The “Reading the Mind in the Eyes” test

Citation for published version:

Digital Object Identifier (DOI):
10.1371/journal.pone.0136521

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
PLoS ONE

Publisher Rights Statement:
© 2015 Baron-Cohen et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
RESEARCH ARTICLE

The “Reading the Mind in the Eyes” Test: Complete Absence of Typical Sex Difference in ~400 Men and Women with Autism

Simon Baron-Cohen¹,²*, Daniel C. Bowen¹, Rosemary J. Holt¹, Carrie Allison¹, Bonnie Auyeung¹,³, Michael V. Lombardo¹,⁴, Paula Smith¹, Meng-Chuan Lai¹,²,⁵,⁶*

¹ Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom, ² CLASS Clinic, Cambridgeshire and Peterborough NHS Foundation Trust, Cambridge, United Kingdom, ³ Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom, ⁴ Department of Psychology and Center of Applied Neuroscience, University of Cyprus, Nicosia, Cyprus, ⁵ Centre for Addiction and Mental Health and Department of Psychiatry, University of Toronto, Toronto, Canada, ⁶ Department of Psychiatry, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan

* sb205@cam.ac.uk (SB-C); mcl45@cam.ac.uk (M-CL)

Abstract

The “Reading the Mind in the Eyes” test (Eyes test) is an advanced test of theory of mind. Typical sex difference has been reported (i.e., female advantage). Individuals with autism show more difficulty than do typically developing individuals, yet it remains unclear how this is modulated by sex, as females with autism have been under-represented. Here in a large, non-male-biased sample we test for the effects of sex, diagnosis, and their interaction. The Eyes test (revised version) was administered online to 395 adults with autism (178 males, 217 females) and 320 control adults (152 males, 168 females). Two-way ANOVA showed a significant sex-by-diagnosis interaction in total correct score (F(1,711) = 5.090, p = 0.024, η² = 0.007) arising from a significant sex difference between control males and females (p < 0.001, Cohen’s d = 0.47), and an absence of a sex difference between males and females with autism (p = 0.907, d = 0.01); significant case-control differences were observed across sexes, with effect sizes of d = 0.35 in males and d = 0.69 in females. Group-difference patterns fit with the extreme-male-brain (EMB) theory predictions. Eyes test-Empathy Quotient and Eyes test-Autism Spectrum Quotient correlations were significant only in females with autism (r = 0.35, r = -0.32, respectively), but not in the other 3 groups. Support vector machine (SVM) classification based on response pattern across all 36 items classified autism diagnosis with a relatively higher accuracy for females (72.2%) than males (65.8%). Nevertheless, an SVM model trained within one sex generalized equally well when applied to the other sex. Performance on the Eyes test is a sex-independent phenotypic characteristic of adults with autism, reflecting sex-common social difficulties, and provides support for the EMB theory predictions for both males and females. Performance of females with autism differed from same-sex controls more than did that of males with autism. Females with autism also showed stronger coherence between self-reported dispositional traits and Eyes test performance than all other groups.
Introduction

Autism spectrum conditions (henceforth “autism”) are diagnosed when an individual experiences social-communication difficulties, alongside unusually narrow interests and strong resistance to change, from early childhood and across their lifespan [1, 2]. The “mindblindness” theory [3] proposed that in autism a “theory of mind” (ToM, or “cognitive empathy”), sometimes referred to as “mind-reading” [4] or “mentalizing” [5], is impaired to varying degrees. ToM entails attribution and recognition of mental states in oneself or others [6] and to use such information to make inferences and predict behaviour. Toddlers with autism are impaired on tests of two precursors of ToM: joint attention and pretend play, both typically understood and produced by age 18 months [7-9]. Children and adults with autism are impaired in tasks assessing first-order false belief (i.e., recognizing that another person holds a belief that is not true), typically understood by 4 years old [10, 11]; second-order false belief (i.e., recognizing that another person holds a belief that another person believes something that is not true), typically understood by age 6 years old [12]; faux pas, typically understood by age 9 years old [13]; and reading subtle mental states from the eye region of the face [14], the voice [15], or in movie clips [16]. A ToM deficit is a parsimonious cognitive explanation for the social-communication difficulties that are ubiquitous in individuals with autism, across development, sex/gender (this is the preferred term for research in this field [17], but for simplicity “sex” is henceforth used in this article), IQ, or specific genetic syndromic forms [3, 11, 17, 18]. The consequence of this is difficulties in imagining the world from another’s perspective and tracking another’s mental states in real time during social interaction.

Individuals with autism also score significantly lower than typically developing individuals on the Empathy Quotient (EQ), a self-report [19, 20] or parent-report [21] measure of empathy. Empathy has at least two components [22]: “cognitive empathy” is synonymous with ToM, whereas “affective empathy” entails experiencing an appropriate emotion in response to another’s mental state (e.g., feeling pity in response to someone’s sadness, or feeling pleasure in response to someone’s happiness). Cognitive empathy is impaired in autism [23], whilst affective empathy likely remains intact [24, 25]. This profile in autism is the mirror image of those with psychopathic/antisocial personality, who often have intact or even superior cognitive empathy (e.g., they may analyse a victim’s mental states in great depth in order to identify their vulnerability) but reduced affective empathy [26-29]. For this reason, psychopaths may not care if their victim is in pain, and may even experience an inappropriate, self-centred emotion (such as pleasure) in response to someone else’s pain. Clinically, individuals with autism commonly report not realising that they have upset someone (a sign of impaired cognitive empathy), but feel remorse when it is pointed out to them that they have a sign of intact affective empathy [19, 30]. Unlike psychopaths, people with autism also commonly become upset if they hear someone else or an animal is suffering, and will stand up against injustice, quick to rush to the defence of the victim [26].

Self-report questionnaires (e.g., the EQ) reflect self-perception/evaluation, but not necessarily cognitive capability. The “Reading the Mind in the Eyes” Test (Eyes test) was therefore developed as a performance-based measure [14, 31]. It is an advanced ToM task involving mental state attribution and complex facial emotion recognition from photographs where only the eye region of the face is available. The Eyes test has been evaluated in over 250 studies to date, and has been found to have good reliability [32, 33]. Individuals with autism consistently perform less well relative to controls [16, 34–38]. The Eyes test also reveals a typical sex difference, with females scoring slightly but significantly higher than males [14, 39, 40].

Building on this typical sex differences, the finding that individuals with autism perform lower on the Eyes test than typically developing males fits a pattern predicted by the extreme-
male-brain (EMB) theory of autism, which extends the Empathizing-Systemizing (E-S) theory of typical sex differences, and hypothesizes that at a cognitive level, characteristics of autism (in comparison to those of typically developing individuals) reflect an "extreme-male" form of specific characteristics typically showing sex difference in the general population [41]. In particular, the EMB theory predicts that on tests of empathy, the pattern of scores will be "typically developing females > typically developing males > people with autism". The EMB theory also predicts that on measures of systemizing (the drive to analyse or construct a rule-based system, and to predict how a system works), the pattern of results will be the opposite. These patterns of results have been found using the EQ and the Systemizing Quotient (SQ) in both adults [20, 42] and children [21]. The idea that the autistic brain may be evolved to "hyper-systemize" [43] can explain the narrow interests shown by individuals with autism and their strong "obsessive insistence on the preservation of sameness" [44]. This is because—like science itself—systemizing involves holding all variables constant whilst varying just one variable at a time, in order to identify a law. Similarly, like science itself, systemizing involves repeating observations in order to establish that the law that has been identified holds true across time and place (excessive "repetitive behaviour"). A "masculinization" of scores on empathy and systemizing tests (i.e., where scores are pushed towards or even beyond the "male range") is correlated with fetal testosterone levels in utero [45–47]. Corresponding with these findings, one of the earliest biological correlates of autism are elevated fetal steroid hormones (not just fetal testosterone but also the precursors to this in the Δ4 sex-steroid pathway, including progesterone, 17α-hydroxyprogesterone, androstenedione, and cortisol) [48].

Although performance on the Eyes test has been shown to be a hallmark of social difficulties in autism, the under-representation of females with autism in previous research makes it difficult to examine whether an individual’s sex further modulates how autism manifests when it comes to social cognitive performance. A large sample-size and a sex-balanced design is needed to examine any hypothesis related to performance comparing males and females with and without autism (e.g., when testing predictions from the EMB theory or other hypotheses about mechanisms and causality, such as sex/gender-differential mechanisms, female protective effect, or better compensation in females, etc. [49]); see [17] for a detailed review. Here we recruited a large, sex-balanced sample to test (1) if there are sex differences and diagnostic differences in performance on the Eyes test, and if there are diagnosis-by-sex interactions, using a two-factorial design [20, 37, 50–52]; and (2) if performance differences between groups conform to predictions from the EMB theory. We were also interested to test if typical sex differences on the Eyes test are simply attenuated, as has been found recently on questionnaire measures [20], or are completely absent. If an absence of sex difference is observed, this might be because on questionnaire measures, females with autism score their empathy traits higher than do males, perhaps associated with unconscious gender stereotypes [53] or a desire to be seen as less impaired. In contrast, performance measures might be less susceptible to such gender stereotypes or social desirability effects, and might reveal cognitive capabilities that are closer to the innate difficulties experienced by individuals with autism.

Materials and Methods

Participants and ethics information

Adult participants over 18 years old were recruited from the following websites: www.autismresearchcentre.com (mainly for individuals with autism) or www.cambridgepsychology.com (mainly for individuals without autism), both hosted by the University of Cambridge, during the period of 2007–2014. Once they had logged onto either site, they consented for their data to be held in the Cambridge Autism Research Database (CARD) for research use, with
ethical approval (reference No. Pre.2013.06) from the University of Cambridge Psychology Research Ethics Committee. This ethics approval allows for retrospective data analyses, including the present study, in which all participant information were anonymized and de-identified prior to analysis.

Participants were selected from all available participants on the CARD (i.e., everyone who had logged onto the website and completed the Eyes test, AQ and EQ). Participants who self-reported a clinical autism diagnosis were asked specific information about the date of their diagnosis, where they were diagnosed, and the profession of the person who diagnosed them. The inclusion criterion for participants in the autism group was a clinical diagnosis of an autism spectrum condition (ASC) according to DSM-IV (any pervasive developmental disorder), DSM-5 (autism spectrum disorder), or ICD-10 (any pervasive developmental disorder) from a recognized specialist clinic by a psychiatrist or clinical psychologist. Such online self/parent-reported diagnoses agree well with clinical diagnoses in medical records [54]. Control group participants were included if they had no diagnoses of ASC, and no first-degree relatives with ASC. For both groups, participants were excluded if they reported a diagnosis of bipolar disorder, schizophrenia, eating disorder, obsessive-compulsive disorder, personality disorder, epilepsy, or an intersex/transsexual condition. Participants with a diagnosis of depression or anxiety were not excluded as these conditions are common in the general population and occur at high rates in adults with autism [1, 55]. In all four groups (males and females, with or without autism), any participants scoring zero on the Eyes test were removed from analyses, as this could reflect an overall difficulty in understanding test instructions. To minimize the control group including individuals with a broader autism phenotype (BAP) [56, 57], only individuals who had registered via the www.cambridgepsychology.com website, which was designed to recruit participants from the general population, were included in the control group. In addition, because individuals with BAP tend to score above the cut-off of 26 on the Autism Spectrum Quotient (AQ) [56], only a random sample of 6% of males and 2% of females who scored above this cut-off were included in the control group, in line with a non-biased sampling from the general population [56]. In addition, the four groups were selected to match on chronological age. The final sample included 395 adults with autism (178 males, 217 females) and 320 control adults (152 males, 168 females). See Table 1 for demographic and clinical information.

Measures

The Eyes test (revised and online version) consists of 36 grey-scale photos of people taken from magazines [14, 31]. These photos are cropped and rescaled so that only the area around the eyes can be seen. Each photo is surrounded by four mental state terms and the participant is instructed to choose the word best describes what the person in the photo is thinking or feeling. Only one of the four items is deemed correct (as judged by consensus from an independent panel of judges in the initial psychometric study). Participants were instructed to select the most appropriate item within 20 seconds for each stimulus (presented in random order). The instructions read, "You are going to see a series of 36 photographs of eyes. Your task is to choose the word, from a choice of 4, that best describes what the person in the picture is thinking or feeling. When you think you have found the answer press '1' if it is the top left word, '9' if it is the top right word, 'Q' if it is the bottom left word and 'I' if it is the bottom right word. Before making your choice, please make sure that you have read all 4 words. You should try to answer as quickly as possible, but without making mistakes. Once you have made a decision the next photograph will appear. If you have not made a decision in 20 seconds, it will automatically carry on to the next photograph.” Responses were coded as correct or incorrect (wrong items selected, or no
response after 20 seconds), giving a maximum total correct score of 36. All participants also completed the AQ [58] and the EQ [19] on the same online platform. For these self-report questionnaires the instructions read, “Below are a list of statements. Please read each statement very carefully and rate how strongly you agree or disagree with it.” All participants took the AQ and EQ before taking the Eyes test. The participants could log into and out of their account at any time so did not necessarily take the tests during the same session. However, the order of the tests was fixed for all participants.

### Statistical analysis

To test for differences in total correct score, we used a two-way analysis-of-variance (ANOVA) with diagnosis (autism vs. control) and sex (male vs. female) as the fixed factors. Main and interaction effects were tested at a critical level of $\alpha = 0.05$. We then compared the performance differences to see if they conformed to the patterns predicted by the EMB theory [41], separately for males and females [50]. Specifically, we tested (1) if there is a typical sex difference between control females and control males (control females > control males); and if yes, (2) whether males with autism show significantly poorer performance than control males; and/or (3) whether females with autism show significantly poorer performance than control females; and lastly, (4) whether any typical sex difference is attenuated in the autism groups.

We also used support vector machine (SVM) classification [59, 60] to test if autism diagnostic status could be accurately classified based on the pattern of responses across the 36 items on the Eyes test (1 = correct, 0 = incorrect). This is to test if patterns of response across items reveal unique information that are not evident in scores that sum correct responses across all 36 items, which would be the case if males and females with autism found different items more difficult. SVM classification was applied to the following pairs of groups: (1) control males and control females, (2) males with and without autism, and (3) females with and without autism. SVM is a supervised multivariate classification method where input data are classified into two classes by identifying a separating hyperplane/decision boundary, which maximizes the margin (i.e., distance from the hyperplane to the nearest data points). The algorithm is initially trained on a subset of the data to find a hyperplane that best separates the input space according to the class labels. Once the decision function is learned from the training set it can be used to predict

### Table 1. Demographic and clinical information, AQ, EQ, and Eyes test performance by group.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Age Mean (SD)</th>
<th>Ethnicity % self-described as “White European”</th>
<th>Rates of depression % reporting diagnoses of depression</th>
<th>AQ Mean (SD)</th>
<th>EQ Mean (SD)</th>
<th>Eyes test score Mean (SD)</th>
<th>Eyes test-EQ correlation Pearson’s r / Spearman’s ρ (2-tailed p values)</th>
<th>Eyes test-AQ correlation Pearson’s r / Spearman’s ρ (2-tailed p values)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males with autism</td>
<td>178</td>
<td>39.88 (11.56)</td>
<td>89.3%</td>
<td>33.1%</td>
<td>36.93 (8.64)</td>
<td>18.56 (11.24)</td>
<td>23.53 (6.64)</td>
<td>0.11 / 0.11 (0.162 / 0.132)</td>
<td>-0.13 / -0.14 (0.095 / 0.062)</td>
</tr>
<tr>
<td>Females with autism</td>
<td>217</td>
<td>39.88 (11.80)</td>
<td>85.3%</td>
<td>40.1%</td>
<td>34.71 (10.97)</td>
<td>24.94 (16.94)</td>
<td>23.45 (7.06)</td>
<td>0.35 / 0.40 (&lt;0.001 / &lt;0.001)</td>
<td>-0.32 / -0.33 (&lt;0.001 / &lt;0.001)</td>
</tr>
<tr>
<td>Control males</td>
<td>152</td>
<td>37.70 (12.62)</td>
<td>75.7%</td>
<td>9.2%</td>
<td>17.53 (6.03)</td>
<td>41.90 (11.53)</td>
<td>25.54 (4.57)</td>
<td>0.12 / 0.10 (0.132 / 0.201)</td>
<td>-0.13 / -0.16 (0.114 / 0.044)</td>
</tr>
<tr>
<td>Control females</td>
<td>168</td>
<td>39.15 (10.82)</td>
<td>84.5%</td>
<td>24.4%</td>
<td>15.25 (6.28)</td>
<td>49.88 (12.00)</td>
<td>27.42 (3.43)</td>
<td>0.11 / 0.14 (0.173 / 0.078)</td>
<td>-0.10 / -0.11 (0.190 / 0.157)</td>
</tr>
</tbody>
</table>

N, number; SD, standard deviation

* Participants were inquired whether they have ever been diagnosed with depression; information about anxiety disorders was not specifically collected in the present dataset.

doi:10.1371/journal.pone.0136521.t001
the class of a new set of test examples. We implemented LIBSVM v3.18 (http://www.csie.ntu.edu.tw/~cjlin/libsvm/) under Matlab R2013b. For the above three within-sample classification tasks, we trained and tested the classifier using a 10-fold cross-validation scheme. To find the classification model we used a radial basis function kernel and a regularization parameter (C) set to the default of 1. During training, class-size imbalance was handled by assigning different weights according to group size. To evaluate performance of the classifier we used the measures of accuracy (percentage of test examples correctly classified) and area under curve (AUC, particularly useful for comparing performance across models derived from imbalanced class-sizes). The performance metrics were tested with 10,000 permutations where the class labels were completely randomized, to evaluate the probability of getting a performance metric higher than the ones obtained during the cross-validation procedure by chance.

Finally, we examined whether the SVM model trained within one sex worked equally well when applied to the other sex, to test if case-control effects are similar or different between sexes. We applied the model classifying males with vs. without autism to the female groups (i.e., out-of-sample data), and vice versa. If case-control effects are different in each sex, a model obtained in one sex should not work well for the other sex; if the case-control effects are similar in both sexes, performance metrics using models obtained from the other sex should be close to the one obtained from within-sample cross-validation.

Results

Overall performance on the Eyes test

Performance scores on the Eyes test, EQ and AQ, as well as participant characteristics, are shown in Table 1. A two-way ANOVA showed a significant main effect of diagnosis (F(1,711) = 47.29, p < 0.001, ηp^2 = 0.062) reflecting better performance in controls than those with autism; and a significant main effect of sex (F(1,711) = 4.284, p = 0.039, ηp^2 = 0.006), reflecting better performance in females than males. However there was a significant sex-by-diagnosis interaction (F(1,711) = 5.090, p = 0.024, ηp^2 = 0.007), indicating that the main effects should not be interpreted directly. This interaction was the result from a significant sex difference between control males and control females (p < 0.001, Cohen’s d = 0.47), but an absence of a sex difference in males and females with autism (p = 0.907, d = 0.01); meanwhile there were significant case-control differences in both males and females, with a smaller effect size in males (p = 0.001, d = 0.35) but a larger effect size in females (p < 0.001, d = 0.69).

Building on the presence of a typical sex difference (control females > control males), diagnostic group differences confirmed predictions from the EMB theory in both males ([control females > control males] AND [control males > males with autism]) and females ([control females > control males] AND [control females > females with autism]) [50]. See Fig 1.

Pearson’s correlation analyses (used here as sample size was large per group and data distribution approximated normality) showed that Eyes test scores were significantly correlated with EQ and AQ scores only in females with autism (Table 1). Fisher’s test [61] showed that the substantial Eyes-EQ correlation in females with autism (r = 0.35) was significantly different from that in males with autism (p = 0.011), control males (p = 0.024), and control females (p = 0.013), all r ≈ 0.1. The substantial Eyes-AQ correlation in females with autism (r = -0.32) was also significantly different from that in males with autism (p = 0.042), control males (p = 0.057), and control females (p = 0.026), all r ≈ -0.1. As data distributions were slightly skewed in the autism groups, non-parametric Spearman’s correlations were also performed, and the same group-differential correlation patterns were confirmed (Table 1).
Group classification based on response patterns on the Eyes test

SVM classification with 10-fold cross-validation showed that the model classified control males vs. control females with modest performance, and significantly better than chance using permutation testing (accuracy 54.7%, $p = 9.99 \times 10^{-5}$; AUC 0.568, $p = 9.99 \times 10^{-5}$). The classification between control males vs. males with autism was better and significantly better than chance (accuracy 65.8%, $p = 9.99 \times 10^{-5}$; AUC 0.678, $p = 9.99 \times 10^{-5}$), and the classification between control females vs. females with autism was even better (accuracy 72.2%, $p = 9.99 \times 10^{-5}$; AUC 0.729, $p = 9.99 \times 10^{-5}$). Using the model obtained from classifying diagnostic status in males to predict diagnostic status in females showed accuracy (65.2%) close to that obtained by the within-male classification using cross-validation (72.2%). Similarly, using the model from females to predict diagnostic status in males showed accuracy (65.2%) close to that obtained by the within-female classification using cross-validation (65.8%). These together suggest that in terms of response pattern, case-control effects are similar across both sexes.

Discussion

In an age-matched, non-male-biased large sample of adults with and without autism, we tested Eyes test performance to establish if there were (1) typical sex differences, (2) differences between diagnostic groups, (3) an interaction of these, and (4) if the pattern of group differences
fitted predictions from the EMB theory. We confirmed previous results showing a typical sex difference [14], with control females scoring higher than control males. We also confirmed previous results showing that, in both males and females, individuals with autism score significantly lower than typically developing controls [16, 34–38]. Furthermore, and importantly, we found that sex and diagnosis significantly modulate each other: the diagnostic effect was much larger in females than in males, and the typical sex difference was completely absent in the autism groups. The patterns of group differences fit predictions from the EMB theory in both males (i.e., [control females > control males] AND [control males > males with autism]) and females (i.e., [control females > control males] AND [control females > females with autism]) [50]. Finally, when examining response patterns, effects of diagnostic group were similar across males and females.

**Difficulty in mental state attribution and complex facial emotion recognition from the eye region is found in both males and females with autism**

Performance on the Eyes test is a reliable [32, 33, 62] phenotypic [16, 34–38] and endophenotypic [35, 63, 64] measure for the cognitive bases of autism. How such phenotypic and endophenotypic characteristics are further modulated by sex has not previously been examined in detail. The present findings of similar performance between adult males and females with autism, and the larger shift in females than in males with autism away from same-sex controls, replicates our earlier findings from a smaller adult sample [37]. In addition, the classification model based on item response patterns obtained within one sex appears to work equally well when applied to the other sex. Together, these findings confirm the validity of the Eyes test as a sex-independent phenotypic measure of autism, indicating mental state attribution and complex facial emotion recognition are aspects of the core social difficulties in both males and females with autism.

Typically developing adults perform just as well on complex emotion/mental state recognition when the eyes alone are the only cue, or when the whole face is available [65]. This suggests that the eyes alone contain sufficient information as the whole face does, during complex emotion recognition. In contrast, individuals with autism show poorer performance on complex emotion/mental state recognition by the eyes alone than by the whole face, and perform significantly worse than age- and IQ-matched controls at identifying complex emotion/mental states from both the eyes alone and the whole face conditions [65]. This suggests that there may be a “language of the eyes” [65–68] that we use ToM to identify other people’s mental state from their eyes, and that individuals with autism are weaker on this.

The difficulties that adults and adolescents with autism have on the Eyes test [14, 36] mirror more basic difficulties earlier in development, using eye direction as a cue that someone is thinking [69], to infer a person’s desires and goals [70], or to infer which object is the intended reference to decode speech during language acquisition [71]. Reduced attention to the eye region of the face is one of the earliest abnormalities in autism developmentally [72–74]. It may not only be an early infancy marker of the later difficulties in ToM/cognitive empathy, but also a contributor to it [75], as it may result in reduced social orientation, social reward and social learning, and increased risk for a later autism diagnosis and social disability [76, 77].

**Group-specific correlation between Eyes test performance and self-report dispositional traits**

In previous studies conducted in the general population and not sex-stratified, Eyes test scores show an inverse correlation with AQ scores [58, 78], and a positive correlation with EQ scores.
When sex-stratified, a university student-based mid-sized study (65 men and 79 women) showed a significant negative AQ-Eyes test correlation only in men but not in women [79]. Interestingly, in the present large-scale study, when analyses were stratified by both sex and diagnosis, the correlation patterns between Eyes test performance and self-reported empathy/autistic traits were found to be both sex- and diagnosis-dependent. Eyes-EQ and Eyes-AQ correlations were significant only in females with autism, but not in males with autism, control males, or control females. The correlations in females with autism were statistically significantly different from those in the other three groups, suggesting stronger coherence between self-reflected/evaluated traits and neuropsychological performance in females with autism. In the general population, self-reported autistic traits are often associated with social cognitive performance in males but not in females, suggesting a more fractionable neurocognitive structure in females, potentially indicative of a “female protective effect” [17]. The reversal was found here for adults with autism, perhaps reflecting a loss of such fractionation in females with autism. Another possibility may be that adult females with autism have heightened self-awareness (possibly affected by socio-cultural contexts [80]), reflected in the significant association between their objective mentalizing performance and their subjective reflection of their personal empathy-related traits. These hypotheses are speculative and await further investigation. Whether this female-specific pattern of heightened association between self-reflected/evaluated traits and social cognitive performance are associated with a plausible “female-phenotype” of autism that has been anecdotally reported (e.g., heightened social awareness and social motivation, better imitation, more camouflaging) [17] is an important focus for further investigation.

Confirmation of EMB theory predictions in both males and females

Predictions from the EMB theory on Eyes test performance were confirmed in both males (i.e., [control females > control males] AND [control males > males with autism]) and females (i.e., [control females > control males] AND [control females > females with autism]). In addition, unlike attenuation of typical sex difference in autism found on self-report traits [20], here we found a complete absence of typical sex difference on the Eyes test. This may partly be because performance measures are less susceptible than self-report questionnaires to implicit gender stereotypes or social desirability effects, and reflect ability-based characteristics. It is also interesting to note that the EMB theory predictions were clearer in females than in males with autism, as shown by the larger case-control differences in the female than the male groups. These might be because the “masculinization” effects/characteristics of autism are more readily observable in females than males with autism, as previously noted in other domains including childhood play [81], brain structural characteristics [50], serum hormonal level [82–84], and anthropometry [84].

Potential uses of the Eyes test

The Eyes test has been used in both neuroimaging and lesion studies, revealing the involvement of the inferior frontal and temporal gyri, and amygdala [85–87]. During repeat fMRI between ages 12–19 years old,activation of the right superior temporal sulcus and right inferior frontal gyrus for the contrast “mental state > control” is a stable pattern of activity in performing the Eyes test [88]. However, partially sex-dependent functional brain correlates during performance on the Eyes test have been found in adolescents with and without autism [36]. Interestingly, neuro-endophenotypic effects are also present, and are again stronger in females [36]. Performance on the Eyes test among typically developing boys and girls shows an inverse association with prenatal testosterone levels [45], and in adulthood shows association with single
nucleotide polymorphisms in *OXTR* [89], *NTRK2, NTRK3*, *HSD17B2, HSD17B4, CYP1B1, CYP7A1, EN2*, and *GABRA6* [90]. Arginine-vasopressin administration in typically developing males reduces performance on the Eyes test, compared to placebo [91]. Similarly, testosterone administration in typically developing females reduces performance on the Eyes test, compared to placebo [92]. Administration of 3,4-methylenedioxymethamphetamine (MDMA, or ecstasy) enhances recognition of positive emotions but impairs recognition of negative emotions on the Eyes test across both general population males and females [93]. On the other hand, oxytocin administration in males with autism improves performance on the Eyes test, compared to placebo [94]. Future studies need to identify sex-common and sex-specific cognitive, neurobiological and psychopharmacological correlates of the Eyes test, and related tasks measuring cognitive vs. affective empathy, in individuals with autism or other atypical social-affective developmental conditions across the lifespan.

Performance on the Eyes test also reveals individual differences within the general population. It has been found to differentiate “high tech” (e.g., surgeons) vs. “high touch” (e.g., psychiatrists) doctors [95]. As mentioned above, it has also been used as a sensitive outcome measure for oxytocin administration studies [96, 97]. Finally, in terms of clinical groups, individuals with psychopathy [98], social anxiety disorder [99], schizophrenia [100], borderline personality disorder [101, 102], or victims of child abuse and neglect [103] all show different patterns of atypical performance on the Eyes test. Women with anorexia have been postulated to include individuals with undiagnosed autism, and show similar impairment on the Eyes test to people with a formal diagnosis of autism [104–106]. How such performance is modulated by sex requires further investigation.

The SVM classification between control males vs. control females was only modest (accuracy 54.7%, AUC 0.568). This shows that performance on the Eyes test, whilst sex-linked, is not wholly determined by one’s sex. One might need a battery of tests to better classify typically developing males vs. females. Even then, we suspect that measures such as the Eyes test would better classify cognitive styles [107] than sex. In addition, both classification performance between control males vs. males with autism (accuracy 65.8%, AUC 0.678) and between control females vs. females with autism (accuracy 72.2%, AUC 0.729), though significant, were not accurate enough to be diagnostic when used in isolation. This is not surprising, as the Eyes test is only related to part of the autism phenotype, and there are other aspects of social cognition, as well as executive, visuo-spatial, and sensori-perceptual features that are linked to the autism phenotype. Thus, we would need a combination of cognitive tasks to improve the accuracy of classifying autism vs. controls, and to potentially help clinical diagnostic procedures [38]. Nevertheless, cognitive performance still reveals individual differences and the heterogeneity of the autism spectrum [108, 109], reflected in the substantial inter-individual variation in the present study (Fig 1). This may be useful in the context of individualized education and medicine.

### Limitations

Some limitations of the present study should be acknowledged. First, Eyes test performance has been associated with verbal IQ [110] or even performance IQ [111]. This may reflect that on this task, mental state recognition requires selecting one out of four mental state words, and a person’s mental state lexicon may itself be related to their verbal or performance IQ. In the present study we have no measure of verbal IQ or language performance, nor performance IQ, so cannot test the extent to which the present findings are influenced by intelligence and linguistic abilities. Future studies taking into account variances explained by measured IQ or linguistic abilities are needed to further clarify group difference patterns seen in the present study.
Second, this study only included (higher-functioning) individuals who could complete online tasks and self-reported their formal clinical diagnoses of autism, and we further excluded those who had significant co-occurring psychiatric conditions. We therefore do not know if the findings would generalize to subgroups with intellectual disabilities, those with significant psychiatric comorbidity, or those without access to the internet or who are unable to volunteer for online research.

Third, it has been shown that Eyes test performance can be influenced by the correspondence between the ethnicity of the participant and that of the stimuli used in the test [112]. Our stimuli are all taken from Caucasian faces and the majority of participants across four groups self-describe as “White European”. However the very small numbers and the wide distribution of participants reporting other ethnicities renders a statistically sound examination of potential moderating effects from ethnicity and culture (and potential effects of familiarity) difficult. The extent to which current findings may be modulated by ethnic, cultural, or familiarity factors remains to be clarified by large-scale cross-ethnic and cross-cultural datasets.

Fourth, we were not able to investigate the complex interactions between affect regulation/emotion disorders (which frequently co-occur with autism) and how sex and autism diagnosis influence Eyes test performance in the present study, because we did not collect data for anxiety disorders (which frequently co-occur with depression), and also because we lack the statistical power to address complex three-way interactions (between emotion disorders, sex, and autism). The extent to which the present findings may be affected by co-occurring affect regulation difficulties is therefore unknown. This is an important topic to investigate in the future, with a better-powered and more comprehensive dataset.

Fifth, there was no independent in-person verification of diagnoses for the majority of the autism groups since participants were recruited online. However, previous studies have shown high levels of agreement between self/parent-reported diagnoses and clinical diagnoses of autism in medical records [54]. In addition, all participants with autism provided the name of the psychiatrist or clinical psychologist who had diagnosed them and the name of the clinic where they were diagnosed, and we have no obvious reason to disbelieve such data.

Finally, the present study is cross-sectional and only about cognitive characteristics. Therefore the study is only at the phenotypic and descriptive level and demonstrates differences and similarities between males and females with ASC. Other hypotheses about mechanisms and causality in relation to sex/gender differences in autism, such as sex/gender-differential mechanisms, female protective effects, or better compensation in females [49] need to be tested with approaches that can reveal developmental mechanisms and etiologies, and that use longitudinal and/or multi-level datasets. In addition, for such hypothesis testing, clinical phenotypes need to be measured comprehensively in order for any cross-sex/gender comparison to be unambiguously interpreted [17]. The present study is not able to test for etiologies or developmental mechanisms, but is a useful foundation for phenotypic characterization of autism, taking sex and gender into account.

**Conclusions**

In a large, non-male-biased adult sample, we confirm that performance (in terms of both accuracy and response pattern) on the Eyes test is a sex-independent phenotypic characteristic of individuals with autism, reflecting sex-common social difficulties. The presence of a typical sex difference in performance on the Eyes test confirms previous reports of a female advantage in cognitive empathy. Performance of females with autism deviated away from same-sex controls to a greater extent than that seen in males with autism. Females with autism also showed stronger coherence between self-reported dispositional traits and Eyes test performance than all
other groups. Finally, the findings provide support to predictions from the EMB theory, in both sexes, in terms of performance accuracy.

**Acknowledgments**

This study was submitted in part fulfilment of the BSc in Psychology by DCB. The research was supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care—East of England (CLAHRC-EoE). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

**Author Contributions**

Conceived and designed the experiments: SB-C DCB RJH CA BA M-CL. Performed the experiments: SB-C DCB RJH CA BA MVL PS M-CL. Analyzed the data: DCB CA MVL PS M-CL. Contributed reagents/materials/analysis tools: SB-C CA BA MVL PS M-CL. Wrote the paper: SB-C DCB RJH CA BA MVL PS M-CL.

**References**


Absence of Typical Sex Difference on the Eyes Test in Autism


