD, typical development. MANCOVA, corrected with FSIQ and age; MANCOVA, corrected with age; $\eta^2$ and $\rho$ values derived from adjusted means derived from the log transformation calculation.

**p < 0.008; ***p < 0.001, two-tailed

164
### TABLE 4. Emotion recognition accuracy rates and error patterns for ASD groups

<table>
<thead>
<tr>
<th>Response %</th>
<th>Egypt</th>
<th>ASD</th>
<th>Finland</th>
<th>Sadness</th>
<th>Anger</th>
<th>Egypt</th>
<th>ASD</th>
<th>Finland</th>
<th>Sadness</th>
<th>Anger</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>77.0</td>
<td>7.4</td>
<td>8.8</td>
<td>6.9</td>
<td>3.7</td>
<td>2.4</td>
<td>64.1</td>
<td>3.1</td>
<td>1.6</td>
<td>9.4</td>
</tr>
<tr>
<td>Neutral</td>
<td>2.0</td>
<td>47.8</td>
<td>4.4</td>
<td>2.0</td>
<td>5.1</td>
<td>2.9</td>
<td>51.1</td>
<td>69.1</td>
<td>7.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Surprise</td>
<td>2.0</td>
<td>3.3</td>
<td>22.1</td>
<td>2.9</td>
<td>2.2</td>
<td>5.4</td>
<td>8.9</td>
<td>7.8</td>
<td>52.3</td>
<td>3.1</td>
</tr>
<tr>
<td>Disgust</td>
<td>2.0</td>
<td>1.1</td>
<td>2.9</td>
<td>38.2</td>
<td>2.2</td>
<td>1.0</td>
<td>4.2</td>
<td>2.7</td>
<td>2.3</td>
<td>68.8</td>
</tr>
<tr>
<td>Sadness</td>
<td>10.8</td>
<td>33.8</td>
<td>24.3</td>
<td>14.7</td>
<td>70.6</td>
<td>15.2</td>
<td>4.7</td>
<td>12.9</td>
<td>7.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Anger</td>
<td>4.9</td>
<td>4.8</td>
<td>5.9</td>
<td>31.4</td>
<td>8.8</td>
<td>66.2</td>
<td>0.4</td>
<td>0.4</td>
<td>16.7</td>
<td>0</td>
</tr>
<tr>
<td>Fear</td>
<td>1.5</td>
<td>2.2</td>
<td>31.6</td>
<td>3.9</td>
<td>7.4</td>
<td>6.9</td>
<td>3.0</td>
<td>3.9</td>
<td>29.7</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. %, Percentages of answers; ASD, autism spectrum disorder. Accuracy rates, where the response matches the presented emotion are presented in boldface across the diagonal; a group's highest error rate in each emotion is underlined.

### TABLE 5. Emotion recognition accuracy rates and error patterns for TD groups

<table>
<thead>
<tr>
<th>Response %</th>
<th>Egypt</th>
<th>TD</th>
<th>Finland</th>
<th>Sadness</th>
<th>Anger</th>
<th>Egypt</th>
<th>TD</th>
<th>Finland</th>
<th>Sadness</th>
<th>Anger</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>87.7</td>
<td>2.6</td>
<td>2.2</td>
<td>0</td>
<td>0.7</td>
<td>0</td>
<td>83.9</td>
<td>0</td>
<td>0</td>
<td>2.4</td>
</tr>
<tr>
<td>Neutral</td>
<td>3.4</td>
<td>68.4</td>
<td>4.4</td>
<td>0</td>
<td>7.4</td>
<td>1.0</td>
<td>9.5</td>
<td>77.7</td>
<td>8.0</td>
<td>0</td>
</tr>
<tr>
<td>Surprise</td>
<td>1.5</td>
<td>5.1</td>
<td>70.6</td>
<td>2.9</td>
<td>11.8</td>
<td>14.2</td>
<td>1.8</td>
<td>3.1</td>
<td>72.3</td>
<td>1.2</td>
</tr>
<tr>
<td>Disgust</td>
<td>0.5</td>
<td>0.4</td>
<td>61.8</td>
<td>0</td>
<td>2.5</td>
<td>3.6</td>
<td>2.7</td>
<td>0</td>
<td>66.7</td>
<td>9.8</td>
</tr>
<tr>
<td>Sadness</td>
<td>6.4</td>
<td>22.4</td>
<td>8.8</td>
<td>4.9</td>
<td>69.1</td>
<td>3.9</td>
<td>0.6</td>
<td>13.8</td>
<td>3.6</td>
<td>2.4</td>
</tr>
<tr>
<td>Anger</td>
<td>0</td>
<td>0</td>
<td>0.7</td>
<td>27.5</td>
<td>1.5</td>
<td>75.0</td>
<td>0</td>
<td>0.4</td>
<td>26.2</td>
<td>2.7</td>
</tr>
<tr>
<td>Fear</td>
<td>0.5</td>
<td>1.1</td>
<td>13.2</td>
<td>2.9</td>
<td>9.6</td>
<td>3.4</td>
<td>0.6</td>
<td>2.2</td>
<td>16.1</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Note. %, Percentages of answers; TD, typical development. Accuracy rates, where the response matches the presented emotion are presented in boldface across the diagonal; a group's highest error rate in each emotion is underlined.

### TABLE 6. Emotion recognition accuracy rates and error patterns in Fear-Surprise blends for ASD and TD groups

<table>
<thead>
<tr>
<th>Response (%)</th>
<th>Egypt</th>
<th>ASD</th>
<th>Finland</th>
<th>Sadness</th>
<th>Anger</th>
<th>Egypt</th>
<th>TD</th>
<th>Finland</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>4.1</td>
<td>4.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>2.3</td>
<td>1.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surprise</td>
<td>17.6</td>
<td>28.8</td>
<td>67.1</td>
<td>46.4</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disgust</td>
<td>4.1</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0.7</td>
<td>0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sadness</td>
<td>11.8</td>
<td>1.3</td>
<td>0.6</td>
<td>0.7</td>
<td>0</td>
<td>0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anger</td>
<td>17.6</td>
<td>0.6</td>
<td>4.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fear</td>
<td>42.4</td>
<td>63.1</td>
<td>28.2</td>
<td>52.1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. %, Percentages of answers; ASD, autism spectrum disorder; TD, typical development. Accuracy rates, where the response matches the presented emotion are presented in boldface; a group's highest error rate in each emotion is underlined.
children labeled anger in their incorrect answers most often as surprise (56.9%), whereas Finnish TD children labeled anger in their incorrect answers most often as disgust (39.3%) ($\chi^2 = 42.3$, df = 5, $p < 0.001$). Egyptian TD children labeled anger more often as surprise, compared to Egyptian children with ASD (56.9% vs 15.9%, $\chi^2 = 51.2$, df = 5, $p < 0.001$).

**Fear-Surprise blended emotions**

Both Egyptian and Finnish children with ASD labeled fear-surprise blends more often as fear than as surprise (Table 6). Egyptian TD children labeled fear-surprise blends most often as surprise, whereas Finnish TD children labeled fear-surprise blends most often as fear. In Egypt, children with ASD labeled fear-surprise blends more often as fear compared to TD children ($\chi^2 = 60.8$, df = 6, $p < 0.001$). In Finland, there were no statistically significant differences between children with ASD and TD in fear-surprise blends ($\chi^2 = 10.5$, df = 5, $p = 0.06$).

**Association between ER and autistic symptoms**

There were no significant associations between ASSQ-R total score and FEFA scales within groups.

**Association between ER and FSIQ**

In the Egyptian ASD group, controlling for age, FSIQ was associated significantly with the FEFA total scale ($r = 0.627$, $p < 0.001$) and with the FEFA neutral expression scale ($r = 0.650$, $p < 0.001$), indicating that Egyptian children with ASD performed better in ER when they had higher cognitive capacity. In Egyptian TD children and in Finnish children with ASD, controlling for age, FSIQ was not associated with ER.

**Discussion**

The aim of our study was to examine basic ER from the eye region and ER error patterns in children with and without ASD. To examine the possible impact of ASD on ER, we compared children with ASD to TD children. To examine the possible cultural impact on ER, we selected an Arab country (Egypt), representing a mainly collectivistic culture, in which some negative emotions possibly should not be expressed, and a Scandinavian country (Finland) as an individualistic culture, in which negative emotions are also acceptable.

In accordance with our hypothesis and meta-analyses (12, 13), we found that both Arab and Scandinavian children with ASD had deficits in ER, compared to their country specific TD peers. In addition to general ER difficulties (i.e., lower FEFA total scale score), we also found that Arab children with ASD specifically recognized the emotions of disgust, surprise, neutral, and blended emotions from the eye region less often correctly than did their country specific TD children, whereas Scandinavian children with ASD recognized happiness less often correctly than did their country specific TD children. Happiness, disgust, and neutral are all mostly recognized from the mouth region (29-31), and thus it may be that children with ASD are not used to recognizing these emotions from the eye area. Our findings support those of Fink et al. (15) and Nuske et al. (16), as eye region ER deficits in ASD appear not to be systematic or universal: Berggren et al. (17) found deficits in happiness recognition in Scandinavian children with ASD, Franco et al. (3) found deficits in happy/surprise and sad/angry definition in multi-ethnic, English children with ASD, and Kuusikko et al. (18) found deficits in general ER in Scandinavian children with ASD.

Regarding our second hypothesis, we did not find cross-cultural differences in ER accuracy between Arab and Scandinavian TD children. However, cross-cultural differences in ER accuracy were significant between Arab and Scandinavian children with ASD: Egyptian children with ASD recognized less correctly emotions in general (FEFA total scale score and blended emotion scale) and separately in FEFA basic emotion scales of surprise, disgust, and neutral as well as in fear-surprise blended emotions than did Finnish children with ASD. In light of our findings, it therefore appears that children with ASD may be more vulnerable to cultural effects that limit the ER in the eye area, specifically in surprise, disgust, and neutral, compared to TD children. Our cross-cultural findings of basic ER from the eye region differ from the findings of Friderson-Hayo et al. (14), who reported insignificant differences in full-facial ER skills among Swedish (n = 19), British (n = 16), and Israeli (n = 20) 5 to 9 year old children with ASD. However, it would be important to have more ER studies conducted in Arab countries in order to confirm our findings: our study is the first cross-cultural ER study conducted in the field of ASD which has examined basic ER from the eye region, and it is to our knowledge only the second cross-cultural study which has examined the cross-cultural aspects of facial ER in ASD.

Our third hypothesis, that TD children from a collectivistic culture (i.e., Arab) would have more positive ER errors than TD children from an individualistic culture (i.e., Scandinavian), was partly confirmed: Arab children with TD confused sadness and anger as surprise, whereas Scandinavian TD children confused sadness as neutral expression and anger as disgust. Arab TD children also labeled fear-surprise blends more often as surprise than did Scandinavian TD children. Scandinavian children labeled emotions (i.e., happiness and sadness) more
often as neutral expressions compared to Arab children, which may show cultural differences in eye contact itself: in Scandinavia, eye contact may be common and mundane, whereas in Arab countries it may be more communicative and significant. That is, Arab children may search for more meaning from eye contact than Scandinavian children may.

In addition, Arab children with ASD labeled sadness and anger in their ER error patterns more negatively than did Arab TD children, but there were no differences between Scandinavian children with ASD and TD in ER error patterns. However, Arab children with ASD biased happiness, sadness, and anger more negatively than did Scandinavian children with ASD, who biased happiness and sadness as neutral expression and anger as disgust.

In addition, we found that all children labeled disgust most often as anger in their incorrect answers, and both Arab and Scandinavian children with ASD labeled fear-surprise blends more often as fear than as surprise: our findings are in line with those of Jones et al. (48) and Wallace et al. (49). However, our results do not support the findings of Eack et al. (47): in our study children with ASD did not confuse happy, sad, and angry faces more often with neutral expressions compared to TD individuals, instead both Scandinavian children with and without ASD biased emotions more often to neutral expression than did Arab children with and without ASD, indicating a cultural difference in confusing emotions with neutral expression. Overall, our ER error pattern results indicate that Arab children with ASD biased emotions more negatively than did Scandinavian children with ASD or Arab TD children, which may tell something about the emotional experiences of Arab children with ASD. ER biases are worthy of further study, because, as noted, children’s ER biases may be more relevant in predicting social behavior than in their recognition accuracy.

In contrast to information from the literature (32, 33), we did not find an association between the parentally reported severity of autistic symptoms and ER in either Egypt or Scandinavia. Our findings suggest that in children with FSIQ in the normal range, ER deficits in ASD may represent the sum of a broader range of variables (e.g., attentional distractibility) associated with ASD than only the severity of autistic symptoms (17).

In line with the literature, we found that facial ER might be more dependent on the intellectual ability of children with ASD than in the case of TD children (34): specifically, higher cognitive capacity in the normal range was associated with better ability in ER in Arab children with ASD. However, we did not find the same association in Scandinavian children with ASD, who were older than the Arab children. Therefore we suggest that ER skills may become more independent of cognitive capacity with increasing age.

**Limitations**

There were several limitations to our study. First, internal reliability was low on FEFA single emotion scales in our study sample. The FEFA single emotion scales are low itemized and the error pattern analyses revealed differences between groups and showed that emotions are seen/interpreted differently by a significant number of individuals: these may partly explain the low alphas. Second, although we did not find any correlation between parentally reported autistic symptoms and ER in our study nor did we find significant difference between Arab and Scandinavian ASD children on ASSQ-R, it must be noted that Arab TD children scored significantly higher than did Scandinavian TD children on ASSQ-R. Arab TD children scored significantly higher than Scandinavian TD children on half of the ASSQ-R items and the frequency of being rated as a “definite abnormality” was especially high (i.e., 18-24% of Arab TD children) on four ASSQ-R items (i.e., literal understanding of metaphorical language; makes naïve and embarrassing remarks; can be with other children but only on his/her terms; Lacks best friend), whereas none of the Scandinavian TD children were rated having “definite abnormality” on any of the ASSQ-R items. The ASSQ is not yet validated in Arab countries, and validation of ASSQ would be of great importance in order to employ it as a tool in cross-cultural comparison studies of ASD symptoms and screening. For now it is unclear whether significant difference on ASSQ-R between Arab and Scandinavian TD children is due to differences in composition of the samples, cultural differences in parent interpretation of the items, or perhaps other factors. Furthermore, validation of ASSQ should also include both parents’ and teacher’s ratings (57). Third, we did not measure the FSIQ of Finnish TD children, but assumed that it would be within the normal range. In Finland, all children attending mainstream education most probably have an IQ within the normal range, as children with lower than average IQ are referred to special education. Based on Programme for International Student Assessment results, the standard deviation in Finnish student performance is among the smallest in Organisation for Economic Co-operation and Development (OECD) countries (59). However, it is possible that Finnish TD children have significantly different (higher or lower) FSIQ compared to other study sample children. Future studies should consider these issues in order to enhance the validity of our findings.
Conclusions
Regardless of their cultural background, children with ASD appear to have deficits in ER from the eye region, compared to our country specific TD peers. The differences between the Arab and Scandinavian children may reflect cultural differences in ER and ER error patterns. Due to the humanitarian crisis in Middle-Eastern countries, there are increasing numbers of immigrants and refugees from different cultural backgrounds in Scandinavia. Clinical workers should recognize the possible cultural impacts on ER when working with children with and without ASD and their families.

Conflict of interest disclosure
The FEFA program is part of the KONTAKT social skills training program published by Hogrefe publishers. Dr. Bölte received royalties for the manual and work books. All other authors report no conflicts of interest.

Funding
This study received financial support from the Alma and K. A. Snellman Foundation, Oulu, Finland; the Emil Aaltonen Foundation, Finland; the Sigrid Jusélius Foundation, Finland and Thule Institute, University of Oulu, Finland. Sven Bölte was supported by the Swedish Research Council (523-2009-7054).

Acknowledgments
The authors thank their participants for their participation in this study, and Katja Jussila, Marko Kielenen, Sirkka-Liisa Linna, Marja-Leena Mattila, David Pauls, and Jukka Rahko for their help in the Finnish data collection.

References