Perception of emotions from facial expressions in high-functioning adults with autism

Daniel P. Kennedy [a,1,*], Ralph Adolphs [a,b,**]

[a] Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA
[b] Division of Biology, California Institute of Technology, Pasadena, 1200 E, California Blvd., MC 228-77, Pasadena, CA 91125, USA

ARTICLE INFO

ABSTRACT

Impairment in social communication is one of the diagnostic hallmarks of autism spectrum disorders, and a large body of research has documented aspects of impaired social cognition in autism, both at the level of the processes and the neural structures involved. Yet one of the most common social communicative abilities in everyday life, the ability to judge somebody's emotion from their facial expression, has yielded conflicting findings. To investigate this issue, we used a sensitive task that has been used to assess facial emotion perception in a number of neurological and psychiatric populations. Fifteen high-functioning adults with autism and 19 control participants rated the emotional intensity of 36 faces displaying basic emotions. Every face was rated 6 times—once for each emotion category. The autism group gave ratings that were significantly less sensitive to a given emotion, and less reliable across repeated testing, resulting in overall decreased specificity in emotion perception. We thus demonstrate a subtle but specific pattern of impairments in facial emotion perception in people with autism.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Impaired social communication is one of the hallmarks of autism spectrum disorders (ASD). It is commonly thought that people with ASD are impaired also in a specific aspect of social communication, the recognition of basic emotions from facial expressions (i.e., happiness, surprise, fear, anger, disgust, sadness). However, the literature on this topic offers highly conflicting findings to date: whereas some studies find clear impairments in facial affect recognition in autism (Ashwin, Wheelwright & Baron-Cohen, 2006; Corden, Chilvers & Skuse, 2008; Dziobek, Bahnemann, Convit & Heekeren, 2010; Law Smith, Montagne, Perrett, Gill, & Gallagher, 2010; Philip et al., 2010; Wallace et al., 2011), others do not (Adolphs, Sears & Piven, 2001; Baron-Cohen, Jolliffe, Mortimore & Robertson, 1997; Neumann, Spezio, Piven & Adolphs, 2006; Rutherford & Towns, 2008). Part of this discrepancy may be traced to the known heterogeneity of ASD, together with differences in the stimuli and tasks used in the various studies; and part may derive from the specific aspects of facial emotion perception that were analyzed in the studies.

A recent and comprehensive review attempted to make sense of this mixed literature (Harms, Martin & Wallace, 2010). The authors suggest that the ability of individuals with an ASD to identify facial expressions depends, in large part, upon several factors and their interactions, including demographics (i.e., subjects' age and level of functioning), the stimuli and experimental task demands, and the dependent measures of interest (e.g., emotion labeling accuracy, reaction times, etc.). Other factors, such as ceiling effects or the use of compensatory strategies by individuals with an ASD, might also obscure true group differences that would have been otherwise found. The authors further make the interesting point that other behaviorally- or biologically-based measures almost invariably demonstrate that individuals with ASDs process faces differently, so perhaps previous studies of facial affect recognition which failed to find group differences used tasks and/or measures that are simply not sensitive enough to detect group differences. Difficult or unfamiliar tasks are more likely to reveal impairment, since they are better able to avoid ceiling effects and, in some cases, are less well-rehearsed and preclude compensatory strategies.

Two distinct methodological approaches have been used to achieve these goals of providing sensitive measures of facial affect recognition. One approach has been to manipulate the stimuli in some way, such as with facial morphing (e.g., Humphreys, Minshew, Leonard & Behrmann, 2007; Law Smith et al., 2010; Wallace et al., 2011). This approach gives the experimenter parametric control of the intensity of the stimuli, and so can assess emotion discrimination at a fine-grained level, but with the important caveat that the morphs

* Correspondence to: 1011 E, 10th Street, Bloomington, IN 47405, USA.
Tel.: +1 812 855 1169.
** Corresponding author. Tel.: +1 626 395 4486; fax: +1 626 793 8580.
E-mail addresses: dpk@indiana.edu (D.P. Kennedy),
radolphs@hss.caltech.edu (R. Adolphs).
[1] Current Address: Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA.

are artificially generated and not necessarily the same as the subtle expressions that one would encounter in the real world. The second main approach is to modify the task demands (e.g., changing task instructions, reducing the length of stimulus presentation, etc.), rather than manipulating the stimuli in any way. By doing so, the experimenter can increase the task difficulty and reduce the likelihood that an explicit, well-rehearsed cognitive strategy is used for decoding the expression, while still using naturalistic stimuli. Here, we took this latter approach. We used a well-validated and widely used set of facial emotion stimuli (Paul Ekman's Pictures of Facial Affect; Ekman, 1976) and obtained detailed ratings of emotion. An additional motivation for using these stimuli is that they provide continuity with a number of prior studies in a wide variety of populations including ASD (Adolphs et al., 2001), patients with brain lesions (Adolphs, Tranel, Damasio & Damasio, 1995; Adolphs, Damasio, Tranel, Cooper & Damasio, 2000), frontotemporal dementia (Diehl-Schmid et al., 2007), Parkinson's disease (Sprengelmeyer et al., 2003), and depression (Persad & Polivy, 1993).

Given that facial expressions are complex and are often comprised of varying degrees of two or more emotions in the real world, participants were asked to determine the intensity levels of each of the 6 basic emotions for every emotional face they were shown (e.g., rate a surprised face on it is intensity (i.e., degree) of happiness, surprise, fear, anger, disgust, and sadness, etc.). In keeping with previous descriptions of this task (e.g., Adolphs et al., 1994,1995), we refer to it as an emotion recognition task, since it requires one to recognize (and rate) the level of a particular emotion displayed by a face. For instance, for one to rate a surprised face as exhibiting a particular intensity of fear requires recognizing that emotion, fear, in the first place. Given that participants are unlikely to have practiced this task during any sort of behavioral intervention they may have been exposed to, we expected this task to reveal group differences, particularly in the overall intensity ratings and the degree of response selectivity (i.e., tuning or sharpness) for particular emotional facial expressions. We also assessed test-retest reliability in a subset of our study sample, to explore whether a less stable representation of emotional expression would be reflected in increased response variability across these testing sessions.

2. Methods

2.1. Participants

Seventeen high-functioning male adults with an ASD and 19 age-, gender- and IQ-matched control participants took part in this experiment. Two ASD participants were excluded because their scores exceeded at least 3 standard deviations (SD), calculated across both groups together, for one or more of their facial expression-emotion judgment categories, resulting in a final sample size of 15 ASD and 19 control participants. Control and ASD groups were well matched on age and verbal, performance, and full-scale IQ (see Table 1a). IQ scores were not available for 3 control participants and 1 ASD participant. DSM-IV diagnosis of an ASD was made by a clinical psychologist following administration of the Autism Diagnostic Observation Scale (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R; when a parent was available; for 12 of 15 participants total; see Table 1b). In order to ensure that our findings could not be explained by possible group differences in facial identity recognition, we also tested participants on the Benton Facial Recognition Test (14/15 ASD participants and 12/19 controls). There were no significant differences between groups [$t(24)$=4.39, p=0.70; ASD=46.5 (4.7); control=45.8 (4.1)]. However, two participants with an ASD scored below the cutoff for "Normal" on Benton Facial Recognition (1 scored in the "Borderline" range and 1 scored in the "Moderate Impairment" range), and 1 control participant scored in the "Borderline" range. The main effects reported below remained significant even after re-running the analyses with these 3 participants excluded.

2.2. Stimuli

Each participant rated the emotional intensity of 36 Ekman faces, each displaying 1 of the 6 basic emotions (i.e., happiness, surprise, fear, anger, disgust, sadness; Ekman, 1976). All images were black and white, and were forward facing.

Table 1
(a) Mean age and IQ scores (verbal, performance, and full-scale) were not significantly different between control and autism groups. Standard deviations are given in the parentheses. (b) Means, standard deviations, and ranges for the autism group on clinical measures (ADOS=Autism Diagnostic Observation Schedule; ADI-R=Autism Diagnostic Interview-Revised; RSB=restricted and stereotyped behaviors).

	Control group (n=19)[*]	Autism group (n=15)[*]	T-test
Age	31.9 years (11.9)	30.5 years (11.5)	$t(32)$=0.36, p=0.72
Verbal IQ	111.5 (10.9)	109.0 (16.8)	$t(28)$=0.49, p=0.63
Performance IQ	107.2 (12.8)	104.9 (12.3)	$t(28)$=0.49, p=0.63
Full scale IQ	110.2 (10.2)	107.1 (11.7)	$t(28)$=0.78, p=0.44
		Mean (SD)	Range
ADOS: communication		3.7 (1.3)	2–6
ADOS: social		7.7 (3.0)	5–14
ADOS: RSB		1.3 (1.3)	0–4
ADI-R A: social		20.8 (5.0)	12–28
ADI-R B: communication		15.8 (3.6)	10–22
ADI-R C: RSB		6.3 (2.9)	2–12

* IQ scores were not available from 3 control participants and 1 ASD participant.
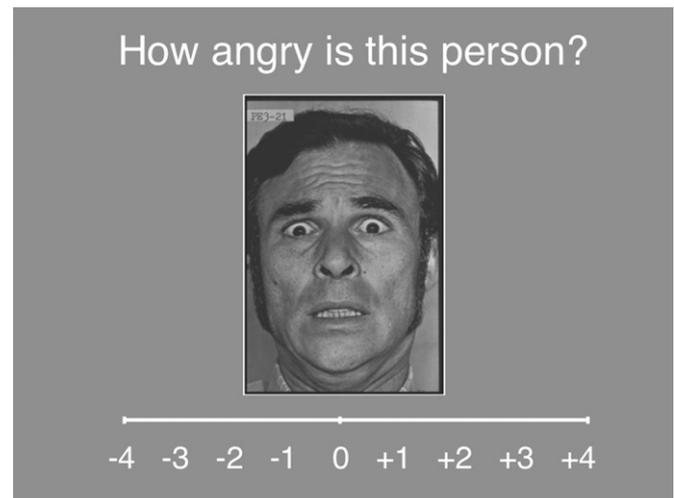

Fig. 1. An example of a trial. Participants were asked to rate the intensity of a particular emotion (in this example, anger) for all 36 emotional faces used in the experiment, each representing 1 of the 6 basic emotions (in this case, fear). The emotion of the rating category and face could be concordant (e.g., rate a happy face on how happy it is) or discordant (such as in the current example). Every face was rated 6 times—once for each emotion category.

An additional 3 faces displaying neutral expressions were also included, but these are considered separately from the other emotional stimuli. The specific images were identical to those used in (Adolphs et al., 1995, 2001), and were chosen on the basis of an independent group of control participants rating these faces as exemplars of each particular emotion category (Ekman, 1976). Stimuli were 336×500 pixels (width $\times$ height) and displayed on an LCD monitor with a resolution of 1024 by 768 pixels.

2.3. Task

Subjects viewed and rated the emotional intensity of the 36 emotional faces and 3 neutral faces, presented in randomized order, a total of 6 times- once for each emotion category (e.g., rating the intensity of happiness in all 39 faces; then rating the intensity of surprised, etc.; see Fig. 1). Therefore, ratings could either be concordant (i.e., the rating category and the primary/intended facial expression are the same emotion) or discordant (i.e., the rating category and primary facial expression are different emotions). Intensity was rated on a Likert scale from −4 to +4, with a +4 indicating that the face very strongly displayed the emotion of the rating category, while a −4 indicated that the face very strongly displayed an emotion opposite to the rating category (a rating of 0 indicated that the face was

neutral for the rating category). Subjects rated each of the 39 faces on a single emotion category before proceeding to the next emotion category. In total, participants made 234 ratings of emotional faces (6 emotion judgment categories × 39 faces). Participants were told to take as much time as they needed, and to answer as carefully as possible using a keyboard with the keys appropriately labeled. The order of face stimuli and the order of the emotion judgment categories were entirely randomized both within and across participants.

In order to investigate test-retest reliability, a small subset of participants from each group (ASD $n=6$; control $n=8$) performed the experiment twice, with test-retest interval of no more than 1 day. Participants were selected to participate in this additional testing solely on the basis of their availability. The mean test-retest interval was not significantly different between groups [$t(12)=1.71$, $p=0.11$].

2.4. Analysis

Data were analyzed using univariate and multivariate mixed effects, repeated-measures ANOVAs, with Group (ASD, controls) and Emotion (i.e., the six intended facial expressions – happy, surprise, fear, anger, disgust, sad) as the factors.

We derived from the raw ratings three dependent measures of interest. First, concordant intensity ratings were those ratings for the concordant face/label ratings, whereas discordant intensity ratings were the mean of all the other five discordant ratings. We also derived a measure of selectivity, in this context, as the difference between a concordant rating (e.g., a happy face rated on happiness) and the mean of the discordant ratings. Note that discordant ratings are not necessarily incorrect, as subtler levels of multiple emotions can typically be perceived in emotional faces (e.g., one can often perceive happiness in surprise).

3. Results

A comprehensive plot of the data is given in Fig. 2, which reveals several findings. Overall, at the mean group level, the pattern of ratings on the different emotion labels across all the different facial emotion stimuli were highly correlated between ASD and control groups (mean Pearson's r across all stimuli $= 0.97$,

$p < 0.00001$; see Fig. 2). Both groups showed a similar pattern in which they assigned the highest intensity for concordant ratings, and displayed similar patterns of "confusion" for particular emotions (e.g., fear-surprise; disgust-anger), though in this context confusion does not necessarily mean the judgment was incorrect. Happy faces were rated with the greatest selectivity in both subject groups (Fig. 2). Neutral faces were rated similarly by the two groups across the 6 emotion judgments [$(F(1,32)=0$; n.s.)], nor was there a Group × Emotion Judgment interaction [$F(5,160)=0.26$, $p=0.94$].

Despite these similarities between subject groups, we also found several important differences. Visual inspection of the pattern of data shown in Fig. 2 suggests reduced selectivity in the ASD group, which was confirmed with a significant main effect of Group [$F(1,32)=6.74$, $p=0.014$]. As described above, selectivity is defined for each facial expression as the difference between the concordant and discordant ratings (i.e., the cell that falls on the diagonal minus the non-diagonals in each row). Although there was also a main effect of Emotion [$F(5,160)=58.01$, $p < 0.0001$], there was no Group × Emotion interaction [$F(5,160)=0.65$, $p=0.48$]. Across the 6 facial emotions, the mean selectivity in the control group was 4.08 (SD$=0.83$), while the mean selectivity in the ASD group was 3.34 (0.81) (see Fig. 3A). Furthermore, selectivity was not significantly associated with full-scale IQ in the ASD group [$r=0.13$, $p=0.66$]. To ensure that these findings could not be accounted for by possible differences in the range of the rating scale used between groups, we z-scored the data with respect to each participant's distribution of ratings so that all subjects' ratings were on a comparable scale. To do so, we subtracted each participant's mean rating (across all the faces) from their rating for a given face, divided by their standard deviation (across all the faces). Even after this z-score
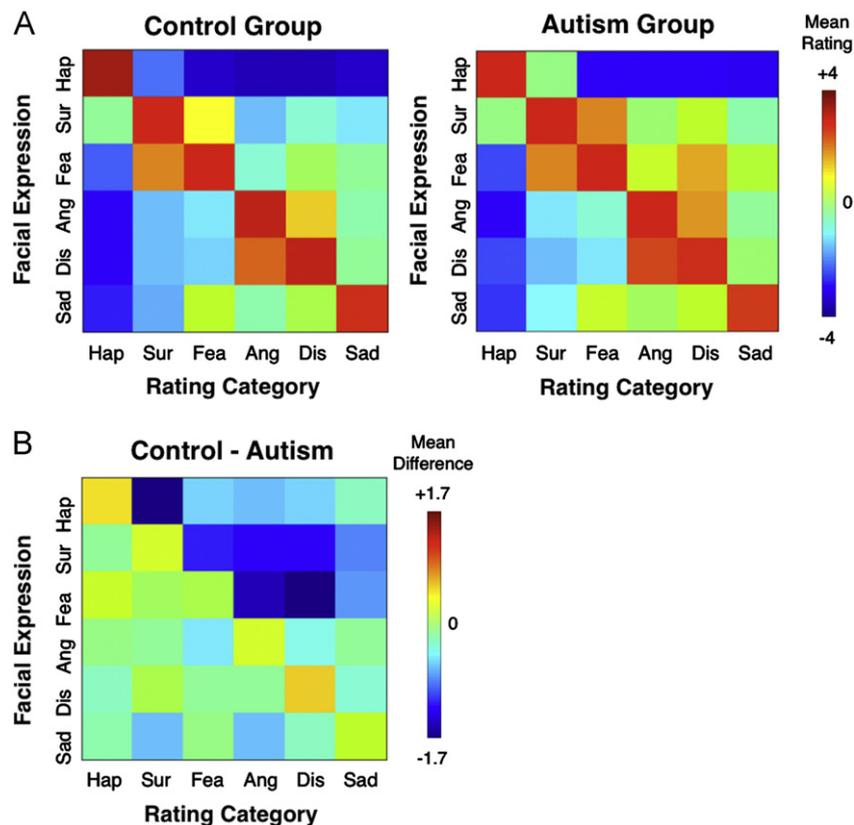

Fig. 2. (A) Autism and control group rating matrices. The intended facial expression is given on the y-axis, and the emotion rating category is given on the x-axis. Concordant ratings fall along the diagonal, and discordant ratings fall along the off-diagonal. (B) The difference between autism and control groups. No individual cell survives FDR correction for multiple comparisons ($q < 0.05$).
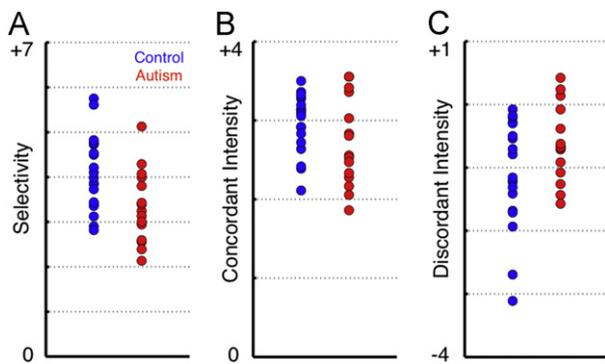
Fig. 3. Scatterplots of emotion selectivity, concordant intensity, and discordant intensity ratings. These plots represent data that are collapsed across all 6 emotion categories.

normalization, the above described main effect remained significant [$F(1,32)=7.05$, $p=0.012$].

To further visualize this reduced selectivity between the emotions exhibited by the autism group, we used a method known as non-metric Multidimensional Scaling (MDS). To carry out this analysis, individual subject similarity matrices were derived by calculating the correlations between the 6 ratings given for each face with the 6 ratings given for every other face (see Adolphs et al., 1994). Next, we fisher z-transformed these correlation matrices, averaged them together for each group separately, converted back to correlation coefficients with the inverse fisher z transformation, and then transformed them into dissimilarity matrices by subtracting the result from 1. MDS then determined the Euclidean distance (corresponding to the perceived similarity) between the 39 faces in 2 dimensions (see Fig. 4), which was chosen following visual inspection of a scree plot showing that 2 dimensions captured most of the variance. The amount of MDS stress, which provides a goodness-of-fit measure of the MDS result, was similar between groups (ASD=0.097; controls=0.094). However, one can observe that the ASD group shows less separation between the different emotion categories, confirming the above finding of reduced emotion selectivity.

Selectivity, as defined above, could be a consequence of two distinct factors, or a combination of both. First, reduced selectivity might result from reduced intensity ratings for the concordant emotion, or second, it might be a consequence of increased "confusion," (i.e., higher intensity ratings for the discordant emotions), neither of which are mutually exclusive. Therefore, we ran a multivariate repeated-measures ANOVA, which revealed no main effect of Group [$F(1,32)=0.53$, $p=0.47$] but a significant interaction between the two dependent variables (concordant and discordant intensity) and Group [$F(1,160)=6.74$, $p=0.014$]. To further explore the contribution of each of these variables to specificity, we next ran 2 univariate 2×6 repeated-measures ANOVAs. In terms of concordant intensity, there was a possible trend toward a significant main effect of Group (ASD/controls) [$F(1,32)=2.45$, $p=0.127$], as well as a significant main effect of Emotion [$F(5,160)=3.20$, $p=0.0088$], but the interaction was not significant [$F(5,160)=0.74$, $p=0.59$], suggesting possibly overall reduced intensity ratings for concordant emotion judgments in individuals with an ASD (mean=2.72, SD=0.56) compared to the control group (mean=2.97, SD=0.37) (Fig. 3B). Second, in terms of discordant ratings, there was a trend toward a significant main effect of Group (ASD/control) [$F(1,32)=3.51$, $p=0.07$], there was a main effect of Emotion [$F(5,160)=115.93$, $p<0.0001$] and a trend toward a Group × Emotion interaction [$F(5,160)=1.97$, $p=0.086$]. The mean intensity rating for discordant emotion judgments (i.e.,
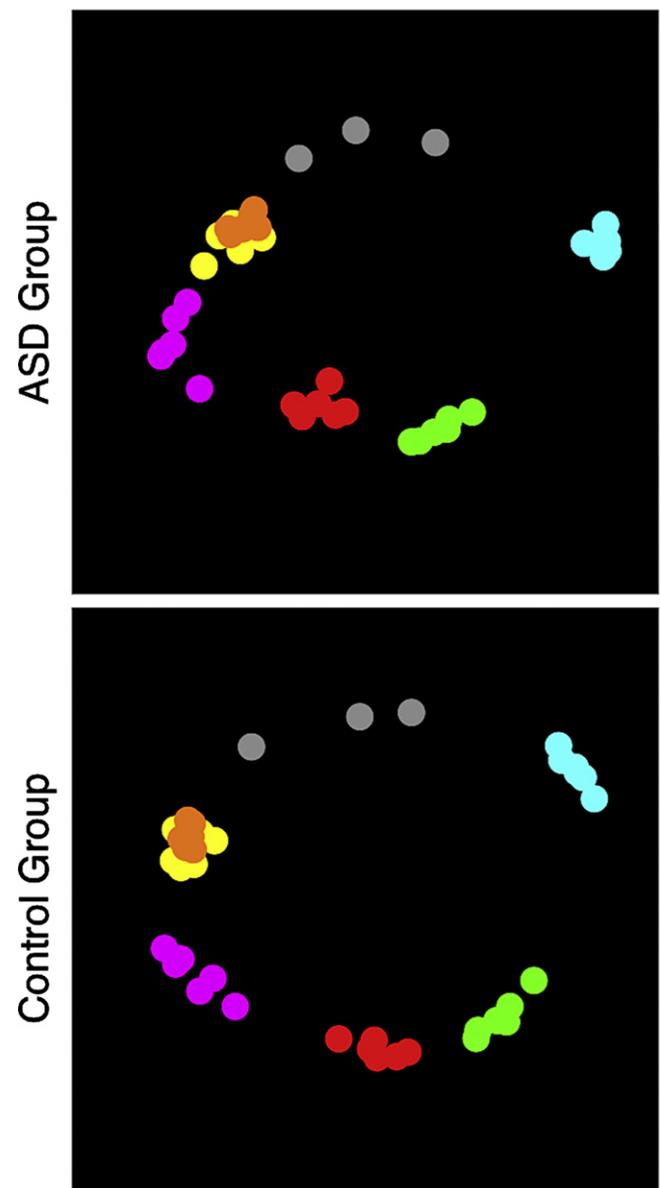

Fig. 4. Multidimensional scaling (MDS) of perceived similarity of the emotional expression in faces. Each colored dot represents a single Ekman face ($n=39$), colored according to the intended expression of the face (happy=cyan; surprised=green; afraid=red; angry=yellow; disgust=orange; sad=magenta; neutral=gray). The Euclidian distance between points represents their perceived similarity. In both groups, the faces belonging to a particular emotion category generally cluster near one another, although the autism group has less separation overall between the different emotion categories (mean pairwise distance between points in the ASD group=0.814; mean pairwise distance in the control group=0.923). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

confusion) in the ASD group (mean=−0.62, SD=0.60) was higher than the control group (mean=−1.11, SD=0.85) (Fig. 3C). Post-hoc tests carried out for each emotion separately revealed higher discordant ratings in the ASD group for happy faces [$t(32)=2.69$, $p=0.01$, uncorrected], surprised faces [$t(32)=2.71$, $p=0.01$, uncorrected], and fear faces [$t(32)=2.09$, $p=0.046$, uncorrected]. Therefore, the reduced selectivity in ASD appears to be a consequence of both reduced concordant ratings as well as increased discordant ratings (perhaps especially for happy, surprise, and fear).

To explore whether the lower selectivity we found in the ASD group might in fact result from lower reliability, we next
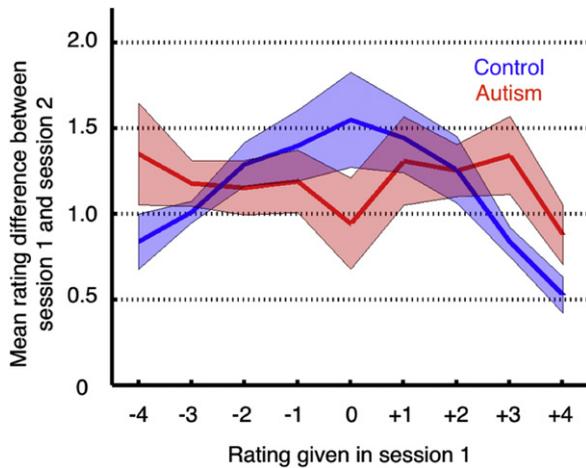
Fig. 5. Test-retest reliability. The autism group exhibited less consistent ratings across a second identical testing session for high intensity ratings (i.e., −4, −3, +3, +4). The y-axis represents the mean of the absolute value of the difference between session 1 and session 2, and this value is plotted as a function of the rating level given in session 1 (x-axis).

examined test-retest reliability in a small subset of participants who completed the entire task a second time (ASD $n=6$; control $n=8$). The dependent measure was the difference between the rating given in the first session and the rating given in the second session, collapsed across all stimuli and ratings. To fully understand whether reliability might be influenced by the magnitude of a rating given in the first place, we carried out a 2 (group; ASD/control) by 9 (initial rating, from −4 to +4) ANOVA, using the raw (non-z-scored) absolute value of the difference between the test and retest. Although there was no overall main effect of Group [$F(1,12)=0.10$, $p=0.76$], there was a significant effect of Initial Rating [$F(8,124)=2.87$, $p=0.007$] and a significant Group-×Initial Rating interaction [$F(8,124)=2.47$, $p=0.018$]. We next grouped the data into high-intensity initial ratings (i.e., −4, −3, +3, +4) and low-intensity initial ratings (i.e., −2, −1, 0, +1, +2). Compared to controls, individuals with ASD had less reliable ratings for those faces that they initially rated as more emotionally intense [$t(12)=3.29$, $p=0.006$] whereas they did not differ from controls in reliability for faces they rated as initially less intense [$t(12)=0.88$, $p=0.40$] (Fig. 5). The mean rating change for high intensity ratings in the ASD group was 1.19 (SD=0.30) and in the control group was 0.80 (SD=0.12). The mean rating change for low intensity ratings in the ASD group was 1.18 (SD=0.42) and in the control group was 1.39 (SD=0.43).

4. Discussion

We found two specific impairments in facial emotion perception in autism, which we believe are likely to be related. First, we found a significantly reduced selectivity in autism, arising from a combination of giving somewhat lower intensity ratings to the intended (concordant) emotion label for a face, and somewhat higher intensity ratings to the unintended (discordant) emotion label for a face. Second, we found significantly decreased test-retest reliability in autism, specifically for those faces that on the first round received the most positive or most negative ratings.

Our interpretation of these two significant effects is that they reflect two sides of the same coin. Whereas controls rate prototypical emotional expressions as exhibiting high intensity on their intended emotion to the relative exclusion of other, unintended, emotions (high selectivity), people with autism have a lower selectivity (i.e., a flatter response profile across for a particular

emotion). This lower selectivity in autism appears to arise not from a highly consistent altered judgment of the faces, but rather from difficulty in consistently mapping the stimuli shown to the emotion categories intended, with a resulting lower test-retest reliability.

It is important to note that the abnormalities we report, while significant, are relatively subtle. Overall, the emotional intensity ratings given by the autism and control groups were highly similar in terms of broad patterns (Fig. 2). The ratings across all faces between the two groups were highly correlated, and there were no differences in ratings to neutral facial expressions, demonstrating that there was no overall group bias for simply ascribing emotions to faces. Our findings suggest that decreased emotional selectivity cannot be accounted for completely by either reduced ratings for concordant emotions alone or by increased ratings for discordant emotions alone. Rather, trends toward a group difference for both of these factors suggest that the combination of the two resulted in reduced emotional selectivity. In addition, exploratory post-hoc tests revealed that the autism group exhibited increased discordant ratings for happy, surprised, and fearful faces, suggesting that these emotions may show reduced selectivity in autism in large part due to their low discordant ratings in the controls. Taken together, we believe the findings admit of three broad conclusions: (1) Adults with ASD, at least when high-functioning and compared to well-matched controls, show relatively intact basic ability to judge emotions on their intended labels, do not show any evidence of a global bias towards one emotion or another, and do not show any evidence of increased difficulty in discriminating faces in general. (2) There is a relatively specific impairment in ASD in selectivity and test-retest reliability for rating facial emotions across multiple emotion categories. (3) The impairment is consistent with an inability to distinguish clearly between emotion categories conceptually, or an inability to map faces onto words for the emotion categories, or both. This last conclusion also points immediately towards important future studies that should be done to probe the source of the impairment in more detail. In particular, it will be essential, across multiple classes of stimuli, to assess whether people with autism have normal emotion concepts to begin with, or whether their concepts for emotions show similarly degraded selectivity as does their facial emotion perception in our experiment.

It is particularly notable that these differences were found in a group of high-functioning, very cognitively-able adults with autism, and with the use of rather simple stimuli and an uncomplicated experimental design. Rather than making the task more difficult by ambiguating the facial expression in some way (e.g., degrading the stimuli, morphing to a more subtle expression) or decreasing the time available to make a judgment, we chose instead to ask participants to make a less well rehearsed and more fine-grained judgment about facial expressions (i.e., it is doubtful that an intervention would have emphasized the identification of the non-primary emotional expression exhibited in a face). In addition, this avoided the problems inherent with using morphed facial expressions or highly degraded stimuli—namely, that they differ from what is encountered in the real world. Furthermore, given that the task was self-paced and that task instructions explicitly emphasized accuracy (in this case, rating the intensity of the emotion as best they could) over response time, the present results cannot be explained by a non-specific group difference in processing speed. This is particularly important, as processing speed and reaction time differences are often observed between autism and neurotypical control groups (Mayes & Calhoun, 2008; Oliveras-Rentas et al., 2012; Schmitz, Daly & Murphy, 2007). One would expect to find larger group differences had the task been timed; however, it would then be

necessary to determine whether or not the results are task-specific or task-general.

It is striking that the autism group demonstrated less reliable ratings over time for the high-intensity emotions. One would predict that high-intensity emotions would yield more stable ratings and low-intensity emotions would be yield less stable ratings—the exact pattern exhibited by the control group. That is, controls benefit in reliability from the intensity of the intended emotion, since strongly concordant or discordant emotion ratings presumably correspond to relatively categorical decisions regarding the emotion shown and hence high reliability over time. The autism group did not benefit from emotion intensity in this way at all, showing essentially a flat curve of reliability with emotional intensity (cf. Fig. 5).

How these findings might translate to real-world situations should also be considered. Under our experimental conditions, although we identified group differences in emotional selectivity, we also observed many similarities between groups, suggesting an overall preserved ability to identify emotional facial expression at a gross level. However, in a more ecologically-valid social context (e.g., during a face-to-face interaction, watching videos of naturalistic social scenes), the task of identifying facial expressions becomes much more difficult. Real-time online processing of dynamic stimuli is required, and expressions need to be integrated simultaneously with other cues, such as language, prosody, and context—all of which have been shown to be abnormal in autism (Paul, Augustyn, Klin & Volkmar, 2005; Tager-Flusberg, 1981; Tudusciuc & Adolphs, 2011; Van Lancker, Cornelius & Kreiman, 1989). Furthermore, the task of identifying emotions from faces is less constrained in the real world, as many more expressions exist, particularly in terms of social emotions (e.g., embarrassed, compassion, etc.). It is likely that the differences identified in this simplified experimental task would be magnified in a live interaction with others, perhaps contributing to the sometimes profound real world deficits in social interaction observed in otherwise highly intelligent individuals with autism.

An important factor that might relate to the present findings of altered perception of facial expressions in autism is altered patterns of gaze and attention to faces. Given that different facial expressions contain key identifying information in different areas of the face (e.g., fear in the eyes, anger in the eyebrows, happiness in the mouth; Smith, Cottrell, Gosselin & Schyns, 2005), the well-documented abnormalities in fixation patterns and attention in autism could play a role in the group differences found here (Klin, Jones, Schultz, Volkmar & Cohen, 2002; Langdell, 1978; Pelphrey et al., 2002). Given that high-functioning individuals, as a group, are known to rely less on information in the eyes when judging facial expressions (Spezio, Adolphs, Hurley & Piven, 2007a,b), and that surprise and fear can be most easily differentiated from other emotions using information in the eye region (Smith et al., 2005), this may account for our findings of higher discordant emotion ratings for surprised and fear faces in the autism group. This is also consistent with Corden et al. (2008), who found in a sample of adults with autism that the ability of an individual to correctly identify faces exhibiting fear directly related to the amount of time spent fixating the eyes in the face. If the pattern of fixation (and attention) to faces can account for group differences, then manipulations that normalize fixation to the eye region might normalize the perception of emotional facial expressions in autism.

How can we reconcile the present results with the mixed reports in the literature? While some studies have identified impairments in high functioning adolescents and adults with autism (Ashwin et al., 2006; Corden et al., 2008; Dziobek et al., 2010; Law Smith et al., 2010; Philip et al., 2010; Wallace et al., 2011), others have not (e.g., Baron-Cohen et al., 1997; Neumann et al., 2006; Rutherford & Towns, 2008; for review, see Harms et al., 2010), even those using the identical stimuli and a near-identical task (Adolphs et al., 2001). These mixed results are likely partly attributable to the enormous heterogeneity of autism spectrum disorders, which will manifest itself in both increased within-study variance (possible resulting in a lack of a group difference) and across-study variance (possibly resulting in a lack of study replication). An additional source of variance identified in the current experiment is within-participant variance, which cannot be assumed to be equivalent in different populations. One way around this variability is to include more measurements per subject along with larger samples, such as a recent study by Jones and colleagues (Jones et al., 2011), which included 99 individuals with autism. Other possible explanations for mixed findings across studies are the differences across tasks and stimuli, and the relationship between the two. For instance, simple labeling of a facial expression using full-intensity faces (such as Ekman stimuli) is a much easier task than labeling a less intense morphed expression, or than rating the intensity of the primary and non-primary emotions contained in a face (i.e., the task used in the current experiment). Finally, the above described sources of study variance also interact with the analysis procedure. In the current study, a significant effect would not have been found if the analysis were limited to comparing only concordant intensity ratings between groups, since it was the pattern of ratings (i.e., the relationship between concordant and discordant ratings) that differentiated groups. Future studies should carefully consider all of these factors when designing behavioral experiments that aim to test whether or not there are differences between autism and control groups, and in particular analyze the relative patterns of responses that participants make.

Finally, we should also highlight several limitations of the current work, which should be improved upon in future studies. First, our sample consisted of a very restricted subpopulation within the autism spectrum – namely, high-functioning male adults. It would be important to extend these results to younger children and lower-functioning individuals, where the impairment in facial emotion perception might be more pronounced, and where developmental and IQ effects can be studied. Similarly, a more complete description would require the inclusion of women with autism. Secondly, our sample size was relatively small (especially in the test-retest experiment). This precludes our ability to confidently resolve relationships between emotion selectivity and clinical or behavioral symptoms. A third concern is whether or not our participants fully understood the directions. This is a significant concern, since one could imagine obtaining similar results of reduced selectivity with participants who did not fully understand the task of assigning emotional intensities to the unintended emotions. However, we believe this is unlikely to account for our results, since all of our participants were high-functioning adults (with a mean full-scale IQ of 107), and expressed to us that they fully understood the directions. Furthermore, all participants gave varied ratings for the non-primary emotion in faces, suggesting that they did understand the task, and the overall pattern of ratings was very similar between autism and controls (see Figs. 2 and 3). Finally, the design of the experiment did not allow us to test whether this reduced selectivity and test-retest reliability is specific to judging the emotional intensity in faces, or whether it extends to different types of non-social categories. In other words, are the effects we observed here domain-general or domain-specific?

In conclusion, the current work provides further evidence for altered perception of emotions from facial expressions in a group of high-functioning adults with autism, with a particular finding: people with autism produce emotion judgments that are less

specific and less reliable. It will be important in future experiments to determine whether individuals with autism have abnormalities in judging emotional intensity in faces because they have abnormal emotion concepts, whether the impairment stems from abnormal attention to salient features of the face, and what the neural substrates of these impairments may be.

Acknowledgements

References

Adolphs, R., Damasio, H., Tranel, D., Cooper, G., & Damasio, A. R. (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *Journal of Neuroscience*, 20(7), 2683–2690.

Adolphs, R., Sears, L., & Piven, J. (2001). Abnormal processing of social information from faces in autism. *Journal of Cognitive Neuroscience*, 13(2), 232–240.

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, 372(6507), 669–672.

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. R. (1995). Fear and the human amygdala. *Journal of Neuroscience*, 15(9), 5879–5891.

Ashwin, C., Wheelwright, S., & Baron-Cohen, S. (2006). Attention bias to faces in Asperger Syndrome: a pictorial emotion Stroop study. *Psychological Medicine*, 36(6), 835–843.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: evidence from very high functioning adults with autism or asperger syndrome. *Journal of Child Psychology and Psychiatry*, 38(7), 813–822.

Corden, B., Chilvers, R., & Skuse, D. (2008). Avoidance of emotionally arousing stimuli predicts social-perceptual impairment in Asperger's syndrome. *Neuropsychologia*, 46(1), 137–147.

Diehl-Schmid, J., Pohl, C., Ruprecht, C., Wagenpfeil, S., Foerstl, H., & Kurz, A. (2007). The Ekman 60 Faces Test as a diagnostic instrument in frontotemporal dementia. *Archives of Clinical Neuropsychology*, 22(4), 459–464.

Dziobek, I., Bahnemann, M., Convit, A., & Heekeren, H. R. (2010). The role of the fusiform-amygdala system in the pathophysiology of autism. *Archives of General Psychiatry*, 67(4), 397–405.

Ekman, P. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. *Neuropsychology Review*, 20(3), 290–322.

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(4), 685–695.

Jones, C. R., Pickles, A., Falcaro, M., Marsden, A. J., Happe, F., Scott, S. K., et al. (2011). A multimodal approach to emotion recognition ability in autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, 52(3), 275–285.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, 59(9), 809–816.

Langdell, T. (1978). Recognition of faces: an approach to the study of autism. *Journal of Child Psychology and Psychiatry*, 19(3), 255–268.

Law Smith, M. J., Montagne, B., Perrett, D. I., Gill, M., & Gallagher, L. (2010). Detecting subtle facial emotion recognition deficits in high-functioning Autism using dynamic stimuli of varying intensities. *Neuropsychologia*, 48(9), 2777–2781.

Mayes, S. D., & Calhoun, S. L. (2008). WISC-IV and WIAT-II profiles in children with high-functioning autism. *Journal of Autism and Developmental Disorders*, 38(3), 428–439.

Neumann, D., Spezio, M. L., Piven, J., & Adolphs, R. (2006). Looking you in the mouth: abnormal gaze in autism resulting from impaired top-down modulation of visual attention. *Social Cognitive and Affective Neuroscience*, 1(3), 194–202.

Oliveras-Rentas, R. E., Kenworthy, L., Roberson, R. B., 3rd, Martin, A., & Wallace, G. L. (2012). WISC-IV profile in high-functioning autism spectrum disorders: impaired processing speed is associated with increased autism communication symptons and decreased adaptive communication abilities. *Journal of Autism and Developmental Disorders*, 42(5), 655–664.

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 35(2), 205–220.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders*, 32(4), 249–261.

Persad, S. M., & Polivy, J. (1993). Differences between depressed and nondepressed individuals in the recognition of and response to facial emotional cues. *Journal of Abnormal Psychology*, 102(3), 358–368.

Philip, R. C., Whalley, H. C., Stanfield, A. C., Sprengelmeyer, R., Santos, I. M., Young, A. W., et al. (2010). Deficits in facial, body movement and vocal emotional processing in autism spectrum disorders. *Psychological Medicine*, 40(11), 1919–1929.

Rutherford, M. D., & Towns, A. M. (2008). Scan path differences and similarities during emotion perception in those with and without autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 38(7), 1371–1381.

Schmitz, N., Daly, E., & Murphy, D. (2007). Frontal anatomy and reaction time in Autism. *Neuroscience Letters*, 412(1), 12–17.

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science*, 16(3), 184–189.

Spezio, M. L., Adolphs, R., Hurley, R. S., & Piven, J. (2007a). Abnormal use of facial information in high-functioning autism. *Journal of Autism and Developmental Disorders*, 37(5), 929–939.

Spezio, M. L., Adolphs, R., Hurley, R. S., & Piven, J. (2007b). Analysis of face gaze in autism using "Bubbles". *Neuropsychologia*, 45(1), 144–151.

Sprengelmeyer, R., Young, A. W., Mahn, K., Schroeder, U., Woitalla, D., Buttner, T., et al. (2003). Facial expression recognition in people with medicated and unmedicated Parkinson's disease. *Neuropsychologia*, 41(8), 1047–1057.

Tager-Flusberg, H. (1981). On the nature of linguistic functioning in early infantile autism. *Journal of Autism and Developmental Disorders*, 11(1), 45–56.

Tudusciuc, O., & Adolphs, R. (2011). Recognition of context-dependent emotion in autism. *International Meeting For Autism Research*

Van Lancker, D., Cornelius, C., & Kreiman, J. (1989). Recognition of emotional-prosodic meanings in speech by autistic, schizophrenic, and normal children. *Developmental Neuopsychology*, 5(2-3), 207–226.

Wallace, G. L., Case, L. K., Harms, M. B., Silvers, J. A., Kenworthy, L., & Martin, A. (2011). Diminished sensitivity to sad facial expressions in high functioning autism spectrum disorders is associated with symptomatology and adaptive functioning. *Journal of Autism and Developmental Disorders*, 41(11), 1475–1486.