Title:

The Role of Inhibitory Control and Social Cognition in a Naturalistic Theory of Mind Test

Noa Keren

Word Count: 9831

Abstract

Background: Deficits in Theory of Mind (ToM), the attribution of mental states to oneself and others, are characteristic of Autism Spectrum Disorders (ASD) and have been linked to inhibitory control. Impairments were demonstrated in siblings of autistic individuals.

Method: Participants were 12-16-year-old from 129 families with at least one twin with ASD and 81 matched control twin pairs from the Social Relationships Study. Differences in performance between the ASD, co-twin and control groups in the Penny Hiding Game (naturalistic ToM), Luria Hand Game (inhibitory control) and false-belief ToM (social cognition) tasks were analysed. Relationships between the tasks were explored.

Results: The ASD group performed significantly ($p<.05$) worse than the control group in the PHG, LHG and false-belief ToM tasks. The co-twin group did significantly worse in the PHG than the control group. Significant correlations ($P<.05$) were found between the PHG and LHG and between PHG and first order ToM.

Discussion: Significant deficits in ToM and inhibitory control were demonstrated in autistic adolescents. The correlations suggested that better ToM performance in a naturalistic test may be associated with mastering inhibitory control. ToM deficits were also observed in typically developing co-twins, supporting the incorporation of ToM deficits in the broad autism phenotype.

Acknowledgements

The author sincerely expresses her gratitude to Professor Francesca Happé, for supervising this project and for the Social Relationships Study team at the Medical Research Council Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry for allowing me to use their collected data.

Introduction

Autism Spectrum Disorder (ASD) is a common neurodevelopmental disorder, characterised by the ‘triad of impairments’ (communication, social interaction and stereotyped behaviour) first described by Wing and Gould (1979). This study aimed to investigate the Penny Hiding Game; a naturalistic test of Theory of Mind and its relationship with executive function, particularly inhibitory control, and social cognition in children with ASD and their co-twins in comparison to typically developing (TD) twins. Twins were chosen due to the high genetic liability of ASD, as twin studies demonstrated heritability of 0.7 for ASD and 0.9 for the broader autism phenotype (Geschwind, 2009). Non-autistic relatives seem to have characteristics of the broader autism phenotype; defined as subclinical manifestations of one or a few of the triad of impairments (Happé et al, 2006a). The fact that association and genetic linkage studies have yet to identify specific loci for ASD may be explained by the vast heterogeneity of both the phenotypic and genotypic aspects of the spectrum (Geschwind, 2009). Some suggest it is time to give up looking for a single explanation or a genetic cause for ASD as the genetic effects seem to be specific to each of the triad of impairments (Happé et al, 2006a). Even within the broader phenotype each trait is highly heritable on its own (Geschwind, 2009). This suggests that a combination of various genes underpins susceptibility to ASD.

Theory of Mind

Baron-Cohen (1995) proposed that individuals with ASD show some degree of ‘mind-blindness’ or lack ToM, reflecting an impaired ability to attribute mental states. They fail to understand their own actions and the consequences thereof in terms of mental representations and subsequently cannot use the concept of mental states to understand and predict other people’s behaviour (Joseph and Tager-Flusberg, 2004). TD children develop ToM skills at around four years of age (Baron-Cohen, 2008).

ToM abilities are usually measured cognitively using false-belief tasks which aim to tap into the individual’s ability to attribute a false-belief in an experimental paradigm (Peskin and Ardino, 2003). A location change false-belief task, based on Wimmer and Perner (1983), involves telling a story using props in which an object’s location changes. Whilst the main character sees the location change, the second character is unaware of it. To succeed, participants must be able to understand that the absent character’s mental representation of the location is different to their own. Another version follows a similar style with an unexpected content task, for example a pencil in a Smarties tube. ToM deficits in ASD have been widely replicated. For instance, Baron-Cohen et al (1985) showed that 80% of the autistic children failed on both tests, which was significantly more than TD controls and children with Down’s syndrome. Ozonoff et al (1991a) demonstrated that children with ASD performed significantly worse on ToM measures compared to TD children.

ToM deficits were also observed in siblings of children with ASD. Dorris et al (2004) used the Reading the Eyes in the Mind Test to examine social cognition in siblings of children with Asperger syndrome and TD children. The siblings group scored significantly lower than the control group matched for age and vocabulary (Dorris et al, 2004). This is consistent with the notion of a broader autism phenotype which includes ToM and social cognition impairments. However, external validity is limited by the small sample. Thus, it calls for replication on a larger scale with siblings of ASD children from the entire spectrum.

Baron-Cohen et al (1985) suggest that a cognitive deficit in ToM could explain the lack of pretend play and social impairment in ASD. This was further supported by Peskin and Ardino (2003) who demonstrated that performance on false-belief tasks significantly related to the ability to play hide-and-seek successfully and to keep a secret. Hence, an association between experimental measures of social cognition and behaviour in a quasi-natural environment is shown. Furthermore, it follows that ToM skills are important for normal social interaction. This was also seen in Fisher and Happé's (2005) study as training in ToM was associated with an improvement in real life behaviour as inferred from teachers' reports. However, this study has somewhat limited external validity due to its small sample of middle class families. A meaningful comparison of social cognition measures with other measure of ToM in a large sample would go some way in further understanding the ToM deficit in ASD.

The Penny Hiding Game (PHG) is another measure of ToM, first devised by Gratch (1964). It is a naturalistic test of deception involving the experimenter or participant hiding a penny in their fist whilst the other tries to guess the penny's location. As deception involves inferring other people's beliefs and intentions, ToM is integral to it. In turn, understanding deception is vital for social interaction. The PHG occurs naturally as an enjoyable game between parents and children, hopefully decreasing motivational problems. It has also been shown to be highly suitable for autistic children as it entails minimal linguistic demands (Baron-Cohen, 1992). Baron-Cohen (1992) used revised coding to differentiate between object occlusion (keeping the penny out of sight) and information occlusion (preventing the experimenter from inferring the penny's location). Although children with ASD succeeded with object occlusion, significantly fewer passed information occlusion compared to TD and children with learning difficulties. Errors of information occlusion were more frequent in the ASD group (e.g. revealing the penny) (Baron-Cohen, 1992). All those who demonstrated information occlusion also displayed object occlusion, but the opposite did not apply, indicating that information occlusion is more developmentally sophisticated (Baron-Cohen, 1992). This failure at deception may be a result of the ToM deficits in autism as the children were oblivious to the experimenter's knowledge and thus, failed to deceive him/her. Although this study benefits from a better coding method, controlled design and structured assessment, the small sample size (15) limits its external validity. A further investigation in a large heterogeneous sample is required. Performance on the PHG was also related to passing a false-belief task. In Baron-Cohen's (1992) sample, those in the ASD group who showed information occlusion also passed the false-belief task whereas the opposite did not apply. This suggests that ToM is needed but is not sufficient to acquire information occlusion (Baron-Cohen, 1992). Oswald and Ollendick (1989) also found that the PHG significantly correlated with a dimension of role playing and with measures of social skills as inferred from teachers' and primary care-givers' reports. This indication that PHG correlated well with real-life social skills is highly noteworthy. Its relationship with social cognition, as measured by the false-belief tasks, is also intriguing and deserves further investigation.

Linking Executive Function and Theory of Mind

Executive function (EF) is associated with the prefrontal cortex and includes; planning, problem solving, set-shifting and inhibition. There have been suggestions that executive dysfunction could be a causal factor of the repetitive behaviours, communicative and social deficits in ASD as they all require the constant evaluation and selection of appropriate responses to stimuli (Joseph and Tager-Flusberg, 2004). Empirical support for executive dysfunction in ASD is strong; Happé et al (2006b) found significant EF impairments in ASD compared to TD children, matched for age and IQ. This complements longitudinal studies which demonstrated an improvement with age from childhood to adolescence, especially in high-functioning ASD (Happé et al, 2006b).

Executive dysfunction is also prevalent in siblings of individuals with ASD. For example, Hughes (1999) demonstrated significant difficulties with EF tasks especially on the Tower of London, indicating problems of planning and flexibility. This may suggest that executive dysfunction is present in the broader autism phenotype.

A relationship between EF and ToM deficits in ASD has been indicated. Its causal nature, however, remains controversial. Two main theories seek to explain the causal direction between ToM and EF. On one hand, Perner and Lang (2000) argue that understanding representations and mastering ToM is necessary for developing EF. One needs to understand his/her own representations before executing a plan (Pellicano, 2007). On the other hand, Russell (1996, 1997) claims exactly the opposite; one needs to acquire self-monitoring abilities before they can develop an understanding of other people's mental states. In other words, EF is the basis for establishing ToM.

Pellicano (2007) looked at the relationship between ToM and EF in autistic children and found that 40% showed both ToM and EF deficits. Striking dissociation emerged as 27% showed impaired ToM with intact EF whilst none showed intact ToM with impaired EF (Pellicano, 2007). No significant differences in chronological age, nonverbal IQ and symptom severity were found. This supports Russell's approach and provides explicit evidence against Perner's. However, these results are limited by the somewhat arbitrary definition of impairment and possible difficulty in equality between the EF and ToM tasks. Although the pattern of association may be clear, it was not a longitudinal study and thus causality cannot be established. Russell's theory is further supported by findings from a training study which demonstrated improvement solely in ToM two months following training in either ToM or EF (Fisher and Happé, 2006). Thus, although those trained in EF did not improve on EF tasks, they did show significant advancement in ToM. This may reflect a methodological limitation if the tasks were not of matched difficulty. External validity is challenged by the small sample and further longitudinal studies are needed to substantiate causation. Nevertheless, these preliminary conclusions further support Russell's theory.

The Luria hand game (LHG) is a simple task developed by Luria et al (1964) to test inhibitory control in patients with frontal lobe lesions. Since then it has also proven useful in testing inhibition in autistic children as it benefits from a simple, linguistic-free structure. The task has two parts; firstly the participant copies the experimenter's hand movement (fist or point) and secondly, the participant has to execute the opposite action (Pellicano, 2007). Hughes (1996) demonstrated that children with ASD were able to imitate the actions but performed significantly worse than verbally-matched controls in the conflict condition. This could be explained by an inability to inhibit direct response to external stimuli, indicating executive dysfunction.

Hughes (1998) speculates that social contact requires children to monitor their own and others' actions and to adjust their behaviour accordingly. Hence, children's emerging EF skills facilitate social interactions. For children with ASD, problems of inhibition may further limit their already small scope for social interaction which may have detrimental implications for their ToM development (Hughes, 1998). Performance on the LHG positively correlated with the level of meta-cognitive awareness in another task suggesting there may be a link between impaired inhibitory control and acquiring ToM skills (Hughes, 1996). Moreover, Hughes et al (1996) reported a correlation between interviewers' ratings of social impairments in parents of children with ASD and executive dysfunction. Although weakened by subjective biases, these findings are suggestive of an association of EF and social skills in the broader autism phenotype. However, a direct comparison of inhibitory control measures and ToM abilities has yet to be done. As both PHG and the LHG are simple, linguistic-free tasks they could lend themselves well to a meaningful comparison between ToM and EF in an ASD sample.

The Present Study

This study aims to investigate the associations of ToM skills in a naturalistic game with social cognition and inhibitory control. It aims to do so in the context of children and adolescents with ASD, their TD co-twins and TD control twins. This design will allow investigation of ToM and EF abilities in ASD compared to TD. Looking at the co-twin group will permit an exploration of the broader autism phenotype. A large sample of 129 twin sets in which at least one twin has ASD (from all abilities on the spectrum) and 81 TD twin sets of participants aged 12-16 years.

The hypotheses of this study are as follows. Firstly, the ASD group will demonstrate worse performance in the PHG, LHG and ToM false-belief tasks than the control group. Secondly, the co-twin group may show some impairment on the three tasks with worse performance than the control group but better than the ASD group, reflecting the deficits in the broader autism phenotype. Thirdly, performance on the PHG will correlate with inhibitory control as measured by the LHG. Lastly, performance on the PHG will correlate with social cognition as measured by the ToM false-belief tasks.

Method

Participants

This study is looking at data collected as part of the ongoing Medical Research Council funded Social Relationships Study (SRS). Participants were recruited to the SRS from the Twins' Early Development Study (TEDS) in 2007 (see figure 1 for the selection process). The TEDS is a large-scale follow-up study of 13,472 families recruited from the Office for National Statistics as families of twins born in England and Wales in 1994-1996. It has been shown to be fairly representative of families with children in the UK (Oliver and Plomin, 2007). Using the Childhood Asperger Syndrome Test (CAST) and the ASD module of the Development and Well-being Assessment (DAWBA; Goodman et al, 2000), 230 families were identified with at least one child who met DAWBA ASD criteria. Five families were also recruited from a National Autistc Society advert and mailout to all UK child psychiatrists. The control sample comprised of 81 families recruited via TEDS with CAST score below twelve. They were matched to the ASD group in terms of gender, zygosity, age and socio-economic status. Participants characteristics (sex, zygosity, age and IQ) are detailed in table 1 accodring to groups.

Table 1 Participants' characteristics: sex and zygosity; number (percentages in each group), Age and IQ (Full-Scale Wechsler Abbreviated Scale of Intelligence [WASI], Verbal and Performance); mean (standard deviation) according to groups.

		ASD	Co-Twin	Control
Sex	**Female**	32 (18%)	48 (64%)	131 (31%)
	Male	151 (83%)	27 (36%)	289 (69%)
Zygosity	**MZ***	49 (27%)	5 (7%)	110 (26%)
	DZ*	129 (70%)	69 (92%)	304 (73%)
	Unknown	5(3%)	1 (1%)	6 (1%)
Age (in months)		162 (8.00)	162 (8.00)	154 (13.00)
Full-scale (WASI) IQ		95.00 (18.00)	104.82 (13.38)	102.48 (15.76)
Verbal IQ		92.63 (17.96)	104.23 (14.17)	100.37 (16.11)
Performance IQ		98.16 (18.08)	104.35 (13.89)	103.83 (14.56)

*MZ=monozygotic DZ=dizygotic

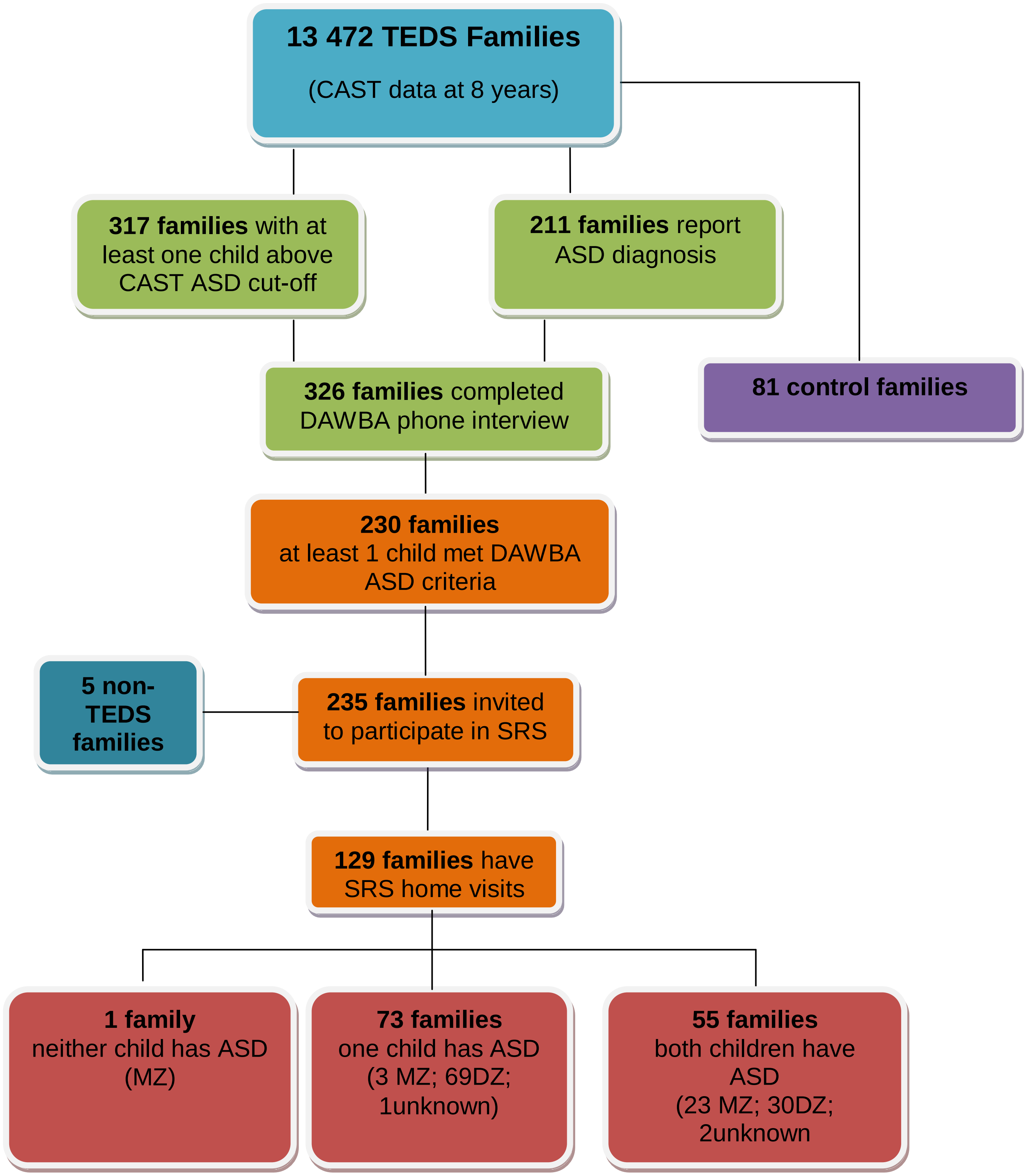


Figure 1 A visual representation of the selection of families from the TEDS family study to participate in the SRS.

Materials and Procedure

The SRS home visits involved the administration of an extensive cognitive battery assessing IQ, executive function, language, ToM and central coherence abilities as well as self/parent questionnaires. The order of the tests was counterbalanced within each twin pair in order to reduce order-effect bias. In addition, families in the ASD group were given gold-standard diagnostic assessments; the Autism Diagnostic Observation Schedule Generic (ADOS-G; Lord et al, 2000) and the Autism Diagnostic Interview-Revised (ADI-R; Le Couteur et al, 2003). All data was double-coded and checked using video recordings of the tasks. This study will look at three tests; the PHG, LHG and the ToM false-belief tasks.

Penny Hiding Game (Gratch, 1964; Baron-Cohen, 1992)

This task involves the experimenter/participant hiding a penny in their fist whilst the other tries to guess its location. The experimenter begins by hiding the penny in one hand behind their back before bringing their hands forward and asking the participant what hand they think the penny is in. This is repeated in the order RLRR. Then, the roles reverse and the participant has to hide the penny. The participant performs a maximum of six trails but the task is discontinued after four trails if no tricks or errors were displayed. During the task the experimenter marks whether they could guess the penny's location or not and notes down any tricks or errors, although emphasis is given to the natural feel of the game.

Errors and tricks are scored as follows:

E1 – participant does not keep both hands out of sight when hiding the penny
E2 – one hand used (hiding or presenting)
E3 – hand(s) are open
E4 – tells you where the penny is (opens hand or verbally tells you)
E5 – display error (penny is hidden but you can tell where it is, e.g. by grip)
T1 – emphasised the wrong hand
T2 – presents two empty hands
T3 – post-decision trick (saying it is the wrong hand, transferring penny)

Luria Hand Game

This task was first devised by Luria et al (1964) and based on Hughes (1996) who used it with ASD participants. There are two parts to this game; initially the participant requires to copy the experimenter's hand gestures (point or first) known as the imitation condition. This is followed by the conflict condition wherein the participant is required to make the opposite gesture to the experimenter. The experimenter makes the gestures in the following order: FPFFPFPPFP in the imitation condition and PFPPFPFFFP in the conflict condition. The task is discontinued if four consecutive correct initial responses are made up to a maximum of ten trails. Scoring includes the participant's initial response as well as any self correction or hesitation

Theory of Mind False-Belief tasks

These tasks test the participant's social cognition. Their ordering depends on the participant's performance (see figure 2 for the detailed order). The first task includes a standard unexpected contents task based on Hogrefe et al (1986) which tests first and second order ToM using a pencil box filled with candles. The experimenter asks for pencils whilst a pencil case is clearly within the participant's sight. First order ToM is tested by asking the participant what their mum/dad would think is in the pencil box.

This is followed by the chocolate story task, adapted from Wimmer and Perner (1983). The story involves Mary and John who receive chocolate from their grandfather which they put in the fridge. John then transfers it to his bag. First order ToM is tested by asking where Mary would think the chocolate is. Later it is revealed that Mary was actually watching John transfer the chocolate without him knowing. Second order ToM is then tested by asking where John would think Mary will look for the chocolate.

Participants who only pass one first order task then do the Sally-David unexpected transfer task (based on Wimmer and Perner, 1983; modelled after Baron-Cohen et al, 1985). This task was not analysed because not all the participants did it. Participants who pass two first order tasks go on to the Trip to the Seaside task (adapted from Bowler, 1992). This task involves a short story about Susan and Tom who are planning to go to Brighton by train. Tom goes to the station early and discovers that all the trains were cancelled. Asking whether Susan knows that the trains have been cancelled tests first order ToM. As the story continues, it is revealed that Tom is going to the bus station to buy tickets whilst Susan hears on the news that the trains are cancelled and goes to Tom's house. Second order ToM is examined when Tom's mum tells Susan he has gone to buy tickets and the experimenter asks "where does Susan think Tom has gone to buy the tickets?".

Throughout the stories comprehension and memory questions are asked and all answers are recorded.

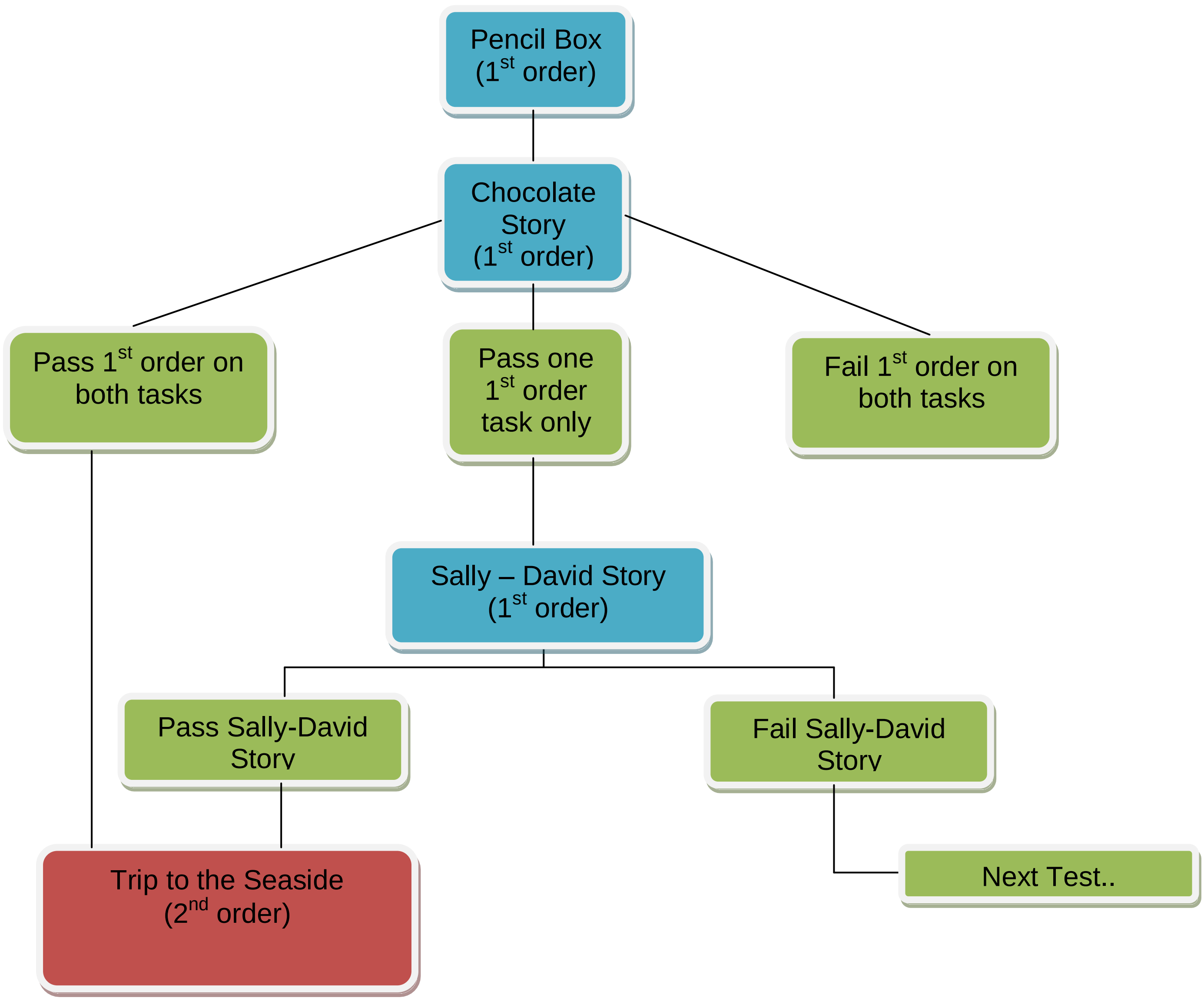


Figure 2 A visual representation of the order in which the ToM false-belief tasks were carried out. Participants only took part in a second order ToM task if they successfully passed two first order tasks (out of a possible three).

Results

Results are presented as follows: firstly differences in performance on the Penny Hiding Game (PHG), Luria Hand Game (LHG) and Theory of Mind (ToM) tasks between the groups are described. Secondly, the associations of the PHG with LHG and ToM tasks are reported.

Statistical Analyses

The test scores' distributions (PHG, LHG and ToM) were highly skewed and so violated parametric assumptions; The Kolmogorov-Smirnov test showed significant results for all tests (PHG errors: D(310)=.48, PHG tricks: D(310)=.51, LHG: D(310)=.45, 1^{st} order ToM: D(310)=.53, 2^{nd} order ToM: D(310)=.53, all at p<.001) after Lilliefors significance correction. Levene's test revealed that the variances across all tasks and groups were significantly non-homogenous (F(2,307)=10.450, F(2,307)=4.671, F(2,307)=15.359, F(2,307)=8.639, F(2,307)=9.171, all at p<.001). Although these tests are less accurate with large samples (Field, 2009), the skewed distributions were confirmed by examining the frequency curves, kurtosis and skewness scores. Log transformations did not 'normalise' the distributions. Therefore, nonparametric tests were carried out to assess whether the differences between the groups were significant. Kruskal-Wallis tests were used to test whether there was a significant difference in task performance between the three groups for all three tasks. Mann-Whitney tests were used to find between which two groups there was a significant difference in performance.

The use of nonparametric tests in some cases led to a large number of tests. This may mean significant results capitalised on chance and the overall probability of Type I errors was beyond 5%. However, corrections for multiple comparisons were not done as Howell (2002) argues that this is not warranted in the case of a priori predictions as were done here with pre-planned analyses. As performance was expected to be intact for some tasks, setting the acceptable α level too low may reduce the power to detect group differences. Hence, exact *p* values are reported so that caution can be taken to interpret the results conservatively.

Penny Hiding Game

The analysis of this task (N=363) focused on the first four trails because the task was discontinued after four trails for most participants, from all groups, as they made no errors or tricks. Performance on this task was examined by looking at the errors and tricks participants made; (detailed in tables 1 and 2 in the Appendix).

Looking at overall performance (Appendix fig.1), display error was the commonest in all groups whilst the commonest trick was emphasising the wrong hand. As participants in all groups progressed through the game they made fewer errors (Appendix fig.2).

There was a significant difference between the groups (H(2)= 32.47, p<.001). In agreement with hypothesis 1, the ASD group made significantly more errors (Mdn=0.00) than the control group (*U*=8713, z=-5.69, p<.001, d=-.33). The co-twin group performed significantly worse than the control group (*U*=4299.5, z=-3.99, p<.001, d=-.27), but there was no significant difference in performance to the ASD group. Hence, the second hypothesis is only partially supported.

The total tricks score significantly differed between the groups (H(2)=14.92, p=.001). Surprisingly, the ASD group made significantly more tricks than the co-twin group (*U*=4377.00, z=-2.49, p=.012, d=-.17) and significantly more than the control group (*U*=10025.50, z=-3.45, p=.001, d=-.20). There were no significant differences between the control and co-twin groups.

Dividing the ASD group into ASD and Broad-Spectrum (according to ADOS, ADI and clinical consensus) showed that the ASD group had a significantly higher mean number of errors in all trials than the Broad-Spectrum group (*U*=1444, z=-2.33, p=.020, d=-.20) as shown in fig. 2 in the Appendix.

In terms of overall performance, a pass/fail score for PHG was calculated whereby those who made no errors passed and those who made one or more errors failed. Figure 3 presents the percentage of 'passers' and 'failers' in each group. In agreement with the first hypothesis; Chi-squared tests demonstrated that the ASD group had significantly more 'failers' than the control group (χ^2(1, N=306)=31.34, p<.001). The control group had significantly more 'passers' than the ASD and co-twin group (χ^2(1, N=306)=15.44, p<.001). The ASD and co-twin groups did not differ significantly.

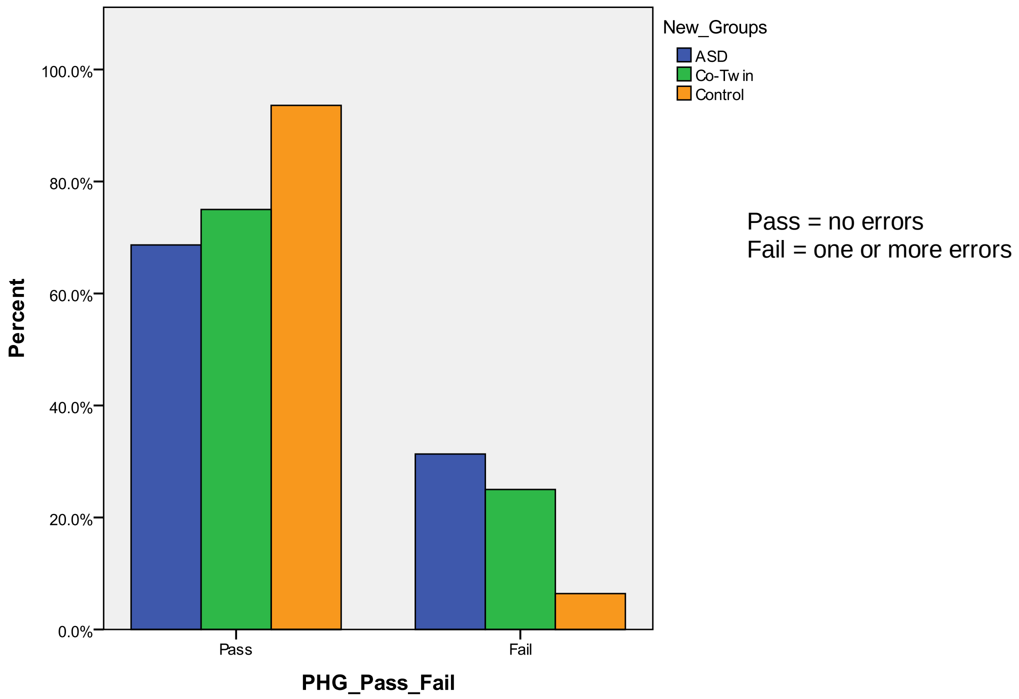


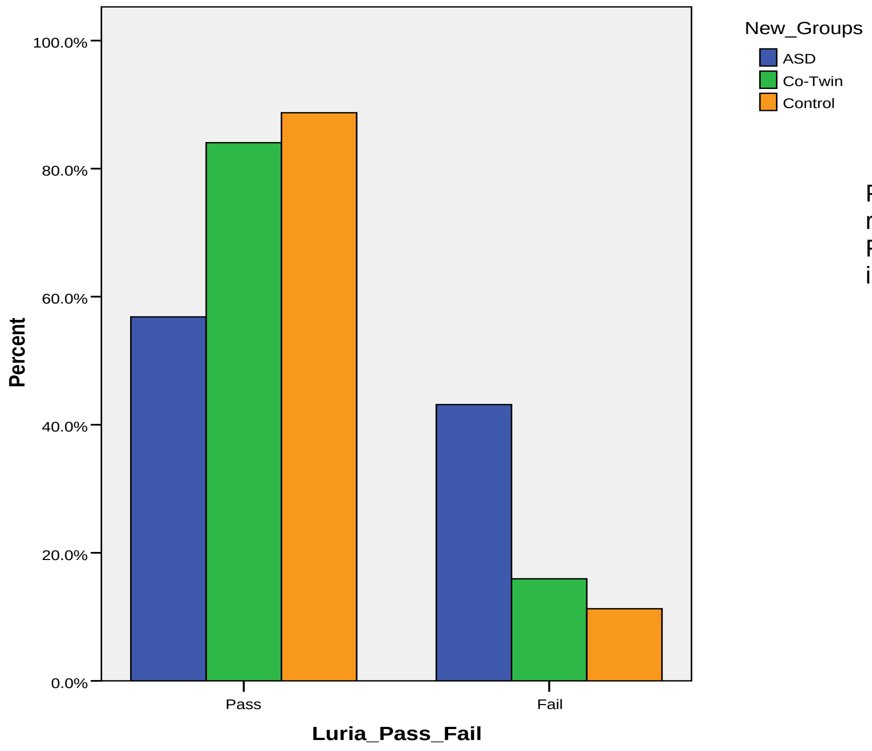
Figure 3 Percentage of each group who passed or failed the Penny Hiding Game

Luria Hand Game

Keeping with the research question regarding inhibitory control and seeing that 89% of participants gained full marks in the imitation condition, the analysis of this task focuses on the conflict condition. It will also look at the first five trails as most participants were successful and so the task was discontinued after the fifth trial. The LHG score (N=357) according to correct initial responses is illustrated in Table 3 in the Appendix.

The correct initial response score (Mdn=5.00) was significantly different between the groups (H(10)=44.43, p<.001). Consistent with hypothesis 1, the ASD group had significantly fewer correct initial responses than both the co-twin and control groups (*U*=3706.00, z=-3.71, p<.001, d=-.25; *U*=6889.00, z=-6.29, p<.001, d=-.371 respectively). There were no significant differences between co-twin and control groups. Within the ASD group, Broad-Spectrum and ASD did not exhibit significant differences in LHG performance.

A pass/fail score was calculated for the LHG whereby 'passers' made no mistakes in their initial responses and those who made one or more mistakes failed (see figure 4). The control group had significantly more 'passers' (89%) than the ASD group (57%) (U=8784.00, z=-5.59, p<.001, d=-.32). The control group also performed significantly better than the co-twin group (U=4318.00, z=-3.92, p<.001, d=-.26). The ASD and co-twin groups did not significantly vary.



Pass = no incorrect initial responses
Fail = one or more incorrect initial responses

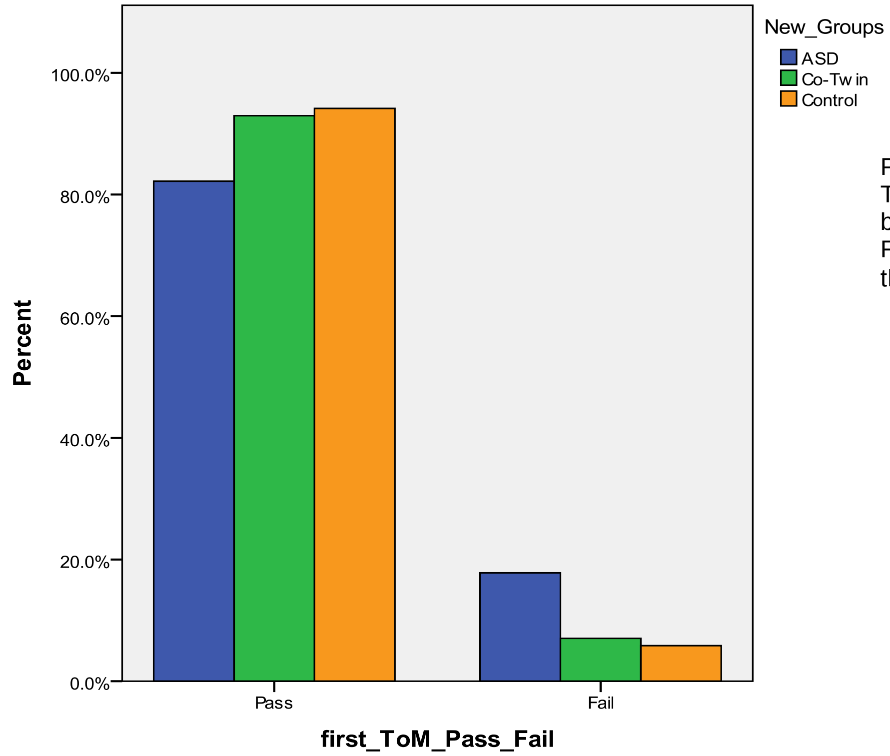
Figure 4 Bar Chart showing the percentage of each group who passed or failed the Luria Hand Game.

Theory of Mind False-Belief Tasks

This task was analysed in two parts; first and second order ToM. First order ToM was based on the chocolate and pencil box tasks whilst second order ToM was based on the chocolate and trip to the seaside tasks. Seeing that most participants were successful, using the memory and comprehension questions to confirm their understanding of the story was unnecessary. Thus, analysis focused on the first and second order questions only. The first order ToM score (N=371; Mdn=4.00) significantly differed between the groups (H(2)=12.66, p=.002) (please see Table 4 in the Appendix for detailed performance). Moreover, the ASD group performed significantly worse than both the co-twin group (*U*=4617.50, z=-2.15, p=.032, d=-.15) and the control group (*U*=9883.50, z=-3.25, p=.001, d=-.19) supporting the first hypothesis. There were no significant differences between the co-twin and control groups. Within the ASD group; the broad-spectrum group performed significantly better than the ASD group (*U*=1583.50, z=-2.71, p=.009, d=-.22).

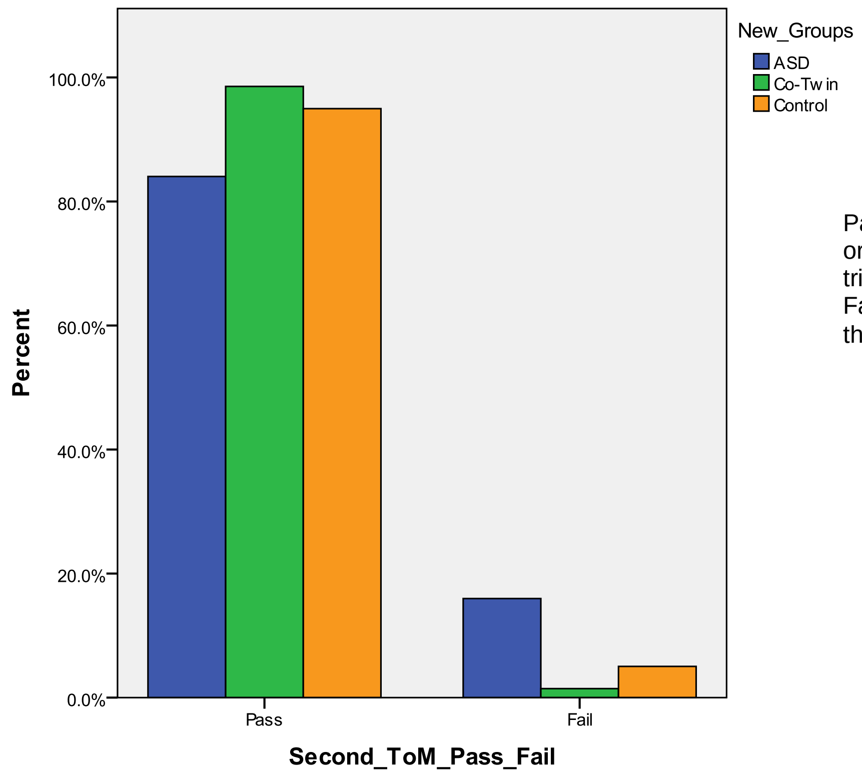
Table 5 in the Appendix details performance on second order ToM tasks (N=372) according to groups. There was significant variation between the groups (Mdn=4.00, H(2)=17.10, p<.001). The ASD group performed significantly worse than both the co-twin (*U*=4245.50, z=-3.13, p=.002, d=-.22) and the control groups (*U*=10187.50, z=-3.15, p=.002, d=-.18). There were no significant differences between the co-twin and control groups or between the ASD and broad-spectrum groups.

A pass/fail score for first order ToM was calculated whereby those who passed both tasks passed overall and those who failed one or more tasks, failed overall (figure 5). The same was done for second order ToM (figure 6). The ASD group had significantly more 'failers' (18%, 16%; first and second order respectively) than both the co-twin group (*U*=4625.00, z=-2.12, p=.039, d=-.14; *U*=4245.50, z=-3.13, p=.002, d=-0.22) and the control group (*U*=9897.00, z=-3.22, p=.002, d=-.19; *U*=10187.50, z=--3.15, p=.002, d=-0.18). The control and co-twin groups did not significantly differ.



Pass = passing both first order ToM tasks (chocolate and pencil box)
Fail = passing one or none of the first order ToM tasks

Figure 5 Bar chart showing the percentage of each group who passed or failed first order theory of mind according to performance on the false belief tasks chocolate and pencil box.



Pass = passing both second order ToM tasks (chocolate and trip to the seaside)
Fail = passing one or none of the second order ToM tasks

Figure 6 Bar Chart showing the percentage of each group who passed or failed second order theory of mind according to performance on the false belief tasks chocolate and trip to the seaside.

Inter-relations of the Penny Hiding Game

Hypotheses 3 and 4 considered whether performance on the PHG was related to performance on the LHG and ToM tasks. To begin with the differences in performance on those two tasks between 'passers' and 'failers' of the PHG was investigated. Table 2 details the characteristics of PHG 'passers' and 'failers' as well as the performance of LHG and ToM 'passers' in the PHG. Those who passed the PHG had a significantly better score on the LHG (*U*=7723.50, z=-2.57, p=.010, d=-.14). PHG 'passers' did significantly better than 'failers' in the first order ToM (*U*=8590.50, z=-2.38, p=.013, d=-.13). However, there was not a significant difference in performance in second order ToM.

Participants who passed the PHG were significantly younger (Mdn=158months, H(1)= 12.17, p<.001) than those who failed. 'Passers' also had significantly higher scores on Full-Scale (WASI) IQ (Mdn=101.5 H(1)= 12.17, p<.001) and performance IQ (Mdn=104, H(1)=6.12, p=.013) but not verbal IQ score.

Table 2 Characteristics of 'passers' and 'failers' of the PHG: number, mean age and full-scale IQ (standard deviation). Performance in the PHG of those who passed the LHG and ToM.

PHG*				LHG*		1st Order ToM*		2nd Order ToM*	
	N	Age (Months) M(SD)	Full-Scale IQ M(SD)	N	%	N	%	N	%
Passers**	300	158(12)	101.21(15.74)	214	84%	264	83%	268	83%
Failers**	74	162(9)	94.40(19.57)	42	16%	54	17%	56	17%

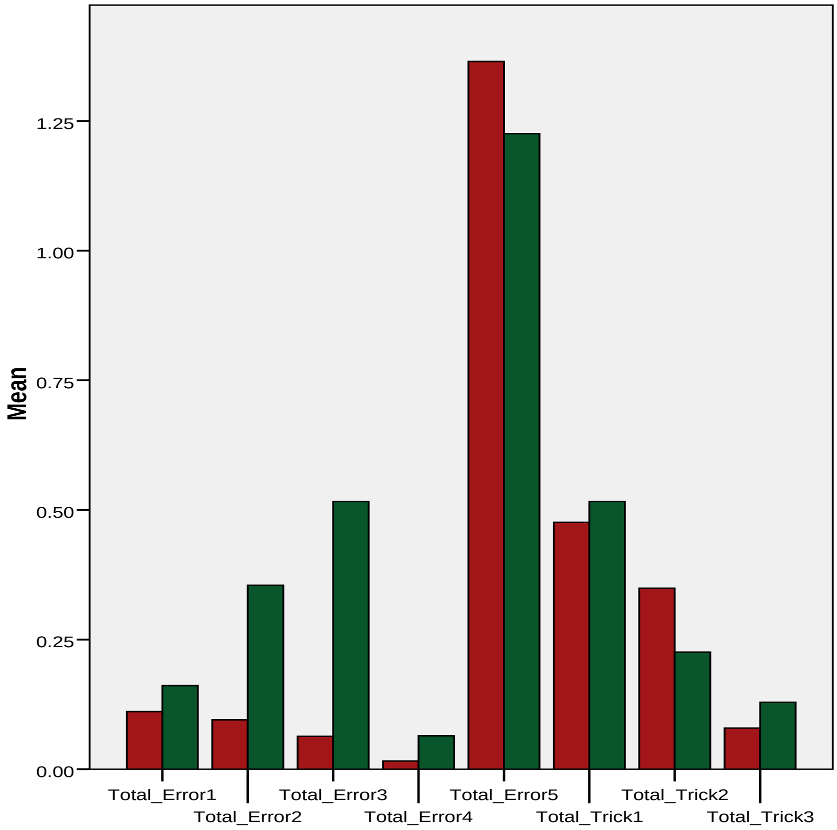
*PHG=Penny Hiding Game, LHG=Luria Hand Game, ToM=Theory of mind task
** Pass=no errors made, Fail=more than one errors made

Further exploration using Spearman's correlation (see Appendix Table 6) revealed significant (1-tailed as direction predicted) weak correlations between PHG and LHG and between PHG and first order ToM. 2% of the variance in ranks in LHG was shared with the PHG, similarly, 2% of the variance in ranks in first order ToM was shared with the variance in PHG ranks. Consistent with hypotheses 3 and 4, the better participants performed in the LHG and first order ToM, the lower number of errors they did on PHG. There was no significant correlation between PHG and second order ToM.

Performance on PHG demonstrated a significant weak correlation with age, full-scale IQ, verbal IQ and performance IQ. 3% of the variance in ranks in age was shared with the PHG (Appendix, Table 7). 1-3% of the variance in ranks in the PHG was shared with IQ variations. Overall, higher IQ score and younger age were associated with better performance in the PHG.

As correlation may be affected by the significant differences in performance between the three groups, further correlations looked at each group separately (Appendix, Table 8). The only correlations to remain significant were in the ASD group; significant weak correlation of PHG with LHG and first order ToM. This correlation remained significant (PHG:r_s=-.17 LHG:r_s=-.23, p<.005) after excluding the broad-spectrum group from the ASD group. The correlation between PHG and full-scale, performance and verbal IQ remained significant only in the ASD group (Appendix, Table 9). Thus, higher IQ correlated with better performance on the PHG in the ASD group only.

Having found a significant correlation between PHG and LHG, performance of 'passers' and 'failers' of the LHG in PHG were also explored. Figure 7 presents the types of errors and tricks in the PHG made by 'passers' and 'failers' of the LHG. Chi-squared tests revealed that LHG 'failers' made significantly more of error 3 (hand(s) are open) than 'passers' (χ^2(4, N=357)=10.03, p=-.006). There were no other significant differences.



Luria_Pass_Fail
Pass
Fail

Luria Hand Game:
Pass = no incorrect initial responses
Fail = one or more incorrect initial responses

Penny Hiding Game:
Error 1 – do not keep both hands out of sight when hiding the penny
Error 2 – one hand used (hiding or presenting)
Error 3 – hand(s) are open
Error 4 – tells you where the penny is (verbally or by opening the hand)
Error 5 – display error (penny is hidden but you can tell where it is eg. by grip

Trick 1 – emphasised the wrong hand
Trick 2 – presents two empty hands
Trick 3 – post-decision trick (saying it is the wrong hand, transferring penny).

Figure 7 Bar chart showing mean errors and tricks by error/trick types in the Penny Hiding Game, according to 'passers' and 'failers' of the Luria Hand Game.

The shared family factors between each twin pair may render the data not entirely independent. In order to test whether the results were robust, analyses were re-run after selecting one twin from each pair in each group. The second analyses revealed the same trends both in terms of group differences and correlations. The correlations remained significant ($p<.005$) as did most group differences except on performance in false-belief ToM.

Discussion

This study aimed to investigate ToM, inhibitory control and social cognition in individuals with ASD, their typically developing (TD) co-twins and controls. Findings demonstrated significant deficits in all three measures in the ASD group compared to TD. For some measures the co-twin group also exhibited impairments compared to controls. The Penny Hiding Game (PHG) showed significant interactions, albeit weak, with Luria Hand Game (LHG) and Theory of Mind (ToM). Further discussion will consider each hypothesis in turn.

Penny Hiding Game

In the PHG, although most participants in all groups performed to a high standard, the ASD group made significantly more errors and had more ‘failers’ than the control group, as predicted in hypothesis one. This is consistent with former studies (Baron-Cohen, 1992, Oswald and Ollendick, 1989) and adds to the substantial empirical base of ToM deficits in ASD. Partially supporting hypothesis two, the co-twin group performed significantly worse than the control group but similarly to the ASD group in ToM. This suggests that ToM deficits form a part of the broader autism phenotype as previously demonstrated using different measures of ToM (e.g. Dorris et al, 2004). The fact that these deficits remained in a naturalistic test supports the value of experimental paradigms to tap into ToM. This may indicate that social interaction impairments are a part of the broader phenotype.

Surprisingly, the findings showed that the ASD group used considerably more tricks than the other two groups. There are a number of potential explanations for this peculiarity. It may be a methodological issue of coding definitions as the commonest trick was emphasising the wrong hand which is similar to error 2 (using one hand). Perseveration in the ASD group could have inflated the tricks score unintentionally, especially as very few tricks were used overall. Although the PHG is interesting and highly appropriate for children with learning disabilities, it may not have been age-appropriate for the TD participants. Embarrassment or boredom could result in behaviour aimed at appeasing the experimenters rather than tricking them. A younger control group may have been more suitable. Future work could develop an age-appropriate naturalistic test of deception that taps into ToM, perhaps looking at bluffs with the aid of new technologies (e.g. constructing real-life situations in a virtual reality setting).

Luria Hand Game

The LHG findings confirmed the first hypothesis as the ASD group's performance was significantly worse than both control and co-twin groups with fewer 'passers' overall. This indication of an executive dysfunction, particularly in inhibitory control (ASD participants could not inhibit copying the experimenter's hand gesture) is consistent with previous findings (Hughes, 1996). Considering that there was no significant difference between control and co-twin groups may point to the simplicity of the task for TD adolescents. It may also suggest that inhibitory control may not be an integral part of the broader phenotype. This contradicts previous studies which found executive dysfunction in siblings of ASD individuals (Hughes, 1999). Perhaps only the more challenging executive functions are impaired in the broader phenotype. Hughes (1999) for instance, reported deficits in planning and cognitive flexibility in siblings of autistic individuals. A hierarchy of functioning may be implied; whilst inhibitory control deficits were not found in the co-twin group, ToM impairments were evident. This is broadly in line with Russell's theory (1996, 1997) wherein establishing EF is necessary prior to acquiring ToM.

False-Belief Tasks

In both first and second order ToM false-belief tasks, although most participants performed well, the ASD group had significantly fewer 'passers' than both the co-twin and control groups. This is consistent with the third hypothesis and contributes to the vast empirical support for social cognition deficits in ASD (e.g Ozonoff et al, 1991a, Pilowsky et al, 2000). In both conditions, there was no significant difference in performance between the control and co-twin groups. Perhaps this indicates that the task was not entirely age-appropriate as previous studies have found ToM deficits in siblings of autistic children (Dorris et al, 2004). Using a different test of ToM such as the 'Strange Stories' task (Happé, 1994b) may be more challenging as it asks participants to infer characters' utterances which are not literally true (white lies, double bluffs) and justify their answers. Such advanced tests may be able to identify subtle ToM deficits in the broader phenotype.

In first order ToM, dividing the ASD group into ASD and broad-spectrum demonstrated that the ASD group did significantly worse. However, in second order ToM there were no significant differences between the two. This might indicate that those with broad-spectrum ASD have less impaired ToM or only exhibit ToM deficits when faced with more challenging situations. The broad-spectrum group may be able to compensate for their ToM deficits in less demanding situations using other skills to infer the correct response. This pattern of findings is similar to that of Ozonoff et al (1991b) which showed that high-functioning, but not Asperger's syndrome, participants exhibited second order false-belief impairments. The fact that a significant proportion of the ASD group still failed even the simpler first order ToM tasks highlights the severity of the ToM deficit in ASD.

Nevertheless, the overall performance of the ASD group in the ToM tasks was better than expected and better than reported in previous studies. Baron-Cohen et al

(1985) showed that 80% of the ASD participants could not attribute a false-belief. The higher level performance in this study may reflect the older age of the participants. Better interventions and an emphasis on ToM training in specialist schools might also contribute to their improved abilities. Future studies ought to further investigate whether ToM abilities translate into better real-life social behaviour. This could be done using a simple design of correlation between false-belief tasks and measures of social behaviour, such as the social score on the ADOS or reports from teachers and parents. Another option is a more complex intervention design whereby effects of teaching ToM are measured in everyday social behaviour. Although the latter was tried in Fisher and Happé's (2005) study, only tentative indications of behavioural improvement following ToM training were possible, since no significant differences were found in pre and post-test real-life behaviour. A larger sample followed over a longer period of time may demonstrate the extension of cognitive improvements into behaviour.

Inter-relations of the Penny Hiding Game

Moving on to the inter-relations of PHG, analyses revealed that those who passed the PHG performed significantly better in the LHG and first order ToM than those who failed the PHG, albeit with effects of small magnitude. Thus, it points to an association between passing or failing the PHG and performance on the LHG and ToM. Interestingly, the association between ToM and PHG was also shown in Fisher and Happé's (2005) study as the group who was trained on false-belief ToM exhibited some generalisation of learning to the PHG. However, this association does not imply causation; longitudinal studies in which one test is manipulated at a time or in separate matched groups will help to elucidate the causative pattern. The lack of a significant difference between PHG 'passers' and 'failers' in second order ToM may reflect the fact that only those who passed two first order tasks moved onto second order tasks. Hence, a more able sample took part in this test which may have excluded more PHG 'failers' than 'passers' and limited the power to detect significant differences.

As expected, PHG 'passers' had significantly higher full-scale (WASI) IQ scores reflecting their increased ability. Surprisingly, PHG 'passers' were significantly younger than 'failers' which may again reflect the age-appropriateness of the PHG. The lower functioning children in the ASD group may have been more motivated to take part playfully than the TD 12-16 year olds.

These associations were further supported by the correlation analyses. Weak, yet significant, correlations between PHG and LHG and between PHG and first order ToM were presented. The better the participants did on the LHG and first order ToM, the fewer errors they made in the PHG, confirming hypotheses three and four. Success in the PHG requires an understanding of ToM and subsequently an ability to understand and participate in the deception involved in the game. Success also entails an ability to inhibit inappropriate reactions, such as exposing the penny, which complements successful deception. Therefore, those who grasped ToM and acquired the EF of inhibitory control were more likely to be successful and make fewer or no errors in the PHG.

Further correlations, taking each group separately in order to account for the significant group differences, revealed that only in the ASD group the correlations above remained significant. This could be explained by the fact that correlations tend to be higher in samples with a wider range (Cohen et al, 2003). The ASD group had the widest range of scores in the PHG with some participants making no errors and others making plenty of errors whereas the other two groups exhibited a near ceiling effect. A more challenging task for the TD participants may have exhibited a wider range of scores which may have yielded significant correlations. A younger control group could have had the same effect.

Further exploration of the association between PHG and LHG yielded some interesting findings. Those who failed the LHG made significantly more of error 3 in the PHG. The LHG requires participants to hold in mind an arbitrary rule and inhibit a prepotent response; thus, high scores reflect good inhibitory control (Pellicano, 2010). Error 3 involves opening the hand(s) which clearly relates to inhibitory control. Those who mastered the skill of inhibition could more easily inhibit their instinctive desire to open the hand and reveal the penny. Developmentally, it has been postulated that early difficulties in regulating one's own behaviour to inhibit socially inappropriate behaviours may have a negative effect on children's social contact and in turn disrupt their ToM development (Pellicano, 2010). Therefore, EF can be a limiting or enabling factor in developing ToM. Thus, not only would executive dysfunction make it harder to play the PHG, those with EF deficits might be developmentally disadvantaged with fewer opportunities to master ToM. It may be hypothesised that the few who failed first order ToM but passed the PHG may have been using these inhibitory control skills to overcome their ToM deficit.

The fact that no significant correlation between PHG and LHG was found in the co-twin and control groups could be explained by their high level of performance. The LHG is a limiting factor for the PHG; without inhibitory control it is much harder to master the PHG. Therefore, as the co-twin and control groups found the LHG rather easy, with over 80% obtaining a full score, the lack of correlation may not be surprising. A more challenging test of inhibitory control such as the Go/NoGo paradigm (Logan et al, 1984) may have yielded a wider range of results in the TD groups and perhaps a significant correlation with the PHG.

Limitations

Although this study was able to detect significant group differences and correlations, it did have some limitations. First and foremost, it must be remembered that all the observed patterns only represent associations and should not be used to imply causation. Longitudinal or intervention studies, as previously suggested, would be better equipped to reach such conclusions. Secondly, during the data analysis it became clear that some of the tests were too easy, especially for the co-twin and control groups. The fact that significant group differences were still found, including some between the co-twin and control group, demonstrates just how severe the deficits in ToM and EF can be, even in the broader autism phenotype. However, future studies ought to consider using more age-appropriate tests or having younger TD groups. Thirdly, the fact that the data were extremely skewed led to the use of nonparametric tests. The ranking of the data overcomes the distribution problems but may reduce the tests' power to find significant results (Field, 2009). Thus, using more sophisticated tests will not only produce a better reflection of abilities in TD participants but also enable the use of more powerful parametric tests. Finally, although the analysis taking only one twin per family per group revealed mostly similar as the original analysis, some were no longer significant. This may be due to the smaller sample sizes which limit the power to detect significant results. Using the cluster command on Stata to account for the family factors may be better for twin data analysis.

Further Research

An interesting avenue for future research would be to investigate the effects of the correlations demonstrated here on real-life behaviour. Some associations between ToM and real-life behaviour have already been established, both for TD children (Peskin and Ardino, 2003) and in autism. Frith et al (1994) used the Vineland Adaptive Behaviour Scales to demonstrate that those who passed the false-belief ToM task showed advantages in social behaviours that required mental states consideration but not on more generally social or advanced daily living skills. A similar design combining PHG, LHG and ToM in comparison with assessment of real-life behaviour would be extremely interesting. Assessment of behaviour can be done using the social behaviour aspects of the ADOS and parent or self report measures. It can also use a similar design to Peskin and Ardino (2003), wherein the experimenter engages in a game with the participant or observes several participants interacting with each other. Aspects of ToM and EF would then be teased out of the behavioural assessment and compared to performance on the experimental tasks. Combining this design with a training intervention may also yield important results for both theory and future interventions.

Conclusion

In conclusion, this study demonstrated significant deficits in theory of mind in individuals aged 12-16 years with ASD. These deficits were evident in a naturalistic test of ToM as well as on social cognition experimental paradigms of ToM. Significant impairments in executive function, particularly in inhibitory control, were also clearly seen. These significantly correlated with performance on the naturalistic ToM task, suggesting that mastering inhibitory control is associated with better ToM performance. Causation could not be established but implications were considered with regard to previous studies and theories. Deficits in ToM were also demonstrated in typically developing co-twins, which adds empirical support to the notion of ToM deficits as part of the broader autism phenotype.

References

Baron-Cohen S (1992) Out of sight or out of mind? Another look at deception in autism. *Journal of Child Psychology and Psychiatry* **33** 1141–1155

Baron-Cohen S, Leslie AM and Frith U (1985) Does the autistic child have a "theory of mind"? *Cognition* **21** 37-46

Baron-Cohen S (2008) *Autism and Asperger Syndrome*. Oxford: Oxford University Press.

Baron-Cohen S. (2009) The Empathizing-Systemizing (E-S) Theory. Annals of the New York Academy of Science **1156** 68-80

Bowler DM (1992) 'Theory of Mind" in Asperger's Syndrome *Journal of Child Psychology and Psychiatry* **33**(5) 877-893

Cohen J, Cohen P, West SG and Aiken LS (2003) *Applied multiple regression/correlation analysis for the behavioural sciences*. 3rd ed. New Jersey: Lawrence Erlbaum Associates

Field A (2009) *Discovering statistics using SPSS*. 3rd Ed. London : SAGE Publications

Fisher N and Happé F (2006) A Training Study of Theory of Mind and Executive Function in Children with Autistic Spectrum Disorders *Journal of Autism and Developmental Disorders* **35**(6) 974-990

Frith U, Happé FGE and Siddons F (1994) Autism and theory of mind in everyday life. *Social developm*ent **3** 108-124

Geschwind DH (2009) Advances in Autism *Annual Review of Medicine* **60** 367–80

Goodman R, Ford T, Richards H, Gatward R, and Meltzer H (2000) The Development and Well-Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. J *Child Psychol Psychiatry* ***41(***5) 645-655.

Gratch G (1964) Response alternation in children: A developmental study of orientations to uncertainty *Vita Humana* **7** 49-60

Happé FGE (1994) An Advanced Test of Theory of Mind: Understanding of Story Characters' Thoughts and Feelings by Able Autistic, Mentally Handicapped, and Normal Children and Adults *Journal of Autism and Developmental Disorders* **24** 129–54.

Happé F, Ronald A and Plomin R (2006a) Time to give up a single explanation for autism. *Nature Neuroscience* **9**(10) 1218-1220

Happé F, Booth R, Charlton R and Hughes C (2006b) Executive function deficits in autism spectrum disorders and attention-deficit/hyperactivity disorder: examining profiles across domains and ages. *Brain and cognition* **61** 25-39

Hogrefe GJ, Wimmer H and Perner J (1986) Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development* **57**(3) 567-582

Howell D (2002) *Statistical methods for psychology* 5th ed. Belmont,CA: Duxbury Press

Hughes C (1996) Control of Action and Thought: Normal Development and Dysfunction in Autism: A Research Note *J. Child PsychoL Psychiat.* **37**(2) 229-236

Hughes C (1998) Finding your marbles: Does preschoolers' strategic behavior predict later understanding of mind? *Developmental Psychology* **34** 1326–1339

Hughes C (1999) Towards a cognitive phenotype for autism: increased prevalence of executive dysfunction and superior spatial span amongst siblings of children with autism. *Journal of child psychology and psychiatry* **40**(5) 705-718

Joseph RM and Tager-Flusberg H (2004) The relationship of theory of mind and executive functions to symptom type and severity in children with autism. *Developmental Psychopathology* **16**(1) 137-155

Le Couteur A, Lord C, and Rutter M (2003) *The Autism Diagnostic Interview-Revised (ADI-R).* Los Angeles, CA: Western Psychological Services.

Logan GD, Cowan WB and Davis KA (1984) On the ability to inhibit simple and choice reaction time responses: a model and a method *Journal of Experimental Psychology: Human Perception and Performance* **10** 276-291

Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, and Rutter M (2000) The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord* **30**(3) 205-223

Luria AR, Pribram KH and Homskaya ED (1964) An experimental analysis of the behavioral disturbance produced by a left frontal arachnoidal endothelioma (meningioma) *Neuropsychologia* **2**(4) 257-280

Oliver BR and Plomin R (2007) Twins' Early Development Study (TEDS): a multivariate, longitudinal genetic investigation of language, cognition and behavior problems from childhood through adolescence. T*win Res Hum Genet* **10**(1) 96-105.

Oswald DP and Ollendick TH (1989) Role taking and social competence in autism and mental retardation *Journal of autism and developmental disorders* **19**(1) 119-127

Ozonoff S, Pennington BF and Rogers SJ (1991a) Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry* **32** 1081–1105

Ozonoff S, Rogers SJ and Pennington BF (1991b) Asperger's syndrome: evidence of an empirical distinction from high-functioning autism. *Journal of Child Psychology and Psychiatry* **32** 1107-1122

Pellicano E (2007) Links between theory of mind and executive function in young children with autism: clues to developmental primacy. *Developmental psychology* **43**(4) 974-990

Pellicano E (2010) Individual Differences in Executive Function and Central Coherence Predict Developmental Changes in Theory of Mind in Autism. *Developmental Psychology* **46**(2) 530-544

Perner J and Lang B (2000) 'Theory of mind and executive function: Is there a developmental relationship' in Baron-Cohen S, Tager-Flusberg H, and Cohen D (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed., pp. 150–181) Oxford, England: Oxford University Press.

Peskin J and Ardino V (2003) Representing the Mental World in Children's Social Behavior: Playing Hide-and-Seek and Keeping a Secret. *Social development* **12**(4) 498-512

Pilowsky T, Yirmiya N, Arbelle S and Mozes T (2000) Theory of mind abilities of children with schizophrenia, children with autism, and normally developing children *Schizophrenia research* **42** 145-155

Russell J. (1996). *Agency: Its role in mental development.* Hove, England: Erlbaum.

Russell J. (1997) 'How executive disorders can bring about an adequate theory of mind' in Russell J (Ed.), *Autism as an executive disorder*. Oxford, England: Oxford University Press pp. 256–304

Wimmer H and Perner J (1983) Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* **13** 103-128.

Wing L, and Gould J (1979) *Severe Impairments of Social Interaction and Associated Abnormalities in Children: Epidemiology and Classification*. Journal of Autism and Developmental Disorders **9** 11-30

Appendix

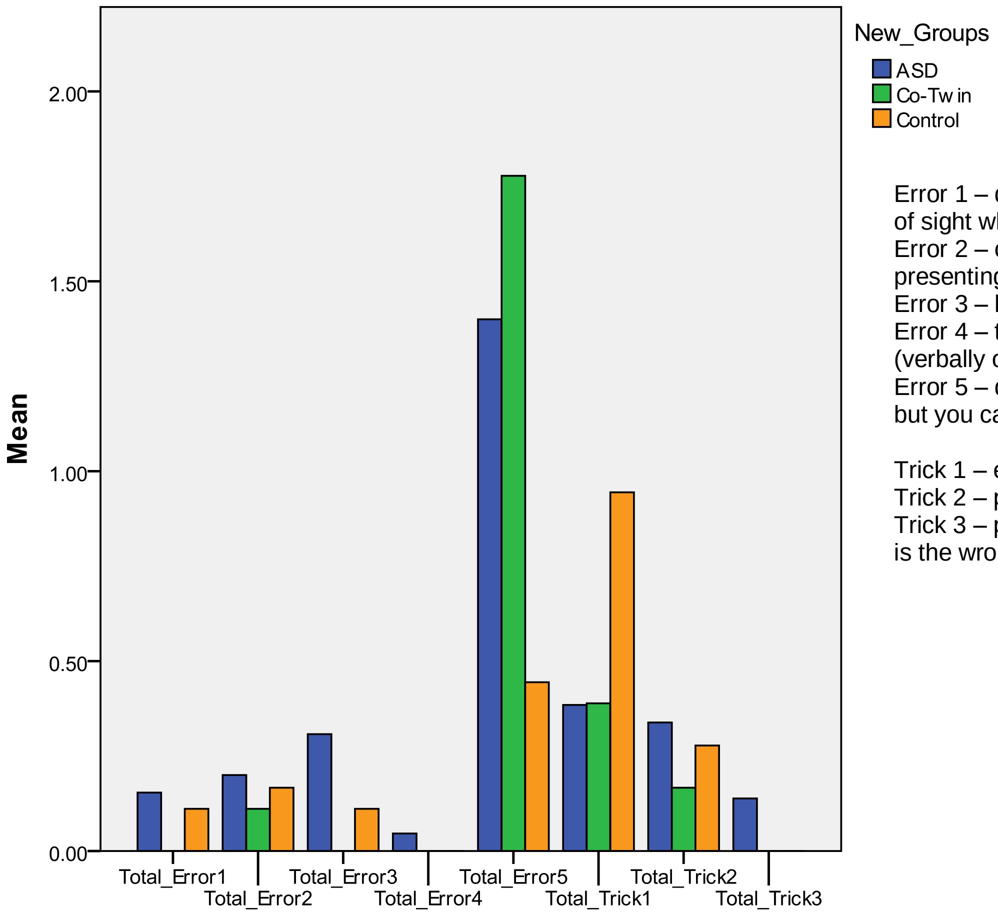
Table 1 Number and percentage of the total errors made by participants in each group in the first four trails of the Penny Hiding Game

Total Errors (PHG)	ASD (N=143)		Co-Twin (N=67)		Control (N=153)	
	Count	%	Count	%	Count	%
.00	103	69%	51	75%	146	94%
1.00	20	13%	8	12%	7	5%
2.00	12	8%	6	9%	2	1%
3.00	8	5%	2	3%		
4.00	4	3%	1	2%	1	1%
5.00	1	1%				
6.00	1	1%				
7.00	1	1%				

Table 2 Number and percentage of the total tricks made by participants in each group in the first four trails of the Penny Hiding Game

Total Tricks (PHG)	ASD (N=143)		Co-Twin (N=67)		Control (N=153)	
	Count	%	Count	%	Count	%
.00	116	77%	62	91%	143	92%
1.00	21	14%	5	7%	8	5%
2.00	11	7%			4	3%
3.00	1	1%	1	2%		
4.00	1	1%			1	1%

Figure 1 Bar chart showing mean errors and tricks in the Penny Hiding Game according to groups by error/trick types.



Error 1 – do not keep both hands out of sight when hiding the penny
Error 2 – one hand used (hiding or presenting)
Error 3 – hand(s) are open
Error 4 – tells you where the penny is (verbally or by opening the hand)
Error 5 – display error (penny is hidden but you can tell where it is eg. by grip

Trick 1 – emphasised the wrong hand
Trick 2 – presents two empty hands
Trick 3 – post-decision trick (saying it is the wrong hand, transferring penny).

Figure 2 Line graph showing the mean number of errors made in each trail from the first to the fourth trail in the Penny Hiding Game, according to groups.

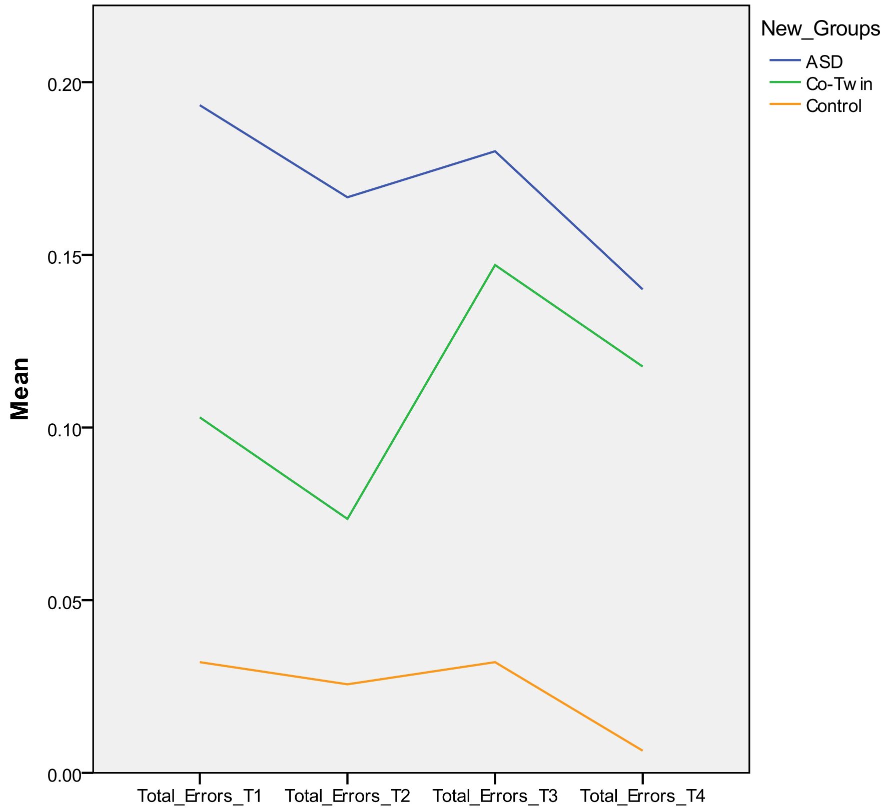


Figure 3 Line graph showing the mean total errors made in each of the first four trials in the Penny Hiding Game in the ASD and Broad Spectrum groups.

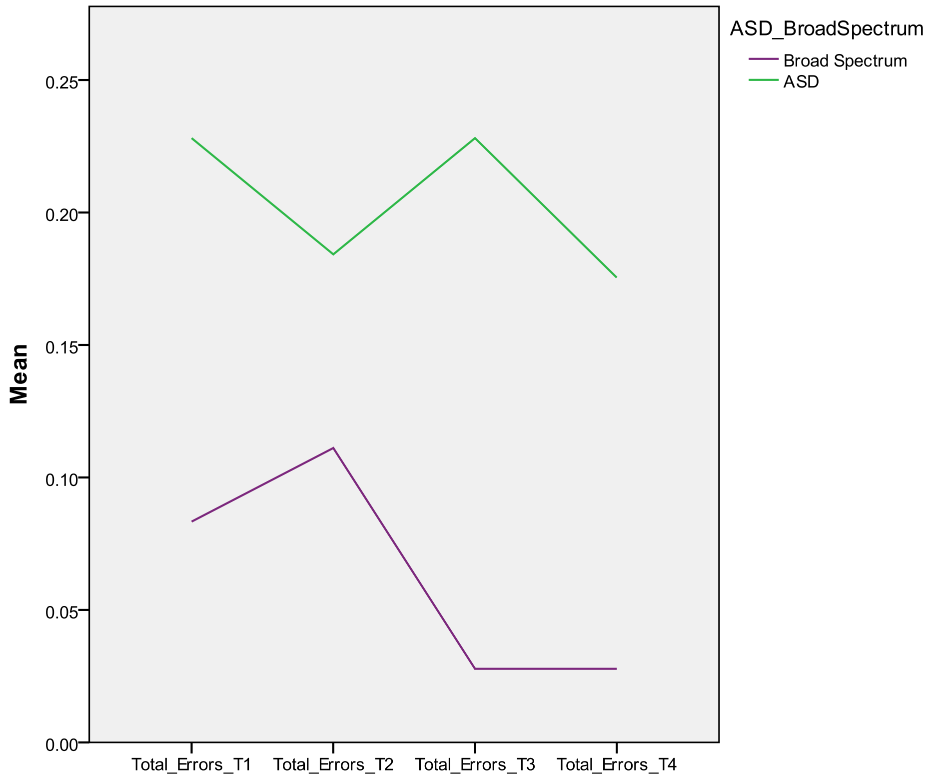


Table 3 Number and percentage (rounded up to the nearest percentage) of correct initial responses in the first five Luria Hand Game trails, according to the three groups

Correct Initial Response	ASD (N=146)		Co-Twin (N=69)		Control (N=142)	
	Count	%	Count	%	Count	%
.00	2	1%				
1.00	4	3%	1	1%		
2.00	5	3%	2	3%		
3.00	17	12%	3	4%	2	1%
4.00	35	24%	5	7%	14	10%
5.00	83	57%	58	84%	126	89%

Table 4 Number and percentage in each group who failed both first order theory of mind tasks, passed only one task or passed both first order theory of mind.

1st Order ToM Tasks	ASD (N=146)		Co-Twin (N=71)		Control (N=154)	
	Count	%	Count	%	Count	%
Fail both	3	2%	0	0	0	0
Pass one	23	16%	5	7%	9	6%
Pass both	120	82%	66	93%	145	94%

Table 5 Number and percentage that failed both second order theory of mind tasks, passed only one task or passed both second order theory of mind within each group.

2nd Order ToM Tasks	ASD (N=144)		Co-Twin (N=69)		Control (N=159)	
	Count	%	Count	%	Count	%
Fail both	2	1%				
Pass one	21	15%	1	1%	8	5%
Pass both	121	84%	68	99%	151	95%

Table 6 Spearman's nonparametric correlations between PHG and the other tasks (LHG and ToM). R_s-squared values for the proportion of shared variance between the ranks of significantly correlated variables.

		LHG score	1st Order ToM	2nd Order ToM
Spearman's rho	**PHG (Total Errors)**	-.15**	-.13**	-.04
R_s^2	**PHG (Total Errors)**	.02	.02	

** Correlation is significant at the 0.01 level (1-tailed).

Table 7 Spearman's nonparametric correlations between PHG and age (months), Full-Scale IQ, verbal IQ and Performance IQ. R_s-squared values for the proportion of shared variance between the ranks of significantly correlated variables.

		Age (months)	Full-scale (WASI) IQ	Verbal IQ	Performance IQ
Spearman's rho	**PHG (Total Errors)**	.18**	-.13**	-.10*	-.16**
R_s^2	**PHG (Total Errors)**	.03	.02	.01	.03

*Correlation is significant at the 0.05 level (1-tailed).
**Correlation is significant at the 0.01 level (1-tailed).

Table 8 Spearman's nonparametric correlations between PHG and LHG and ToM according to groups

			LHG score	1st Order ToM	2nd Order ToM
ASD			-.17*	-.22**	.05
Co-Twin	Spearman's rho	PHG (total errors)	.16	.14	-.18
Control			.01	.07	-.05

*Correlation is significant at the 0.05 level (1-tailed)
**correlation is significant at the 0.01 level (1-tailed)

Table 9 Spearman’s nonparametric correlations of the PHG with age (months), Full-scale (WASI) IQ, verbal and performance IQ according to groups

			Age (months)	**Full-scale (WASI) IQ**	**Verbal IQ**	**Performance IQ**
ASD			.12	-.24(**)	-.18(*)	-.24(**)
Co-Twin	Spearman’s rho	PHG (total errors)	-.02	.12	.20	-.09
Control			.06	-.04	-.05	-.05

*correlation is significant at the 0.05 lever (1-tailed)
**correlation is significant at the 0.001 lever (1-tailed)