Social Perception in Autism Spectrum Disorders: Impaired Category Selectivity for Dynamic but not Static Images in Ventral Temporal Cortex

Jill Weisberg¹, Shawn C. Milleville¹, Lauren Kenworthy¹,², Gregory L. Wallace¹, Stephen J. Gotts¹, Michael S. Beauchamp³ and Alex Martin¹

¹NIMH, Laboratory of Brain and Cognition, Bethesda, MD 20850, ²Children’s National Medical Center, Center for Autism Spectrum Disorders, Rockville, MD 20850 and ³Department of Neurobiology and Anatomy, University of Texas Medical School at Houston, Houston, TX 77030, USA

Address correspondence to Jill Weisberg, San Diego State University Research Foundation, Laboratory for Language and Cognitive Neuroscience, 6495 Alvarado Rd, Suite 200, San Diego, CA 92120, USA. Email: jweisberg@projects.sdsu.edu

Studies of autism spectrum disorders (ASDs) reveal dysfunction in the neural systems mediating object processing (particularly faces) and social cognition, but few investigations have systematically assessed the specificity of the dysfunction. We compared cortical responses in typically developing adolescents and those with ASD to stimuli from distinct conceptual domains known to elicit category-related activity in separate neural systems. In Experiment 1, subjects made category decisions to photographs, videos, and point-light displays of people and tools. In Experiment 2, subjects interpreted displays of simple, geometric shapes in motion depicting social or mechanical interactions. In both experiments, we found a selective deficit in the ASD subjects for dynamic social stimuli (videos and point-light displays of people, moving geometric shapes), but not static images, in the functionally localized lateral region of the right fusiform gyrus, including the fusiform face area. In contrast, no group differences were found in response to either static images or dynamic stimuli in other brain regions associated with face and social processing (e.g. posterior superior temporal sulcus, amygdala), suggesting disordered connectivity between these regions and the fusiform gyrus in ASD. This possibility was confirmed by functional connectivity analysis.

Keywords: Asperger’s syndrome, autism, fusiform gyrus, MRI/fMRI, social cognition

Introduction

A core feature of autism spectrum disorders (ASDs) is impaired social interaction, but the precise nature and the neural basis of this impairment remain unclear. It is well established that a rich assortment of social information is gleaned simply by observing human motion (Blake and Shiffrar 2007). However, while typically developing (TD) individuals do so effortlessly, several lines of evidence indicate that from a very young age, individuals with ASD may lack this ability. For example, when viewing whole-body point-light animations, TD individuals implicitly infer physical characteristics such as gender, age, and identity (Cutting and Kozlowski 1977; Kozlowski and Cutting 1977; Mather and Murdoch 1994; Sumi 2000; Troje 2002), as well as higher order characteristics concerning affect, motivation, and intention (Walk and Homan 1984; Dittrich et al. 1996; Pollick et al. 2001; Clarke et al. 2005). These inferences about physical and higher order characteristics occur quickly and spontaneously, even though point-light displays lack information about color, form, and texture. Children and adolescents with ASD are sometimes impaired at point-light human action recognition (Blake et al. 2003; Koldewyn et al. 2010). They may fail to distinguish between animate and inanimate objects as readily as TD children (Rutherford et al. 2006; Congiu et al. 2010) and prefer viewing moving geometric patterns to videos of other children (Pierce et al. 2011). In addition, when describing the animations in which simple geometric shapes in motion depict social interaction, children and adolescents with autism and Asperger’s syndrome produce less appropriate descriptions than their TD counterparts (Abell et al. 2000; Klin 2000; Campbell et al. 2006; Klin and Jones 2006; Salter et al. 2008) and include fewer words referring to mental states (Klin 2000; Castelli et al. 2002). These atypical behavioral patterns appear to emerge early, as demonstrated by infants with ASD spending more time focused on non-social than social physical contingencies (Klin et al. 2009).

In TD adults, a predominantly right hemisphere network, including the lateral part of the fusiform gyrus (which includes the fusiform face area, FFA; Kanwisher et al. 1997; and fusiform body area; Peelen and Downing 2005; Schwarzoese et al. 2005), the posterior superior temporal sulcus (pSTS), the ventromedial prefrontal cortex, and the amygdala, is spontaneously engaged by depictions of animate entities, relative to inanimate objects (see Blakemore 2008; Adolphs 2009; Martin 2009; Frith and Frith 2010 for recent reviews). Evidence indicates that performing tasks that require higher order social inferences, for example, when interpreting moving geometric shapes as social interactions (modeled after Heider and Simmel 1944), much of this same neural circuit becomes active (Castelli et al. 2000, 2002; Martin and Weisberg 2003; Schultz et al. 2003; Wheatley et al. 2007).

We exploited these findings to conduct a detailed investigation of the integrity of the neural system mediating social cognition in high-functioning adolescents with ASD by comparing brain activity associated with viewing stimuli from two contrasting conceptual domains: animate entities (people) and inanimate, mechanical objects. In Experiment 1, subjects viewed human forms and man-made, manipulable objects (i.e. tools) presented as static pictures, as full color videos, and as moving point-light displays, allowing assessment of the integrity of the neural system mediating human form and motion perception across varying stimulus formats (Beauchamp et al. 2002, 2003). To investigate responses to higher order concepts, subjects in Experiment 2 viewed animated vignettes of moving geometric shapes depicting social or mechanical action (Martin and Weisberg 2003). Thus, we were able to assess neural responses during inferences based solely on the motion of identical geometric forms.
Materials and Methods

Subjects

Twenty-four high-functioning adolescents with ASD (3 females) and 19 TD adolescents (1 female) participated in Experiment 1. Twenty-five ASD adolescents (3 female) and 24 TD adolescents (2 females) participated in Experiment 2. Of these, thirteen ASD and 11 TD subjects participated in both experiments. Data from 1 ASD subject common to both experiments was excluded because his standardized IQ measure did not meet the criterion of IQ ≥ 85, and data from 1 ASD subject in Experiment 1, 7 in Experiment 2, and 6 TD subjects in Experiment 2 were excluded from analysis due to excessive movement during scanning (see the Data analysis section). Demographic and clinical characteristics of the remaining 22 ASD subjects in Experiment 1, 17 in Experiment 2, and the age- and IQ-matched TD subjects for each experiment (TD n = 19 in Experiment 1 and n = 18 in Experiment 2) are described in Table 1. All subjects were recruited from the Washington DC metropolitan area. An experienced clinician verified that each subject in the ASD group met the DSM-IV (American Psychiatric Association 2000) clinical criteria for ASD (i.e. autistic disorder, Asperger’s disorder, or pervasive developmental disorder—not otherwise specified). In addition, scores from the Autism Diagnostic Interview (ADI or ADI-R; LeCouteur et al. 1989; Lord et al. 1994) and/or Autism Diagnostic Observation Schedule (ADOS; Lord et al. 2000), as administered by a trained and research-reliable clinician, indicated that all participating ASD subjects met research criteria for “broad ASD” according to guidelines established by the NICHD/NIDCD Collaborative Programs for Excellence in Autism (Lainhart et al. 2006). Because the ADI and ADOS do not provide an algorithm for Asperger’s syndrome, Lainhart et al. developed research criteria that include an individual on the broad autism spectrum if s/he meets the ADI cut-off for “autism” in the social domain and at least one other domain, or meets the ADOS cut-off for the combined social and communication score. All subjects (and their parents, if minors) gave informed consent under a protocol approved by the NIMH IRB and were paid for participation.

Stimuli and Experimental Design

Experiment 1

Subjects viewed depictions of people and tools in 3 formats: static images, full color video clips of moving images, and moving point-light displays (see Beauchamp et al. 2002, 2003 for details). Thus, we presented a total of 6 types of visual stimuli. Examples of human actions include jumping jacks, climbing up or down stairs, and soccer kicks; examples of tool stimuli moving in their characteristic manners include a chef’s knife, hammer, paintbrush, and scissors. The video clips were recorded against a gray background using human actors, and the tools moved realistically but without a visible means of manipulation (Beauchamp et al. 2002). Static images consisted of frames selected from the video stimuli that best suggested the type of movement performed and showed the human or tool in mid-action. The point-light stimuli were created using a motion capture system recording small reflective targets placed on the human actor or tool, illuminated by camera-mounted light-emitting diodes (Beauchamp et al. 2003). Each stimulus was overlaid with a white central fixation cross and presented on a gray background (Fig. 1A).

We used a rapid event-related experimental design and recorded responses with Presentation software (Neurobehavioral Systems, Inc.) running on a Dell Inspiron 8500 notebook computer. Across the experiment, half of the experimental trials depicted people and half depicted tools, equally represented across the 3 image formats (still, video, point-light). For each stimulus type, 50 trials were presented across 4 separate imaging runs (an average of 12.5 trials per stimulus type per run), with each run lasting 5 min. Stimuli were presented for 2500 ms (video clips and point-light trials were displayed at 30 frames/s) followed by a white fixation cross on a gray background for 500 ms. The sequence optimizer program OptSeq2 was used to randomize stimulus presentation order and calculate jitter for null trials (fixation baseline). Each imaging run began and ended with a fixation period (16 and 22 s, respectively) and included 37 s of null trials (average duration per null trial = 1059 ms, range = 1000–8000 ms) randomly interspersed among experimental trials. The order of runs was counterbalanced across subjects. Stimuli were rear-projected onto a screen positioned at the foot of the scanning bed and visible via a mirror attached to the head coil.

To ensure that subjects attended to each image during scanning, they were instructed to press one button for depictions of people and another button for tools as quickly and accurately as possible, using a response device held in the left hand. Subjects had one practice trial with each stimulus type prior to scanning to ensure that they understood the task instructions.

Experiment 2: Vignette Task

Sixteen animated vignettes used simple geometric shapes in motion to depict either social (8 vignettes) or mechanical (8 vignettes) action (see Martin and Weisberg 2003 for details). The social vignettes elicited interpretations of the moving shapes as social interactions such as playing baseball, dancing, sharing, scaring, and swimming (Fig. 1B); the mechanical vignettes elicited interpretations of mechanical interaction or objects such as billiards, a cannon, and a conveyor belt. The numbers and types of shapes were similar in the social and mechanical vignettes. In addition, subjects viewed 16 control vignettes which used the same geometric stimuli as the meaningful

Table 1

<table>
<thead>
<tr>
<th>Subject group characteristics for each experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Experiment 1</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>n (males:females)</td>
</tr>
<tr>
<td>Age in years</td>
</tr>
<tr>
<td>Full-scale IQ</td>
</tr>
<tr>
<td>Social interaction</td>
</tr>
<tr>
<td>Verbal communication</td>
</tr>
<tr>
<td>Non-verbal communication</td>
</tr>
<tr>
<td>Repetitive behaviors</td>
</tr>
<tr>
<td><strong>ADOS</strong></td>
</tr>
<tr>
<td>Communication</td>
</tr>
<tr>
<td>Social interaction</td>
</tr>
<tr>
<td><strong>Experiment 2</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>n (males:females)</td>
</tr>
<tr>
<td>Age in years</td>
</tr>
<tr>
<td>Full-scale IQ</td>
</tr>
<tr>
<td>Social interaction</td>
</tr>
<tr>
<td>Verbal communication</td>
</tr>
<tr>
<td>Non-verbal communication</td>
</tr>
<tr>
<td>Repetitive behaviors</td>
</tr>
<tr>
<td>ADOS</td>
</tr>
<tr>
<td>Communication</td>
</tr>
</tbody>
</table>

Numbers in parentheses indicate standard deviations.
*ADOS scores were not available for one participant in each experiment.
stimuli were presented using QuickTime (Apple, Inc.) and responses recorded using Presentation software. In each run, 4 blocks presented meaningful vignettes (2 social, 2 mechanical), alternating with 4 control blocks (2 random motion, 2 static). An 8 s fixation period was added to the beginning and end of each run. Run order was counterbalanced across subjects and conditions were counterbalanced across runs with the constraint that each run began with a control vignette. To ensure that they understood the task instructions, subjects were trained and received practice with meaningful (2 social, 1 mechanical), and control vignettes (2 random motion, 1 still picture). If they chose an incorrect answer for a meaningful vignette, the vignette and the correct answer were explained and a second round of practice occurred as necessary. Practice vignettes were not used during scanning.

A subset of the subjects ($n = 8$ TD, 10 ASD) returned to the laboratory approximately 3 months after scanning for additional behavioral testing. At this time, the subjects viewed each meaningful vignette they had seen during scanning and were asked to first state what they thought it was about and then select the phrase that best described each from the response screen, but without pressure to respond within a limited time period (i.e. during the 4 s period allotted during scanning). Each vignette was then shown again and the subjects were asked to describe it during viewing. The audio from these sessions was recorded and transcribed for scoring (see the Data Analysis section).

**Experiment 2: Localizer Task**

To functionally localize object category-responsive brain regions, we collected 2 additional imaging runs during which subjects viewed alternating blocks of black and white photographs of faces, tools, and phase-scrambled versions of these pictures (Fig. 1C). Subjects were instructed to respond as quickly and accurately as possible by button press (held in the left hand) each time the exact same picture repeated (i.e. a repetition detection task). Each face and tool block contained a total of 10 photographs, with 3 or 4 repetitions per block, and the phase-scrambled blocks contained 5 images, with 1 or 2 repetitions per block. Each stimulus was presented for 500 ms, followed by a 1500 ms interstimulus interval containing a fixation cross. Each 3’16” run contained six 20 s blocks of photographs (3 of faces, 3 of tools), and six 10 s blocks of scrambled pictures (3 of scrambled faces, 3 of scrambled tools), plus an 8 s fixation period at the beginning and end of each run. Object categories alternated throughout each run, with each followed by a baseline block of scrambled pictures from the same category. The first block contained faces in one run and tools in the other, with run order counterbalanced across subjects. Stimuli were presented and responses collected using Presentation software. To familiarize subjects with the repetition detection task, they were given a practice run with stimuli not used during scanning.

**Imaging Parameters**

Magnetic resonance imaging (MRI) data for Experiments 1 and 2 were collected in separate sessions on a GE 3 T scanner using an 8-channel...
phased-array head coil. The average time between sessions was 101 days. In each session, acquisition of a high-resolution anatomical scan (MPRAGE 128 axial slices, 1.2 mm thick, field of view (FOV) = 24 cm, acquisition matrix = 256 x 256) was followed by gradient-echo echo-planar scans (repetition time = 2 s, echo time = 30 ms, flip angle = 90°, FOV = 22 cm for Experiment 1 and 24 cm for Experiment 2, acquisition matrix = 64 x 64, 31 contiguous slices, 3.8 mm thick, number of brain volumes = 150 for Experiment 1, 152 for the vignette task in Experiment 2, and 98 for the localizer task in Experiment 2).

**Data Analysis**

**Behavioral Data**

Category decision accuracy rates (percentage correct trials) collected during Experiment 1 scanning sessions were submitted to a Group (ASD, TD) x Category (human, tool) x Format (static, video, point-light) repeated-measures analysis of variance (ANOVA). Accuracy rates collected during the Localizer task in Experiment 2 were analyzed with a Group (ASD, TD) x Category (Face, Tool) repeated-measures ANOVA. Data for 2 ASD subjects were lost due to technical difficulties. One additional ASD subject’s data were excluded from this analysis due to near-chance task performance (<61% accuracy).

Vignette task response data collected during scanning for each subject were scored as correct or incorrect, and incorrect responses were further classified according to the type of incorrect response selected (possible but wrong, or unlikely and wrong). The mean scores (as a percent of total trials for each condition) for each response type were submitted to separate Group (ASD, TD) x Category (social, mechanical) ANOVAs. For data for 2 ASD subjects were lost due to technical difficulties. Vignette descriptions collected during the post-scan behavioral testing session were scored for 1) phrases denoting intentionality, 2) number of pronouns used, and 3) length of description (number of clauses). The intentionality score reflected the type of deliberate actions described, with zero points scored if no deliberate action was described: 1 = solo action; 2 = action with others; 3 = action in response to another’s action; 4 = action with reference to mental states; and 5 = action with the explicit goal of affecting another’s mental state (Abell et al. 2000; Castelli et al. 2000). Subject’s mean scores for each vignette type were submitted to a separate Group (ASD, TD) x Vignette Type (social, mechanical) ANOVA for each measure.

**Imaging Data**

All MRI data were processed and analyzed using the AFNI Software Package (Cox 1996). After pre-processing and prior to group analyses, each individual’s data were transformed into the stereotactic space of Talairach and Tournoux (1988). We registered each individual’s anatomical scan to the TT_N27 template in AFNI using 12-parameter-affine registration, and report coordinates converted via the Lancaster transformation (Lancaster et al. 2007; Laird et al. 2010) in the Montreal Neurological Institute (MINI) space (Evans et al. 1992). During pre-processing, the first 4 volumes of each time series were discarded to allow for equilibrium magnetization and all remaining volumes were slice-time corrected and registered to one volume of the time series collected nearest in time to the acquisition of the high-resolution anatomical scan. The motion parameters output by the volume registration algorithm were examined to determine any time points for which the square root of the sum of squares of the derivatives exceeded 1 (equal to a translation of 1 mm in any one translational direction, or a combination of translations of 0.577 mm in all 3 translational directions; Jones et al. 2010). When this occurred, we censored that time point plus the immediately preceding and following time points from further statistical analysis.

After spatial smoothing (4.5 mm full-width at half-maximum Gaussian filter), we implemented multiple regression analysis, with regressors of no interest including the 6 parameter motion estimates from the volume registration algorithm and polynomial regressors to account for baseline shifts and linear drifts in each scan.

For Experiment 1, we calculated the response to each stimulus type compared with the fixation baseline. For each of the 6 regressors of interest (one for each stimulus type), we fit tent basis functions (representing the response to each stimulus) to the fMRI signal at each half volume of echo-planar acquisition (1 s resolution) over a 12 s window, with no assumptions about the shape of the hemodynamic response, resulting in an estimate of the response to a single stimulus type for each second. The resulting deconvolution provided voxel-wise 12 s time series with a 1 s resolution for each stimulus type. The response magnitude to each stimulus type was calculated by summing the beta weights of the regressors from seconds 2 to 8 of the response at each voxel, resulting in one estimate per subject per stimulus type (see Beauchamp et al. 2002 for details). We performed a mixed-effects ANOVA on the summed beta weights with Group (TD, ASD), Category (human, tool), and Stimulus Format (static, video, point-light) as fixed factors and with subjects as a random factor. Unless otherwise stated, all regions in this report were significant at a voxel-wise false discovery rate (FDR) corrected threshold, designated as q.

In addition, to provide an unbiased assessment of the response to the dynamic stimuli (video and point-light displays), we identified regions of interest (ROIs) that showed a category effect based solely on the static stimuli. To accomplish this, we created a mask consisting of voxels exceeding a threshold of q < 0.05 for the contrast of static people vs. static tools, to serve as a within-experiment, category localizer. For each subject, we calculated the average beta weight within each independently identified ROI in this mask for each of the 4 dynamic conditions (human video, human point-light, tool video, tool point-light) and submitted those values to mixed-effects ANOVAs.

For the face–tool localizer scans in Experiment 2, we estimated the response to each condition vs. its corresponding baseline with one regressor of interest for each category (faces, tools), and the baseline for the regression model comprised of scrambled pictures. Voxel-wise response estimates for each stimulus type (i.e. beta weights) for each subject were submitted to a mixed-effects Group (TD, ASD) x Category (faces, tools) ANOVA. In the resulting statistical map, we identified voxels that surpassed a statistical threshold of q < 0.05 for the effect of Group, Category, or an interaction between them. Clusters of contiguous voxels (unrestricted by cluster size because we applied FDR correction) identified by this analysis served as masks to extract time series from the functional data of each subject in regions where the condition of interest (i.e. faces, tools, social, mechanical) elicited a response significantly greater than the baseline. Thus, clusters identified by the localizer task served as ROIs to extract time series from the vignette data. This allowed us to examine responses to the social vignettes in regions that were independently identified as showing a preference for faces. To calculate voxel-wise estimates of responses to the social and mechanical vignettes, we included one regressor of interest for each vignette type (12 time points for each vignette), leaving all remaining time points (including static shapes) as the baseline for the model. For each subject, we calculated the average response to each vignette type within each face–tool localizer cluster and entered the results into a mixed effects Group (ASD, TD) x Vignette Type (social, mechanical) ANOVA.

To assess group differences in functional connectivity, we pre-processed the functional data from the vignette task by removing signal related to motion artifact, ventricles, and white matter (see Jo et al. 2010 for details) along with their respective temporal derivatives (Fox et al. 2005). It has become increasingly clear that analyses of group differences in functional connectivity are particularly sensitive to differences in head motion (e.g. Power et al. 2012). Following time point censoring as described above, we evaluated group differences in motion using root mean square (RMS) head position change (Power et al. 2012) and a sharp motion index similar to mean motion in motion using root mean square (RMS) head position change (Power et al. 2012).
and then this matrix was reapplied to all subsequent TRs in the scan. This is equivalent to starting the head position where it left off just before the cut TRs, allowing subsequent motion events to occur relative to the realigned position. Results of these analyses confirmed that the groups did not differ on either measure of motion prior to, or following, data scrubbing (Mann–Whitney U-test, RMS measure; \(P = 0.575\) before and \(P = 0.729\) after time point censoring for the localizer data; \(P = 0.544\) before and \(P = 0.380\) after for the vignette task data. Mann–Whitney U-test, sharp motion measure; \(P = 0.77\) before and \(P = 0.77\) after for the localizer data; \(P = 0.95\) before and \(P = 0.113\) after censoring for the vignette task data). We then applied a smoothing filter, scaled the data to a mean of 100, and performed multiple regression as previously described to calculate residual fluctuations (remove task effects) for each individual in Experiment 2 (see Jones et al. 2010, Method 3, for details). We created 2 residual time series: one contained only time points from the social animation task blocks and the other contained only time points from the mechanical animation task blocks. We created seed regions by averaging each residual time series across the voxels within a 5 mm sphere located within the right lateral fusiform gyrus, centered on the local maxima for the main effect of Category from the localizer task (faces > tools), and obtained Pearson’s correlation coefficients between these seed ROIs and all other brain voxels (using 3dTcorr1D in AFNI). For each individual, we then averaged the correlation coefficients for the 2 residual time series (social, mechanical), applied a Fisher-Z transform, and identified differences between groups using 3dttest in AFNI. Significant differences \((P < 0.05)\) identified in the right pSTS and amygdala were further examined by extracting the average correlation across a 5 mm spherical ROI centered on the peak \(t\)-values and submitting these ROI-averaged correlation coefficients for each individual to independent sample between-group \(t\)-tests.

**Results**

**Experiment 1: Behavioral Data**

The TD and ASD groups were highly accurate (>90% correct) with no effect of group \((F < 1)\) or interactions with group \((all \(F S < 2\)) for the category decision task in Experiment 1. Performance was comparable for both human and tool stimuli (main effect of Category, \(F < 1)\). Thus, the task was no more difficult for individuals with ASD than for TD individuals, regardless of whether the stimuli depicted people or tools, or were static or dynamic.

**Experiment 1: Imaging Data**

The mixed-effects whole-brain analysis revealed a strong effect of Category \((q < 0.05)\) (Table 2), with a preference for depictions of people in bilateral regions of the lateral fusiform gyrus, pSTS and the amygdala, stronger in the right hemisphere than the left. This analysis also revealed heightened activity for viewing images of people relative to tools in the hippocampus, bilaterally, and in the right cuneus and precuneus. A separate, largely left lateralized network of regions displayed the opposite preference, responding more to tools than to people, with clusters of heightened activation located in the ventral temporal cortex in the medial part of the fusiform gyrus (bilaterally, but stronger on the left) and in the left parahippocampal gyrus, as well as in the posterior region of the middle temporal gyrus, posterior intraparietal sulcus, and ventral premotor cortex.

The main effect of Group was not significant in any region. There was, however, a Group x Category interaction in one region: the lateral portion of the right fusiform gyrus \((q < 0.05)\), local maxima located at \(x = 41, y = -49, z = -15\). In this region, the response in the TD group was stronger when viewing depictions of people than when viewing tools \((people > tools, t_{18} = 5.44, P < 0.001)\), but there was no such category preference in the ASD group \((P > 0.05)\). Moreover, although no region showed a significant 3-way Group x Category x Stimulus Format interaction using a corrected \(P\)-value, this 3-way interaction was noted within the right lateral fusiform cluster at an uncorrected threshold (peak voxel, \(F = 3.04, P < 0.05, x = 50, y = -59, z = -21\)).

<table>
<thead>
<tr>
<th>Region</th>
<th>Size (mm^3)</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>(t)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>People &gt; tools</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R precuneus</td>
<td>36,990</td>
<td>20</td>
<td>-56</td>
<td>16</td>
<td>7.76</td>
</tr>
<tr>
<td>R extrastriate body area</td>
<td>26,109</td>
<td>48</td>
<td>-72</td>
<td>7</td>
<td>9.48</td>
</tr>
<tr>
<td>Superior temporal sulcus</td>
<td>5589</td>
<td>16</td>
<td>-32</td>
<td>-3</td>
<td>7.51</td>
</tr>
<tr>
<td>L hippocampus</td>
<td>4023</td>
<td>45</td>
<td>-43</td>
<td>-26</td>
<td>6.52</td>
</tr>
<tr>
<td>R hippocampus</td>
<td>3106</td>
<td>-9</td>
<td>-32</td>
<td>-2</td>
<td>5.70</td>
</tr>
<tr>
<td>R amygdala</td>
<td>2268</td>
<td>22</td>
<td>-7</td>
<td>15</td>
<td>5.20</td>
</tr>
<tr>
<td>L lateral fusiform gyrus</td>
<td>1026</td>
<td>-42</td>
<td>-40</td>
<td>-21</td>
<td>5.11</td>
</tr>
<tr>
<td>L amygdala</td>
<td>945</td>
<td>-3</td>
<td>-1</td>
<td>15</td>
<td>4.05</td>
</tr>
<tr>
<td>Tools &gt; people</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L posterior intraparietal sulcus</td>
<td>10,179</td>
<td>-35</td>
<td>-44</td>
<td>52</td>
<td>5.02</td>
</tr>
<tr>
<td>R medial fusiform gyrus</td>
<td>8721</td>
<td>-29</td>
<td>-52</td>
<td>-17</td>
<td>8.76</td>
</tr>
<tr>
<td>R medial fusiform gyrus</td>
<td>3537</td>
<td>32</td>
<td>-61</td>
<td>-10</td>
<td>6.11</td>
</tr>
<tr>
<td>L inferior frontal gyrus</td>
<td>1917</td>
<td>-42</td>
<td>27</td>
<td>20</td>
<td>4.61</td>
</tr>
<tr>
<td>L extrastriate body area</td>
<td>1431</td>
<td>-32</td>
<td>38</td>
<td>-8</td>
<td>5.08</td>
</tr>
<tr>
<td>L parahippocampal gyrus</td>
<td>1647</td>
<td>-35</td>
<td>-24</td>
<td>-23</td>
<td>5.73</td>
</tr>
<tr>
<td>L frontal gyrus(^*)</td>
<td>270</td>
<td>-58</td>
<td>4</td>
<td>15</td>
<td>2.69</td>
</tr>
</tbody>
</table>

Note: All regions active at \(q < 0.05\), except \(* q = 0.065\).
Experiment 2: Behavioral Data

Analysis of the behavioral data from the face-tool localizer task indicated that responses were highly accurate and at or near ceiling for both subject groups (percent accuracy, TD = 99.6%, ASD = 97.2%).

Subjects interpreted the mechanical vignettes more accurately than the social vignettes (mechanical vignettes = 82.6%, social vignettes = 75.5%; P < 0.05). In addition, TD subjects were more accurate than ASD subjects (TD = 86.8%, ASD = 71.3%; P < 0.001). There was also a significant Group × Vignette Type interaction (F(1,31) = 6.30; P < 0.05) reflecting the fact that whereas the TDs interpreted the social and mechanical vignettes equally accurately (86.8% for both vignette types), ASDs were more accurate with the mechanical (78.3%) than social vignettes (64.2%; P < 0.05). Interpretation of these results is complicated by the fact that the ASD group failed to choose an answer much more frequently than the TD group (ASD = 23.8%, TD = 6.6%; P < 0.001). Moreover, whereas failures to respond were equally divided among vignette types for TD subjects (social vignettes = 6.3%; mechanical vignettes = 6.9%), the ASD group failed to respond disproportionately more often for social (30%) than for mechanical vignettes (17.5%; P < 0.05). Analysis of the data collected outside the scanner approximately 3 months after the scanning session suggested that this group difference in accuracy was largely due to the time constraint imposed during scanning. Specifically, when the ASD subjects returned for testing (n = 10), their accuracy rate was near ceiling (>95%) and equivalent to that of TD individuals who also returned for testing (n = 8) (Fs < 1). Verbal descriptions of the vignettes were scored based on measures of intentionality, number of pronouns used, and number of phrases uttered. Separate Group × Vignette Type ANOVAs for each measure indicated that the social vignettes were consistently described with higher levels of intentionality, more pronouns, and longer utterances than mechanical vignettes (P < 0.001 for each measure). Importantly, neither the group effects nor interactions were significant for any measure, indicating that the ASD and TD groups produced qualitatively similar verbal descriptions for the meaningful vignettes (all Fs < 2).

Experiment 2: Imaging Data

The whole brain mixed-effects ANOVA of the functional localizer scans (Group [ASD, TD] × Category [faces, tools]) revealed no main effect of group or interaction with group, indicating that when attentively viewing photographs of faces and tools, brain activity in the ASD group was comparable to that of the TD group. As expected, several regions showed a main effect of Category (F(1,33) = 13.17, q < 0.05) (Table 4). Consistent with the results of Experiment 1, nearly every ASD subject (15 of 17) demonstrated enhanced activity for faces, relative to tools, in the lateral part of the right fusiform gyrus (mean peak coordinates [SD], x = 43 [5.81], y = −52 [5.47], z = −18 [3.63]; Fig. 2 and Table 5), and again, peak coordinates were highly similar to those of the TD subjects (x = 43 [4.62], y = −54 [8.81], z = −18 [3.81]), with no significant differences in the location of the FFA (Group × Coordinate ANOVA, F < 1 for main effect of Group and interaction). Also consistent with Experiment 1, a network of predominantly right hemisphere regions showed heightened responses to faces relative to tools.
to tools (including the lateral fusiform gyrus, pSTS, and amygdala), and a separate, predominantly left hemisphere network showed the opposite preference (medial fusiform and parahippocampal gyrus, inferior frontal gyrus, inferior parietal lobule).

We next evaluated responses to the social and mechanical vignettes within the face-responsive ROIs identified by the main effect of Category in the localizer task. In most regions, we found a greater response to the social than mechanical vignettes, including in the right pSTS (Group [TD, ASD] × Vignette Type [social, mechanical] ANOVA, main effect of Vignette Type, $F_{1,33} = 46.52$, $P < 0.001$) and the right amygdala ($F_{1,33} = 3.62$, $P = 0.066$) (Fig. 5). There was no effect of Group or a Group × Vignette Type interaction in these regions.

However, we did find a significant interaction between Group and Vignette Type in the lateral fusiform gyrus in the
right hemisphere ($F_{1,33} = 4.49$, $P < 0.05$; Fig. 6). Although faces elicited a markedly stronger response than tools in this region across both groups during the localizer task, only the TD group showed an increased response for social, relative to mechanical vignettes ($F_{1,7} = 2.216$, $P < 0.05$). Vignette Type had no modulatory effect on responses in the ASD group ($I < 1$). No other face-prefering region (and no tool-prefering regions) identified by the localizer task demonstrated such an interaction. Thus, as in Experiment 1, the only region in Experiment 2 to show both a category-related response and an interaction with diagnosis was the face-prefering right lateral fusiform gyrus.

Functional connectivity analysis with task effects removed from the vignette task data revealed group differences in the strength of connections between the right lateral fusiform gyrus and other brain regions involved in social cognition. Specifically, the right lateral fusiform region that showed a preference for photographs of faces, relative to tools, showed weaker connections with right pSTS ($t_{33} = 2.34$, $P < 0.05$; TD mean $r = 0.35$, SE = 0.04, ASD mean $r = 0.24$, SE = 0.05) and with the right amygdala ($t_{37} = 2.47$, $P < 0.05$; TD mean $r = 0.39$, SE = 0.04, ASD mean $r = 0.24$, SE = 0.05) in the ASD group, relative to the control group.

### Discussion
To examine the neural systems supporting object perception and social cognition in adolescents with and without autism, we compared responses to stimuli from different conceptual domains using multiple stimulus formats and tasks. The results suggest that object perception systems are largely intact in autism. Specifically, no group differences were found in response to static images of people (Experiment 1), faces (Experiment 2), or tools (Experiments 1 and 2). In addition, no group differences were observed in response to videos of tools in motion, to point-light displays of tools in motion, or to abstract depictions of mechanical interactions represented by moving geometric shapes. Comparable selectivity across groups for static photographs of people and faces in our studies is consistent with a growing body of evidence that the face perception network—including the right lateral fusiform gyrus—responds normally to neutral faces in ASD. Reports of reduced activation in the right lateral fusiform gyrus are often related to the use of implicit tasks, passive viewing, and reduced attention, each of which can result in absent or reduced responses in ASD (Schultz et al. 2000; Pierce et al. 2001; Hubl et al. 2003; Deelley et al. 2007; Humphreys et al. 2008; Scherf et al. 2010). In contrast, tasks such as ours that actively engage ASD subjects and require focus on the face often result in normal face selectivity (Hadjikhani et al. 2004, 2007; Pierce et al. 2004; Dalton et al. 2005; Bird et al. 2006; Kleinhans et al. 2008, 2009). The lack of group differences in regions selective for depictions of tools in Experiments 1 and 2 is consistent with reports of normal neural responses in ASD groups in regions preferring other non-social stimuli such as houses (Schultz et al. 2000; Bird et al. 2006; Humphreys et al. 2008; Kleinhans et al. 2008; Scherf et al. 2010).

However, the results of both experiments converged to implicate dysfunction in the right lateral fusiform gyrus of subjects with ASD when confronted with dynamic stimuli. Specifically, whereas TD subjects showed a strong category preference in this region for dynamic social stimuli (videos of people in motion, point-light displays of human actions, and abstract depictions of social interactions represented by moving geometric shapes, relative to similar depictions of tools or mechanical action), no preference was evident in the ASD group. Thus, in the ASD group, dynamic stimuli elicited an aberrant response pattern in a distinct region specialized for perception of animate beings, consistent with the idea that this region may play a prominent role in at least some aspects of impaired social functioning in ASD (Schultz et al. 2003).

It was not the case, however, that this region of the fusiform gyrus responded weakly to dynamic stimuli in ASD subjects. Rather, it responded strongly but without the expected category preference for animate/social stimuli. This lack of category-related activity for dynamic social stimuli, but not for static photographs, suggests that activity in this region is not...
Figure 5. Experiment 2: Responses to the localizer and vignette tasks in the right amygdala and right pSTS. (A) Activation maps show the main effect of Category (faces vs. tools, $q < 0.05$) from the face localizer Group × Category ANOVA, in the right amygdala (circled, top) and right pSTS (circled, bottom). Regions in red were more active for viewing photographs of faces than tools and regions in blue showed the opposite preference. Histograms in (A) show group-averaged hemodynamic responses to faces and tools within the circled amygdala (top) and pSTS (bottom) regions identified by the localizer task. Asterisks indicate a significant effect of Vignette Type ($P < 0.001$ in pSTS; $P < 0.05$ in amygdala). Crosshatch indicates a trend in the same direction (social > mechanical, $P = 0.066$) for the ASD group.

Figure 6. Experiment 2: Responses to the face localizer and vignette tasks in the right lateral fusiform gyrus. (A) As in Figure 5, activations show the main effect of Category (faces vs. tools, $q < 0.05$) for the localizer Group × Category ANOVA. Circle indicates the FFA ROI in the right lateral fusiform gyrus, shown on a coronal slice. The histogram below shows the group-averaged hemodynamic responses to faces and tools during the localizer task within the circled region. (B) Group-averaged hemodynamic responses during the vignette task extracted from the FFA region identified by the localizer task. Asterisk indicates a significant effect of Vignette Type ($P < 0.05$).
modulated appropriately by one or more other nodes in the social network specialized for processing dynamic depictions of animacy (e.g., pSTS and/or the amygdala; Bonda et al. 1996; Grossman et al. 2000; Beauchamp et al. 2002, 2003; Martin and Weisberg 2003; Pelphrey et al. 2003; Schultz 2005; Wheatley et al. 2007; see Blake and Shiffrar 2007; Adolphs 2009 for reviews). Consistent with this idea, abnormal selectivity formoving animate and social stimuli in the lateral region of the right fusiform gyrus occurred within the context of typical response patterns for dynamic stimuli in the right pSTS and amygdala. This finding, in turn, supports the notion of disrupted connectivity in ASD (see Müller et al. 2011 for recent review).

Our finding is similar to that reported by Kana et al. (2009), who found no dysfunction in pSTS or amygdala when subjects with autism made social inferences to moving geometric shapes (though they did find a difference, relative to TDs, in a more anterior part of STS). Also consistent with our findings is a recent report of abnormal responses in the lateral region of the fusiform gyrus to video clips of naturalistic scenes containing dynamic faces, coupled with no group differences in pSTS when ASDs and TDs were compared (Scherf et al. 2010).

Several studies have reported aberrant eye movement patterns in ASD to social/animate stimuli, although others have not (see Boraston and Blakemore 2007 for review). Is this a possible explanation for the lack of a category preference for dynamic social stimuli within the fusiform gyrus? The specificity of our findings within the FFA (normal selectivity for static, but not for dynamic, social stimuli, coupled with normal selectivity for both static and dynamic stimuli in all other parts of the social cognition network) strongly mitigates this possibility. Even if gaze patterns differed between groups only for dynamic, but not for static stimuli, one would expect those gaze patterns to affect responses in other social brain regions, such as the amygdala and pSTS, as well as in ocular motor control regions, such as the frontal and supplementary eye fields. However, our groups did not differentially activate any of these areas, including the area most strongly associated with viewing biological motion (i.e., the posterior region of the STS). In addition, although aberrant eye gaze patterns have been reported in ASD when stimuli depict complex social situations involving multiple characters (Klin et al. 2002; Speer et al. 2007; Norbury et al. 2009; Riby and Hancock 2009; von Hofsten et al. 2009), no group differences in gaze patterns were found when dynamic, but isolated, characters were depicted (as in our Experiment 1; Speer et al. 2007), or for moving shapes (as in our Experiment 2; von Hofsten et al. 2009). Thus, the available evidence suggests that differences in eye gaze patterns are unlikely to account for our findings.

Taken together, the patterns of category-related activity detected in our ASD subjects suggest that there are abnormal connections between the region of the ventral temporal cortex associated with identifying social stimuli and areas in temporal cortex associated with biological motion processing (and possibly other regions integral to social cognition). Our connectivity analysis supported this possibility, revealing that, after regressing out the effects of task, activity in the right lateral fusiform region was more weakly correlated with activity in the right pSTS and the right amygdala in the ASD group, compared with the TD group. This is consistent with reports of a significantly reduced correlation between the pSTS and the extrastriate cortex in individuals with ASD during rest, suggesting reduced synchronization of neural activity between these regions (Lai et al. 2010; Paakki et al. 2010). Our findings are also consistent with studies reporting reduced correlations of slow fluctuating neural activity between the fusiform gyrus and the amygdala (Kleinmans et al. 2008; Dziobek et al. 2010; Lai et al. 2010), and the finding that fractional anisotropy in the right fusiform gyrus and pSTS was correlated with social and communication diagnostic scores in autism (Cheung et al. 2009).

Our results contribute to a growing body of evidence that face perception per se is intact in high-functioning individuals with autism and suggest that other regions common to face and body perception and social cognition systems fail to influence responses in the fusiform gyrus, resulting in reduced category selectivity for complex, dynamic, and more abstract stimuli. In two experiments that used multiple static and dynamic stimulus formats, viewing static photographs of people, faces, and tools led to category-related responses in the lateral region of the fusiform gyrus and FFA of subjects with autism that were largely indistinguishable from those of TD individuals with regard to location, overall response magnitude, and category effect. However, in the group with autism, activity in this region failed to distinguish social from non-social stimuli depicted in videos, point-light displays, and vignettes of moving geometric shapes as it did in the TD group. Category preferences for these stimuli did persist in other regions of the social cognition network, such as pSTS and the amygdala, where dynamic stimuli elicited heightened responses to social, compared with non-social objects in ASD. Dysfunctional connections such as those identified here may prevent these structures from modulating responses in the lateral region of the fusiform gyrus and FFA, perhaps resulting in impaired spontaneous social inferences (Senju et al. 2009). Disrupted connectivity can account for many facets of autism, not the least of which is the heterogeneity of the disorder itself. Poor communication between the pSTS, amygdala, and fusiform gyrus would likely cause the dysfunctional response pattern we observed, though irregular contact with prefrontal regions may conspire as well (Castelli et al. 2002; Bookheimer et al. 2008; Koshino et al. 2008; Jones et al. 2010). Further studies in carefully characterized subject groups are required to elucidate the precise nature of disrupted connectivity and how connections within and between different regions vary throughout development (see Gotts et al. 2012 for an unbiased, whole-brain analysis of functional connectivity in ASD).

**Funding**

This research was supported by the National Institute of Mental Health, Division of Intramural Research.

**Notes**

We thank Rachel Caravella, Jennifer Silvers, Laura Case, Laura Dolan, and Madeline Harms for their help with data collection and analysis, and most importantly, all of the participants and their families. **Conflict of Interest:** None declared.
References


