



HHS Public Access

Author manuscript

J Autism Dev Disord. Author manuscript; available in PMC 2016 September 01.

Published in final edited form as:

J Autism Dev Disord. 2015 September ; 45(9): 2721–2733. doi:10.1007/s10803-015-2436-3.

Making inferences: Comprehension of physical causality, intentionality, and emotions in discourse by high-functioning older children, adolescents, and adults with autism

Kimberly E. Bodner,

Department of Psychological Sciences, University of Missouri, 3 McAlester Hall, Columbia, Missouri, 65211, Phone: (573) 884-8109, Fax: (573) 882-7710

Christopher R. Engelhardt,

Department of Health Psychology, Thompson Center for Autism & Neurodevelopmental Disorders, University of Missouri, 205 Portland Street, Columbia, MO, 65211, Phone: 573-882-1923, Fax: 573-884-1151

Nancy J. Minshew, and

University of Pittsburgh School of Medicine, Department of Psychiatry, 3811 O'Hara Street, Suite 300 Webster Hall, Pittsburgh, PA 15213-2593, Phone: (412) 246-5460, Fax: (412) 246-5470

Diane L. Williams

Speech-Language Pathology, Rangos School of Health Sciences, Fisher Hall 409, Duquesne University, 600 Forbes Avenue, Pittsburgh, PA 15282, Phone: (412) 396-4217, Fax: (412) 396-4196

Kimberly E. Bodner: kimberly.bodner@mail.mizzou.edu; Christopher R. Engelhardt: engelhardtc@health.missouri.edu; Nancy J. Minshew: minshewnj@upmc.edu; Diane L. Williams: williamsd2139@duq.edu

Abstract

Studies investigating inferential reasoning in autism spectrum disorder (ASD) have focused on the ability to make socially-related inferences or inferences more generally. Important variables for intervention planning such as whether inferences depend on physical experience or the nature of social information have received less consideration. A measure of bridging inferences of physical causation, mental states, and emotional states was administered to older children, adolescents, and adults with and without ASD. The ASD group had more difficulty making inferences, particularly related to emotional understanding. Results suggest that individuals with ASD may not have the stored experiential knowledge that specific inferences depend upon or have difficulties accessing

Correspondence to: Diane L. Williams, williamsd2139@duq.edu.

Affiliation at time of study: Kimberly E. Bodner, NIH Autism Center of Excellence, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; Christopher R. Engelhardt, Thompson Center for Autism & Neurodevelopmental Disorders and Department of Health Psychology, University of Missouri, Columbia, Missouri, USA; Nancy J. Minshew - NIH Autism Center of Excellence, University of Pittsburgh School of Medicine and Departments of Psychiatry and Neurology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; and Diane L. Williams - NIH Autism Center of Excellence, University of Pittsburgh School of Medicine and Department of Speech-Language Pathology, Duquesne University, Pittsburgh, PA, USA.

Change in author's affiliation: Kimberly Bodner is now at affiliation: Department of Psychological Sciences, University of Missouri, Columbia, Missouri, USA.

Conflict of Interest

The authors declare that they have no conflict of interest.

relevant experiences due to linguistic limitations. Further research is needed to tease these elements apart.

Keywords

autism spectrum disorder; inference; theory of mind; emotion; language

Introduction

Individuals with autism spectrum disorder (ASD) have difficulty with comprehension of the meaning of both spoken and written discourse that affects their ability to function socially and academically (Loukusa et al. 2007; Arciuli et al. 2013; Huemer & Mann 2010; Ricketts 2011). Successful comprehension of discourse depends not only on interpretation of the linguistic forms but also on the integration of that interpretation within the communicative context (Brown et al. 2013; Leinonen & Kerbel 1999; Sperber & Wilson 2002). Furthermore, the temporal nature of spoken language and the physical limits of written language make it impractical for the speaker or writer to explicitly state every fact or idea needed to comprehend the intended message. Therefore, the ability to *make inferences*, or to fill in gaps using your own world knowledge, is an essential skill for comprehension of discourse (Snyder & Caccamise 2010).

Individuals with ASD who have acquired a high level of spoken and written language skills are thought to have persistent difficulty with the cognitive process of inferencing, resulting in a tendency to interpret utterances literally and to make other types of pragmatic errors during social interactions (Loukusa et al. 2007). This assumption has been supported by the performance of verbal, relatively high-functioning individuals with ASD on standardized behavioral measures, such as the items from the Test of Language Competence – Expanded Edition (TLC-E; Wiig & Secord 1989). Items from the Making Inferences subtest of the TLC-E assess bridging inferences which represent four different script types: situational, personal, instrumental, and combined (Wiig & Secord 1989). Each item is first read aloud by the examiner and then presented in text to the participant, with the participant selecting an answer from four written choices, suggesting that it is primarily an assessment of making inferences from written text. Studies which have used this instrument to assess inferencing skills have generally reported a relative deficit in inferencing for individuals with ASD (Dennis et al. 2001; Lewis et al. 2007; Minshew et al. 1995).

However, the results from other studies suggest that individuals with ASD with well-developed verbal skills do not have an overall problem with making inferences but have particular difficulty with inferences about *social* information. In several studies, stories describing physical events have been used as a control task because the individuals with ASD were described as having no difficulty in making inferences about this type of information (Happé 1994; Kaland et al. 2005). Similarly, in another study that used a textual, two-sentence vignette paradigm, adolescents with Asperger syndrome were reported as having no difficulty making causal and predictive inferences, even though they had difficulty making inferences about intentionality (Le Sourn-Bissou et al. 2009). The results

of these studies suggest that verbal, relatively high-functioning individuals with ASD may not have difficulty making inferences per se but may have difficulty making inferences about more abstract information, particularly social information such as intentionality or mental states.

A functional magnetic resonance imaging (fMRI) study of verbal, adults with autism, with cognitive ability in the average range, has provided some evidence of a neurological basis for the difficulty with inferencing in ASD (Mason et al. 2008). The results of this study indicated that there is an inefficiency of processing in the neural network related to making bridging inferences during a comprehension task from written text about physical, mental, and emotional states (Mason et al.). Given the inefficiencies in neural processing, making inferences may be a particularly challenging task even for verbal, cognitively able individuals with ASD, especially as the processing demands increase either because of the type of information being processed or because of the conditions under which the processing occurs.

The Role of Context

Consistent with the assumption that inferencing about social information is what is affected in ASD, the cognitive process of inferencing has primarily been studied in relation to theory of mind (ToM), a specific form of inferencing about the intentions or mental states of others. One of the primary tools that has been used in these investigations is the Strange Stories task (Happé 1994) which was developed as a more challenging test of ToM for older, verbally-able individuals with ASD. The Happé task presents subjects with linguistically and socially complex stories about everyday experiences that represent a wide array of mental states (i.e., sarcasm, pretending, lies, bluffing, irony, etc.) with control scenarios that evaluate physical causation. The stories are simultaneously read aloud to the subjects and presented in text. A number of studies using the Strange Stories task have demonstrated deficits in the ability of verbal children, adolescents, and adults with ASD with cognitive abilities in the average range to make inferences about mental states (e.g. Brent et al. 2004; Happé 1994; Jolliffe & Baron-Cohen 1999; Kaland et al. 2005). However, the results of these studies were not clear, indicating that verbally-able individuals with ASD could make some mental state inferences but these inferences were contextually inappropriate, suggesting that the problem is not necessarily one of inferencing about mental states but of coherence or the integration of contextual information (Happé 1994; Jolliffe & Baron-Cohen 1999; Kaland et al. 2005). The individuals with ASD in these studies were described as giving correct physical state answers for some of the stories in which a mental state (ToM) response was expected, suggesting that the participants with ASD may have been able to draw an inference but failed to focus on the expected elements of the story (Happé 1994; Kaland et al. 2005).

The suggestion that contextual integration is what is challenging for individuals with ASD rather than inferencing per se was supported by the results of a study with children from three different clinical groups, one of which was a small number (10) of children, ages 6 to 10 years, with high-functioning autism (Norbury & Bishop 2002). In that study, stories were read aloud to the participants and questions relating to three types of inferences (literal, text-connecting, and gap-filling) were asked. The results indicated that participants with

pragmatic language impairment, specific language impairment, and autism all had more difficulty answering literal and inferential questions than the age-matched peers with typical development. The children with autism had relatively more difficulty making inferences than other children with linguistic impairments and the children with more behavioral symptoms of autism had poorer inferencing. An error analysis indicated that the problem for the children from all three clinical groups, including the children with autism, was not in making inferences per se but in making inferences that related to the context of the story (Norbury & Bishop, 2002).

Differential responsiveness to contextual demands by verbal adults with ASD with cognitive abilities in the average range as compared to age- and IQ-matched adults with typical development was also evident in the Mason et al. (2008) functional imaging study mentioned above. In that study, three different types of information were included in the bridging inferences: a) physical causation, b) mental states, and c) emotional states. Given that the task was designed to be successfully performed by the participants, the behavioral performance on the three conditions did not differ for either the ASD or the control group of individuals with typical development. Despite a lack of difference in the behavioral performance, the brain activation data for the ASD group differed from that of the control group particularly in one very interesting way. The activation pattern for the ASD group was highly similar for all three types of inferences, whereas the pattern differed for the controls with typical development by condition. That is, the data of the control group indicated a sensitivity to the differing demands of the three text conditions that was not evident in the data for the ASD group (Mason et al. 2008).

Rationale for the Current Study

The nature of inference making in ASD also lacks clarity because difficulty with making inferences, even for social information, has not been a universal finding, particularly when more indirect behavioral measures have been used. For example, a study that used priming and reading times as a measure of making implicit bridging inferences based on textual, two-sentence vignettes reported no difference in adolescents with ASD (who had good word reading accuracy but relatively poorer text comprehension) for either physical or social information at an automatic level of inference (Saldaña & Frith 2007). It should be noted that the items from that study only required a yes/no response and were, by design, relatively easy with high rates of accuracy for both the ASD and control groups.

In summary, it is not clear if individuals with ASD have difficulty with comprehension of spoken and written discourse related to a more general problem with making inferences about information that is implicit in the situation, because of a specific problem with making inferences about the thoughts of others (ToM), or because of a problem with integration of context. Further understanding of the source of these comprehension difficulties is important so that it is clearer what underlying cognitive skills should be targeted when working clinically with individuals with ASD.

Need for an Appropriate Measure

The studies that have suggested that the problem may be one of contextual integration have primarily used the Happé Strange Stories task (e.g., Happé 1994; Jolliffe & Baron-Cohen 1999; Kaland et al. 2005). The Happé stories were designed to interrogate comprehension of different types of social language (e.g., sarcasm, pretending, lies, bluffing, etc.) and were not specifically designed to study the cognitive process of making inferences. As such, they tend to be quite lengthy, requiring the listener to maintain in working memory large amounts of detailed information and to relate this information to previously obtained world knowledge, particularly understanding of social situations. Given the substantive demands for contextual integration, it is not surprising that the poor performance of individuals with ASD on this measure has been interpreted as indicating a difficulty in this area rather than a clear indication of difficulty with making inferences.

Therefore, as a first step toward bringing some clarity to understanding the source of difficulty with comprehension of discourse in ASD, we wanted to more clearly assess the process of inference. A novel measure, the Pittsburgh Inference Test (PIT), was developed for use with verbal, older children, adolescents and adults with ASD. For this measure, we kept the type of inference (bridging) that was required the same across all the test items which allowed us to vary the type of information (physical causation, mental states, and emotional states) as the salient factor. Bridging inferences were chosen because they are considered valid measures of this cognitive process and have been frequently used in investigations in both typical and atypical populations (Graesser et al. 1994; Singer 2013). The short, two to four-sentence format of the bridging inference also allows investigation of the process of inferencing in textual discourse by presenting (in written form while being read aloud to the participant) a limited amount of information, avoiding the problem of other possible interfering factors when the participants must attend to and process large amounts of orally presented information. The three different types of information (physical, mental, emotional) allowed us to examine if it was the cognitive skill of drawing an inference that was difficult or if it was the type of information (visible/experiential) vs. internal states that was challenging for individuals with ASD. Comparing two different types of internal states would provide information on whether it was abstract information in general or more specific types of abstract information (mental thinking vs. emotional reaction) that was potentially challenging and was consistent with recent work suggesting that these two types of theory-of-mind (cognitive vs. affective) are dissociable (Shamay-Tsoory, 2011).

Predicted Results

The predictions were as follows: a) If drawing an inference in general is the problem, then the individuals with ASD would have poor performance across all of the items despite information content; however, b) if making inferences about abstract information is what is difficult, physical causation would be less challenging than mental and emotional states for individuals with ASD based on the assumption that these individuals have experiential knowledge about physical situations and less understanding of ToM; finally, c) if a specific type of deficit in ToM, or making inferences about the thoughts of others in general or emotion related content, is the problem, then the individuals with ASD would give fewer appropriate responses to items that incorporated an interpretation of the type of thoughts the

characters were thinking. Therefore, examination of the performance of individuals with ASD on the PIT will provide information as to whether the cognitive skill of drawing an inference is difficult overall or whether the *type* of information about which the inference is being made is an important factor to impaired performance.

Methods

Participants

The participants were 86 older children and adolescents and adults with ASD and 65 age- and ability-matched typically developing controls (TD) who were all between the ages of 10 and 45 years. The group with ASD consisted of 37 older children and adolescents (between 10 – 16 years) and 49 adults (between 17 – 45 years), and the TD group consisted of 16 older children and adolescents and 49 adults. The two groups (ASD and TD) were group matched for age, gender, socioeconomic status (SES: Hollingshead 1975), and Full Scale IQ, Verbal IQ, and Performance IQ as assessed by the Wechsler Abbreviated Scale of Intelligence (WASI: Wechsler 1999). One participant received the Wechsler Adult Intelligence Scale. Four adult participants with ASD did not report SES. All participants had full scale IQs greater than 85, were able to communicate in complete spoken sentences, did not have attention or behavioral problems that prevented them from completing testing, did not have any associated or causative genetic, metabolic, or infectious conditions, were in good medical health, and had no history of seizures, birth injury, or head trauma. See Table 1 for participant information by diagnostic group.

The diagnosis of autism for participants with ASD was established using two structured research diagnostic instruments, the Autism Diagnostic Observation Schedule-Generic (ADOS-G: Lord et al. 2000) and the Autism Diagnostic Interview-Revised (Lord et al. 1994), and confirmed by expert clinical opinion. All ASD participants met criteria for autism on the ADI and for autism or spectrum disorder on the ADOS (25 met ASD cut-offs and 61 met autism cutoffs on the ADOS). No ADI scores were available for four adult participants due to lack of suitable informants, but all four had life long histories and current manifestations that were consistent with an ASD diagnosis.

The control participants were recruited from the community in response to advertisements. TD participants were screened by telephone questionnaires, interviews, and psychometric evaluations. Participants with TD were excluded if found to have a family history (in parents, siblings, and offspring) of autism, developmental cognitive disorders, learning disabilities, affective disorders, anxiety disorders, schizophrenia, obsessive-compulsive disorder, or other neurologic or psychiatric disorders thought to have a genetic component.

All participants were recruited and assessed by an autism research center at a major university. The data for this study were collected as part of a larger subject characterization battery. Recruitment and data collection procedures were approved by the Institutional Review Boards at two major universities. Written informed consent was obtained from participants and/or guardians prior to testing.

Assessment Instrument

To create the items for the PIT, the stimulus items from the Mason et al. (2008) functional imaging study of ToM processing were used as initial models. Thirty 2- to 4-sentence stories (28 for testing with two for practice) that presented typical life situations followed by a verbal question that implicitly invited the participant to make an inference were created. The test consisted of two types of items. The first type was designed to elicit responses that described physical relationships (7 questions). The second type (internal) was designed to elicit items that required inferences about mental or emotional states (ToM) (21 questions); however, it was possible that the respondents could provide an answer that described a physical relationship instead. For example, one internal story states, “Andy was only 2 years old. He was sitting in his mother’s lap when a big dog ran up and licked him on the cheek. Andy’s eyes got really big, and he started to cry.” The examiner then asks the participant, “Why did Andy do that?” using an open-ended questioning format. This allows the participant to produce a range of response types. For example, the participant may provide responses that incorporate an understanding of internal states, such as: “Andy was scared of the dog” or “Andy was surprised/startled by the dog” (both correct emotional ToM responses). Alternatively, the participant may provide responses that are technically correct but do not provide the expected ToM aspect because they refer to physical rather than mental or emotional states. For example, responses such as “Because the dog licked him” (correct physical response). Even when the participant responds incorrectly, information may be gathered as to their inferential abilities. For example, a response such as “Andy is allergic to the dog” is incorrect and also indicates that the respondent made an inference about a physical state. This latter feature was decided as important to include based on previous results reported by Norbury and Bishop (2002), Happé (1994), and Joliffe and Baron-Cohen (1999) in which the participants in those studies were described as providing responses that indicated that an inference had been made but that these inferences were inappropriate to the story context.

The stories were written so that they could be easily understood by children and adults with at least a fourth grade reading level (assessed through the Flesch-Kincaid Grade Level). The number of words in each story ranged from 22 to 38 words ($M = 31.8$). The number of sentences in each story ranged from 2 to 4 sentences ($M = 3.03$). The grade equivalent of each story ranged from 2.3 to 4.9 grade ($M = 3.7$), and reading ease ranged from 76.4 to 94.3 ($M = 86.7$). [However, it should be noted that during administration the stories are read out loud to the participants to be consistent with previous work in this area (e.g., Brent et al. 2004; Happé 1994; Kaland et al. 2005) and to limit the effect of reading ability on the measure.] All of the stories were narrative in form with named individuals engaged in the described events. The names of the characters in the story were taken from the Social Security online database of popular baby names to ensure the names would be familiar to participants who were United States residents (Social Security Online 2005).

Procedures

Test Administration and Scoring—The PIT was administered as part of a battery of neuropsychological tests by trained research assistants as follows. Each participant was presented with a stimulus book that contained one story printed on each page. The examiner

read each story aloud to the participant and then asked the corresponding question. The examiner recorded the participant's response verbatim or circled one of the sample answers if the participant provided a common response. The examiner began with two practice stories and provided feedback and additional opportunities to respond if needed until the participant demonstrated understanding of the testing process. The examiner did not tell the participant how to answer the questions or give examples of correct answers. It was only required that the participant be able to provide relevant responses to the questions that followed the stories. Then the examiner administered test questions 1 – 28 and recorded each answer verbatim. The examiner queried a response if it was unclear, if the response only repeated elements of the story, or if the participant initially answered "I don't know." Only one query of "Tell me more." or "What do you mean?" was given per question if needed to clarify an ambiguous response.

The responses for each story were scored as correct or incorrect and then categorized as a physical or ToM response. For the 21 internal stories, ToM responses were further categorized by type: emotion-ToM response or other-ToM response. In addition to physical and ToM responses, participants could simply repeat the story, have a nonsensical/other response, or choose not to respond at all. These latter types of responses were always queried once, and if repeated, they were scored as incorrect. To minimize systematic error due to rater biases, steps were taken to make the scoring of verbal responses as objective as possible by providing clear and detailed descriptions of potential responses. In addition, a scoring guide was developed to provide common responses and their corresponding appropriate scores for each story on the PIT.

The total number of correct responses and incorrect responses were tallied. Correct responses were weighted (as described below) to indicate the assumed difficulty level of responses: a correct physical response received 1 point (sum = weighted physical total) and a correct ToM response received 2 points (sum = weighted ToM total). Incorrect responses of any type were given 0 points. The weighted physical total and weighted ToM total were added together for an overall total weighted score. Next, for ToM responses given by the respondent, the number of correct and incorrect emotion-ToM and correct and incorrect other-ToM responses were tallied to obtain raw scores in each sub-category. These were not included in the total weighted score, but were important to more specifically characterize ToM inference making abilities.

Correlational Analyses—We also examined the performance of individuals with ASD in relation to commonly used measures of ToM. If the PIT was evaluating similar underlying cognitive and linguistic constructs, the performance of the individuals with ASD on the PIT should correlate with their performance on other measures that require making an inference about mental states. Participants were administered three well known measures of first and second order ToM: *Sally and Anne* (Baron-Cohen et al. 1985); *John and Mary* (Perner & Wimmer 1985); and, *Peter and Jane* (Bowler 1992). Similar to Bowler (1997), an aggregate ToM score was tabulated by summing the number of correct belief, reality, and memory questions from each of the three ToM tasks (potential maximum score of 9 total points). Seven individuals did not have these three ToM measures that were administered within the PIT testing session and, therefore, were excluded from this specific analysis.

Participants also completed the *Reading the Mind in the Eyes Test-Revised* (Adult or Child Version; Baron-Cohen et al. 2001). Participants viewed only the eyes of an individual and were asked to determine what the person was thinking or feeling by choosing one of four presented words. Adult participants completed 36 sets of eyes and were provided with a word definition booklet if needed. A child version of the test was administered to participants 15 years of age and under. Child participants completed a set of 28 sets of eyes, and the examiner read each word aloud.

Performance on the PIT was also examined in relation to performance on the *Test of Language Competence—Expanded* (TLC-E; Wiig & Secord 1989), a standardized assessment of metalinguistic abilities including making inferences that has been previously used in research in this area (e.g., Dennis et al. 2001; Minshew et al. 1995). The TLC-E consists of four subtests that sample metalinguistic abilities including the understanding of Ambiguous Sentences (participants select two different meanings for an ambiguous sentence from four printed choices), Making Inferences (examiner reads two statements that provide incomplete information about a single event and the participant chooses two of four possible explanations), Recreating Sentences (participants are orally and visually given three single words that were supposedly spoken by people in a scene and asked to use the words to construct a sentence that could have been used in the pictured situation), and Figurative Language (participants are asked to tell in their own words what a person meant when saying an expression in a given situation; participants then choose which of four expressions was closest in meaning to the conversational statement). Raw scores for each subtest were calculated according to the procedures in the test manual and were used for all statistical analyses because the age range for standardized scores for this measure is 18 years and a number of the adult participants had chronological ages above that level. All four subtests were combined into a sum of subtests raw score for each participant. Of note, raw scores from subtests 2 (Making Inferences) and 4 (Figurative Language) are most closely related to the inference making constructs examined by the PIT, and were included in subsequent analyses.

Assessment of Reliability—Inter-rater reliability was calculated for the PIT to ensure that all testers objectively scored test responses in the same way. Approximately 10% of tests ($n=17$) were randomly sampled and scored by two experienced examiners. The observed agreement between the two raters was nearly unanimous (Cohen's kappa = .99). All of the TLC-E and ToM protocols were rescored by a second tester and any scoring or calculation errors were corrected.

Results

Analytical Approach

Because of the increased backlash against using null hypothesis significance tests (NHSTs) as a vehicle for statistical inference (Anderson 1997; Cumming 2012; Cumming 2014; Kirk 2003; Wagenmakers 2007), we do not report these tests or the p values associated with them. Rather, we report Bayes factors (BFs; see Hoijtink et al. 2008; Jeffreys 1961; Kass & Raftery 1995) to state evidence in favor of or against statistical models, an approach that has

been advocated repeatedly (Berger & Berry 1988; Edwards et al. 1963; Gallistel 2009; Kass 1993; Morey et al. 2014; Myung & Pitt 1997; Raftery 1995; Rouder et al. 2009; Wagenmakers 2007). This approach differs from traditional NHSTs because Bayes factors permit a method of model comparison in which models including main effects and interactions are pitted against models that systematically exclude them. Bayesian analysis was therefore chosen because it can simultaneously address our hypotheses and allow evidence to be considered continuously rather than dichotomously.

In the sections that examine PIT outcome measures – overall weighted total scores, physical scores, other-ToM scores, and emotion-ToM scores – we use the general linear model in which main effects and interactions are assessed (Table 2 contains descriptive statistics for all PIT outcome measures). Nineteen models were assessed for each PIT outcome: the null model in which there are no effects; a model including group diagnosis only; a model including Verbal IQ only; a model including age only; three additive models in which two of the three main effects only are included; an additive model in which only the three main effects are included; ten models including all possible combinations of the selective presence or absence of the 2-way interactions (with the constraint that the terms that comprise an interaction term also appear as main effects in the model); and a full model including the three main effects, all 2-way interactions, and the 3-way interaction. Verbal IQ, a general measure of the verbal ability of the participants, was included in these analyses in lieu of Full Scale or Performance IQ given that the responses on the PIT were verbal ones and that previous research has suggested that verbal ability is an important variable to examine when investigating inferential skills in ASD (Norbury & Bishop 2002).

For each analysis, we report the best-fitting model and the model testing for an invariance of a particular PIT outcome despite including group diagnosis in the model. In consideration of space limitations, the results of all model comparisons are not reported here but are available upon request. The results below have the following interpretation: when group diagnosis is included in the best-fitting model, this finding can be interpreted as evidence that group diagnosis is needed to model the data; when group diagnosis is not included in the best-fitting model, this result can be interpreted as evidence that group diagnosis is not needed to model the data.

The computations used to calculate the BFs here can be found in a previous report (Rouder et al. 2012). All BFs were calculated in Morey and Rouder's BayesFactor package for R using the generalTestBF function (Morey & Rouder 2014). BFs are easily interpretable. They are reported in ratios, such as 5-to-1, in favor of a model that includes a parameter (or parameters) relative to a model in which that parameter (or parameters) has been removed. These ratios should be interpreted as the extent to which beliefs about the models should be updated in light of data. Bayesian analysts must also place prior distributions on model parameters. In line with the recommendations by Rouder and colleagues (2012), we adopt a default prior for this purpose, where the effect size under the alternative has a point mass at zero and small effect sizes are more likely to be observed than large effect sizes. We set the scale parameter r of the prior to 0.50 because we expected small-to-medium effects. This scale parameter corresponds to an expected effect size of $\rho = 0.24$. We find this prior to be reasonable.

We also report standardized effect sizes and 95% confidence intervals (CIs), in line with the recommendations from the American Psychological Association (2010), when appropriate. Unbiased Cohen's d (Cumming 2012) for independent t tests was calculated using the pooled within-groups standard deviation as the standardizer; the 95% confidence intervals for Cohen's d were derived from approximations for the noncentral t distribution (Algina & Keselman 2003; Cumming 2012; Cumming & Fidler 2009; Rosnow & Rosenthal 2009). Effect size r and its 95% CI are reported for all bivariate correlations.

Weighted Total Scores

Recall that weighted total scores were calculated as the sum of the physical responses (one point each) and the ToM responses (two points each). The best-fitting model included all three main effects, the diagnosis x age interaction, and the age x VIQ interaction. This model was preferred over the null model by a factor of 7.3×10^6 -to-1. The model with group diagnosis only also was preferred over the null model by a factor of 125-to-1.

Individuals with ASD had lower scores on the PIT compared to TD individuals, $d_{\text{unb}} = 0.63$ [0.29, 0.96] (see Table 2). Additionally, weighted total scores were positively associated with age, $r = .23$ [.07, .38], and VIQ, $r = .37$ [.22, .50]. Visual inspection of the diagnosis x age interaction indicated that TD individuals had higher weighted total scores than individuals with ASD at younger but not older ages (see Figure 1); visual inspection of the age x VIQ interaction suggested that individuals with higher VIQ scores had higher weighted total scores than individuals with lower VIQ scores at younger but not older ages (see Figure 2 for a graphical depiction of this interaction).

Subscale Scores

In accordance with our predictions and with previous research, further examination of the physical and ToM (other vs. emotional) raw scores is warranted to investigate potential contributions of the *type* of inferential response to overall performance on the PIT.

Correct Physical Scores—The model with diagnosis only was the best-fitting model, and was preferred to the null model by a factor of 182-to-1. The model with group diagnosis was preferred to the model with VIQ only, by a factor of 992-to-1, and was preferred to the model with age only, by factor of 1,026-to-1.

Correct Other-ToM Responses—The model with a main effect of age and VIQ, and an age x VIQ interaction was found to be the best-fitting model. Importantly, the null model was preferred to the model with only diagnosis by a factor of 5.6-to-1. This result suggests substantial evidence in favor of the hypothesis that the number of correct other-ToM responses is invariant to a diagnosis of ASD. The number of correct other ToM responses were positively associated with age, $r = .20$ [.04, .35], and VIQ, $r = .26$ [.11, .41]. Visual inspection of the age x VIQ interaction suggested that individuals with higher VIQ scores had higher other-ToM responses than individuals with lower VIQ scores at younger but not older ages (see Figure 3 for a depiction of this interaction).

Correct Emotion-ToM Responses—The best-fitting model for this outcome included a main effect of VIQ, but, as predicted, a main effect of diagnosis. This model was preferred to the null model by a factor of 1,307-to-1. Moreover, this model was preferred to a model with only VIQ (i.e., dropping the diagnosis term) by a factor of 31-to-1. This pattern of results indicates substantial evidence in favor of the hypothesis that individuals with ASD have greater difficulty with emotion ToM responses compared to TD individuals, $d_{\text{unb}} = 0.57 [0.24, 0.90]$ (see Table 2).

Bivariate Analyses

Pearson's correlations and 95% CIs were calculated to investigate the relationship between inference making abilities on the PIT and metalinguistic ability (TLC-E) for participants with ASD (see Table 3). Correlations were calculated using the 3 raw scores on the TLC (sum of subtests and subtests 2 & 4) and PIT responses (see previous model comparisons). TLC-E raw scores were used because there are no norms for individuals over the age of 18 years.

In general, the PIT scores (weighted total, other-ToM, and emotion-ToM) were moderately to highly correlated with the TLC-E scales. However, the correlation between PIT physical responses and TLC-E scores were markedly lower. Results from the TLC-E subtest analyses were similar, with the exception of the relationship between physical responses and the TLC-E subtest 2 scores, which increased slightly in magnitude. Taken together, these results suggest that the higher order language skills of the individuals with ASD were related to their abilities to make ToM inferences on the PIT but were not related to their ability to make inferences about physical events.

The relationships between inference abilities on the PIT (weighted total scores, physical scores, other-ToM scores, and emotion-ToM scores) and well known ToM tasks (ToM Aggregate Score and Reading Mind in the Eyes) were also investigated (see Table 4 for bivariate correlation coefficients and 95% CIs). For participants with ASD, correlations between PIT variables – with the exception of the PIT physical scores – and scores on the Reading the Mind in the Eyes task were moderate to strong. Similar results were found between the PIT scores and ToM aggregate scores.

Correlations between the PIT and Reading the Mind in the Eyes were further investigated by examining Child and Adult groups separately because they were administered different versions of the Mind in the Eyes task. Adults with ASD showed large correlations between two PIT subscales (Weighted total, and Emotion ToM scores) and the Reading Mind in the Eyes task. The correlations between the PIT subscales and the Reading the Mind in the Eyes task were smaller among children with ASD compared to adults with ASD.

Discussion

Model comparison approaches using Bayes factors suggested that overall PIT performance was relatively poorer in individuals with ASD, lending further support to the notion that individuals with ASD have a general problem with drawing inferences (Loukusa et al. 2007; Arciuli et al. 2013; Huemer & Mann 2010; Ricketts 2011). However, overall PIT

performance in individuals with ASD increased as a function of age. The level of language skills also affected PIT performance as a function of age, with more verbally-able individuals scoring higher than less verbally-able individuals, especially when those individuals were younger in age. The improvement in overall performance by the individuals with ASD with age may result from a developmental increase in language skills, but it may also reflect a difference in the experiential level for the adults with ASD. Therefore, both the level of language and experiential knowledge are potentially important factors in whether or not an individual with ASD will be able to draw an inference. Of particular significance, however, is the finding that the ability to make emotion-related inferences did not improve with age in the ASD group, suggesting that this is a continued area of difficulty for individuals with ASD over the course of their lifespan despite improvements in language abilities and more life experience.

Difficulties in ascertaining inferences related to physical causation in individuals with ASD appeared to be due to the participants providing inaccurate physical responses, indicating an understanding, though incorrect, of the physical nature of the scenarios and some attempt to draw conclusions from the provided information. The current findings endorse the assumption that, at times, deficits in discourse processing for individuals with ASD may be related to difficulty with an ability to integrate world knowledge within a specific context or situation. Our findings are in contrast to previous studies reporting no ASD-related difficulties in making physical inferences in participants with ASD ages 8 to 45 years (Happé 1994) and participants with Asperger syndrome ages 10 to 20 years (Kaland et al. 2005). However, in both of those studies, the physical scenarios were used either as a screening tool (Happé 1994) or to check for possible comprehension deficits (Kaland et al., 2005). The physical stories in the Happé study were relatively easy with all the participants (those both with and without ASD) reported as performing at ceiling. Kaland and colleagues used lengthy stories with greater detail, which may have provided the cues necessary to ascertain correct physical inferences. The current findings endorse the importance of using both social and non-social information when assessing the drawing of inferences in individuals with ASD.

The correlation of the performance on the PIT and the TLC-E suggests that the PIT was measuring similar skills as measured by this standardized test of metalinguistic abilities. However, the PIT had the added element of measuring inferencing about physical events that was not provided by the standardized measure. Therefore, the PIT provides a more complete picture of the inferencing abilities of individuals with ASD than provided by previous investigations that have used only this standardized measure (Dennis et al. 2001; Lewis et al. 2007; Minshew et al. 1995).

Adults with ASD displayed a marked relationship between the PIT and Adult version of the Reading the Mind in the Eyes, especially for emotion-related inferences. The relationships between the Adult Version of the Mind in the Eyes and the remaining PIT subscales (other-ToM and physical inference) were dramatically smaller. Therefore, the PIT is a sensitive enough measure to tease apart deficits in emotion vs. other inference-making abilities noted in individuals with ASD. Of note, and somewhat unexpectedly, the correlations between the PIT and Child Version of the Reading the Mind in the Eyes task were observed to be small

in magnitude. However, upon further review, these findings might be explained by the development of more complex inference-making abilities over time in this group.

The results of the current study suggest that difficulty in drawing an inference in and of itself is not a specific cognitive impairment that is characteristic of ASD. Although the group with ASD was relatively more impaired with making inferences than the group with typical development, participants with ASD were able to make inferences about physical and mental states especially with increases in language ability and/or age. Difficulty in making inferences may be reflective of a more generalized underlying difficulty with information processing mechanisms consistent with a *complex information processing model* of ASD (Minshew et al. 1997; Williams et al. 2006). That is, as suggested by the results of the Mason and colleagues (2008) fMRI study of bridging inferences in ASD, the individuals with ASD may be accomplishing the process of drawing an inference with a more inefficient neural network than that of the age and ability-matched controls with typical development, relying more heavily on their language skills and experiential knowledge to compensate for this inefficiency in cognitive processing. This view is supported by numerous other studies of cognitive and linguistic processing that suggest an underlying problem with the formation and/or functioning of neural processing networks in ASD (for review see Groen et al. 2008).

The interpretation of the results of the current study as indicating that individuals with ASD use language as a bootstrap for drawing an inference is based upon earlier research that has suggested a relationship between language and the development of theory-of-mind in children with ASD (Tager-Flusberg & Joseph, 2005). However, in the current study, the level of language of the individuals with ASD was important not only for making inferences related to theory-of-mind but for inferential thinking more generally.

The age-related effects in the behavioral data from the current study are consistent with research that has reported differences in the neurofunctional patterns that underlie cognitive and linguistic processing of children and adults with ASD. For example, a recent fMRI study that compared brain activation during the processing of literal and ironic text of older children and adults with autism found differences between these two age groups that suggested positive effects in brain function that appeared to be related to increases in semantic and experiential knowledge that occur with age (Williams et al. 2013). The current study lends support to the assumption that, despite persistent underlying neurofunctional differences, the continued acquisition of semantic and experiential knowledge can have positive effects on the functioning of verbal, relatively-able individuals with ASD.

A particularly interesting finding from the current study was the difficulty that the individuals with ASD had with making inferences about emotional states, a challenge that did not diminish with age or improvements in language ability. This finding is consistent with previous work that has proposed that affective theory-of-mind (making inferences about the emotional states of others) is dissociable from cognitive theory-of-mind (making inferences about the intentionality or mental states of others) with the former thought to require a more elaborate neural network than the latter (Shamay-Tsoory 2011; Sebastian, 2012).

Limitations

The current study used a large, well-characterized population of individuals with ASD who had Verbal IQs of 80 or above; therefore, the results are most applicable to this verbally capable population of individuals with ASD. Although the measurement tool that was used, the PIT, is a simple test that is successful in investigating complex aspects of discourse processing in individuals with ASD, it may be overly simplistic for TD participants. This line of reasoning is supported by an ostensible ceiling effect in the TD group. That is, TD individuals uniformly responded with a correct answer to all story stems, thus complicating investigations of age effects in this population. An evaluation of the types of responses (ToM) in each group revealed more pronounced variations between individuals with ASD and with TD. Generally, TD participants answered each question correctly, while ASD participants gave correct answers but had less correct physical and emotion-based ToM responses. Potentially significant variations in the responses of TD individuals may be masked by a ceiling effect of the test. Future studies may work to remedy this ceiling effect by adding more stories and/or more difficult stories to the assessment.

Future Directions

Whereas the results of this study provided evidence that both children and adults with ASD had difficulty with making inferences, we did not directly relate their abilities in this cognitive skill to their ability in the comprehension of discourse. Further work in this area should investigate this relationship as well as including more direct assessment of contextual integration, another potentially important contributor to the processing of discourse that may be affected in ASD.

Given the age effects obtained in the current study, an instrument such as the PIT may be useful for longitudinal or cross-sectional studies in which children and adults with ASD are compared to further examine the developmental progression of comprehension in general and inference making more specifically. The PIT could also be utilized in treatment efficacy and effectiveness studies in order to evaluate potential improvements in discourse processing as the result of ASD interventions. Finally, measurement of drawing inferences from various types of social and non-social information may be clinically useful, identifying specific areas that could be the target of intervention for improving comprehension of discourse in both academic and social situations.

Conclusions

In conclusion, the current study extends the literature by reporting not only inference making difficulties in individuals with ASD, but more importantly identifies relevant types of inference making deficits (e.g. emotion related) in this population. More encouraging, are the reported improvements related to age and linguistic level in some types of inference making abilities, though these do not appear to extend to emotion-related inferences.

Acknowledgments

We acknowledge the support of the National Institute of Child Health and Human Development (NICHD) [HD055748, an Autism Center of Excellence, to N.J.M.]; and, the National Institute on Deafness and other Communication Disorders (NIDCD) [K23DC006691 to D.L.W.]. We are grateful to the participants and families

who generously gave of their time and effort to this study. We acknowledge the contribution of Amanda Brening, Kelsey Woods, and Maureen McNiff in the development of the PIT. KEB would like to thank Denis McCarthy for his guidance and feedback on the development of the psychometric properties of the PIT. We thank Rob Mason for his contribution of stimuli that were important in the developmental process.

References

- Algina J, Keselman HJ. Approximate confidence intervals for effect sizes. *Educational and Psychological Measurement*. 2003; 63:537–553.
- American Psychological Association. *Publication manual of the American Psychological Association*. 6. Washington, DC: Author; 2010.
- Anderson, D. A few quotes regarding hypothesis testing. 1997. Retrieved from tiny.cc/nhstquotes
- Arciuli J, Stevens K, Trembath D, Simpson IC. The relationship between parent report of adaptive behavior and direct assessment of reading ability in children with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*. 2013; 56(6):1837–1844.
- Baron-Cohen S, Leslie AM, Frith U. Does the autistic child have a “theory of mind”? *Cognition*. 1985; 21:37–46. [PubMed: 2934210]
- Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I. The “Reading the Mind in the Eyes” Test Revised Version: A study with normal adults, and adults with Asperger Syndrome or high-functioning autism. *J Child Psychol Psychiatr*. 2001; 42:241–251. [PubMed: 11280420]
- Berger JO, Berry DA. Statistical analysis and the illusion of objectivity. *American Scientist*. 1988; 76:159–165.
- Bowler D. “Theory of Mind” in Asperger’s Syndrome. *Journal of Child Psychology and Psychiatry*. 1992; 33:877–893. [PubMed: 1378848]
- Bowler D. Reaction times to mental state and non-mental state questions in false belief tasks by high-functioning individuals with autism. *European Child & Adolescent Psychiatry*. 1997; 6:160–165. [PubMed: 9383650]
- Brent E, Rios P, Happé F, Charman T. Performance of children with autism spectrum disorder on advanced theory of mind tasks. *Autism*. 2004; 8:283–299. [PubMed: 15358871]
- Brown HM, Oram-Cardy J, Johnson A. A meta-analysis of the reading comprehension skills of individuals on the autism spectrum. *Journal of autism and developmental disorders*. 2013; 43(4): 932–955. [PubMed: 23054199]
- Cumming, G. *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge; 2012.
- Cumming G. The new statistics: Why and how. *Psychological Science*. 2014; 25:7–29. [PubMed: 24220629]
- Cumming G, Fidler F. Confidence Intervals: Better answers to better questions. *Zeitschrift fuer Psychologie/Journal of Psychology*. 2009; 217:15–26.
- Dennis M, Lazenby AL, Lockyer L. Inferential language in high-function children with autism. *Journal of Autism and Developmental Disorders*. 2001; 31:47–54. [PubMed: 11439753]
- Edwards W, Lindman H, Savage LJ. Bayesian statistical inference for psychological research. *Psychological Review*. 1963; 70(3):193–242.
- Gallistel CR. The importance of proving the null. *Psychological Review*. 2009; 116(2):439–453. [PubMed: 19348549]
- Graesser AC, Singer M, Trabasso T. Constructing inferences during narrative text comprehension. *Psychological review*. 1994; 101(3):371. [PubMed: 7938337]
- Groen WB, Zwiers MP, van der Gaag RJ, Buitelaar JK. The phenotype and neural correlates of language in autism: An integrative review. *Neuroscience and Biobehavioral Reviews*. 2008; 32:1416–1425. [PubMed: 18562003]
- Happé FGE. 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*. 1994; 24:129–154. [PubMed: 8040158]
- Hoijtink, H.; Klugkist, I.; Boelen, P. *Bayesian evaluation of informative hypotheses that are of practical value for social scientists*. New York: Springer; 2008.

- Hollingshead, AB. Four factor index of social status. Yale University, Department of Sociology; New Haven, CT: 1975.
- Huemer SV, Mann V. A comprehensive profile of decoding and comprehension in autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2010; 40(4):485–493. [PubMed: 19915968]
- Jeffreys, H. *Theory of probability*. Oxford, UK: Oxford University Press; 1961.
- Jolliffe T, Baron-Cohen S. The strange stories test: a replication with high-functioning adults with autism or Asperger syndrome. *Journal of Autism and Developmental Disorders*. 1999; 29:395–404. [PubMed: 10587886]
- Kaland N, Møller-Nielsen A, Smith L, Mortensen EL, Callesen K, Gottlieb D. The strange stories test: A replication study of children and adolescents with Asperger syndrome. *European child & adolescent psychiatry*. 2005; 14(2):73–82. [PubMed: 15793686]
- Kass RE. Bayes factors in practice. *The Statistician*. 1993; 42:551–560.
- Kass RE, Raftery AE. Bayes factors. *Journal of the American Statistical Association*. 1995; 90:773–795.
- Kirk, RE. The importance of effect magnitude. In: Davis, SF., editor. *Handbook of research methods in experimental psychology*. Malden, MA: Blackwell; 2003. p. 83-105.
- Leinonen E, Kerbel D. Relevance theory and pragmatic impairment. *International Journal of Language & Communication Disorders*. 1999; 34(4):367–390. [PubMed: 10884907]
- Le Sourn-Bissaoui S, Caillies S, Gierski F, Motte J. Inference processing in adolescents with Asperger syndrome: Relationship with theory of mind abilities. *Research in Autism Spectrum Disorders*. 2009; 3(3):797–808.
- Lewis FM, Murdoch BE, Woodyatt GC. Communicative competence and metalinguistic ability: Performance by children and adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*. 2007; 37:1525–1538. [PubMed: 17665298]
- Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavone PC, et al. The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*. 2000; 30:205–223. [PubMed: 11055457]
- Lord C, Rutter M, LeCouteur AL. Autism diagnostic interview revised. A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*. 1994; 24:659–685. [PubMed: 7814313]
- Loukusa S, Leinonen E, Kuusikko S, Jussila K, Mattila ML, Ryder N, Ebeling H, Moilanen I. Use of context in pragmatic language comprehension by children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders*. 2007; 37:1049–1059. [PubMed: 17072751]
- Mason RA, Williams DL, Kana RK, Minshew NJ, Just MA. Theory of mind disruption and recruitment of the right hemisphere during narrative comprehension in autism. *Neuropsychologia*. 2008; 46:269–280. [PubMed: 17869314]
- Minshew NJ, Goldstein G, Siegel DJ. Speech and language in high-functioning autistic individuals. *Neuropsychology*. 1995; 9:255–261.
- Minshew NJ, Goldstein G, Siegel DJ. Neuropsychological functioning in ASD: Profile of a complex information processing disorder. *Journal of the International Neuropsychological Society*. 1997; 3:303–316. [PubMed: 9260440]
- Morey, RD.; Rouder, JN. BayesFactor: Computation of Bayes factors for common designs. R package version 0.9.7. 2014. <http://CRAN.R-project.org/package=BayesFactor>
- Morey RD, Rouder JN, Verhagen J, Wagenmakers E-J. Why hypothesis tests are essential for psychological science: A comment on Cumming 2014. *Psychological Science*. n.d; 25:1289–1290. [PubMed: 24604147]
- Myung IJ, Pitt MA. Applying Occam’s razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin and Review*. 1997; 4(1):79–95.
- Norbury CF, Bishop DV. Inferential processing and story recall in children with communication problems: a comparison of specific language impairment, pragmatic language impairment and

- high-functioning autism. *International Journal of Language and Communication Disorders*. 2002; 37:227–251. [PubMed: 12201976]
- Perner J, Wimmer H. “John thinks that Mary thinks that...” Attribution of second-order beliefs by 5–10 year old children. *Journal of Experimental Child Psychology*. 1985; 39:437–471.
- Raftery AE. Bayesian model selection in social research. *Sociological Methodology*. 1995; 25:111–164.
- Ricketts J, Jones CRG, Happé F, Charman T. Reading comprehension in autism spectrum disorders: The role of oral language and social functioning. *Journal of Autism and Developmental Disorders*. 2013; 43:807–816. [PubMed: 22843036]
- Rosnow RL, Rosenthal R. Effect sizes: Why, when, and how to use them. *Zeitschrift fuer Psychologie/ Journal of Psychology*. 2009; 217:6–14.
- Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*. 2009; 16(2):225–237. [PubMed: 19293088]
- Saldaña D, Frith U. Do readers with autism make bridging inferences from world knowledge? *Journal of Experimental Child Psychology*. 2007; 96:310–319. [PubMed: 17196973]
- Sebastian CL, Fontaine NM, Bird G, Blakemore SJ, De Brito SA, McCrory EJ, Viding E. Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Social Cognitive and Affective Neuroscience*. 2012; 7:53–63. [PubMed: 21467048]
- Shamay-Tsoory SG. The neural bases for empathy. *The Neuroscientist*. 2011; 17:18–24. [PubMed: 21071616]
- Singer M. Validation in reading comprehension. *Current Directions in Psychological Science*. 2013; 22(5):361–366.
- Social Security Online. Popular names by birth year: popularity in 2005. 2008 May. Retrieved from Social Security Administration website: <http://www.ssa.gov/OACT/babynames/>
- Snyder, L.; Caccamise, D. Comprehension processes for expository text: Building meaning and making sense. In: Nippold, M.; Scott, CM., editors. *Expository discourse in children, adolescents, and adults*. New York, NY: Psychology Press; 2010. p. 13-39.
- Sperber D, Wilson D. Pragmatics, modularity and mind-reading. *Mind & Language*. 2002; 17:3–23.
- Tager-Flusberg, H.; Joseph, RM. How language facilitates the acquisition of false-belief understanding in children with autism. In: Astington, JW.; Baird, JA., editors. *Why language matters for theory of mind*. New York, NY: Oxford University Press; 2005. p. 298-318.
- Wagenmakers EJ. A practical solution *n* to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*. 2007; 14(5):779–804. [PubMed: 18087943]
- Wechsler, D. Wechsler abbreviated scale of intelligence. San Antonio, TX: Psychological Corporation; 1999.
- Wiig, EH.; Secord, W. Test of Language Competence-Expanded Edition (TLC-E). San Antonio, TX: Psychological Corporation; 1989.
- Williams DL, Cherkassky VL, Mason RA, Keller TA, Minshew NJ, Just MA. Brain function differences in language processing in children and adults with autism. *Autism Research*. 2013; 6:288–302. [PubMed: 23495230]
- Williams DL, Goldstein G, Minshew NJ. Neuropsychologic functioning in children with autism: Further evidence for disordered complex information processing. *Child Neuropsychology*. 2006; 12:279–298. [PubMed: 16911973]

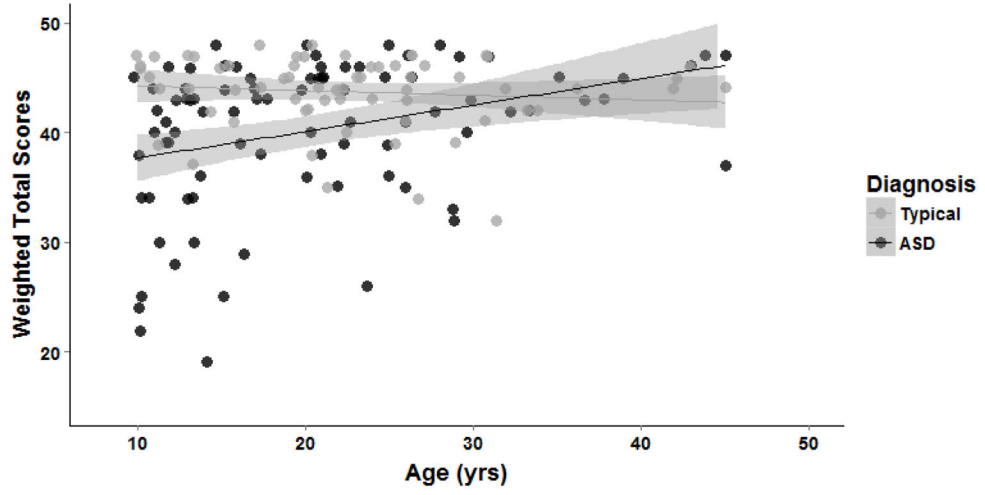


Figure 1. Association between age and PIT weighted total scores by group diagnosis

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

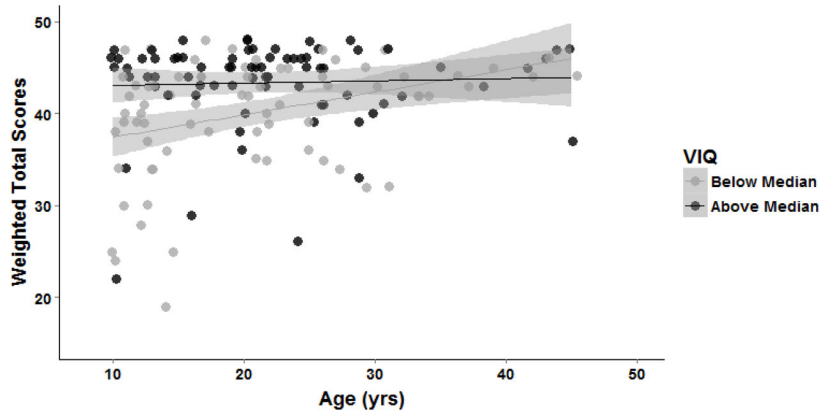


Figure 2. Association between age and PIT weighted total scores by low and high VIQ scores

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

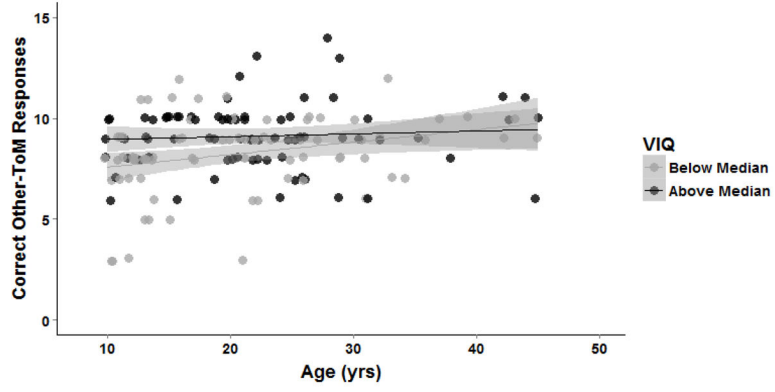


Figure 3. Association between age and number of correct other-ToM responses by low and high VIQ scores

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Sample Characteristics

Variable	ASD (n = 86)		Non-ASD (n = 65)	d_{unb} [95% CI]
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)		
Age (years)	20.56 (9.07)	22.63 (8.38)		0.23 [-0.09, 0.56]
Verbal IQ*	110.21 (13.52)	111.69 (8.43)		0.13 [-0.2, 0.45]
Performance IQ*	109.79 (12.17)	112.95 (8.05)		0.3 [-0.03, 0.63]
FSIQ*	111.2 (12.45)	114.0 (8.16)		0.26 [-0.07, 0.59]
SES**	43.97 (17.08)	37.96 (18.47)		0.34 [0.01, 0.67]
ADOS Total	11.28 (3.18)	-		-
Gender(M/F)	73/13	57/8		-

Note:

* Based on the WASI (Psychological Corporation, 1999);

** One participant with ASD and 6 participants without ASD declined to report SES

Table 2

ASD and non-ASD PIT Scores

Variable	ASD (n = 86)	Non-ASD (n = 65)	d_{unb} [95% CI]
	<i>M(SD)</i>	<i>M(SD)</i>	
Weighted Total Scores	40.24(6.69)	43.72(3.37)	0.63 [0.29, 0.96]
Physical Scores	8.66(1.73)	9.78(1.72)	0.65 [0.31, 0.98]
Other-ToM Scores	8.66(2.14)	8.72(1.36)	0.03 [-0.29, 0.36]
Emotion-ToM Scores	7.13(2.14)	8.25(1.68)	0.57 [0.24, 0.90]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3
Correlations and 95% CIs between PIT and TLC-E Raw Scores in Individuals with ASD (n = 86)

	Weighted Total (PIT)	Physical (PIT)	Other ToM (PIT)	Emotion ToM (PIT)
Total Raw (TLC)	0.61 [0.46, 0.73]	0.06 [-0.15, 0.27]	0.54 [0.38, 0.68]	0.38 [0.19, 0.55]
Subtest 2 (TLC)	0.58 [0.42, 0.71]	0.21 [-0.01, 0.4]	0.46 [0.27, 0.61]	0.37 [0.17, 0.54]
Subtest 4 (TLC)	0.72 [0.60, 0.81]	0.03 [-0.19, 0.24]	0.66 [0.52, 0.77]	0.46 [0.27, 0.61]

Correlations and 95% CIs between PIT responses and ToM Measures in Individuals with ASD

Table 4

	Weighted Total (PIT)	Physical (PIT)	Other ToM (PIT)	Emotion Tom (PIT)
ToM RME (n=86)	0.45 [0.26, 0.60]	-0.13 [-0.33, 0.09]	0.35 [0.15, 0.52]	0.40 [0.21, 0.57]
Tom Aggregate (n = 79) [^]	0.49 [0.30, 0.64]	0.14 [-0.08, 0.35]	0.38 [0.18, 0.56]	0.33 [0.11, 0.51]
Adult Only RME (n = 53)	0.48 [0.24, 0.66]	-0.11 [-0.37, 0.17]	0.18 [-0.09, 0.43]	0.50 [0.27, 0.68]
Child Only RME (n = 33)	0.20 [-0.15, 0.51]	-0.16 [-0.48, 0.19]	0.20 [-0.15, 0.51]	0.21 [-0.15, 0.51]

Note:

[^] 7 individuals with ASD were not administered all three ToM tasks.

RME = Reading Mind in the Eyes.