



Measuring Individual Differences in Cognitive, Affective, and Spontaneous Theory of Mind Among School-Aged Children with Autism Spectrum Disorder

Melody Altschuler¹ · Georgios Sideridis^{2,3} · Shashwat Kala⁴ · Megan Warshawsky⁵ · Rachel Gilbert² · Devon Carroll^{2,6} · Rebecca Burger-Caplan¹ · Susan Faja^{2,3,6,7}

Published online: 3 July 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

The present study examined individual differences in theory of mind (ToM) among a group of 60 children (7–11 years-old) with autism spectrum disorder (ASD) and average intelligence. Using open-ended and structured tasks to measure affective ToM, cognitive ToM, and spontaneous social attribution, we explored the nature of ToM and assessed whether ToM predicts the phenotypic heterogeneity in ASD through structural equation modeling. Affective ToM uniquely predicted social symptom severity, whereas no ToM types predicted parent reported social functioning. Our findings suggest that differentiating among theoretical components is crucial for future ToM research in ASD, and ToM challenges related to reasoning about others' emotions may be particularly useful in distinguishing children with worse social symptoms of ASD.

Keywords Theory of mind · Affective functioning · Social cognition · Autism spectrum disorder · Symptom severity

Introduction

There is a longstanding theoretical connection between autism spectrum disorder (ASD) and theory of mind (ToM), the capacity to attribute mental states to oneself and others (Premack and Woodruff 1978). Many studies have provided data consistent with the hypothesis that social difficulties in ASD are explained by a decreased ability to use mental

state concepts to interpret and predict one's own and other people's behavior (Baron-Cohen et al. 2000; Tager-Flusberg 2007). Although a "single cause" explanation for all the features of ASD is now considered implausible (Happé et al. 2006), the theoretical influence of ToM on social abilities in ASD remains highly prominent in the literature, with more recent work examining the potential combined contribution

✉ Susan Faja
susan.faja@childrens.harvard.edu

Melody Altschuler
melody.altschuler@yale.edu

Georgios Sideridis
georgios.sideridis@childrens.harvard.edu

Shashwat Kala
shashwat.kala.18@dartmouth.edu

Megan Warshawsky
megan.warshawsky@tufts.edu

Rachel Gilbert
rachel.gilbert@childrens.harvard.edu

Devon Carroll
devon.carroll@childrens.harvard.edu

Rebecca Burger-Caplan
rebecca.burger-caplan@yale.edu

¹ Yale School of Medicine, Yale Child Study Center, New Haven, CT, USA

² Division of Developmental Medicine, Boston Children's Hospital, Boston, MA, USA

³ Harvard Medical School, Boston, MA, USA

⁴ Dartmouth College, Hanover, NH, USA

⁵ Tufts University, Medford, MA, USA

⁶ Department of Psychiatry, Boston Children's Hospital, Boston, MA, USA

⁷ Laboratories of Cognitive Neuroscience, Boston Children's Hospital, 1 Autumn Street, BCH3178, Boston, MA 02215, USA

of ToM with other factors such as executive function (Jones et al. 2018).

Traditionally, ToM is assessed by asking young children to reflect on a naive story character that holds a “false belief.” False belief understanding is essential for social interactions, including both “first-order” false beliefs, which indicate what children think about real events, and more advanced “second-order” false beliefs, which refer to what children think about other people’s thoughts (Perner and Wimmer 1985). Preschoolers with ASD consistently show deficits in false belief reasoning compared to typically developing children (Baron-Cohen et al. 2000; Senju 2012). However, the sensitivity of traditional false belief paradigms is limited when used with school-aged and cognitively unimpaired children with ASD, who usually perform as well as typically developing children on these tasks (Scheeren et al. 2013; but see Senju et al. 2009).

Within ToM, there are distinct theoretical types with varying complexity, as well as different ways of measuring the broad construct of ToM (Apperly 2012). Advanced ToM consists of many abilities, including higher-order false belief understanding (Perner and Wimmer 1985), emotion recognition (Baron-Cohen et al. 1997), and social understanding (Baron-Cohen et al. 1999). A recent factor analysis found that advanced ToM abilities, unlike basic ToM, include social reasoning, reasoning about ambiguity, and recognizing transgressions of social norms (Osterhaus et al. 2016). These advanced ToM abilities are conceptually distinct from the more basic type of ToM measured by traditional false belief reasoning tasks.

Recent work has begun to examine advanced ToM abilities in school-aged and cognitively unimpaired children with ASD. Some studies have demonstrated that these children perform equally well on advanced ToM tasks as children without ASD (Begeer et al. 2016; Scheeren et al. 2013), while other studies continue to find deficits (Baribeau et al. 2015; Salter et al. 2008; White et al. 2009). However, across studies, the definition used to characterize advanced ToM is inconsistent, and there is variability in the tasks used. For example, advanced ToM has been defined and measured as the ability to interpret the emotional cues in pictures of eyes in some studies (Baribeau et al. 2015; Baron-Cohen et al. 1997), but as the ability to reason about others’ thoughts, feelings, and behaviors in response to social stories in other studies (Mazza et al. 2017; Scheeren et al. 2013; White et al. 2009). Moreover, most studies have used a single measure to assess advanced ToM (Baribeau et al. 2015; Salter et al. 2008; Shimoni et al. 2012), rather than multiple measures within the same sample. The mixed literature suggests that future work should clarify which components of ToM are most impacted in ASD.

One possible reason for the inconsistent findings is that past studies did not adequately distinguish performance on

tasks measuring distinct theoretical components of ToM. Research has suggested that affective and cognitive aspects of ToM are somewhat separable. Shamay-Tsoory and Aharon-Peretz (2007) found a distinction between cognitive ToM (knowledge about others’ beliefs) and affective ToM (knowledge about others’ emotions), which are thought to involve overlapping but dissociable brain systems. Another possible reason for inconsistencies in the literature is that laboratory-based ToM tasks vary in their sensitivity to real-life task demands. In order to be successful in real-life social situations, it is necessary to spontaneously identify relevant social information (e.g., relationships, intentions) before using social cognitive problem solving skills (Klin et al. 2003). Yet, only a few studies have examined spontaneous social attribution skills by eliciting open-ended narratives in children with ASD (Salter et al. 2008; Shimoni et al. 2012), and none simultaneously measured cognitive and affective ToM.

In addition to variation in theoretical ToM subdomains that may account for the mixed findings in the literature, there are many methodological differences across studies. The majority of ToM measures present information verbally, while fewer present information visually. Individual differences in language ability have been found to relate to ToM performance in both typically developing children and children with ASD (Milligan et al. 2007), suggesting performance may be mediated by verbal ability. Children with strong verbal skills may be able to “hack” false belief reasoning through use of cognitive skills (Happé 1995). Age differences may also account for some of the conflicting findings in the literature, as some studies suggest that ToM is delayed rather than deficient in ASD (Burger-Caplan et al. 2016; Steele et al. 2003). In addition, the majority of ToM measures are scored as either pass or fail, while fewer allow for open-ended responses that may provide a more sensitive measurement of ToM ability. Examining different methodological components of ToM tasks is important in order to separate ToM ability from factors related to task demands, such as how much verbal reasoning and spontaneous attention to key details of the stimuli are required. A related consideration is the sample size of previous work. Given known heterogeneity within ASD, if ToM is not universally impaired among individuals with ASD, smaller samples may include differing proportions of children with ToM impairments. Research with larger samples may clarify mixed results.

Many studies have examined the extent to which ToM challenges contribute to the wide range of social difficulties observed in ASD, both in terms of the severity of specific ASD symptoms (e.g., Jones et al. 2018) and broader social functioning in daily life (e.g., Burger-Caplan et al. 2016). Symptoms represent specific social communication deficits, whereas social functioning represents

overall behavior across many domains, including interpersonal relationships, adaptive function, and leisure time (Yager and Ehmann 2006). ToM deficits have been linked to higher levels of ASD symptoms in individuals with ASD (Jones et al. 2018; Lerner et al. 2011; Shimoni et al. 2012). Other studies have shown that higher ToM predicts better communicative and socially adaptive functioning, as well as fewer broad social problems (Berenguer et al. 2018; Bishop-Fitzpatrick et al. 2017; Burger-Caplan et al. 2016; Joseph and Tager-Flusberg 2004; Mazza et al. 2017; Shimoni et al. 2012). In contrast, some research suggests that ToM ability is unrelated to social communication difficulties in ASD (Cantio et al. 2016; Pellicano 2013). More research using detailed ToM batteries is needed to clarify whether distinct components of ToM predict symptom severity and social functioning in ASD.

Present Study

Given that ToM is thought to be a multifaceted construct with ToM types that vary in level of complexity (Apperly 2012), the present study examines whether theoretical distinctions are present among school-aged children (7–11 years-old) with ASD and average to above average intelligence (IQs of 85 or above). A substantial limitation of past research on ToM in ASD, particularly among school-aged children, is that few studies have examined the advanced ability to reason about others' emotions in social situations as well as the ability to spontaneously interpret others' ambiguous actions. Tasks that more closely mirror the unfolding of spontaneous social interpretations in everyday life, and that capture the advanced ToM skill of reasoning about the emotional experiences of others (Osterhaus et al. 2016), may represent distinct aspects of the ToM profiles of school-aged children with ASD. To date, no studies have simultaneously examined cognitive, affective, and spontaneous ToM in ASD—a theoretical limitation addressed by the present study.

First, we tested whether theoretical distinctions (cognitive, affective, and spontaneous) exist within ToM. We hypothesized that cognitive, affective, and spontaneous ToM would represent separate factors with distinct patterns of variation. Because language ability accounts for some individual differences in ToM performance (Milligan et al. 2007), we also included tasks that present information both verbally and visually. Second, we tested whether ToM predicted social symptoms and functioning. Consistent with literature in school-aged children with ASD (e.g., Jones et al. 2018), we predicted that better performance on traditional ToM tasks would relate to better social functioning. Yet, given recent work related to advanced aspects of ToM (Osterhaus et al. 2016), we

Table 1 Sample characteristics (N=60)

| | M | Range (SD) |
|--|--------|----------------|
| <i>Age (in months)</i> | 110.60 | 85–144 (16.56) |
| <i>ADOS-2 calibrated severity scores</i> | | |
| Total | 8.92 | 5–10 (1.43) |
| Social affect | 8.28 | 5–10 (1.44) |
| Restricted repetitive | 9.12 | 4–10 (1.12) |
| <i>ADI-R raw scores</i> | | |
| Social | 18.12 | 9–29 (5.04) |
| Verbal communication | 15.87 | 8–25 (4.26) |
| Restricted and repetitive behavior | 8.35 | 3–15 (2.77) |
| <i>Vineland-2</i> | | |
| Adaptive behavior composite | 84.75 | 69–105 (8.06) |
| Socialization domain | 80.88 | 61–112 (9.78) |
| Daily living skills domain | 88.08 | 73–110 (9.55) |
| Communication domain | 91.20 | 74–110 (9.23) |
| <i>WASI-2</i> | | |
| Full Scale IQ | 106.53 | 85–135 (12.93) |
| Verbal Comprehension Index | 104.67 | 79–137 (13.19) |
| Perceptual Reasoning Index | 107.20 | 69–141 (14.42) |

ADOS-2 Autism Diagnostic Observation Schedule, Second Edition

ADI-R Autism Diagnostic Interview, Revised

Vineland-2 Vineland Scales of Adaptive Behavior, Second Edition

WASI-2 Wechsler Abbreviated Scale of Intelligence, Second Edition

also explored whether advanced ToM would be more sensitive to individual differences in social symptoms and real-life social behavior than traditional false belief ToM measures.

Method

Participants and Procedure

Sixty children (4 females) between the ages of 7–11 years-old with ASD and IQs of 85 or above participated (see Table 1). Children were recruited from the hospital's recruitment registry and other community sources (events, clinics, word of mouth). Exclusionary criteria included severe sensory or motor impairments that limited the ability to complete the test battery, colorblindness, insufficient English fluency for valid completion of standardized measures, medical disorders that impact the central nervous system, prolonged prenatal substance exposure, and a history of seizures or use of anticonvulsant medications. The hospital's human subjects division approved all study procedures, and all parents consented for their children to participate.

All participants had a previous ASD diagnosis, which was confirmed using the Autism Diagnostic Observation Schedule, Second edition (ADOS-2; Lord et al. 2012), the

Autism Diagnostic Interview, Revised (ADI-R; Rutter et al. 2003) according to Collaborative Programs of Excellence in Autism (CPEA) criteria (see Sung et al. 2005 for details), and the Diagnostic and Statistical Manual of Mental Disorders, Fifth edition (DSM-5; American Psychiatric Association 2013) criteria for ASD.

Basic exclusion criteria were screened by phone prior to enrollment. Then, the Vineland Scales of Adaptive Behavior, Second edition (Vineland-2; Sparrow et al. 2005), ADI-R, ADOS-2, Wechsler Abbreviated Scale of Intelligence, Second edition (WASI-2; Wechsler 2011), and a colorblindness screening were administered. Based on the results, eligibility was determined. A battery of ToM measures was administered in a fixed order during two additional visits. During the second visit, a parent or guardian was asked to complete questionnaires, including measures of social functioning.

Materials

ToM Measures

Given that the literature is limited on how to best measure individual differences in ToM in school-aged children with average to above average IQ, multiple measures were used to assess different types of ToM. All ToM stimuli (pictures, text, audio, and video recordings) and task instructions were presented via computer on E-prime 2.0 software.

In the First-Order False Belief (FOFB) Videos task (see Appendix 1), children answered questions after watching two videos. In the Location Change False Belief video (Saxe 2009; Wimmer and Perner 1983), children inferred the knowledge and belief of a person regarding the location of an object that was moved to a new location while he was absent. Moral judgment of his behavior (false belief or naughty) was also obtained. In the Unexpected Contents False Belief video (Perner et al. 1987), children viewed a familiar container with unexpected contents and were asked about the belief of a naive person regarding the contents. Percent correct was calculated by dividing the number of correct answers to test questions across the two videos by the total number of test questions (7). Four participants did not complete this task due to time constraints.

In the Theory-of-Mind Test (TOM Test; Muris et al. 1999), children answered questions about a series of cartoons and audio stories designed to test both basic and more advanced aspects of ToM. The task contained 38 items and three subscales: TOM level 1 (20 items), TOM level 2 (13 items), and TOM level 3 (5 items). TOM Test level 1 measures affective ToM, such as recognition of others' affective mental states and understanding of social scenarios and emotions. TOM Test level 2 measures first-order false belief. TOM Test level 3 measures second-order false belief. The experimenter scored the child's responses, based on a

Table 2 Inter-rater reliability

| Variable | <i>r</i> |
|------------------------|----------|
| TOM test total correct | |
| TOM test level 1 | 0.93 |
| TOM test level 2 | 0.99 |
| TOM test level 3 | 1.0 |
| SAT variables | |
| Number of propositions | 1.0 |
| Pertinence Index | 0.89 |
| Salience Index | 0.70 |
| Cognition Index | 0.87 |
| Affect Index | 0.97 |
| Problem-Solving Index | 0.98 |
| Person Index | 0.98 |

scoring sheet with common correct and incorrect responses provided by Muris et al. (1999). The responses flagged for review during administration were resolved post-administration by consensus coding and reviewed by the last author. Percent correct scores were calculated by dividing the total number of correct answers by the total number of questions for each level. One participant did not complete this task due to time constraints. Seven videos (12%) were scored by a second scorer. Intraclass correlation coefficients indicated excellent reliability (see Table 2).

In the Social Attribution Task (SAT; Klin 2000), children viewed animated geometrical figures that are commonly interpreted as enacting a social scene (Heider and Simmel 1944). Children were asked to describe the meaning of the animation in an open-ended manner and then with verbal cues about its social nature. The SAT video clips, instructions, and coding scheme were identical to those used by Klin (2000). The first narrative was obtained after viewing the full animation twice. Then, a series of meaningful segments were presented one at a time (narratives 2 through 7). For narratives 1 through 7, children were asked “what happened there?” For narratives 8 through 10, they were asked “what kind of person is the big triangle/small triangle/small circle?” Explicit questions followed (narratives 11 through 17), such as “why did the big triangle go into the house?” All responses were recorded for transcription and coding.

To quantify spontaneous ToM, five index scores were derived from the narratives: salience, cognition, affect, person, and problem-solving (Klin 2000)¹. Propositions² in

¹ All indices were identical to Klin (2000) except the animation index, which was not included in the present study because pilot coding revealed insufficient agreement between coders.

² A proposition, defined as a verb plus its complement, served as the narrative unit for scoring all index scores obtained from narratives 1 through 7.

Table 3 Descriptive statistics and bivariate correlations among ToM variables

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|--------------------------------|------|--------|--------|--------|--------|-------|--------|--------|--------|
| Cognitive ToM: false belief | | | | | | | | | |
| 1. TOM test level 2 | – | 0.53** | 0.33 | 0.41** | 0.12 | –0.04 | 0.09 | 0.44** | 0.47** |
| 2. FOFB videos | | – | 0.36** | 0.44** | 0.20 | –0.14 | 0.05 | 0.34 | 0.40** |
| 3. TOM test level 3 | | | – | 0.37** | 0.10 | –0.15 | –0.01 | 0.13 | 0.48** |
| Affective ToM | | | | | | | | | |
| 4. TOM test level 1 | | | | – | 0.46** | 0.09 | 0.38** | 0.36** | 0.42** |
| 5. SAT affect | | | | | – | 0.15 | 0.52** | 0.48** | 0.14 |
| Spontaneous social attribution | | | | | | | | | |
| 6. SAT cognition | | | | | | – | 0.23 | –0.04 | 0.09 |
| 7. SAT salience | | | | | | | – | 0.33 | 0.11 |
| 8. SAT problem-solving | | | | | | | | – | 0.22 |
| 9. SAT person | | | | | | | | | – |
| n | 59 | 56 | 59 | 59 | 60 | 60 | 60 | 60 | 60 |
| M | 0.58 | 0.72 | 0.46 | 0.75 | 0.02 | 0.11 | 0.34 | 0.32 | 2.03 |
| Minimum | 0.15 | 0.00 | 0.00 | 0.40 | 0.00 | 0.00 | 0.05 | 0.00 | 0 |
| Maximum | 0.92 | 1.0 | 1.0 | 1.0 | 0.19 | 0.33 | 0.75 | 0.80 | 6 |
| SD | 0.17 | 0.28 | 0.33 | 0.14 | 0.04 | 0.09 | 0.15 | 0.18 | 1.45 |

** $p < .01$

narratives 1 through 7 that were rated as non-pertinent to the broad social theme of the animation (inconstant attributions, vague references, misattributions, and irrelevant attributions) were eliminated for the coding of the salience, cognition, and affect indices. The cognition and affect indices measure the frequency of cognitive mental state terms (e.g., knowledge, desire, or belief) and affective mental state terms (e.g., jealousy or embarrassment), respectively. Each index is scored by dividing the number of cognitive or affective terms by the total number of propositions in narratives 1 through 7, such that higher scores indicate a higher frequency of spontaneous mental state terms. The salience index measures the ability to impose a social interpretation to the animations and is scored by summing the total number of social elements identified (from a total of 20 pre-identified themes) divided by 20. Higher salience index scores indicate a higher frequency of spontaneous social attribution. The person index measures the capacity to create personality attributions about the shapes and is scored hierarchically based on level of sophistication using an ordinal scale, wherein 0 represents no personality attributions (e.g., describes the shapes in physical terms only) and 6 represents attributions of sophisticated psychological characteristics. Finally, the problem-solving index measures the ability to answer explicit questions about the animation. It is scored by summing the number of correct responses (from a total of 10 items) divided by 10, with higher scores representing an increased ability to make salient social attributions when asked about the social nature of the animation.

To maximize inter-rater reliability (Klin 2000): (1) three coders were trained on SAT scoring before coding the

transcripts included in this study, and frequent meetings were held to learn explicit scoring guidelines and clarify coding issues, (2) coders followed a procedural sequence for coding each transcript, and (3) examples of frequent terms encountered in SAT narratives were included in the manual with their corresponding codes. Twenty-nine (48%) of transcripts were coded by two different coders. Intraclass correlation coefficients for SAT index scores and the total number of propositions indicated moderate to excellent reliability (see Table 2). A summary of the descriptive statistics for ToM measures and simple bivariate correlations between measures is shown in Table 3.

Measures of Social Symptoms and Social Function

The ADOS-2 Module 3 provided a measure of ASD social symptom severity using revised algorithms that included items from the Social Affect scale (Hus et al. 2014). Calibrated severity scores are derived from these items. Higher scores reflect more severe ASD-related symptoms adjusted for ADOS-2 module and age.

The Vineland-2 (Sparrow et al. 2005) interview was administered to caregivers by a clinician and determines a child's current level of adaptive functioning with normative scores relative to same-aged peers. The Socialization domain captures interpersonal relationships, play and leisure time, and coping skills. Higher scores reflect better social functioning.

Analytic Approach

Data were analyzed by use of Structural Equation Modeling (SEM) to assess both measurement and structural models (Bentler 2007; Joreskog 1973). SEM models employ a simultaneous equation approach to define unobservable constructs using observed behaviors (measured using items/exercises, etc.). The goal of the model is to minimize the discrepancy between a population covariance matrix [implied variance–covariance matrix $\Sigma(\theta)$] and the sample's variance–covariance matrix $[\Sigma]$. Model fit is evaluated by means of an omnibus asymptotic Chi square test, which tests the hypothesis that $[\Sigma(\theta) = \Sigma]$, with good model fit being reflected by a non-significant Chi square value. Given power-related concerns (excessive power given large samples), additional fit-related information comes from absolute, incremental, and parsimonious fit indices such as the Comparative Fit Index (CFI), the Normative Fit Index (NFI), and the Adjusted Goodness of Fit Index (AGFI), for which acceptable values are greater than 0.95 (Hu and Bentler 1999). Unstandardized residuals (i.e., RMSEA) below 0.05 are indicative of excellent fit, whereas values between 0.05 and 0.08 indicate acceptable model fit. Evaluation of the residuals provides the most unbiased estimate of model fit (MacCallum et al. 1996).

Power for the measurement and structural models was evaluated using a Monte Carlo simulation³ and the recommendations of Muthén and Muthén (2006) using Mplus 8.0. For the measurement model, latent variables were standardized with a mean of zero and variance equal to one. Factor loadings were specified at 0.50 and residual variances at 0.75. Between factors correlation was posited at 0.50. Results using 1,000 replicated samples of $n=60$ participants indicated that power for identifying significant factor loadings equal to or greater than 0.50 ranged between 87.1% and 90.9%. Coverage ranged between 93.2% and 93.9%. Structural paths were specified to a standardized value of 0.50 as per Cohen (1992) conventions of medium effect size. Power for those structural paths was 84.3%. Thus, the sample size of 60 participants would suffice to identify both the proper measurement model as well as clinically important structural paths. These power-related findings agreed with a recent simulation in which power and parameter stability were observed with SEM models having 50–70 participants (Sideridis et al. 2014). In addition to the use of inferential statistics, the significance of structural paths was further verified by simulating the population distribution of a slope using 10,000 replicated b-values using the mean and variance estimates of the sample. Evidence of significant effects

would be manifested with point estimates and 95% confidence intervals that do not contain the value of zero. These findings are described in the section “[Structural Model](#).”

Results

Measurement Model

A two-factor correlated model was posited to assess the cognitive and affective domains of ToM. As shown in Fig. 1, a two-factor correlated model provided good fit to the data. Unstandardized residuals (i.e., RMSEA; Steiger and Lind 1980) were less than 0.01 and descriptive fit indices, such as the CFI and TLI, had values equal to unity (Bentler 1990). Furthermore, the omnibus Chi square test was not significant, suggesting “exact fit” of the data to the model (MacCallum et al. 1996). The correlation between cognitive and affective domains was 0.6, suggesting good convergence but also the presence of discriminant construct validity. More on the model's fit is shown in Appendix 2 by use of absolute and incremental fit indices. This model was contrasted to a model positing a unidimensional structure. Using a Chi square difference test, results pointed to the superiority of the 2-factor correlated model over its unidimensional counterpart [$\Delta_{\text{Chi-square}}(1) = 17.379^4$, $p < .001$], with the latter not fitting the data well (e.g., RMSEA = 0.157, CFI = 0.861). A second alternative involved contrasting the 2-factor correlated solution to a 3-factor model in which a domain of spontaneous ToM was included. This model did not provide a good fit to the data, as the RMSEA was far beyond the 5% cutoff value (i.e., 0.137), fit indices were unacceptable (CFI = 0.755, TLI = 0.633), and two factor loadings were not significant for the spontaneous ToM factor. Further results again pointed to the superiority of the 2-factor correlated solution as model fit, as the 3-factor model was significantly worse by use of a Chi square difference test [$\Delta_{\text{Chi-square}}(19) = 48.316^5$, $p < .001$].

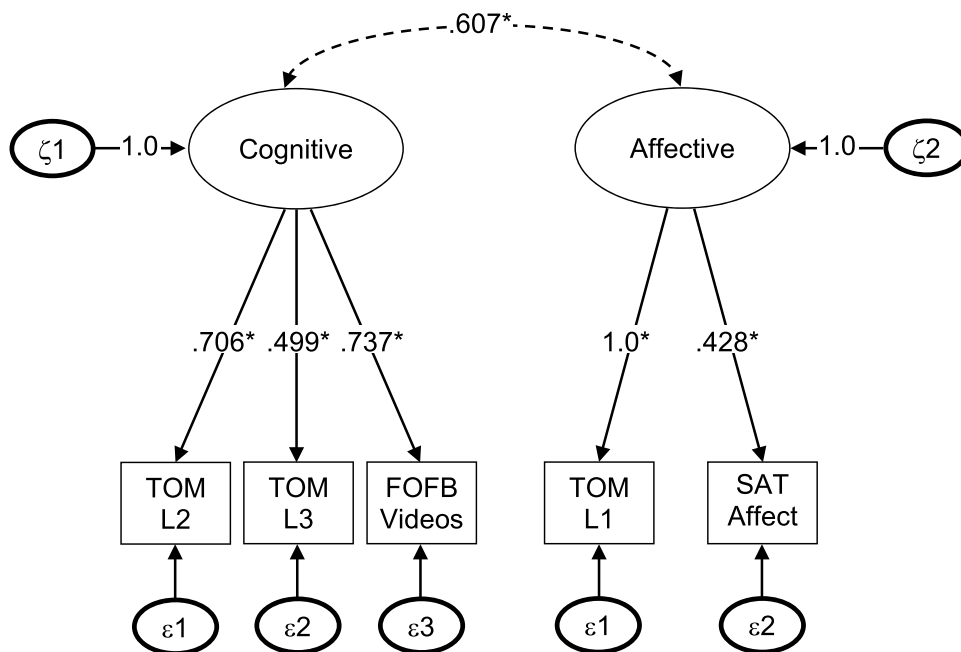
The roles of age and verbal IQ were assessed by relating both variables with the latent factors of cognitive and affective ToM. Both covariates exerted null effects on ToM. Specifically, affective ToM was unrelated to age ($r = .306$, $p = .19$) and verbal IQ ($r = -.133$, $p = .50$). Similarly, cognitive ToM was unrelated to age ($r = -.088$, $p = .39$) and verbal IQ ($r = -.110$, $p = .37$).

³ The Monte Carlo simulation routines are available from the second author upon request.

⁴ For a critical value of 3.84 Chi square units.

⁵ For a critical value of 30.144 Chi square units.

Fig. 1 Two-factor correlated solution for the measurement of cognitive and affective ToM. Residual values (1-squared factor loadings) are not shown for parsimony

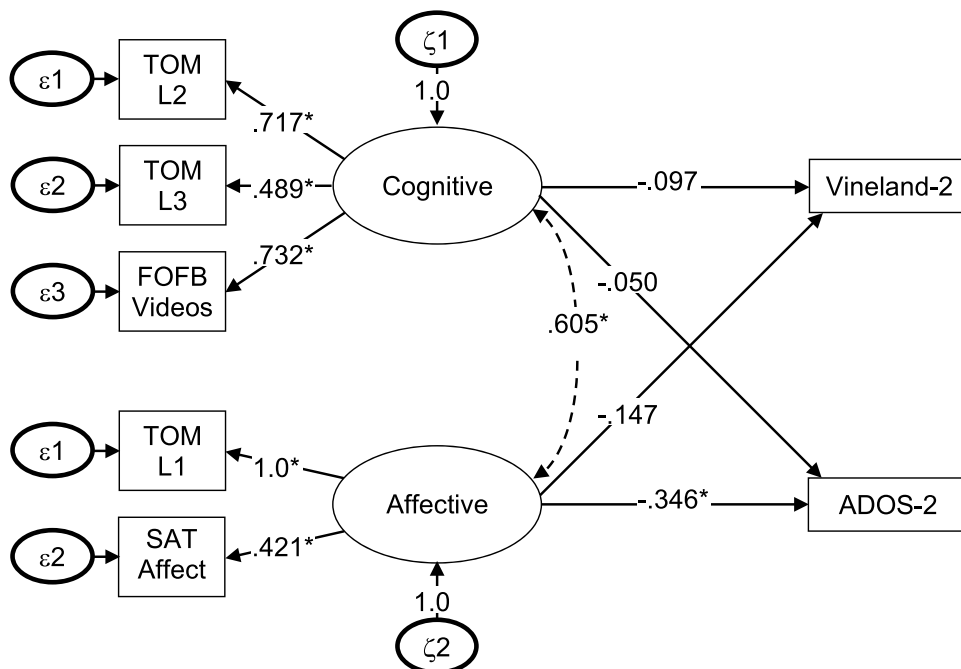


Structural Model

Following identification of the optimal simple structure for the measurement of cognitive and affective ToM, a structural model was posited in which social outcomes, including social symptoms (ADOS-2 Social Affect calibrated severity score) and functioning (Vineland-2 Socialization score), were predicted by cognitive and affective ToM (see Fig. 2). Results indicated that social symptoms were predicted significantly by affective ToM ($b = -346, p < .05$).

No other structural path exceeded conventional levels of significance. For the significant structural path, bootstrapping involved simulating the population distribution of b-values. Using 10,000 replicated paths, results indicated that the bias between the population and sample estimates was equal to 0.002. Furthermore, the 95% confidence intervals did not contain zero, suggesting that the structural path between ADOS-2 and affective ToM was different from zero (see Fig. 3). Better affective ToM performance predicted less severe social symptoms.

Fig. 2 Structural equation modeling (SEM) predicting the cognitive and affective aspects of ToM using the Vineland-2 and ADOS-2 instruments. A (*) indicates two-tailed significance at $p < .05$



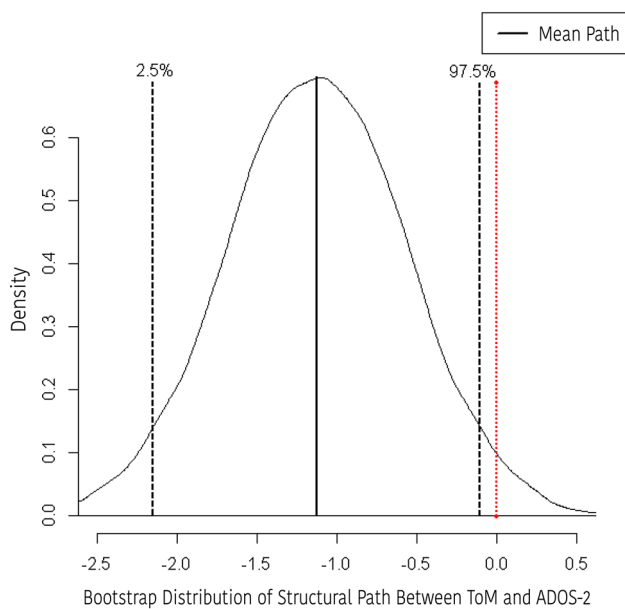


Fig. 3 Simulated population distribution of path coefficient between affective ToM and the ADOS-2 using 10,000 replicated values using mean and variance estimates from the structural model. The value of zero was not included in the 95% confidence interval of the population distribution, suggesting that the path was significantly different from zero, agreeing with the inferential statistical finding (i.e., $p < .05$). The simulation was run using Mplus, 8.0

Discussion

The present study examined individual differences in ToM in a large sample of cognitively unimpaired school-aged children with ASD. Our ToM battery measured cognitive, affective, and spontaneous ToM with a variety of different task demands, including presentation of social information visually and verbally and with structured and open-ended questions. Our results demonstrate that affective ToM has explanatory power for understanding the phenotypic heterogeneity in social symptoms of ASD and that theoretical constructs within ToM matter.

The first aim of the present study was to examine the nature of the construct of ToM in school-aged children with ASD. We examined three potential theoretical distinctions within ToM—cognitive, affective, and spontaneous. A two-factor model significantly fit the data and provided support for latent cognitive and affective factors. As hypothesized, traditional false belief variables related across visual and verbal task demands. Likewise, affective ToM variables related across structured and open-ended task demands. However, the model that included a third latent factor of spontaneous ToM, assessed by asking open-ended questions about the ambiguous actions of shapes, was not supported. This suggests that spontaneous ToM may not be a specific theoretical type of ToM, but rather, a ToM methodology that

is more consistent with real-life social attribution demands because of its open-ended and implicit nature. As with the SAT, spontaneous ToM tasks may contain both cognitive and affective elements. Taken together, our results identify two theoretically distinct types of ToM present in ASD—cognitive and affective—and demonstrate that they can be measured across structured, open-ended, visual, and verbal task demands. The present study indicates the importance of explicitly evaluating and describing affective and cognitive ToM components in ASD.

The other focus of the present study was to examine the explanatory power of ToM with respect to individual differences in social outcomes in ASD. Only affective ToM predicted social symptom severity. In contrast, cognitive ToM did not predict symptom severity, which supports the theory that individuals with ASD may use compensatory strategies to infer others' cognitive mental states (Happé 1995; Senju 2012). The finding that only affective ToM predicts social symptom severity may explain the inconsistent findings in the ToM literature in ASD and has many theoretical and clinical implications. Our results highlight that affective ToM, which captures the advanced ToM factor of social reasoning about the emotional experiences of others (Osterhaus et al. 2016), is sensitive to subtle individual differences in social function. The SAT affect index measures spontaneous affective attributions to shapes in a video, while affective questions on the TOM Test assess understanding emotional situations, that certain stimuli lead to behavioral and emotional responses, and that the social context of situations leads to different emotions. Although Muris et al. (1999) call the scale “ToM precursors,” our results indicate that affective ToM remains elusive for some children with ASD and may represent a more clinically relevant type of ToM than cognitive ToM. This adds to the literature on emotion recognition deficits in ASD (e.g., Fridenson-Hayo et al. 2016; Trevisan and Birmingham 2016) by specifying that children with worse social impairment have a particular challenge reasoning about affective components of mental states. It is also consistent with the theory that ToM challenges in ASD extend beyond mental state reasoning to include the fundamental processing of social stimuli (Klin et al. 2003). Clinically, these findings indicate that affective ToM assessments may help identify social challenges that contribute to symptom severity of children with high verbal reasoning skills who are able to perform well on more rule-bound, traditional false belief tasks.

Our results also indicate affective ToM difficulties are uniquely predictive of social symptoms that are characteristic of ASD, but not social functioning more broadly. This suggests that individual differences in ToM are most closely tied to the specific social challenges associated with ASD, rather than general social skills as would be expected for children without ASD. In contrast to the argument that social

cognition contributes to social functioning (Beauchamp and Anderson 2010), these results support the model within the ASD literature that suggests social cognition contributes specifically to symptoms (Frith 1989). However, it is possible that laboratory-based ToM tasks are more likely to relate to a laboratory-based social measure, which may capture different social behaviors than assessments asking parents to report on enduring real-world social behaviors.

Several limitations of the present study should be addressed in future research. The ToM battery was designed to assess a wide range of ToM abilities, but floor and ceiling effects may have constricted the range and could dampen true effects for some measures. Yet, the variability in performance across the entire battery is also a strength, as the battery captured different components of ToM that vary in level of difficulty for children with ASD. Future work will benefit from the use of batteries that include an even wider range of measures and task demands in order to capture individual differences and replicate our finding of affective and cognitive factors within ToM among school-aged children with ASD. As well, given that the two processes of emotion perception and ToM are conceptually linked but poorly understood (Mitchell and Phillips 2015), future studies that examine both affective ToM and basic emotion processing skills in children with ASD will shed light on the extent to which higher-level integration and inference of social-emotional information is impaired, beyond basic emotion perception. Future work is also needed to explore how other variables, such as motivation or executive functioning, contribute to ToM performance. Finally, research with a larger number of females and using a comparison group without ASD or with a clinical condition thought to have impaired ToM will provide useful information about the generalizability of our findings.

In sum, through use of a large battery of ToM measures, this study demonstrates that the emotional aspect of interpreting others' mental states can be distinguished from traditional cognitive aspects of ToM and uniquely predicts phenotypic heterogeneity in ASD. Finding two factors within ToM provides valuable theoretical insight into the ToM profile of school-aged children with ASD. Furthermore, to our knowledge, this is the first demonstration that reasoning about the affective component of others' ambiguous mental states and social scenarios uniquely predicts social symptom severity in ASD. These results specify the centrality of affective social reasoning as a critical component of ToM in school-aged children with ASD. Clinically, it may be important to develop assessments and treatments that target affective ToM difficulties and other related challenges with reasoning about the social and emotional world.

Acknowledgments Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number K99/R00HD071966. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors would like to thank Dr. Ami Klin for helping us learn the SAT coding scheme, Dr. Kate White for her comments and advice on previous drafts of the manuscript, the staff and students who assisted with collecting and scoring these measures, and to particularly thank the children and families who participated in this study. This work represents the honors thesis of the first author, which was conducted at Bates College and supported by the Bates Student Research Fund, Summer Research Fellowship, and Hoffman Research Support Grant. Preliminary results were first presented at the International Meeting for Autism Research in San Francisco, CA in May 2017.

Author Contributions MA contributed to the design of the study, contributed to the analyses and interpretation of results, participated in the acquisition and coding of the data, helped to develop the SAT coding manual, and drafted the manuscript; GS conducted the analyses and contributed to interpretation of results; SK and MW participated in the acquisition and coding of the data; RG participated in the design and coordination of the study and acquisition of the data; DC and RB helped to develop the SAT coding manual and provided consultation to clarify coding issues; SF conceived of the study, participated in its design and coordination, contributed to data acquisition and analysis, helped to develop the SAT coding manual, and helped to draft the manuscript. All authors read, reviewed, and approved the final manuscript.

Funding This study was funded by K99/R00HD071966, Bates Student Research Fund, Bates Summer Research Fellowship, and Bates Hoffman Research Support Grant.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from parents of all individual participants included in the study and all children provided their assent.

Appendix 1: First-Order False Belief (FOFB) Videos Task

Location Change False Belief Video

Script from TED Talk by Saxe (2009)

Here is a pirate. His name is Ivan. And you know what pirates really like? Pirates really like cheeseburgers. So, Ivan has this cheeseburger, and he says, "Yum yum yum yum yum! I really love cheeseburgers." And Ivan puts his

burger over here, on top of the pirate chest. And Ivan says, “You know what? I need a drink with my lunch.” And so, Ivan goes to get a drink. And while Ivan is away, the wind comes, and it blows the burger down onto the grass. And now, here comes the other pirate. This pirate is called Trevor. And Trevor also really loves cheeseburgers. So, Trevor has a cheeseburger and he says, “Yum yum yum yum yum! I love cheeseburgers.” And he puts his cheeseburger over here on top of the pirate chest.

Which cheeseburger is Trevor’s? (Comprehension Question 1).

And which cheeseburger is Ivan’s? (Comprehension Question 2).

So now Trevor goes off to get a drink. Ivan comes back and he says, “I want my cheeseburger.”

Which one do you think Ivan is going to take? (Test Question 1).

Let’s see. [Ivan is shown taking Trevor’s cheeseburger].

Why did he take that one? (Test Question 2).

Is Ivan being mean and naughty for taking Trevor’s sandwich? (Test Question 3).

Why do you think that? (Test Question 4).

Unexpected Contents False Belief Video

What do you think is in this box? (Comprehension Question 1).

If we asked your parents, what do you think they would say is in here? (Comprehension Question 2).

If we asked your friend, what do you think they would say is in the box? (Comprehension Question 3).

[Contents are revealed and explicitly labeled].

What’s inside the box? (Comprehension Question 4).

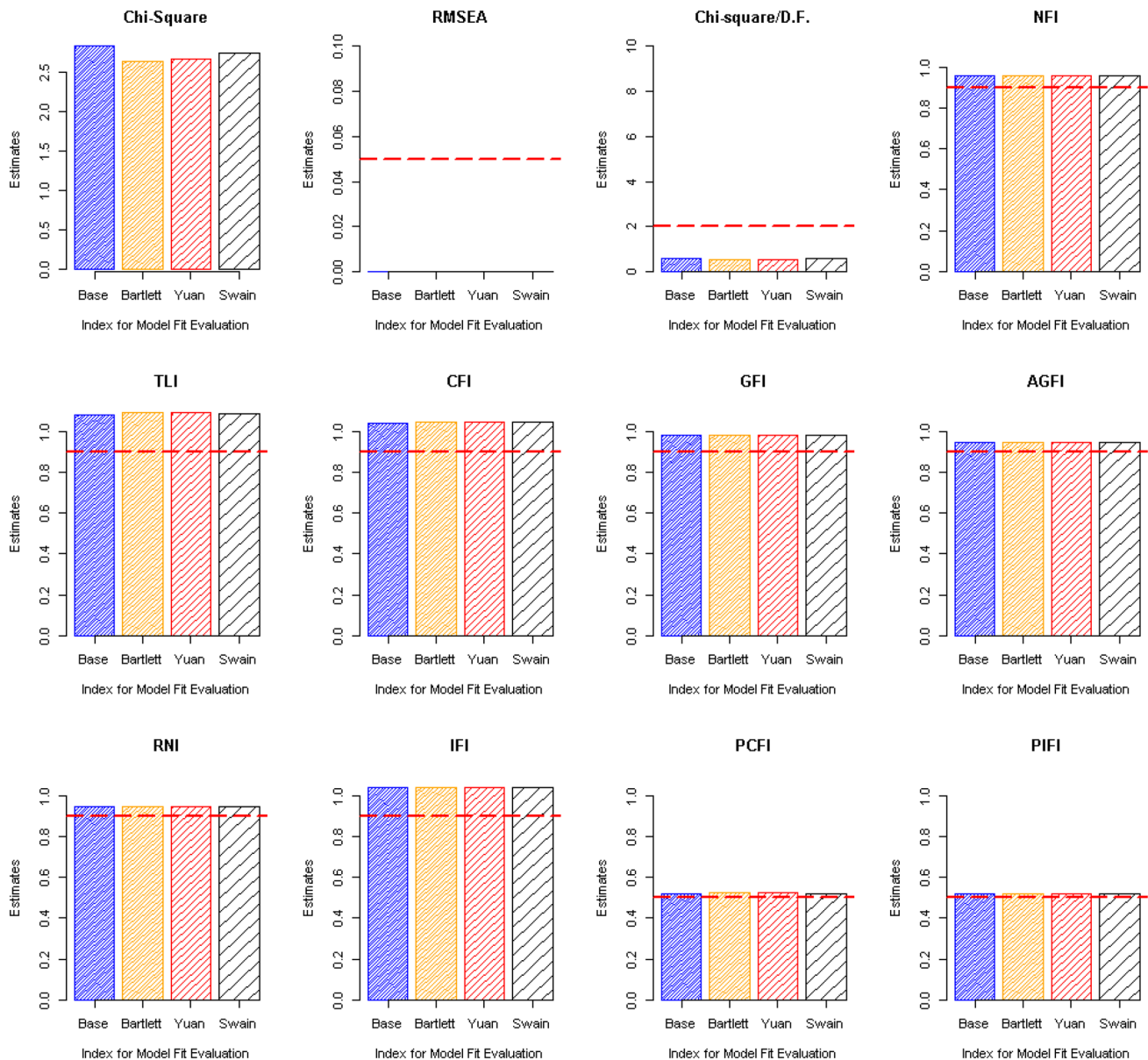
Now, if I asked you again, what would you say is in there? (Test Question 1).

If your parent came in now, and we asked what they think is in the box, what would they say? (Test Question 2).

If we asked your friend, what do you think they would say is in the box? (Test Question 3).

Appendix 2: Additional Fit Indices

Several more fit indices were utilized in order to test model fit using three recommended corrections for small sample sizes (i.e., Bartlett’s, Yuan’s, and Swain’s). As shown above, RMSEA values were close to zero, the Chi square to D.F. ratio was well below the recommended cutoff of 2.0 units, all descriptive fit indices had values greater than 0.90 generally, and 0.95 specifically, and all parsimonious fit indices had values greater than 0.50. Consequently, model fit for the 2-factor correlated model was adequate using all available means. Supplemental fit statistics were estimated using an R-function we developed for that purpose. For detailed analyses of more than 20 fit indices, please contact the second author.



References

- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, D.C.: American Psychiatric Association.
- Apperly, I. A. (2012). What is “theory of mind”? Concepts, cognitive processes and individual differences. *The Quarterly Journal of Experimental Psychology*, *65*(5), 825–839.
- Baribeau, D. A., Doyle-Thomas, K. A. R., Dupuis, A., Iaboni, A., Crosbie, J., McGinn, H., et al. (2015). Examining and comparing social perception abilities across childhood-onset neurodevelopmental disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*(6), 479–486.
- Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry*, *38*(7), 813–822.
- Baron-Cohen, S., O’Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders*, *29*(5), 407–418.
- Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D. J. (2000). *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed.). New York: Oxford University Press.
- Beauchamp, M. H., & Anderson, V. (2010). SOCIAL: An integrative framework for the development of social skills. *Psychological Bulletin*, *136*(1), 39.
- Begeer, S., Bernstein, D. M., Abfal, A., Azdad, H., Glasbergen, T., Wierda, M., et al. (2016). Equal egocentric bias in school-aged children with and without autism spectrum disorders. *Journal of Experimental Child Psychology*, *144*, 15–26.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246.

- Bentler, P. M. (2007). Covariance structure models for maximal reliability of unit-weighted composites. In S. Lee (Ed.), *Handbook of computing and statistics with applications: Handbook of latent variable and related models* (Vol. 1). New York: Elsevier.
- Berenguer, C., Miranda, A., Colomer, C., Baixauli, I., & Roselló, B. (2018). Contribution of theory of mind, executive functioning, and pragmatics to socialization behaviors of children with high-functioning autism. *Journal of Autism and Developmental Disorders*, *48*(2), 430–441.
- Bishop-Fitzpatrick, L., Mazefsky, C. A., Eack, S. M., & Minshew, N. J. (2017). Correlates of social functioning in autism spectrum disorder: The role of social cognition. *Research in Autism Spectrum Disorders*, *35*, 25–34.
- Burger-Caplan, R., Saulnier, C., Jones, W., & Klin, A. (2016). Predicting social and communicative ability in school-age children with autism spectrum disorder: A pilot study of the Social Attribution Task, Multiple Choice. *Autism*, *20*(8), 952–962.
- Cantio, C., Jepsen, J. R. M., Madsen, G. F., Bilenberg, N., & White, S. (2016). Exploring ‘The autisms’ at a cognitive level. *Autism Research*, *9*(2), 1328–1339.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155.
- Fridenson-Hayo, S., Berggren, S., Lassalle, A., Tal, S., Pigat, D., Bölte, S., et al. (2016). Basic and complex emotion recognition in children with autism: Cross-cultural findings. *Molecular Autism*, *7*(1), 52.
- Frith, U. (1989). A new look at language and communication in autism. *British Journal of Disorders of Communication*, *24*(2), 123–150.
- Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, *66*(3), 843–855.
- Happé, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. *Nature Neuroscience*, *9*(10), 1218–1220.
- Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*(2), 243–259.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55.
- Hus, V., Gotham, K., & Lord, C. (2014). Standardizing ADOS domain scores: Separating severity of social affect and restricted and repetitive behaviors. *Journal of Autism and Developmental Disorders*, *44*(10), 2400–2412.
- Jones, C. R. G., Simonoff, E., Baird, G., Pickles, A., Marsden, A. J. S., Tregay, J., et al. (2018). The association between theory of mind, executive function, and the symptoms of autism spectrum disorder. *Autism Research*, *11*(1), 95–109.
- Joreskog, K. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences*. New York: Seminar Press.
- Joseph, R. M., & Tager-Flusberg, H. (2004). The relationship of theory of mind and executive functions to symptom type and severity in children with autism. *Development and Psychopathology*, *16*(01), 137–155.
- Klin, A. (2000). Attributing social meaning to ambiguous visual stimuli in higher functioning autism and Asperger syndrome: The social attribution task. *Journal of Child Psychology and Psychiatry*, *41*(7), 831–846.
- Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: Lessons from autism. *Philosophical Transactions of the Royal Society of London B*, *358*(1430), 345–360.
- Lerner, M. D., Hutchins, T. L., & Prelock, P. A. (2011). Brief report: Preliminary evaluation of the theory of mind inventory and its relationship to measures of social skills. *Journal of Autism and Developmental Disorders*, *41*(4), 512–517.
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule-2nd edition*. Los Angeles: Western Psychological Corporation.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149.
- Mazza, M., Mariano, M., Peretti, S., Masedu, F., Pino, M. C., & Valenti, M. (2017). The role of theory of mind on social information processing in children with autism spectrum disorders: A mediation analysis. *Journal of Autism and Developmental Disorders*, *47*(5), 1369–1379.
- Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta analysis of the relation between language ability and false belief understanding. *Child Development*, *78*(2), 622–646.
- Mitchell, R. L. C., & Phillips, L. H. (2015). The overlapping relationship between emotion perception and theory of mind. *Neuropsychologia*, *70*, 1–10.
- Muris, P., Steerneman, P., Meesters, C., Merckelbach, H., Horselenberg, R., van den Hogen, T., et al. (1999). The TOM test: A new instrument for assessing theory of mind in normal children and children with pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *29*(1), 67–80.
- Muthén, B., & Muthén, L. (2006). *Mplus user’s guide*. Los Angeles: Muthén and Muthén.
- Osterhaus, C., Koerber, S., & Sodian, B. (2016). Scaling of advanced theory of mind tasks. *Child Development*, *87*(6), 1971–1991.
- Pellicano, E. (2013). Testing the predictive power of cognitive atypicalities in autistic children: Evidence from a 3-year follow-up study. *Autism Research*, *6*(4), 258–267.
- Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds’ difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, *5*(2), 125–137.
- Perner, J., & Wimmer, H. (1985). “John thinks that Mary thinks that... attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, *39*(3), 437–471.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526.
- Rutter, M., Le Couteur, A., & Lord, C. (2003). *Autism diagnostic interview-revised*. Los Angeles: Western Psychological Services.
- Salter, G., Seigal, A., Claxton, M., Lawrence, K., & Skuse, D. (2008). Can autistic children read the mind of an animated triangle? *Autism*, *12*(4), 349–371.
- Saxe, R. (2009). How we read each other’s minds. [Video file]. Retrieved from https://www.ted.com/talks/rebecca_saxe_how_brains_make_moral_judgments.
- Scheeren, A. M., de Rosnay, M., Koot, H. M., & Begeer, S. (2013). Rethinking theory of mind in high-functioning autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *54*(6), 628–635.
- Senju, A. (2012). Spontaneous theory of mind and its absence in autism spectrum disorders. *The Neuroscientist*, *18*(2), 108–113.
- Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. *Science*, *325*(5942), 883–885.
- Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: A lesion study. *Neuropsychologia*, *45*(13), 3054–3067.
- Shimoni, H. N., Weizman, A., Yoran, R. H., & Raviv, A. (2012). Theory of mind, severity of autistic symptoms and parental correlates in children and adolescents with Asperger syndrome. *Psychiatry Research*, *197*(1), 85–89.
- Sideridis, G., Simos, P., Papanicolaou, A., & Fletcher, J. (2014). Using structural equation modeling to assess functional connectivity in the brain: Power and sample size considerations. *Educational and Psychological Measurement*, *74*(5), 733–758.

- Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland adaptive behavior scales: Second edition (Vineland II), survey interview form/caregiver rating form*. Livonia: Pearson Assessments.
- Steele, S., Joseph, R. M., & Tager-Flusberg, H. (2003). Brief report: Developmental change in theory of mind abilities in children with autism. *Journal of Autism and Developmental Disorders*, *33*(4), 461–467.
- Steiger, J. H., & Lind, J. C. (1980, May). *Statistically-based tests for the number of common factors*. Paper presented at the Annual Meeting of the Psychometric Society in Iowa City, IA.
- Sung, Y. J., Dawson, G., Munson, J., Estes, A., Schellenberg, G. D., & Wijsman, E. M. (2005). Genetic investigation of quantitative traits related to autism: Use of multivariate polygenic models with ascertainment adjustment. *American Journal of Human Genetics*, *76*(1), 68–81.
- Tager-Flusberg, H. (2007). Evaluating the theory-of-mind hypothesis of autism. *Current Directions in Psychological Science*, *16*(6), 311–315.
- Trevisan, D. A., & Birmingham, E. (2016). Are emotion recognition abilities related to everyday social functioning in ASD? A meta-analysis. *Research in Autism Spectrum Disorders*, *32*, 24–42.
- Wechsler, D. (2011). *Wechsler abbreviated scale of intelligence-2nd edition*. San Antonio: Psychological Corporation.
- White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the strange stories: Revealing mentalizing impairments in autism. *Child Development*, *80*(4), 1097–1117.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.
- Yager, J. A., & Ehmann, T. S. (2006). Untangling social function and social cognition: A review of concepts and measurement. *Psychiatry*, *69*(1), 47–68.