Cognition in Males and Females with Autism: Similarities and Differences

Meng-Chuan Lai1*, Michael V. Lombardo1, Amber N. V. Ruigrok1, Bhismadev Chakrabarti1,2, Sally J. Wheelwright1, Bonnie Auyeung1, Carrie Allison1, MRC AIMS Consortium1, Simon Baron-Cohen1

1 Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom, 2 Centre for Integrative Neuroscience and Neurodynamics, School of Psychology and Clinical Language Sciences, University of Reading, Reading, United Kingdom

Abstract

The male bias in autism spectrum conditions (ASC) has led to females with ASC being under-researched. This lack of attention to females could hide variability due to sex that may explain some of the heterogeneity within ASC. In this study we investigate four key cognitive domains (mentalingual and emotion perception, executive function, perceptual attention to detail, and motor function) in ASC, to test for similarities and differences between males and females with and without ASC (n = 128 adults; n = 32 per group). In the mentalizing and facial emotion perception domain, males and females with ASC showed similar deficits compared to neurotypical controls. However, in attention to detail and dexterity involving executive function, although males with ASC showed poorer performance relative to neurotypical males, females with ASC performed similarly compared to neurotypical females. We conclude that performance in the social-cognitive domain is equally impaired in male and female adults with ASC. However, in specific non-social cognitive domains, performance within ASC depends on sex. This suggests that in specific domains, cognitive profiles in ASC are modulated by sex.

Introduction

Autism spectrum conditions (ASC) affect more males than females, with a ratio of 4.3 to 1 reported in earlier studies [1]. This male bias has led to females with ASC being under-studied. This neglect may have implications not only for our understanding of ASC but also for clinical practice and how we diagnose ASC. For example, it could be that the male bias in ASC is partly due to the current diagnostic criteria being more aligned to presentations that are more apparent in males. In other words, we may be under-diagnosing females due to their ‘non-male-typical’ profile, or their ability to better camouflage/compensate for their difficulties [2–9].

Recent epidemiological surveys have reported a lower male bias (2.0–2.6:1) in the prevalence of ASC in the general population [10–12], even among those who are high-functioning (1.7:1) [10]. A recent prospective report of infant-siblings at high-risk for developing autism showed only a small (1.63:1) male bias in those later diagnosed at 3 years of age, and that a large proportion of individuals accounting for this reduction in male bias were females at higher-functioning levels [13]. These lines of evidence suggest increasing awareness of females on the autistic spectrum, especially of those without intellectual disability. There is thus an immediate need for better understanding of females with ASC [14,15], particularly with respect to their similarities and differences compared to males with ASC.

Previous studies of sex differences in core behavioral features of autism are inconsistent, in samples that range widely in age and IQ. A few studies find no sex differences in core features [16–19], but others do show sex differences [15,20–23]. At the biological level however, females with ASC show different profiles in serum biomarkers [24], genetics [25–27], neuroanatomy [28], and early brain overgrowth [29–32]. However, we still know little about cognition, with the exception of a small number of studies showing possible differences in executive function and visuospatial processing. For example, Nydén and colleagues reported that, in a clinic sample, girls (mean age 9.8 years, 5 with ASC, 12 with ADHD) performed worse than boys (mean age 10.1 years, 5 with ASC, 12 with ADHD) on the Tower of London test (measuring planning ability) [33]. Koyama and colleagues found that on a standardized IQ test, girls (mean age 8.2 years) with high-functioning ASC performed better than boys (mean age 9 years) with high-functioning ASC on processing speed, coding and symbol search, but that boys were better on block design [34]. Bolte and colleagues found that adolescent females with high-functioning ASC outperformed males with high-functioning ASC on the Trail Making Test (measuring set-shifting), but the opposite was observed in their unaffected siblings [35]. Lemon et al. reported that girls (aged 6 to 16 years) with high-functioning ASC performed worse than typical girls on a stop signal task (measuring response inhibition), but boys with high-functioning ASC...
performed comparably to typically developing boys and girls, who did equally well [36]. The present study extends our earlier report of the behavior of high-functioning adult males and females with ASC [15] to the level of cognition.

Cognitive features of ASC range from differences in basic sensory perceptual processing to high-level complex social cognition [37]. One way to illuminate the full picture is to apply a wide array of cognitive tasks [38,39]. As one of the first studies to formally compare adult males and females with and without ASC we opt for a more parsimonious approach, selecting the most widely used tasks in four key domains of cognition closely tied to ASC characteristics: (i) mentalizing and emotion perception, (ii) executive function, (iii) perceptual attention to detail, and (iv) motor function. Based on the emerging evidence of potential sex differences within autism, we test if the effect of autism is dependent on sex within each of these cognitive domains. If confirmed, this would be expressed as a significant sex-by-diagnosis interaction in a two-way factorial design (factor 1: sex, factor 2: diagnosis).

Materials and Methods

Ethics Statement

Informed written consent was obtained for all participants in accordance with procedures approved by the Suffolk Research Ethics Committee.

Participants

Thirty-three neurotypical male, 35 neurotypical female, 45 ASC male and 38 ASC female right-handed Caucasian English-speaking adults were recruited through the UK Medical Research Council Autism Imaging Multicentre Study (MRC AIMS) consortium (full details are described elsewhere [15]). The inclusion criteria included being aged between 18 to 49 years, without intellectual disability (IQ ≥70), and participants in the ASC group had a formal clinical diagnoses of autistic disorder or Asperger’s disorder, based on DSM-IV [40] or ICD-10 [41] criteria, from a psychiatrist or clinical psychologist in the UK National Health Service.

Exclusion criteria (for all groups) included a history of or a current psychotic disorder, substance-use disorder, severe head injury, syndromic genetic disorder associated with autism (e.g. fragile X syndrome, tuberous sclerosis), intellectual disability (i.e., IQ <70), hyperkinetic disorder, Tourette’s disorder, any other medical condition significantly affecting brain function (e.g. epilepsy), and/or current use of antipsychotic medication. The neurotypical groups did not have an ASC diagnosis themselves nor was it present in their family history.

The ASC participants were further selected based on their ADI-R scores [42]. To be included, they had to reach the diagnostic algorithm cut-offs for ‘autism’ but were permitted to score one point below threshold in one of the three core symptom domains, to allow for possible underestimation of early developmentally atypical behaviours in the recall of caregivers whose children were now adults; these all followed our earlier studies and rationale for inclusion [15,43–46]. Thirty-three male and 29 female adults scored above the threshold. Another three women, although not having ADI-R data available (because their childhood caregivers were not available to be interviewed), were also included for the following reasons: One scored above the cut-off for ‘autism spectrum’ on the Autism Diagnostic Observation Schedule (ADOS) [47], one previously received a diagnosis using the Adult Asperger Assessment (AAA) [48] which had incorporated caregiver reports on childhood behaviors, and one had been diagnosed by an expert clinician with assessments that included a comprehensive childhood developmental history. After matching for age, IQ and sample size across groups, the final cohort for analysis consisted of 32 participants per group.

Measures

**Mentalizing and emotion perception.** Impaired social-emotional-communication is the cardinal feature of ASC [37,49,50]. This can be viewed as stemming from two different aspects of atypical functioning: secondary and primary intersubjectivity [51]. For cognition related to secondary intersubjectivity, theory of mind (ToM) or mentalizing deficits have been found across the life span, from the classical first-order ToM deficit in children [52], to complex ToM deficits revealed in moral judgments [53] and spontaneous ToM [54] in adults with ASC. Here we used the ‘Reading the Mind in the Eyes’ test (Eyes Test) [55] to investigate mentalizing ability. The Eyes Test, comprising 36 items, requires the individual to infer mental status solely from the information of a person’s eyes and the immediate surrounding areas of the face in gray-scale photographs. Both the correct score and reaction time (RT) were taken as outcome measures. However, due to its high cognitive demands, total correct score was considered more informative. RT was positively skewed; and for parametric methods to be applicable it was log-transformed to approximate a normal distribution.

Cognition related to primary intersubjectivity involves processes supporting dyadic interaction such as face processing, emotion perception and social motivation. In particular, facial emotion recognition is frequently reported to be atypical in ASC. For example, adolescents and adults with ASC demonstrate atypical processing of basic negative facial emotions [56–58], subtle expressions of fear [59], sadness [60], disgust [61], as well as more complex emotional states [62]. In this study, an online version of the Karolinska Directed Emotional Faces Test (KDEF Test) was used to measure basic emotion perception. The KDEF Test is a 140-item basic emotion recognition task using stimuli from the KDEF database of photographs of basic emotions [63], comprising seven sets of 20 color faces presenting six basic emotions (happy, sad, angry, fear, disgusted, surprised) and a neutral expression, with stimuli presented in random order. Participants were asked to choose one from seven responses to identify the emotion of the face stimuli, by pressing a key from 1–7 on the keyboard, and were instructed to go as quickly and as accurately as possible. Due to potential ceiling effects in accuracy, we opted for reaction time but calculated accuracy-adjusted reaction time (aaRT = mean reaction time/accuracy) for each emotion, to take into account performance information on both [64]. The seven aaRTs were all positively skewed and were therefore log-transformed to approximate a normal distribution. Owing to the presence of a small number of right-tail outliers even in these log-transformed aaRTs, winsorizing was further performed as a trial by recoding all outliers to the score that fell on two standard deviations above the mean. [Note: Outliers (all/Extreme) were identified by each group. There were in total 0/0 for happy, 2/0 for sad, 4/3 for angry, 5/1 for fear, 5/1 for disgusted, 6/2 for surprised and 3/0 for neutral faces. There was no individual who was an outlier on all or most emotions. This indicated that the outlier was poor in recognizing particular emotion(s) but not generally slow/inaccurate.] This procedure, however, did not change any of the outcomes of statistical comparisons so results from the non-winsorized aaRTs will be reported. The outlier aaRT contains information of the emotion recognition ability on a specific emotion but not generally all emotions, reflecting the
per participant’s actual ability on a particular emotion. Winsorizing may reduce the extent of violation to assumptions for parametric tests, but as a consequence the information about particularly poor performance on a specific emotion, which is very informative for group comparisons, may be lost. We therefore report the results from the non-winsorized data. This is also because the F-test is robust to violations to assumptions [65].

Executive function. Executive dysfunction is a (non-specific) feature of ASC [66–72]. Many aspects of executive function, including planning, set-shifting, inhibition, generativity and self-monitoring, have been reported as impaired in people with ASC [73]. However there is inconsistency associated with experimental designs, IQ and co-occurring conditions such as hyperkinetic disorder or Tourette’s syndrome [74]. Relatively consistent findings point to difficulties in planning, set-shifting and inhibition of a prepotent response [75].

In the present study we used an online version of the Go/No-Go task to test inhibition and signal detection. Participants were instructed to press the ‘left key’ (‘Q’ on the left side of the keyboard) using the left hand when seeing a bold arrow pointing to the left presented on the screen, the ‘right key’ (‘P’ on the right side of the keyboard) using the right hand when seeing an arrow pointing to the right, and to not respond when seeing an arrow pointing upward. A total of 300 stimuli (110 left, 110 right, and 80 upward arrows) were presented randomly. Reaction time and responses for all 300 items were recorded. Results were first explored by calculating the classic commission error (pressing ‘left’ or ‘right’ when the stimuli is upward and should be ignored) and omission error (making no response when ‘left’ or ‘right’ should be pressed). Due to the highly skewed distribution, these error rates were rank-transformed to approximate a normal distribution prior to analysis. The performance was then re-analyzed within the framework of signal detection theory (SDT) [76] to estimate two major parameters: sensitivity ($d' = \frac{\zeta_{Hit} - \zeta_{Miss}}{\sqrt{2}}$) where $\zeta_{Hit}$ is the corresponding $\zeta$ value in the normal distribution for the probability of Hit (i.e., signal present and the response is ‘present’), and $\zeta_{Miss}$ is the same for False Alarm (i.e., signal absent but the response is ‘present’) and criterion ($C = -.5 \times (\zeta_{Hit} + \zeta_{Miss})$). Sensitivity $d'$ indicates the participant’s ability to discriminate signal from noise, and criterion $C$ quantifies how liberal (i.e., $C > 0$) or conservative (i.e., $C < 0$) the response strategy (bias) may be. Both $d'$ and C were normally distributed so no further transformations were performed.

Two language-related executive functions were assessed. Phonological memory (i.e., working memory in the auditory domain) was tested using the Non-Word Repetition task [77], consisting of 28 non-words. Participants were asked to listen carefully to a non-word (spoken in a British English accent) and repeat it immediately. Their utterance was audio-recorded and coded by a trained native British researcher using strict criteria: all vowels, consonants and accents in the repeated utterance needed to be exactly the same as the stimulus for the item to be coded as correct. Number of correct items was treated as the outcome measure. Second, the word generativity (F-A-S) task required the participant to produce as many words beginning with the letter ‘F’ as possible in one minute. Names, tense changes, plurals, and other aspects of the task performance [89]. First, average response time from all 12 items (including both correct and failure items) were used as the outcome measure, in order to account for both accuracy and response time; this has been commonly adopted in previous studies [81,90–92]. Second, to purely assess performance speed on correct items, mean response time for correct items only were taken as the outcome measure; also in the following statistical modeling, accuracy was included as a covariate to reduce the influence of accuracy on response time, as suggested by White & Saldana [89].

Dexterity. Although not part of the current diagnostic criteria, motor clumsiness was regarded as a diagnostic feature by Hans Asperger in his seminal report on ‘autistic psychopathy’ [93]. Various motor anomalies (e.g. gross motor, fine motor, coordination, gait and posture, movement imitation) and dyspraxia have been associated with ASC [94–101]. Rigorous meta-analysis also suggests that motor coordination deficits are a pervasive feature of ASC [102]. However there is still a debate as to whether there is intrinsic atypical motor development or if the motor clumsiness reflects executive dysfunction.

Here we explored motor function using the Purdue Pegboard Test [78], a reliable and well-validated standard test for dexterity, involving both gross movement of arms, hands and fingers and fingertip dexterity. The test consists of four subtests: (i) ‘right-hand’: the participant is asked to insert small pins into holes on the board for 30 seconds using only their right hand, and the number of pins successfully placed in the holes is scored; (ii) ‘left-hand’: the same procedure is repeated using the left hand; (iii) ‘alternative/both-hands’: the same procedure is used but the participant is instructed to pick up and place pins in two rows of holes using both hands alternately; and (iv) ‘assembly’: using both hands alternatively, the participant is asked to assemble sequences of pins, collars and washers for 60 seconds, and the number of all items successfully placed in the holes is recorded.

Perceptual attention to detail. At the perceptual level, people with ASC have been reported to show a preference for, and superior attention to detail [79] on visuospatial tasks [80,81]. This islet of superiority has been interpreted as reflecting weak central coherence (WCC) [82,83], superior low-level processing in perceptual modalities (i.e., enhanced perceptual function) [84,85], or superior discrimination (i.e., enhanced discrimination and reduced generalization) [86,87].

Here we used the adult version of the Embedded Figures Test (EFT) [88] to investigate this domain of cognition. Similar to a previous study [81], we used ‘Form A’ which consisted of 12 figures in fixed order plus an additional practice item, each depicting a complex design and a simple shape which was hidden in the complex design. The participant was first shown and asked to study the complex design for no more than 15 seconds, then shown the simple shape (meanwhile the complex design was covered) for no more than 10 seconds. Timing (using a stopwatch) started when the complex design was shown again to the participant (meanwhile the simple shape was covered) and s/he was asked to identify the simple shape with a stylus pen. Time was noted (but not stopped) once the participant said s/he found the simple shape. If the answer was correct, the noted time was recorded. If the answer was incorrect then s/he was asked to find it again, and the final time was recorded when the identification was correct. Participants were given an upper limit of 120 seconds, and failure to find the simple shape within this allotted time was scored as a failure and the response time for the item was recorded as 120 seconds.

Two analysis strategies were employed, accounting for different aspects of the task performance [89]. First, average response time from all 12 items (including both correct and failure items) were used as the outcome measure, in order to account for both accuracy and response time; this has been commonly adopted in previous studies [81,90–92]. Second, to purely assess performance speed on correct items, mean response time for correct items only were taken as the outcome measure; also in the following statistical modeling, accuracy was included as a covariate to reduce the influence of accuracy on response time, as suggested by White & Saldana [89].
Statistical Analysis

All analyses were performed under a two-way factorial analysis of variance framework, with sex and diagnosis as the two fixed factors, each with two levels. Due to potential interdependence among the outcome variables within each domain or outcome measures from the same task, multivariate analysis of covariance (MANCOVA) was first conducted for each domain or outcome measures from the same task, followed by individual ANCOVA if the overall MANCOVA yielded significant results. The alpha level for MANCOVA was adjusted using the Bonferroni method by the total number of MANCOVA performed (i.e., 0.05/5 = 0.01). The alpha level for the post-hoc ANCOVAs was corrected for multiple comparisons using the Bonferroni method within each MANCOVA (i.e., 0.05/number of ANCOVAs performed). For each ANCOVA, if significant, four planned comparisons were performed between (i) males with and without ASC, (ii) females with and without ASC, (iii) neurotypical males and females, and (iv) males and females with ASC; alpha level was set at 0.05/4 = 0.0125 by Bonferroni correction.

Age was positively correlated with most of the reaction time measures, and IQ positively correlated with most accuracy measures. Therefore for all analyses age and FIQ were included as covariates. As a supplement, we also performed all tests without covarying age and FIQ, and found that the group difference patterns and the statistical significance remained the same before and after including the covariates. Therefore, results from models covarying with age and FIQ are reported because they more accurately reflect standard practices in the literature [103,104]. All statistical analyses were performed with the PASW Statistics version 18 (SPSS Inc., Chicago, IL, USA). A brief summary of descriptive statistics for raw scores in all cognitive tasks was provided in Table S1.

Results

Participant Characteristics

All groups (MC: neurotypical control male adults; MA: male adults with ASC; FC: neurotypical control female adults; FA: female adults with ASC) were matched on age and full-scale IQ. For subscales, FC scored higher than MC on verbal IQ and MC scored higher than FA on performance IQ under a non-corrected threshold of p<0.05. As reported earlier [15], for ADI-R MA scored marginally higher than FA in the RSB domain but they had the same severity on social reciprocity and communication domains. MA scored significantly higher than FA in ADOS scores even after correction for multiple comparisons. See Table 1.

Mentalizing and Emotion Perception

Our first analysis was a MANCOVA with nine outcome measures from the Eyes Test and KDEF Test (Table 2) as dependent variables, sex and diagnosis as fixed factors, and age and FIQ as covariates. There was a significant main effect of diagnosis but not of sex, nor was there a significant interaction. Post-hoc univariate two-way factorial ANCOVAs showed that the main effect of diagnosis was significant across sexes forCommission error in either sex (in males $F_{1,122} = 4.958, p = 0.022$, in females $F_{1,122} = 0.686, p = 0.412$), and neutral ($F_{1,122} = 23.976, p < 0.001$) faces; see Figure 1. Post-hoc planned comparisons confirmed the presence of significant simple effects of diagnosis across both sexes for all variables, except non-significance in females for disgusted, surprised and neutral faces [panels F, G, H]. Furthermore, there were no sex differences in the neurotypical group.

To further explore if any particular emotion was specifically difficult for people with ASC, a mixed-model factorial ANCOVA was performed treating emotion as the between-subject factor (seven levels), sex and diagnosis as the between-subject factors, and age and FIQ as covariates. Results showed a significant emotion-by-diagnosis interaction ($F_{5,149} = 3.854, p = 0.009$; $df$ was corrected using Greenhouse-Geisser estimates of sphericity due to a significant violation of the assumption of sphericity, Mauchly’s test $p < 0.001$), but no emotion-by-sex or emotion-by-sex-by-diagnosis interactions. Contrasts revealed that people with ASC were particularly slower than controls in identifying fear compared to all other emotions: happy ($F_{1,122} = 7.341, p = 0.008$), sad ($F_{1,122} = 8.468, p = 0.004$), angry ($F_{1,122} = 7.988, p = 0.006$), disgusted ($F_{1,122} = 7.638, p = 0.007$), surprised ($F_{1,122} = 3.609, p = 0.06$), and neutral ($F_{1,122} = 6.826, p = 0.01$) faces; see Figure 2.

Executive Function

The Go/No-Go data for one male ASC participant was not recorded due to website failure, so there were only 31 male participants included in this analysis. We first performed an exploratory analysis on the conventional outcome measures of commission and omission errors. A MANCOVA treated the two rank-transformed error rates as dependent variables, sex and diagnosis as fixed factors, and age and FIQ as covariates. There was a significant main effect of diagnosis but not of sex, nor a significant interaction. Post-hoc univariate two-way factorial ANCOVAs showed that the main effect of diagnosis (i.e., people with ASC had more errors) was evident for both measures. There was no significant main effect of sex or sex-by-diagnosis interaction on either ANCOVA. See Table 3. Post-hoc planned comparisons showed that the simple effect of diagnosis was significant across sexes for omission error (in males $p = 0.001$, in females $p = 0.010$, but not for commission error in either sex (in males $p = 0.022$, in females $p = 0.235$). See Figure 3, panels A and B.

Data were re-analyzed with measures derived from SDT. A MANCOVA treated the log-transformed mean RT for all ‘go’ trials, sensitivity $d’$ and criterion $C$ as dependent variables, sex and diagnosis as fixed factors, and age and $d’$ as covariates. Echoing the exploratory analysis, there was a significant main effect of diagnosis but not of sex, nor was there a significant interaction. Post-hoc univariate ANCOVAs showed that the main effect of diagnosis were only evident for log-transformed mean RT for ‘go’ responses and $d’$, but not for $C$. There was no significant main effect of sex or sex-by-diagnosis interaction on any ANCOVA. See Table 3. This indicated that in general people with ASC were slower in response to stimuli and less sensitive in discriminating signal from noise, yet their response style/bias was similar to controls. Post-hoc comparisons showed that the simple effect of diagnosis was significant in males but not in females for ‘go’ RT (in males $p = 0.001$, in females $p = 0.458$) and $d’$ (in males $p = 0.001$, in females $p = 0.003$); note that the sex difference of these diagnostic differences did not reach a significant sex-by-diagnosis interaction. See Figure 3, panels C and D.

Another MANCOVA was conducted for the language-related executive functions, treating Non-Word Repetition (NWR) and word generativity (F-A-S) task scores as the dependent variables, sex and diagnosis the fixed factors, and age and $d’$ as the covariates. There was a marginal significant main effect of sex but not of diagnosis, nor was there a significant interaction. Post-hoc ANCOVAs showed that the main effect of sex was only
evident for F-A-S (males better than females) but not for NWR scores, and confirmed no main effect of diagnosis or interaction effect in either of them. See Table 3.

### Table 1. Demographic and behavioral characteristics.

<table>
<thead>
<tr>
<th></th>
<th>MC (N = 32)</th>
<th>MA (N = 32)</th>
<th>FC (N = 32)</th>
<th>FA (N = 32)</th>
<th>Statistics²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD) [range]¹</td>
<td>Mean (SD)</td>
<td>Mean (SD) [range]¹</td>
<td></td>
</tr>
<tr>
<td>Age (Years)</td>
<td>28.7 (5.9)</td>
<td>27.0 (7.2)</td>
<td>27.6 (6.3)</td>
<td>28.1 (8.2)</td>
<td>ns</td>
</tr>
<tr>
<td>Verbal IQ</td>
<td>111.0 (12.2)</td>
<td>112.5 (14.4)</td>
<td>118.3 (10.1)</td>
<td>114.5 (15.4)</td>
<td>FC&gt;MC (p = .030)</td>
</tr>
<tr>
<td>Performance IQ²</td>
<td>118.3 (11.5)</td>
<td>112.2 (15.3)</td>
<td>116.4 (9.4)</td>
<td>110.2 (17.0)</td>
<td>MC&gt;FA (p = .019)</td>
</tr>
<tr>
<td>Full IQ³</td>
<td>116.3 (11.8)</td>
<td>113.7 (15.1)</td>
<td>119.7 (8.4)</td>
<td>114.1 (15.5)</td>
<td>ns</td>
</tr>
</tbody>
</table>

**Perceptual Attention to Detail**

An ANCOVA was performed for the **mean RT for all items in EFT**, with sex and diagnosis as fixed factors, and age and FIQ as covariates. We noted significant main effects of sex and diagnosis, and a marginal interaction. FIQ explained most of the variance in the model ($F_{(1,122)} = 137.400$, $p < 0.001$, $\eta^2_p = 0.530$). Planned

### Table 2. MANCOVA and post-hoc ANCOVAs for the Eyes Test and the KDEF Test.

<table>
<thead>
<tr>
<th></th>
<th>MANCOVA</th>
<th>Main effect of Diagnosis</th>
<th>Main effect of Sex</th>
<th>Interaction effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F_{(9,114)}$</td>
<td>$p$</td>
<td>$V$</td>
<td>$F_{(9,114)}$</td>
</tr>
<tr>
<td>Eyes Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>7.773</td>
<td>&lt;.001</td>
<td>.380</td>
<td>1.315</td>
</tr>
<tr>
<td>log(RT)</td>
<td>3.26</td>
<td>.074</td>
<td>.026</td>
<td>1.52</td>
</tr>
<tr>
<td>KDEF log(aaRT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Happy</td>
<td>38.95</td>
<td>&lt;.001</td>
<td>.199</td>
<td>0.79</td>
</tr>
<tr>
<td>Sad</td>
<td>24.51</td>
<td>&lt;.001</td>
<td>.167</td>
<td>1.25</td>
</tr>
<tr>
<td>Angry</td>
<td>11.43</td>
<td>.001</td>
<td>.086</td>
<td>1.41</td>
</tr>
<tr>
<td>Fear</td>
<td>24.41</td>
<td>&lt;.001</td>
<td>.167</td>
<td>0.60</td>
</tr>
<tr>
<td>Disgusted</td>
<td>15.80</td>
<td>&lt;.001</td>
<td>.115</td>
<td>7.53</td>
</tr>
<tr>
<td>Surprised</td>
<td>17.19</td>
<td>&lt;.001</td>
<td>.123</td>
<td>1.76</td>
</tr>
<tr>
<td>Neutral</td>
<td>14.27</td>
<td>&lt;.001</td>
<td>.105</td>
<td>0.06</td>
</tr>
</tbody>
</table>

V: Pillai's Trace; $\eta^2_p$: effect size partial eta-squared.

doi:10.1371/journal.pone.0047198.t002
comparisons indicated that adult males with ASC performed worse than typical males ($p = 0.001$), which was not the case between the female groups ($p = 0.828$). Sex differences were observed in the control groups ($p < 0.001$) but not in the ASC groups ($p = 0.046$). See Table 4 and Figure 4, panel A. Similar to what was found in a previous study [89], this all-item mean RT was highly correlated with accuracy (Spearman’s $\rho = -0.93$, $p < 0.001$) so should be viewed as mainly reflecting information about accuracy, which itself was distributed far from normality and could not be analyzed adequately within the current framework.

To explore performance speed, a second ANCOVA was conducted for the mean RT for correct items only, with the same fixed factors and covariates but included accuracy (log-transformation of the reflected accuracy, i.e., subtracted from 2, to approximate normality) as an additional covariate to reduce the influence of different accuracy on performance speed [89]. Two males and one female with ASC were excluded because they failed all 12 items. We found a significant main effect of sex but not for diagnosis; there was no significant interaction. FIQ still explained all 12 items. We found a significant main effect of sex but not for males and one female with ASC were excluded because they failed

**Motor Function**

A MANCOVA treated the four outcome measures of the Purdue Pegboard Test (i.e., ‘right-hand’, ‘left-hand’, ‘alternative/both-hands’ and ‘assembly’ scores) as dependent variables, sex and age and FIQ as covariates. There were significant main effects of sex, diagnosis, and a sex-by-diagnosis interaction. Post-hoc ANCOVAs indicated that the significance was driven by the assembly score with significant effects of sex, diagnosis, and an interaction; none of these were significant for the other three measures. See Table 5. Planned comparisons indicated that for the assembly subtest, males with ASC performed worse than typical males ($p < 0.001$), which was not the case between the female groups ($p = 0.200$). A sex difference was observed in both the control ($p < 0.001$) and ASC groups ($p = 0.171$). See Table 4 and Figure 4, panel B.

Sex-by-diagnosis Interactions are Domain-dependent

**Mentalizing and emotion perception.** Regardless of sex, adults with ASC showed impaired mentalizing and basic facial emotion recognition abilities relative to neurotypical adults. This is not surprising given this cognitive domain critically underlies the socio-communication difficulties by which ASC is defined [40,105]. In contrast, we found a striking sex difference in current interactive behaviors on the ADOS, with milder interpersonal autistic features in women with ASC compared to men with ASC [15]. These results strengthen the arguments for superficial camouflaging of social-communication difficulties in females with ASC [2,4,15,106].

The deficit in advanced mentalizing in adults with ASC replicates previous studies [39,55,62,107]. For basic emotion recognition, they also showed difficulty in recognizing emotion faces across all basic emotions. This is in contrast with previous studies that either found impairments on negative but not positive emotions [56–61], or reported comparable behavioral performance on basic facial emotion recognition in adolescents and adults with ASC [108–111]. At least two other reasons could potentially contribute to this observed difficulty, beyond basic emotion recognition ability per se: a generally slower stimulus-response reaction, and/or difficulties in processing faces. We did not have a measure of simple reaction time on a neutral stimulus-response task. However, we did measure the ‘go’ RT in the Go/No-Go task, which indicated stimulus-reaction response in a simple non-social task. Here we found that people with ASC performed slower, and this ‘go’ RT was moderately correlated with all KDEF Test RTs. As an additional test, we corrected each emotion RT according to the ‘go’ RT for individual participants (i.e., dividing each emotion RT by the individual’s ‘go’ RT), and the pattern of group differences remained the same, with a significant main effect of diagnosis across all emotions (happy: $p < 0.001$, sad: $p = 0.001$, angry: $p = 0.014$, fear: $p < 0.001$, disgust: $p = 0.004$, surprised: $p = 0.001$, neutral: $p < 0.009$). This additional test suggests that, even after controlling for generally slower stimulus-reaction response, adults with ASC are still slower in recognizing all basic emotion faces relative to neurotypical adults. Thus, rather than inferring that individuals with ASC are generally slower irrespective of task, the fact that these deficits remain even after accounting for general ‘slowness’ argues for some specificity in deficits within the KDEF test.

To what extent this impairment may be accounted for by difficulties in face processing per se is unclear. Such an influence may exist given earlier reports of face processing difficulties in people with ASC [112–120]. However, a recent report also shows independent facial identity vs. facial emotion processing in children with ASC [121]. Whether atypical facial emotion processing in ASC is primary or secondary to other social deficits in ASC (e.g. face processing difficulties, low interest/motivation in social interaction, and/or reduced scanning of other’s eyes) requires more investigation [122].

We also found that adults with ASC (both males and females) were particularly impaired in identifying fear, compared to other emotion faces. This is partially in line with previous reports on the difficulty in processing fear for people with ASC [56–59], and may implicate atypical amygdala function [123–125]. Another expla-
nation for specific difficulties with fear may be due to a lack of attention to the eye region of the face. Prior work by Adolphs and colleagues [125] showed that bilateral amygdala lesions prevent recognition of fear in faces due to a lack of attention to the eyes. When attention is re-focused to the eyes, fear recognition returns to normal levels. Thus, given individuals with ASC tend to look less at the eyes [126], the deficit observed here for fear may reflect a lack of attention to the most salient feature of a fear face (the eyes).

Executive function. Within the domain of executive function we examined simple response inhibition, phonological memory, generativity and motor coordination/sequencing. Phonological memory and (word) generativity were not impaired in ASC. However, we found reduced sensitivity to signal detection (which is also reflected in more omission errors) in both sexes with ASC (the impairment being slightly smaller in females). We also found a slight impairment in simple response inhibition (reflected in commission errors) in ASC. In addition, on the dexterity subtest of assembly (which requires motor coordination, inhibition and planning on top of basic motor speed) there was a significant sex-by-diagnosis interaction: males with ASC were impaired relative to typical males, whilst females with or without ASC performed equally well. This difference is hard to account for purely in terms of gross or fine motor deficits, since a diagnostic group difference was not observed on any of the other simple dexterity subtests. Additionally, after controlling for basic motor speed, the group difference pattern on the assembly subtest remained the same.

In sum, we replicated earlier reports that, in ASC, phonological working memory and word generativity are intact [37,73,75], but found deficits in both sexes within ASC in simple response inhibition and in sensitivity to signals (less evident in females). Most interestingly, we found a strong sex-by-diagnosis interaction on the dexterity subtest involving motor executive function. This has implications that should be tested in the future: this might predict greater levels of dyspraxia in males with ASC, relative to females. Indeed, in a large sample of children and adolescents, females with ASC have been found to have less fine motor impairment [23].

Sex differences within ASC in executive function have been reported in children and adolescents [33–36]. Although performance across ages is difficult to compare due to developmental change, the message from the present and earlier reports is that sex differences within ASC exist in certain aspects of executive functions. The lower sensitivity to signal detection, alongside comparable response strategies in ASC, is of additional interest. This observation is solely based on the Go/No-Go task. Whether this lower signal sensitivity extends to other sub-domains of executive function or other cognitive domains will be important to establish.

Finally, our participants with ASC performed similarly to neurotypical adults on two language-related executive functions: phonological working memory and word generativity. The former has been reported to be impaired in adolescents with ASC with concurrent language impairment [127], but the latter is inconsistent in ASC [73,128]. Our null finding is most likely to be attributable to sample characteristics - they were all high-functioning adults without current language impairments, so executive functions subserving language processing were likely to be unimpaired. This again reflects the considerable heterogeneity within ASC.

Perceptual attention to detail. Previous reports of EFT performance in ASC are inconsistent [89]. This may be due to different methodological and analytical strategies, as well as heterogeneity in central coherence within ASC [89]. On both outcome measures, i.e., mean RT for all items (reflecting mainly accuracy) and mean RT for correct items adjusted by accuracy (reflecting mainly processing speed), we replicated the typical male advantage [129]. Furthermore in the ASC group, the trend level significance (p<0.046) of male advantage on mean RT for all items corresponds to previous reports of better performance in males than females with ASC on the Block Design task in children [34] and adolescents [35]. On the other hand, we did not replicate the ASC superiority (over neurotypical controls) noted by some early studies [80,81]. The results instead are in line with a well-powered study in children showing no such superiority on EFT performance [89].
functioning adults do not have basic motor impairments. ASC and lower IQ [99,130]. Hence it is plausible that these high-
usually associated with the more severe behavioral phenotype in
This is likely to be due to the fact that we recruited high-
and alternative/both-hands scores on the Purdue Pegboard Test). This may be due to different sample character-
when controlling for basic motor speed (reflected by right-hand or

dysfunction arises. We did not find a female advantage on the assembly subtest in
In contrast to previous findings [102] we
Motor function. In contrast to previous findings [102] we failed to find diagnostic group differences in general motor clumsiness on simple dexterity subtests (i.e., right-hand, left-hand and alternative/both-hands scores on the Purdue Pegboard Test). This is likely to be due to the fact that we recruited high-functioning adults and hardly any of them had a history of motor delay. Neurological comorbidities, including motor clumsiness, are usually associated with the more severe behavioral phenotype in ASC and lower IQ [99,130]. Hence it is plausible that these high-functioning adults do not have basic motor impairments. However, we did find a sex-by-diagnosis interaction on the assembly subtest, which involves a certain degree of motor coordination, planning and inhibition, in addition to basic motor skills. Poor performance in assembly but not other simpler subtests may thus reflect motor executive dysfunction. This was only found in males, but not females, with ASC. This warrants a detailed investigation of how the sex-specificity of motor executive dysfunction arises.

We did not find a female advantage on the assembly subtest in typical controls, in contrast to some previous reports [131], even when controlling for basic motor speed (reflected by right-hand or left-hand scores). This may be due to different sample characteristics (e.g., no laborer volunteers participated in the present study).

Cross-domain Summary
Each cognitive domain was investigated by a small number of tasks, hence a cross-domain summary must be preliminary. The results nevertheless provide initial evidence that both cognitive similarities and differences between the sexes exist in adults with ASC and with average or above-average IQ. Males and females who have comparable childhood autistic symptoms currently show similar cognitive deficits in social domains, but are different in certain non-social domains.

Regarding difficulties in social cognition, the hallmark of ASC, males and females share the same level of deficit. This may be interpreted within the context that males and females with ASC are currently diagnosed using the same behavioral criteria, and one major component of it (social-communication difficulties) is closely tied to social cognition. On the other hand, in cognitive domains not directly related to diagnosing ASC, the pattern appears different. For visuospatial attention to detail and certain aspects of executive function, when there is a diagnostic group difference in performance, it is usually accompanied by a disordinal sex-by-diagnosis interaction: the effect of the diagnosis is seen in males but not in females. These interactions mostly occur when there is a sex difference (males outperforming females) in the particular domain being investigated, and are more prominent in males with ASC than females with ASC.

Regarding difficulties in social cognition, the hallmark of ASC, males and females share the same level of deficit. This may be interpreted within the context that males and females with ASC are currently diagnosed using the same behavioral criteria, and one major component of it (social-communication difficulties) is closely tied to social cognition. On the other hand, in cognitive domains not directly related to diagnosing ASC, the pattern appears different. For visuospatial attention to detail and certain aspects of executive function, when there is a diagnostic group difference in performance, it is usually accompanied by a disordinal sex-by-diagnosis interaction: the effect of the diagnosis is seen in males but not in females. These interactions mostly occur when there is a sex difference (males outperforming females) in the particular domain being investigated, and are more prominent in males with ASC than females with ASC. This heterogeneity regarding how sex and ASC diagnosis interact warrants large-sample studies that include a wider range of IQ and age and equal-sized males and females.

Table 3. MANCOVAs and post-hoc ANCOVAs for the executive function tasks.

<table>
<thead>
<tr>
<th>MANCOVA</th>
<th>Main effect of Diagnosis</th>
<th>Main effect of Sex</th>
<th>Interaction effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F_{(2,120)}$</td>
<td>$p$</td>
<td>$V$</td>
</tr>
<tr>
<td>Go/No-Go rank-transformed error rates</td>
<td>9.490</td>
<td>&lt;.001</td>
<td>.137</td>
</tr>
<tr>
<td>ANCOVA</td>
<td>$F_{(1,121)}$</td>
<td>$p$</td>
<td>$\eta^2_p$</td>
</tr>
<tr>
<td>Commission</td>
<td>6.51</td>
<td>.012</td>
<td>.051</td>
</tr>
<tr>
<td>Omission</td>
<td>17.25</td>
<td>&lt;.001</td>
<td>.125</td>
</tr>
<tr>
<td>MANCOVA</td>
<td>$F_{(2,119)}$</td>
<td>$p$</td>
<td>$V$</td>
</tr>
<tr>
<td>Go/No-Go re-analysis</td>
<td>7.688</td>
<td>&lt;.001</td>
<td>.162</td>
</tr>
<tr>
<td>ANCOVA</td>
<td>$F_{(1,121)}$</td>
<td>$p$</td>
<td>$\eta^2_p$</td>
</tr>
<tr>
<td>log(RT) for ‘Go’ responses</td>
<td>7.33</td>
<td>.008</td>
<td>.057</td>
</tr>
<tr>
<td>SDT d'</td>
<td>12.75</td>
<td>.001</td>
<td>.095</td>
</tr>
<tr>
<td>SDT C</td>
<td>0.84</td>
<td>.360</td>
<td>.007</td>
</tr>
<tr>
<td>MANCOVA</td>
<td>$F_{(2,121)}$</td>
<td>$p$</td>
<td>$V$</td>
</tr>
<tr>
<td>Language related tasks</td>
<td>0.681</td>
<td>.508</td>
<td>.011</td>
</tr>
<tr>
<td>ANCOVA</td>
<td>$F_{(1,122)}$</td>
<td>$p$</td>
<td>$\eta^2_p$</td>
</tr>
<tr>
<td>F-A-S</td>
<td>0.19</td>
<td>.668</td>
<td>.002</td>
</tr>
<tr>
<td>NWR</td>
<td>1.29</td>
<td>.258</td>
<td>.010</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0047198.t003

Regarding RT for all items, males with ASC performed worse than typical males, but females with and without ASC performed comparably. This sex difference within ASC in EFT accuracy suggests that performance in visuospatial attention to detail may characterize men with ASC, but not women with ASC. The lack of superiority (or even worse performance) in ASC also challenges a prediction from the WCC theory. However, whether EFT really disentangles global from local processing is unclear, so interpretations should be made cautiously [89].
In sum, we observed that for high-functioning adults, the expression of ASC is modulated by sex particularly for non-social cognition, which echoes recent reports showing sex differences within ASC in executive function and visuospatial processing in children and adolescents [33–36]. These results have implications both for informing our understanding of mechanisms, and for clinical practice. Mechanistically, it sheds light on how ASC develops and manifests differentially by sex. Clinically, it suggests that sex-specific cognitive assessments may be useful. For instance, social-cognitive tasks may be helpful in identifying ASC-related difficulties for both sexes. However, when interpreting results from executive function and visuospatial processing tasks, sex-specificity should be taken into account. Our findings also suggest that higher-level motor function should be regularly assessed for people with high-functioning ASC. The other clinical implication is that we may need both sex-general and sex-specific intervention strategies. For example, social-cognitive enhancement may be helpful for both sexes, yet for males remediation for the non-social domains may be particularly important.

Limitations and Future Directions

The findings have several limitations. First, we only tested adults with ASC who did not have intellectual disability or other common comorbidities (e.g. epilepsy, hyperkinetic disorder).
Both EFT RT for all items (reflecting mainly accuracy, panel A) and assembly subtest score in the Purdue Pegboard Test (panel C) showed a significant interaction between sex and diagnosis. Males with ASC on average performed worse than neurotypical males, but females with ASC performed equally well as neurotypical females. EFT RT for correct items (reflecting purely processing speed, panel B) showed only a main effect of sex that overall females were on average slower than males. Convention of the graphs is the same as that in Figure 1.

doi:10.1371/journal.pone.0047198.g004

Table 4. ANCOVAs for EFT scores.

<table>
<thead>
<tr>
<th>Main effect of Diagnosis</th>
<th>Main effect of Sex</th>
<th>Interaction effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANCOVA</td>
<td>F(1,122) p g^2</td>
<td>F(1,122) p g^2</td>
</tr>
<tr>
<td>Mean RT for all items</td>
<td>5.54 .020 .043</td>
<td>22.94 &lt;.001 .158</td>
</tr>
<tr>
<td>ANCOVA</td>
<td>F(1,118) p g^2</td>
<td>F(1,118) p g^2</td>
</tr>
<tr>
<td>Mean RT for correct items</td>
<td>1.15 .286 .010</td>
<td>4.06 .046 .033</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0047198.t004
Whether (and how) these co-occurring conditions may further modulate cognition in ASC differentially for males and females is unknown. Therefore the findings may not generalize to other subgroups with ASC. Secondly, not all cognitive domains key to ASC in adults were examined in detail. For example, we did not assess advanced language function, particularly semantics and pragmatics, which are likely to be atypical in high-functioning adults with ASC; certain sub-domains of executive function commonly reported to be impaired in ASC (i.e., set-shifting and planning) were also not specifically examined. A Navon task [136] may have been more sensitive to global-local attention bias than the EFT, and the spontaneous ToM [54] and moral judgment ToM tasks [53] inform different aspects of mentalizing in high-functioning adults. Thirdly, to fully address the relationship between sex and ASC it is best to have tasks showing the largest effect sizes for typical sex differences, such as those measuring visuospatial abilities of targeting, mental rotation, and aggression [129,137]. Future larger-scale research should utilize a more comprehensive battery to pinpoint the fine-grained sex differences in various cognitive domains. Fourthly, the distribution of age varies substantially in our study. Although the influence of age on cognitive performances should have been well controlled by age-matching across groups and by including age as a covariate, we did not examine potential age-by-sex, age-by-diagnosis, or age-by-sex-by-diagnosis interactions. These types of nested interactions between age and the factors of interest would require investigation in much larger cohorts. Lastly, due to the limited sample size per subgroup, there is the potential that we had limited power to detect small effect sizes in the sex-by-diagnosis interactions and these smaller effects may be deemed important in further studies with increases in sample size. Therefore, the null findings should be interpreted with caution and should not be considered unequivocal proof of a lack of sex difference regarding how cognitive characteristics are affected by ASC.

### Supporting Information

**Table S1** A brief summary of the descriptive statistics for raw scores in all the cognitive tasks.

**Table 5. MANCOVA and post-hoc ANCOVAs for Purdue Pegboard Test scores.**

<table>
<thead>
<tr>
<th>MANCOVA</th>
<th>Main effect of Diagnosis</th>
<th>Main effect of Sex</th>
<th>Interaction effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Main effect</td>
<td>p</td>
<td>pV</td>
</tr>
<tr>
<td>MANCOVA</td>
<td>$F_{(1,123)}$</td>
<td>8.280</td>
<td>$.001</td>
</tr>
<tr>
<td>ANCOVA</td>
<td>$F_{(1,122)}$</td>
<td>6.31</td>
<td>.013</td>
</tr>
<tr>
<td>Right-hand</td>
<td>$\eta^2$</td>
<td>0.040</td>
<td>.002</td>
</tr>
<tr>
<td>Left-hand</td>
<td>$\eta^2$</td>
<td>0.040</td>
<td>.002</td>
</tr>
<tr>
<td>Both-hands</td>
<td>$\eta^2$</td>
<td>0.040</td>
<td>.002</td>
</tr>
<tr>
<td>Assembly</td>
<td>$\eta^2$</td>
<td>0.040</td>
<td>.002</td>
</tr>
</tbody>
</table>

**References**


