Original Article

Absence of Sex Differences in Mental Rotation Performance in ASD

Melanie S Rohde, Department of Psychiatry and Psychotherapy, University Hospital of Cologne, Kerpener Str. 62, 50294 Cologne, Germany; melanie.rohde@uk-koeln.de

Alexandra L Georgescu, Department of Psychiatry and Psychotherapy, University Hospital of Cologne, Kerpener Str. 62, 50294 Cologne, Germany and Social Neuroscience Group, University College London Institute of Cognitive Neuroscience, 17 Queen Square, London WC1N 3AR; a.georgescu@ucl.ac.uk

Kai Vogeley, Department of Psychiatry and Psychotherapy, University Hospital of Cologne, Kerpener Str. 62, 50294 Cologne, Germany; kai.vogeley@uk-koeln.de

Rolf Fimmers, Institut für Medizinischen Biometrie, Informatik und Epidemiologie (IMBIE), Sigmund-Freud-Straße 25, 53105 Bonn; rolf.fimmers@ukb.uni-bonn.de

Christine M Falter, Department of Psychiatry and Psychotherapy, University Hospital of Cologne, Kerpener Str. 62, 50294 Cologne, Germany; christine.falter@cantab.net

Abstract

Mental Rotation (MR) is one of the most investigated cognitive functions showing consistent sex differences. The “Extreme Male Brain” hypothesis (Baron-Cohen, 2002; EMB) attributes the cognitive profile of individuals with Autism Spectrum Disorder (ASD) to an extreme version of the male cognitive profile. Previous investigations focused almost exclusively on males with ASD with only limited implications for affected females. The current study is the first testing a sample of 12 female adults with high-functioning ASD compared to 14 males with ASD, 12 typically-developing (TD) females, and 14 TD males employing a computerized version of MR (Shepard & Metzler 1971). Reaction time and accuracy served as dependent variables. Their linear relationship with degree of rotation allows separation of rotational aspects of the task, indicated by slopes of the psychometric function, and non-rotational aspects, indicated by intercepts of the psychometric function. While the typical and expected sex difference for rotational task aspects was corroborated in TD individuals, no comparable sex difference was found in ASD individuals. ASD and TD individuals did not differ in MR performance. This finding does not support the EMB hypothesis of autism.

1. Introduction

Females and males differ **in cognitive abilities**. However, the extent to which this difference applies has been a matter of research for decades. One of the most investigated domains of cognitive differences between typically-developed (TD) females and males with a consistent advantage for males is the ability to rotate objects in mind, a function termed mental rotation (MR; Aleman, Bronk, Kessels, Koppeschaar & van Honk, 2004; Astur, Tropp, Sava, Constable & Markus, 2004; Brosnan, Walker & Collomosse, 2010; Falter, Plaisted & Davis, 2008a; Linn & Petersen, 1985; Shepard & Metzler 1971; Tapley & Bryden, 1977, Voyer, Voyer & Bryden, 1995).

Autism Spectrum Disorders (ASD) are pervasive developmental disorders with an unbalanced sex ratio of approximately four times as many males than females affected (Werling & Geschwind, 2013). Individuals with ASD are characterized by impairments in social interaction and communication as well as restricted interests and stereotyped behaviour (APA, 2013). The so-called Extreme Male Brain Hypothesis (EMB) of autism (Baron-Cohen, 2002; Baron-Cohen, Knickmeyer, Belmonte, 2005) rests on sex differences which are putatively found between females and males. EMB suggests ranking females and males along an empathizing domain in which females are thought to excel and along a systemising domain in which males are thought to excel (Baron-Cohen, 2002; Baron-Cohen et al., 2005). Empathising refers to the ability of understanding and feeling others' emotions and reacting adequately, while systemising refers to the ability of understanding, analysing and predicting rule-governed systems (Baron-Cohen, 2002). EMB theory defines ASD as an extreme version of the so-called "male brain" or cognitive system with minimised empathic but emphasised systemising abilities, predicting superior performance of individuals with ASD in cognitive tasks in which males are assumed to outperform females. Typically, such sex differences in performance are most reliably found in Mental Rotation (MR) tasks (see debate in Falter, Plaisted & Davis, 2008b and Knickmeyer, Baron-Cohen, Auyeung & Ashwin, 2008).

The original **MR** task version by Shepard and Metzler (1971) was further developed (Falter, Arroyo, & Davis, 2006) to the current computerised version in which participants are **presented with pairs of** three-dimensional figures, which are either different (mirrored) or the same but viewed from one of seven rotational

perspectives (0°-120°). **Their task was to decide as quickly and accurately as possible if the figures are the same or not. Therefore**, the **figures** must be kept in working memory and mentally rotated until a matching perspective is found (same trials) or not (different trials). Reaction time (RT) and accuracy (ACC) **as dependant variables** are measured and described as a psychometric function of degree of rotation. **RT and ACC can be depicted separately in a coordinate system, allowing to differentiate slopes and intercepts of psychometric function. Slopes of individual psychometric function have been operationalised as rotational aspects of the task (i.e. the function of mentally rotating a figure). Intercepts of the psychometric functions have been operationalised as non-rotational aspects** (i.e. working memory, stimulus encoding, comparison and decision making processes, response preparation and execution; Gill, O'Boyle & Hathaway, 1998; Hooven, Chabris & Ellison, 2004). Shallower slopes and **lower** intercepts for RT **and higher intercepts for ACC** indicate better performance.

Although still subject to scientific discussion, **some findings can be derived for sex differences in TD individuals and group differences between TD and ASD individuals. S**ex differences in TD individuals have usually been found in slopes, i.e. the rotational aspect of MR (Brosnan, Walker et al., 2010; Falter et al., 2006; Kozaki & Yasukouchi, 2009; Zapf, Glindemann, Vogeley & Falter, 2015). This is in contrast to studies comparing individuals from the two diagnostic groups of ASD and TD individuals **which show** differences in intercepts, i.e. non-rotational aspects of MR (Falter et al., 2008a; Pearson, Marsh, Hamilton & Ropar, 2014). Both studies by Falter et al. (2006; 2008a) used exactly the same task version, thus offering a direct comparison of performance differences between TD females and males with those between individuals with ASD and TD individuals. **EMB would predict** superior performance of individuals with ASD over TD individuals in the same domain **(slopes or intercepts)** in which TD males outperform TD females. **Consequently, as argued by Falter et al. (2008b), findings of group differences in different domains thus challenge EMB theory.**

However, some other studies using different **MR** task versions revealed differences between TD females and males in intercepts instead of slopes (Brosnan, Daggar & Collomosse, 2010; Brosnan, Walker et al., 2010; Hooven et al., 2004). Also, while several studies showed superior performance of individuals with ASD over TD individuals (Falter et al., 2008a; Hamilton, Brindley & Frith, 2009; Soulières,

Zeffiro, Girard & Mottron, 2011), other studies failed to find a difference in MR performance between individuals with ASD and TD individuals (Beacher, Radulesco, Minati, Baron-Cohen, Lombardo, Lai et al., 2012; Conson, Mazzarella, Frolli, Esposito, Marino, Trojano et al., 2013; Silk, Rinehart, Bradshaw, Tonge, Egan, O'Boyle et al., 2006). Overall, a meta-analysis showed only an insignificant advantage for individuals with ASD over TD individuals for intercepts and none for slopes (Muth, Höhnekopp, & Falter, 2014). However, even though not reflected in MR performance, studies employing functional magnetic resonance imaging (fMRI) revealed decreased neural activation among individuals with ASD in regions which are associated with working memory and executive functions (Silk et al., 2006). **Since those functions are assumed to be expressed in intercepts (Hooven et al., 2004), Silk et al. (2006) provide another hint for the assumption that that any observed differences between individuals with ASD and TD individuals in the MR task reside in intercepts.** Importantly, samples of individuals with ASD in most of the studies mentioned above were exclusively male or included only very few females with ASD. It is unclear to what extent EMB theory would predict a performance difference between males and females with ASD. Following the logic of testosterone-mediated MR performance differences, it would be consequent to assume a sex difference within ASD that is found on the same dimension of MR (whether intercepts or slopes) as the sex difference found in a TD sample.

Thus, the aim of the current study was to establish whether sex differences exist in MR performance in ASD and to clarify whether they mirror sex differences found in a TD control sample. Inclusion and scrutiny of females with ASD in cognitive tasks is urgently needed as recent studies emphasize that females with ASD might show different performances in some cognitive domains (Lehnhardt, Falter, Gawronski, Pfeiffer, Tepest, Franklin et al., 2016).

Employing the same validated computerised MR version in the current study as previously (Falter et al., 2006, 2008a; Zapf et al., 2015) allows comparison across studies and a wider generalisation of results. In addition, IPT (Intuitive Physics Test; Baron-Cohen, Weelwright, Scahill, Lawson & Spong, 2001) **indexing systemising abilities was performed as suggested and previously established by Brosnan and colleagues (2010). Since systemizing abilities are the crucial domain where males are thought to excel females in EMB, IPT was tested for correlations with MR performance. The ratio of the right second to fourth digit** (2D:4D) was

measured as proxy of prenatal testosterone levels (Manning, Kilduff & Trivers, 2001) **to test for correlations with MR performance. In normal populations, men were shown to have lower 2D:4D ratios than females (Lutchmaya, Baron-Cohen, Raggatt, Knickmeyer & Manning, 2004; Manning, Stewart, Bundred & Trivers, 2004) and individuals with ASD tend to have even lower 2D:4D ratios than normal populations (Manning et al., 2001). EMB directly links higher prenatal testosterone exposure to higher systemizing abilities as well as lower empathy and to a higher risk for developing ASD, but this suggestion has been discussed controversially (see Hönekopp, 2012, for a review). The current study replicates 2D:4D ratio measurement to reevaluate previous findings showing that MR performance and digit ratios are unrelated (Falter et al., 2006, 2008a). Also, with respect to the heterogeneity of ASD and the broad range of degrees of severity, only participants with ASD as per diagnostic label of ICD10 and an IQ of at least one SD under mean as well as AQ scores over 26 were included. A broad age range was chosen to allow comparability to other studies such as Brosnan, Daggar et al. (2010). Longitudinal data is lacking, but several studies suggest that visuospatial performance or executive functions in ASD do not depend on age (Guy, Mottron, Berthiaume & Bertone, 2016; Sachse, Schlitt, Hainz, Ciaramidaro, Schirman, Walter et al., 2013; Ozonoff, Cook, Coon, Dawson, Joseph, Klin et al., 2004) Finally, several comorbidities are observed to be over-represented in ASD, among them depression (Magnuson & Constantino, 2011; see Masi, DeMayo, Glozier & Guastella, 2017 for a recent review). EMB does not specify co-morbidities as a limitation to the theory, but in order to control for possible influences on performance, depression was assessed and patients with co-morbidities other than depression were not considered as participants.**

2. Method

Participants

34 individuals with ASD and 26 TD individuals were recruited. Ethics approval was granted by the Ethics Committee at the Medical Faculty, University of Cologne, and

written informed consent was obtained before any testing. **Age criterion for ASD participants was 20-55 years. ASD and TD participants were matched pairwise and matching allowed a maximum discrepancy of 6 years.** Three individuals with ASD had to be excluded due to random response behaviour **with less than 50% correct answers in rotations with 0°** and five had to be excluded because they did not meet the age criterion required to achieve group matching. Hence, 26 participants with ASD (12 females, 14 males) and 26 TD participants (12 females, 14 males) were included into the final analysis. Average ASD female age was 41.3 years and average ASD male age was 44.2 years. Average TD female age was 38.1 years and average TD male age was 44.3 years (see Table 1). All ASD participants were diagnosed and recruited in the Autism Outpatient Clinic at the Department of Psychiatry of the University Hospital of Cologne in Germany. **All participants were tested individually. The experiment was performed in a separate testing room at the Center for Psychiatry of the University Hospital of Cologne in Germany. It was equipped with facilities to perform paper-pencil tasks as well as a laptop with a 19'-screen to conduct the MR task. All of them** performed the WIE (Wechsler Intelligenz-Test für Erwachsene; Aster, Neubauer & Horn, 2006). ASD and TD groups as well as males and females were matched with respect to age, verbal IQ and performance IQ (all *p*'s>0.05, largest *t*=-.617). **Furthermore, BDI and AQ were employed.** BDI (Becks Depression Inventory; Beck, Brown, Steer, 2013) is a self-reporting inventory and consists of 21 multiple choice questions designed to measure degree of depression. AQ (Autism-Spectrum Quotient; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) is a self-reporting questionnaire **which** consists of 50 questions revealing traits **that** are usually associated with autism. **ASD and TD group differed in BDI (t (49) = 4.080, p < .001) but only trends and no correlation with ACC or RT occurred (smallest p = .054).** As expected, AQ scores between ASD and TD groups differed significantly (t (49) = 16.247, p < .001). IPT (Baron-Cohen et al., 2001) tests a participant's systemising abilities with 20 multiple-choice questions which should be answered in 10 minutes. The more correct answers participants give the better their intuitive physical understanding and the stronger their systemising tendencies. In addition, the length of the participant's right second and fourth digit **from the basal crease to the tip of the digit** was measured (2D:4D) **with a ruler directly from the participant's hand after the testing** as proxy of prenatal testosterone exposure because higher levels

of testosterone during ontogeny might be associated with higher risk for developing ASD (Manning et al., 2001) **yet unrelated to MR performance (Falter et al., 2006, 2008a)**. Four individuals with ASD and three TD individuals refused measurement so a total number of 22 individuals with ASD and 23 TD individuals entered this part of the analysis.

Table 1.
Demographic data for female and male ASD and TD participants.

			Minimum	Maximum	*Mean*	*SD*
ASD	Females *(N=12)*	Age*	20.17	48.83	41.33	8.71
		VIQ	81	130	104.33	15.20
		PIQ	67	126	101.42	17.93
		BDI	2	28	11.83	7.15
		AQ	27	48	41.75	5.50
		IPT	6	14	9.25	2.34
		2D:4D	0.93	1.04	1.00	0.04
	Males *(N=14)*	Age*	29.08	51.58	44.23	6.42
		VIQ	88	135	115.43	14.21
		PIQ	80	141	112.36	15.25
		BDI	1	27	11.43	8.78
		AQ	34	47	40.86	3.46
		IPT	4	15	7.64	2.84
		2D:4D	0.95	1.04	1.00	0.03
TD	Females *(N=12)*	Age*	22.33	46.67	38.06	8.19
		VIQ	90	129	112.58	12.06
		PIQ	86	134	106.67	13.96
		BDI	0	14	6.33	5.02
		AQ	6	29	14.64	7.06
		IPT	6	14	9.50	2.35
		2D:4D	0.97	1.10	1.02	0.04
	Males *(N=14)*	Age*	25.50	57.67	44.34	8.13
		VIQ	94	137	112.86	13.79
		PIQ	73	132	110.50	17.39
		BDI	0	9	2.31	2.87
		AQ	12	28	18.29	5.24
		IPT	6	16	10.54	2.57
		2D:4D	0.96	1.07	1.01	0.03

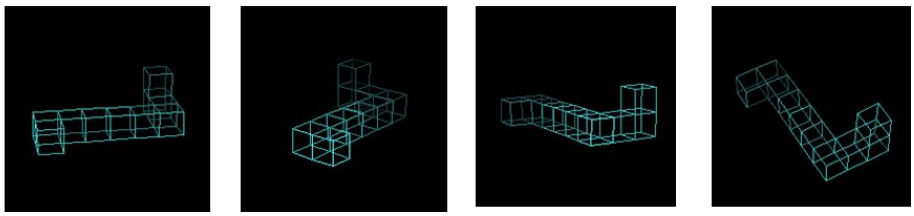
Note: ASD = Autism Spectrum Disorder; TD = Typically Developing; VIQ = Verbal IQ; PIQ = Performance IQ; BDI = Becks Depression Inventory; AQ = Autism-Spectrum Quotient; IPT = Intuitive Physics Test; 2D:4D ratio = second to fourth digit ratio.
*in years

Stimuli and Design

Participants conducted a computerised version of the Shepard & Metzler (1971) MR task (for specificities see Falter et al., 2006). They were presented with **pairs** of three-dimensional figures and had to judge as fast and accurately as possible whether the two objects were the same or different (i.e. mirrored). The figures were viewed from a range of rotational angles: 0°, 20°, 40°, 60°, 80°, 100° to 120°. The software Presentation (Version 17.0, 2014) was used for stimulus presentation and **recording reaction time and number of correct answers for each participant which served as RT and ACC data for the final analysis.** The experiment was conducted on a laptop with a 19'-screen. Before the experiment started, participants were given instructions and performed a test run of a minimum of 10 random trials taken from the original experiment until the experimenter was satisfied that the instructions were understood. After that, the experiment proper was started. The stimuli were light blue three-dimensional figures presented on a black background. Their size was approximately 4° of visual angle in width and height and each figure was constructed of 10 cubes (see Shepard & Metzler, 1971, for construction rules). In each trial, two stimuli were shown simultaneously on the right and left side of the centre of the screen with a distance of approximately 10° visual angle. Half of the trials were combinations of "same" (rotated) figures and the other half were "different" (mirrored and rotated) figures resulting in a total number of 448 trials presented in two blocks. Each combination of rotation angle and each pair of identical or mirrored objects was only shown once and in pseudo-randomised order. Participants were advised to record their decision by pressing one of two corresponding keyboard buttons with their index fingers. After key press the figures disappeared. In case of a wrong answer an acoustical feedback was given. The screen remained black for 500 ms before the next trial started.

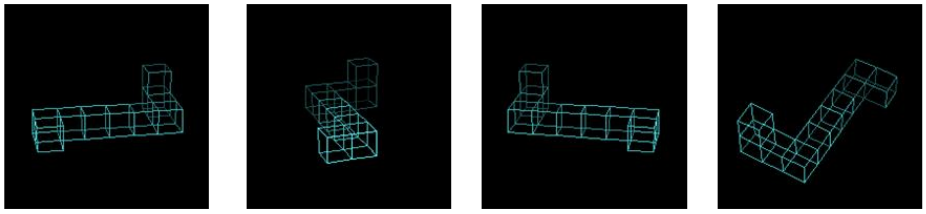
After data collection was finished, slopes and intercepts of scores of accurate responses (ACC) and reaction times (RT) were calculated as a measurement of mental rotation abilities and statistically analysed to detect sex and group differences.

Figure 1. Above, pictures 1-4 show examples for identical objects ("same" trials) from different rotation angles (0°-120°). In each trial, two of them were presented pairwise and participants were advised to signal their judgment by pressing the corresponding key. Below, pictures 1-4 show examples for mirrored objects from different rotation angles (0°-120°) which were presented pairwise.



1 2 3 4

Examples for "same" trials (presented pairwise).



1 2 3 4

Examples for "different" trials (presented pairwise).

3. Results

Slopes and intercepts of accurate responses (ACC) and reaction times (RT) were calculated and analysed separately using mixed ANOVAs with one between-participants factor (GROUP: ASD vs. TD) and one within-participants factor (CONDITION: same vs. different). The expected main effect of CONDITION was confirmed for all dependent variables, ACC intercepts ($F(1, 50) = 43.992$, $p = .000$, $\eta^2 = .468$) and RT intercepts ($F(1, 50) = 94.447$, $p = .000$, $\eta^2 = .654$) as well as in ACC slopes ($F(1, 50) = 21.569$, $p = .000$, $\eta^2 = .301$) and RT slopes ($F(1, 50) = 15.357$, $p = .000$, $\eta^2 = .235$) indicating same figures being easier to compare than different figures. No interaction (smallest $p = .500$) and no main effect of GROUP was found (**ACC intercepts: 94.686 ± .798 (mean ± std), [93.083; 96.289] (95% confidence limits), p = .132; RT intercepts: 3084.807 ± 181.015, [2721.228; 3448,386], p = .583; ACC slopes: -.161 ± .013, [-.187; -.134], p = .794; RT slopes: 21.579 ± 2.477, [16.603; 26.554], p = .336**) which suggests that there was no performance difference between the TD and ASD group in this study. In order to seek for sex differences within each

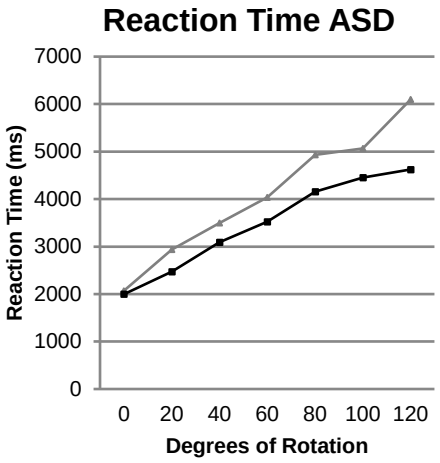
diagnostic group, we performed separate mixed ANOVAs with one between-participants factor (SEX: female vs. male) and one within-participants factor (CONDITION: same vs. different) for both diagnostic groups separately. Besides the same effect of CONDITION as given above, there was a main effect of SEX in the TD group for RT slope ($F(1, 24) = 5.196$, $p = .032$, $\eta^2 = .178$). Concerning intercepts there were only trends for significance for RT in the TD group ($F(1, 24) = 3.126$, $p = .090$, $\eta^2 = .115$) and ACC in the ASD group ($F(1, 24) = 3.077$, $p = .092$, $\eta^2 = .114$). **No interaction was found (smallest *p* = .057) and no other main effects of SEX in either the TD or ASD group occurred (ACC intercepts: 95.771, 93.560 ± 1.027, 1.188 (mean ASD, TD ± std), [93.650, 91.107; 97.891, 96.012] (95% confidence limits), p = .092, .298; RT intercepts: 3000.156, 3154.542 ± 283.003, 221.813, [2416.066, 2696.743; 3584.246, 3612.341], p = .490, .090; ACC slopes: -.156, -.164 ± .018, .020 [-.193, -.205; -.120, -.123], p = .365, .820; RT slopes: 24.337, 19.483 ± 4.583, 1.782, [14.879, 15.805; 33.795, 23.160], p = .332, .032**). With respect to 2D:4D ratios there was only a trend of a group difference between ASD and TD groups ($t(43) = -1.805$; $p = .078$) and there were no SEX differences **in** 2D:4D ratios within both diagnostic groups (smallest $p = .362$). Neither GROUP nor SEX differences were found for IPT (smallest $p = .123$). For the ASD group, correlations between IPT and ACC slopes for same trials ($r = -.628$; $p = .001$) as well as ACC intercepts for different trials ($r = -.750$; $p < .001$) were significant (Bonferroni-Holm corrected alpha-values). No significant correlations occurred for the TD group (smallest $p = .266$).

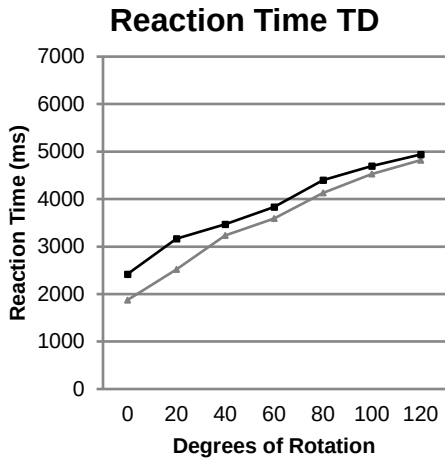
Table 2.
Mean (SD) RT and ACC scores for ASD and TD separately for females and males.

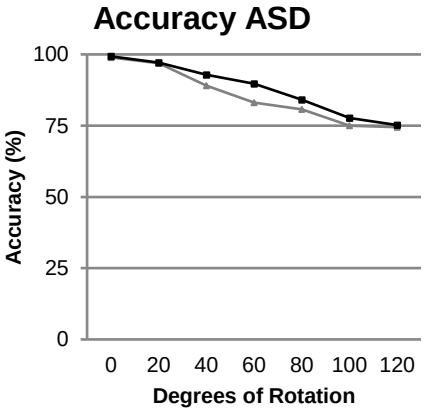
		RT				ACC			
		Intercept		Slope		Intercept		Slope	
		same	different	Same	different	same	different	same	different
ASD	Females	2183.82 (482.01)	4213.06 (1902.49)	31.76 (27.95)	25.98 (32.25)	98.90 (3.84)	89.03 (9.02)	-0.22 (0.17)	-0.06 (0.08)
	Males	2090.29 (1140.78)	3513.44 (2257.67)	23.07 (16.25)	16.53 (18.39)	100.85 (2.46)	94.29 (10.44)	-0.21 (0.15)	-0.13 (0.07)
TD	Females	2054.38 (633.44)	3470.3 (1223.95)	24.54 (7.45)	22.55 (16.28)	99.73 (2.38)	89.92 (8.91)	-0.21 (0.15)	-0.11 (0.12)
	Males	2609.78 (1121.73)	4483.7 (1634.03)	20.61 (8.98)	10.23 (7.51)	97.31 (4.26)	87.28 (11.44)	-0.21 (0.11)	-0.12 (0.15)

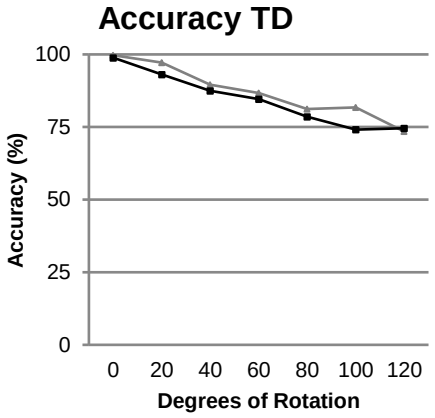
Note: SD = Standard Deviation; RT = Reaction Time; ACC = Accuracy; ASD = Autism Spectrum Disease; TD = Typically Developing.

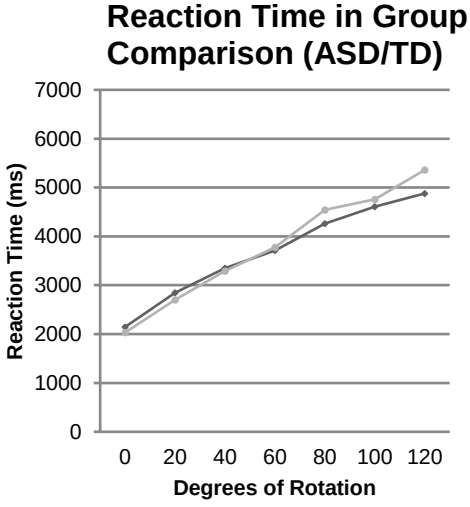
Figure 2. Comparison of MR reaction time (RT) and accuracy (ACC) scores between males (black squares) and females (grey triangles) for the ASD and TD groups separately (upper four panels) and comparison of RT and ACC scores between ASD and TD (bottom panels). Dark grey diamonds represent the TD group; light grey circles represent the ASD group.











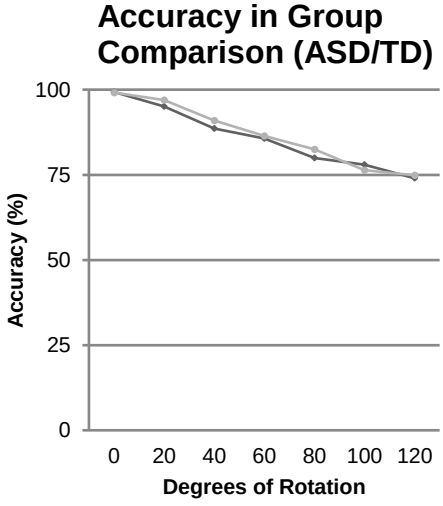
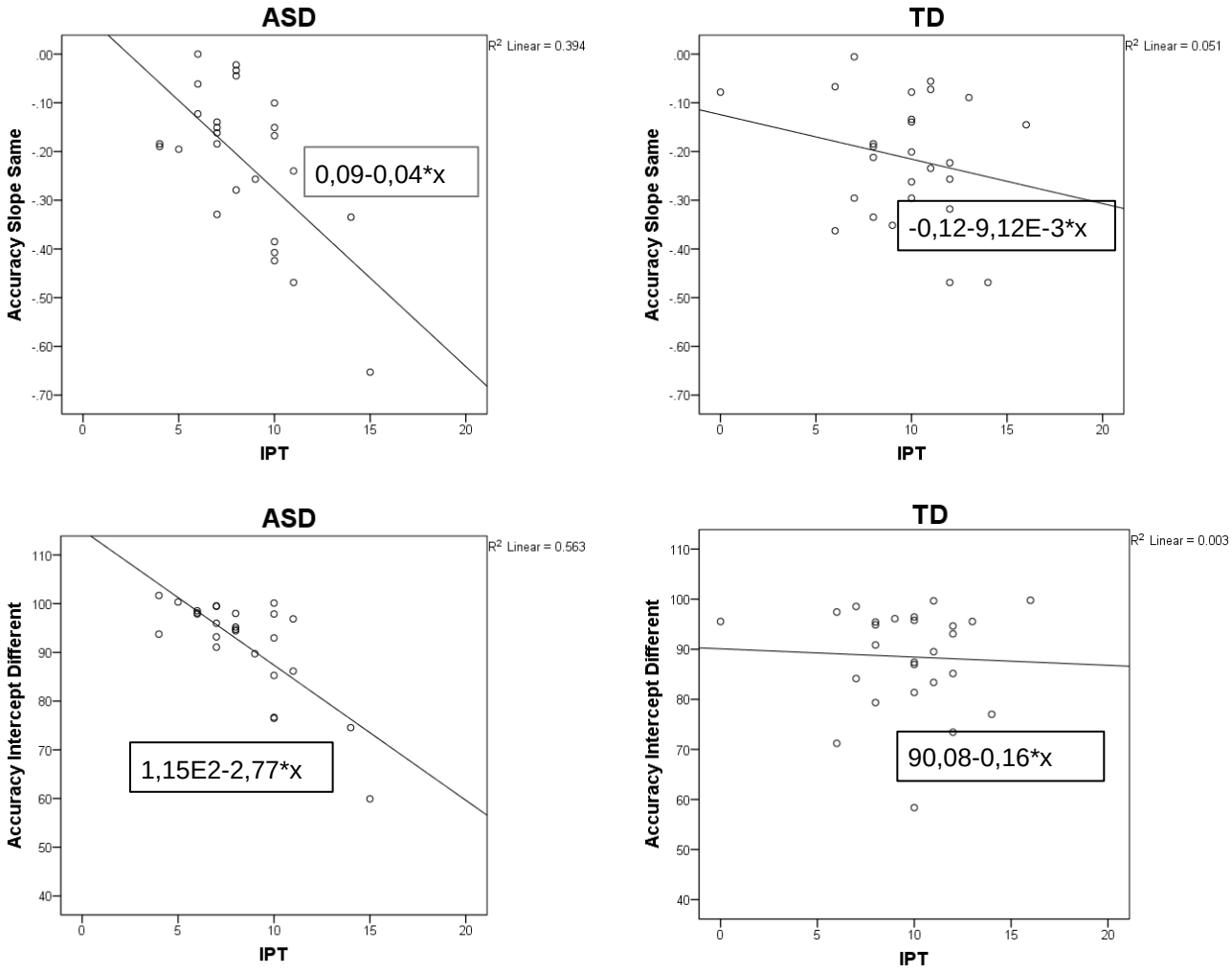


Figure 3. Comparison of correlations between IPT and accuracy slopes for same trials and accuracy intercepts for different trials for ASD and TD group.



4. Discussion

Stimulated by recent evidence for differences in the cognitive profile between females and males with ASD (Lehnhardt et al., 2016), the aim of the current study was to test whether typically found sex differences **between TD females and males** (Aleman et al., 2004; Astur et al., 2004; Brosnan, Walker et al., 2010; Falter et al., 2008a; Linn & Petersen, 1985; Tapley & Bryden, 1977, Voyer et al., 1995) in a classical MR task (Shepard & Metzler 1971) would also be reflected in performance differences between females and males with ASD. **To our knowledge, this is the first study which directly compares MR performance between ASD females and males.** Furthermore, we sought to clarify in which aspects **of the MR task** TD and

ASD females and males differ. **This issue has been addressed before (e.g. Falter et al., 2008a; Brosnan, Daggar et al., 2010; Brosnan, Walker et al., 2010) but discussed controversially. In accordance with previous findings (e.g. Brosnan, Daggar et al., 2010; Brosnan, Walker et al., 2010; Falter et al., 2006; Zapf et al., 2015) males in the current study were found to outperform females.** Although **originally** designed as a visuo-spatial task, MR performance requires both visuo-spatial **skills per se** (i.e. mentally rotating objects in space called rotational skills) **represented by slopes** as well as non-rotational aspects (such as figure comparison, decision making, response preparation, working memory etc.) **represented by intercepts**. Despite of a few reports of sex differences in non-rotational task aspects (Brosnan, Daggar et al., 2010; Hooven et al., 2004) the majority of studies to date found sex differences to reside only in rotational task aspects (Brosnan, Walker et al., 2010; Falter et al., 2006; Falter et al., 2008a; Kozaki & Yasukouchi, 2009; Zapf et al., 2015). **As corroborated by the current findings, there is now converging evidence for sex differences in TD to reside in the domain of rotational task aspects.** This bears implications on EMB theory as argued by Falter et al. (2008a, 2008b; see discussion in Knickmeyer et al., 2008). If, as predicted by EMB theory, individuals with ASD show an extreme version of male cognitive skills patterns, then superior performance would be expected in the same domain in which typical sex differences are found.

However, a few studies previously showed differences in non-rotational aspects **of the MR task** (Brosnan, Daggar et al., 2010; Brosnan, Walker et al., 2010; Hooven, et al., 2004). **In these studies**, it has been argued that testosterone facilitates MR abilities, which is reflected in male superiority over females in non-rotational task aspects. For example, Hooven et al. (2004) found in a sample of TD males that higher circulating testosterone was associated with lower error rates and faster RT, but only for non-rotational and not for rotational task aspects. Similarly, Brosnan, Daggar et al. (2010) tested TD females and males and found a relation between a proxy for circulating testosterone, indicated by daytime, and non-rotational MR aspects. Note though that studies investigating testosterone levels must be interpreted with caution given various factors are known to influence testosterone levels in females and males such as daytime (see Valdez, Reilly et Waterhouse, 2008 for a summary), exercise (Hulmi, Ahtiainen, Selänne, Volek, Häkinnen, Kovanen et al., 2008), menstrual cycle (Celec, Ostatnikova, Putz & Kudela, 2002), or

natural fluctuations (Courvoisier, Renaud, Geiser, Paschke, Gaudy & Jordan, 2013). Also, intercepts themselves are potentially suspect to arousal levels as noted by Falter et al. (2006). As argued, videos showing either sexual or dentistry content used in the study by Hooven et al. (2004) to manipulate testosterone levels could have given rise to different arousal levels in high- compared to low-testosterone males leading to differences in intercepts that might not primarily be caused by testosterone.

Overall, the picture drawn by studies on testosterone is inconsistent (see Falter et al., 2006 for a summary) and lacks further research. Many studies found no association between testosterone and spatial abilities in general (e.g., McKeever, Rich, Deyo & Conner, 1987) or, more specifically, between testosterone and mental rotation (Alexander, Swerdloff, Wang, Davidson, McDonald, Steiner et al., 1998; Falter et al., 2008a; Halari, Hines, Kumari, Mehrotra, Wheeler, Ng et al., 2005). Although testosterone was not the focus of our study we could neither show the reported differences in non-rotational task aspects nor did we find systemising to correlate with MR, although systemising has previously been shown to correlate with a proxy for circulating testosterone and better performance in non-rotational task aspects (Brosnan, Daggar et al., 2010). **Instead,** we replicated previous findings that fetal testosterone indexed by 2D:4D ratios (Manning et al., 2001) is not related to systemising (Voracek & Dressler, 2006) or MR performance (Falter et al., 2006, 2008a) **even if implicated by EMB theory**. Taken together, the lack of relationship between MR performance and both systemising and as 2D:4D ratios in the current study **as well as a increasing number of studies corroborating the idea that typical sex differences observed in MR usually reside in rotational task aspects (Brosnan, Walker, et al., 2010; Falter et al., 2006; Kozaki & Yasukouchi, 2009; Zapf et al., 2015)** add to the growing and converging body of evidence against cognitive profile predictions of EMB theory (see Falter, et al., 2008b).

Contrary to previous research (Brosnan, Daggar et al., 2010; Zapf et al., 2015), we did not find the reported sex differences in IPT which might be due to less statistical power given a smaller number of participants in the current study. Interestingly, Brosnan, Dagger et al. (2010), found systemising to correlate only with non-rotational aspects of MR, specifically accuracy intercepts for same and different responses, and they found sex differences in the same aspect of MR. In contrast, systemising in our study correlated with rotational and non-rotational parts of MR,

specifically slopes of accuracy for same trials and intercepts of accuracy for different trials, and we also found sex differences in rotational components of MR. This corresponds to the findings by Zapf et al. (2015) who also found a correlation between IPT and rotational aspects of MR. Consequently, systemising tends to be rather associated with rotational than non-rotational aspects of MR. Besides that, a closer view on the correlations described above and the correlations found by Zapf et al. (2015) shows a negative correlation between systemising and MR, indicating that better systemising abilities go along with worse performance in MR, while with respect to EMB one would expect a positive relationship between systemising and superior visuo-spatial performance.

Besides that, same and different comparisons yielded significantly different performance in line with the findings of the original Shepard & Metzler (1971) paradigm, indicating that "different" trials are more difficult because mental rotation of the figures needs to be **completed** to reach the conclusion that they do not match. For "same" trials, rotation **only** needs to be performed up to the degree of match (for discussion see Brosnan, Walker et al., 2010; Pearson et al., 2014). Nevertheless, there was no interaction of "same" versus "different" trials with sex, showing that this distinction is irrelevant for the question of sex differences in MR performance.

Compared to the numerous studies on MR in TD individuals, studies on MR in ASD are rather rare. In **a** previous meta-analysis, **it was** found that individuals with ASD do not have a generalised deficit in visuo-spatial abilities. MR performance **varied** between studies and **showed** no overall meta-analytic effect between groups, neither for slopes nor for intercepts (Muth et al., 2014). There was only a weak population effect size for performance differences in MR between ASD and TD located in intercepts and no overall superior performance of individuals with ASD in RT slopes (Muth et al., 2014).

Finally, we sought to investigate possible sex differences in MR performance within a group of individuals with ASD **and interpreted the results in the context of EMB theory**. Given that female ASD performance is located between TD female and male performance, there is no ceiling effect of MR which might hide sex differences among ASD **in our study**. Contrary to EMB predictions we could not verify superior performance of ASD compared to TD individuals. EMB does not explicitly differ between ASD females and males, but following its logic, MR performance of ASD

females **should** be better than the performance of TD males and equivalent to the performance of ASD males. However, this could not be underpinned by this study.

There is **now** emerging evidence for a female ASD cognitive profile which is different from the cognitive profile that males with ASD show (Frazier, Georgiades, Bishop & Hardan, 2014; Lai, Lombardo, Pasco & Ruigrok, 2011; see Lehnhardt et al., 2016, for a recent summary). While ASD males show superior performance in visuo-spatial tasks, ASD females outperform their male counterparts in terms of processing speed (Bölte, Duketis, Poustka & Holtmann, 2011; Koyama, Kamio, Inada & Kurita, 2009; Lehnhardt et al., 2016) and executive functions (Lehnhardt et al., 2016) while they perform worse in terms of working memory (Nydén, Hjelmquist & Gillberg, 2000). Those functions are required in different aspects of MR but performance differences between ASD females and males in certain cognitive domains might be covered by an equal general performance in MR. For example, executive functions and working memory are both encoded in intercepts (Hooven et al., 2004) but while females excel in one domain, they **might not** in the other domain so **speculatively, overall performance in an MR task might not differ. Although such a scenario might give a partial explanation for the results of the current study, it would not fully explain why MR as the most robust test for sex differences in TD does not reveal sex differences in ASD. However, since there is no test comparable to MR in terms of sex differences, it might be challenging but necessary to find a different approach to reevaluate sex differences in ASD in future research.**

In conclusion, this study adds to the controversial discussion on MR performance among TD and ASD and uncovers new findings on female ASD performance in MR. Importantly, it contributes to the growing body of studies showing that females with ASD do not just display an exaggerated male profile but instead need better characterisation and further research attention in the future.

References

Aleman A, Bronk E, Kessels RP, Koppeschaar HP, van Honk J. A single administration of testosterone improves visuospatial ability in young women. Psychoneuroendocrinology. 2004;29(5):612-7.

Alexander GM, Swerdloff RS, Wang C, Davidson T, McDonald V, Steiner B, et al. Androgen-behavior correlations in hypogonadal men and eugonadal men. II. Cognitive abilities. Horm Behav. 1998;33(2):85-94.

American Psychological Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington, VA: American Psychiatric Publishing; 2013.

Aster, M., Neubauer, A., & Horn, R. (2006). *Wechsler Intelligenztest für Erwachsene (WIE). Deutschsprachige Bearbeitung und Adaptation des WAIS-III von David Wechsler.* Harcourt Test Services: Frankfurt.

Astur RS, Tropp J, Sava S, Constable RT, Markus EJ. Sex differences and correlations in a virtual Morris water task, a virtual radial arm maze, and mental rotation. Behav Brain Res. 2004;151(1-2):103-15.

Baron-Cohen S. The extreme male brain theory of autism. Trends Cogn Sci. 2002;6(6):248-54.

Baron-Cohen S, Knickmeyer RC, Belmonte MK. Sex differences in the brain: implications for explaining autism. Science. 2005;310(5749):819-23.

Baron-Cohen S, Wheelwright S, Scahill V, Lawson J, Spong A. Are intuitive physics and intuitive psychology independant? A test with children with Asberger syndrome.: Journal of Developmental and Learning Disorders; 2001. p. 47-78.

Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E. The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. J Autism Dev Disord. 2001;31(1):5-17.

Beacher FD, Radulescu E, Minati L, Baron-Cohen S, Lombardo MV, Lai MC, et al. Sex differences and autism: brain function during verbal fluency and mental rotation. PLoS One. 2012;7(6):e38355.

Beck AT, Brown GK, R.A. S. Beck-Depressions-Inventar-FS (BDI-FS). Manual. Deutsche Bearbeitung von Sören Kliem & Elmar Brähler. Frankfurt am Main: Pearson Assessment.; 2013.

Bölte S, Duketis E, Poustka F, Holtmann M. Sex differences in cognitive domains and their clinical correlates in higher-functioning autism spectrum disorders. Autism. 2011;15(4):497-511.

Brosnan M, Daggar R, Collomosse J. The relationship between systemising and mental rotation and the implications for the extreme male brain theory of autism. J Autism Dev Disord. 2010;40(1):1-7.

Brosnan M, Walker I, Collomosse J. The effect of explicitely varying the proportion of "same" and "different" responses on sex differences in the Shepard and Metzler mental rotation task. European Journal of Cognitive Psychology; 2010. p. 172-89.

Bryden MP, George J, Inch R. Sex differences and the role of figural complexity in determining the rate of mental rotation. Percept Mot Skills. 1990;70(2):467-77.

Celec P, Ostatnikova D, Putz Z, Kudela M. The circalunar cycle of salivary testosterone and the visual-spatial performance. Bratisl Lek Listy. 2002;103(2):59-69.

Conson M, Mazzarella E, Frolli A, Esposito D, Marino N, Trojano L, et al. Motor imagery in Asperger syndrome: testing action simulation by the hand laterality task. PLoS One. 2013;8(7):e70734.

Courvoisier DS, Renaud O, Geiser C, Paschke K, Gaudy K, Jordan K. Sex hormones and mental rotation: an intensive longitudinal investigation. Horm Behav. 2013;63(2):345-51.

Falter CM, Arroyo M, Davis GJ. Testosterone: activation or organization of spatial cognition? Biol Psychol. 2006;73(2):132-40.

Falter CM, Plaisted KC, Davis G. Visuo-spatial processing in autism--testing the predictions of extreme male brain theory. J Autism Dev Disord. 2008(a);38(3):507-15.

Falter CM, Plaisted KC, Davis G. Male brains, androgen, and the cognitive profile in autism: convergent evidence from 2D:4D and congenital adrenal hyperplasia. J Autism Dev Disord. 2008(b);38(5):997-8.

Frazier TW, Georgiades S, Bishop SL, Hardan AY. Behavioral and cognitive characteristics of females and males with autism in the Simons Simplex Collection. J Am Acad Child Adolesc Psychiatry. 2014;53(3):329-40.e1-3.

Gill HS, O'Boyle MW, Hathaway J. Cortical distribution of EEG activity for component processes during mental rotation. Cortex. 1998;34(5):707-18.

Guy J, Mottron L, Berthiaume C, Bertone A. A Developmental Perspective of Global and Local Visual Perception in Autism Spectrum Disorder. J Autism Dev Disord. 2016.

Halari R, Hines M, Kumari V, Mehrotra R, Wheeler M, Ng V, et al. Sex differences and individual differences in cognitive performance and their relationship to endogenous gonadal hormones and gonadotropins. Behav Neurosci. 2005;119(1):104-17.

Hamilton AF, Brindley R, Frith U. Visual perspective taking impairment in children with autistic spectrum disorder. Cognition. 2009;113(1):37-44.

Hönekopp, J. Digit ratio 2D:4D in relation to autism spectrum disorders, empathizing, and systemizing: a quantitative review. Autism Res. 2012 Aug;5(4):221-30.

Hooven CK, Chabris CF, Ellison PT, Kosslyn SM. The relationship of male testosterone to components of mental rotation. Neuropsychologia. 2004;42(6):782-90.

Hulmi JJ, Ahtiainen JP, Selänne H, Volek JS, Häkkinen K, Kovanen V, et al. Androgen receptors and testosterone in men--effects of protein ingestion, resistance exercise and fiber type. J Steroid Biochem Mol Biol. 2008;110(1-2):130-7.

Knickmeyer RC, Baron-Cohen S, Auyeung B, Ashwin E. How to test the extreme male brain theory of autism in terms of foetal androgens? J Autism Dev Disord. 2008;38(5):995-6; author reply 7-8.

Koyama T, Kamio Y, Inada N, Kurita H. Sex differences in WISC-III profiles of children with high-functioning pervasive developmental disorders. J Autism Dev Disord. 2009;39(1):135-41.

Kozaki T, Yasukouchi A. Sex differences on components of mental rotation at different menstrual phases. Int J Neurosci. 2009;119(1):59-67.

Lai MC, Lombardo MV, Pasco G, Ruigrok AN, Wheelwright SJ, Sadek SA, et al. A behavioral comparison of male and female adults with high functioning autism spectrum conditions. PLoS One. 2011;6(6):e20835.

Lehnhardt FG, Falter CM, Gawronski A, Pfeiffer K, Tepest R, Franklin J, et al. Sex-Related Cognitive Profile in Autism Spectrum Disorders Diagnosed Late in Life: Implications for the Female Autistic Phenotype. J Autism Dev Disord. 2016;46(1):139-54.

Linn MC, Petersen AC. Emergence and characterization of sex differences in spatial ability: a meta-analysis. Child Dev. 1985;56(6):1479-98.

Lutchmaya S, Baron-Cohen S, Raggatt P, Knickmeyer R, Manning JT. 2nd to 4th digit ratios, fetal testosterone and estradiol. Early Hum Dev. 2004;77(1-2):23-8.

Manning JT, Baron-Cohen S, Wheelwright S, Sanders G. The 2nd to 4th digit ratio and autism. Dev Med Child Neurol. 2001;43(3):160-4.

McKeever WF, Rich DA, Deyo RA, Conner RL. Androgens and spatial ability: failure to find a relationship between testosterone and ability measures.: Bulletin of the Psychonomic Society; 1987. p. 465-8.

Magnuson KM, Constantino JN. Characterization of depression in children with autism spectrum disorders. J Dev Behav Pediatr 2011, 32: 332–340.

Manning, J. T., Stewart, A., Bundred, P. E., & Trivers, R. L. (2004). Sex and ethnic differences in 2nd to 4th digit ratio of children. Early Human Development, 80, 161–168.

Masi, A., DeMayo, MM., Glozier, N., Guastella, AJ. An Overview of Autism Spectrum Disorder, Heterogeneity and Treatment Options

Muth A, Hönekopp J, Falter CM. Visuo-spatial performance in autism: a meta-analysis. J Autism Dev Disord. 2014;44(12):3245-63.

Nydén A, Hjelmquist E, Gillberg C. Autism spectrum and attention-deficit disorders in girls. Some neuropsychological aspects. Eur Child Adolesc Psychiatry. 2000;9(3):180-5.

Ozonoff S, Cook I, Coon H, Dawson G, Joseph RM, Klin A, et al. Performance on Cambridge Neuropsychological Test Automated Battery subtests sensitive to frontal lobe function in people with autistic disorder: evidence from the Collaborative Programs of Excellence in Autism network. J Autism Dev Disord. 2004;34(2):139-50.

Pearson A, Marsh L, Hamilton A, Ropar D. Spatial transformations of bodies and objects in adults with autism spectrum disorder. J Autism Dev Disord. 2014;44(9):2277-89.

Presentation [Computer Software] (2014). Retrieved from www.neurobs.com.

Sachse M, Schlitt S, Hainz D, Ciaramidaro A, Schirman S, Walter H, et al. Executive and visuo-motor function in adolescents and adults with autism spectrum disorder. J Autism Dev Disord. 2013;43(5):1222-35.

Shepard RN, Metzler J. Mental rotation of three-dimensional objects. Science. 1971;171(3972):701-3.

Silk TJ, Rinehart N, Bradshaw JL, Tonge B, Egan G, O'Boyle MW, et al. Visuospatial processing and the function of prefrontal-parietal networks in autism spectrum disorders: a functional MRI study. Am J Psychiatry. 2006;163(8):1440-3.

Soulières I, Zeffiro TA, Girard ML, Mottron L. Enhanced mental image mapping in autism. Neuropsychologia. 2011;49(5):848-57.

Tapley SM, Bryden MP. An investigation of sex differences in spatial ability: mental rotation of three-dimensional objects. Can J Psychol. 1977;31(3):122-30.

Valdez P, Reilly T, Waterhouse J. Rhythms of mental performance.: Mind, Brain, and Education; 2008. p. 7-16.

Voracek M, Dressler SG. Lack of correlation between digit ratio (2D:4D) and Baron-Cohen's "Reading the Mind in the Eyes" test, empathy, systemising, and autism-spectrum quotiens in a general population sample.: Journal of Clinical and Experimental Neuropsychology; 2006. p. 1481-91.

Voyer D, Voyer S, Bryden MP. Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. Psychol Bull. 1995;117(2):250-70.

Werling DM, Geschwind DH. Sex differences in autism spectrum disorders. Curr Opin Neurol. 2013;26(2):146-53.

Zapf AC, Glindemann LA, Vogeley K, Falter CM. Sex differences in mental rotation and how they add to the understanding of autism. PLoS One. 2015;10(4):e0124628.