



Published in final edited form as:

*Autism*. 2023 February ; 27(2): 472–488. doi:10.1177/13623613221103699.

## Theory of mind in naturalistic conversations between autistic and typically developing children and adolescents

Diana Alkire<sup>1,2</sup>, Kathryn A. McNaughton<sup>1,2</sup>, Heather A. Yarger<sup>1,2</sup>, Deena Shariq<sup>1,2</sup>, Elizabeth Redcay<sup>1,2</sup>

<sup>1</sup>Neuroscience and Cognitive Science Program, University of Maryland, College Park

<sup>2</sup>Department of Psychology, University of Maryland, College Park

### Abstract

Successful social interactions are assumed to depend on theory of mind (ToM)—the ability to represent others’ mental states—yet most studies of the relation between ToM and social-interactive success rely on non-interactive tasks that do not adequately capture the *spontaneous* engagement of ToM, a crucial component of everyday social interactions. We addressed this gap by establishing a novel observational rating scale to measure the spontaneous use of ToM (or lack thereof) within naturalistic conversations (*conversational ToM*; *cToM*). In 50 age- and gender-matched dyads of autistic (AUT) and typically developing (TD) youth aged 8–16 (three dyad types: AUT-TD, TD-TD, AUT-AUT), we assessed *cToM* during 5-minute, unstructured conversations. We found that ratings on the *cToM* Negative scale, reflecting ToM-related violations of neurotypical conversational norms, were negatively associated with two forms of non-interactive ToM: visual-affective and spontaneous. In contrast, the *cToM* Positive scale, reflecting explicit mental state language and perspective-taking, was not associated with non-interactive ToM. Furthermore, autistic youth were rated higher than TD youth on *cToM* Negative, but the groups were rated similarly on *cToM* Positive. Together, these findings provide insight into multiple aspects of ToM in conversation and reveal a nuanced picture of the relative strengths and difficulties among autistic youth.

### Introduction

Everyday life is saturated with social interactions, and leveraging these interactions to form high-quality relationships is critical for mental and physical wellbeing (Hedley et al., 2018; Holt-Lunstad et al., 2010). It is widely assumed that social-interactive success depends on theory of mind (ToM)—the ability to represent the mental states of others and oneself—an assumption that drives the influential “mindblindness” account of social impairment in autism (Baron-Cohen et al., 1985). While empirical findings indicate broad, distal relations between ToM and social outcomes (Bosacki & Astington, 1999; Caputi et al., 2017; Devine & Hughes, 2013; Peterson et al., 2016), the specific mechanisms driving these relations are

Correspondence: Diana Alkire, Department of Psychology, University of Maryland, College Park, MD 20742, diana.r.alkire@gmail.com.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

unclear. That is, how does ToM ability or propensity impact real-world behaviors that in turn influence social-interactive success?

Answering this question with a useful degree of specificity is hampered by the discrepancy between standard ways of measuring ToM in the laboratory and the nature of everyday social interactions. Standard ToM tasks range widely in format, but they share a common feature: the subject plays the role of an observer rather than a participant in a social exchange (Begeer et al., 2010; Schilbach et al., 2013). As such, the behaviors directly elicited by observational tasks—labeling mental states, or verbally predicting or explaining a character’s actions, based on passive observation—do not easily map onto the behaviors that typically occur throughout everyday social interactions. Furthermore, most laboratory tasks include explicit prompts to make ToM inferences, for example, asking participants to reason about a fictional character’s motivations in a social vignette (e.g., Happé, 1994). Such prompts rarely occur in everyday interactions; thus, performance on these standard tasks may not correspond with the tendency to represent the mental states of one’s real-world interaction partners.

Reliance on explicit, non-interactive measures is especially problematic when it comes to characterizing ToM abilities in autism. Autistic individuals who pass first-order false belief tasks nevertheless struggle with everyday mindreading (Frith, 1994; Peterson et al., 2009), a finding that has prompted the development of several more advanced ToM measures such as the Strange Stories (Happé, 1994) and the Reading the Mind in the Eyes Test (Baron-Cohen et al., 2001). Yet, even performance on advanced ToM tasks is not consistently related to social impairment in autism (Alkire et al., 2021; Barendse et al., 2018; Usher et al., 2015), suggesting that autistic individuals may solve explicit ToM tasks through compensatory strategies (i.e., verbal or domain-general reasoning) that do not serve them in naturalistic, interactive contexts (Happé, 1995; Scheeren et al., 2013). Furthermore, the unstructured nature of everyday social situations (as opposed to highly structured laboratory tasks) appears to exacerbate ToM difficulties in autism (Ponnet et al., 2008; Roeyers et al., 2001). Thus, to more accurately characterize real-world ToM and its role in interactive success, particularly for autistic individuals, there is a need for more direct measures of the application of ToM within unstructured social interactions.

Naturalistic conversation is an ideal unstructured format in which to measure applied ToM. Conversation becomes an increasingly prominent feature of peer interactions during middle childhood to adolescence, an important period for social development in general (Raffaelli & Duckett, 1989; Rubin et al., 2007). Crucially, successfully navigating a conversation inherently requires ToM (Roth & Spekman, 1984; Sperber & Wilson, 1987). As such, the application of ToM within conversation, hereafter termed *conversational ToM* (cToM), represents a key aspect of interactive social cognition and, potentially, its impairment in autism.

cToM is apparent in a variety of communicative acts. While some of these acts are overt indications of ToM, such as explicitly referring to a conversation partner’s mental states, other speech acts *imply* that the speaker has or has not taken the listener’s perspective into account. Failing to engage in this perspective-taking can result in statements that are under-

or over-informative, irrelevant, ambiguous, or disorganized (Grice, 1975), at least from a neurotypical perspective. Conversely, speakers engage in cToM when they clarify potential ambiguities, distinguish between common and privileged ground (i.e., information that is or is not mutually shared; Clark & Marshall, 1981; Raczaszek-Leonardi et al., 2014), and repair misunderstandings (Schegloff, 2004). These properties of everyday conversation offer a promising avenue for investigating how ToM supports social interaction. However, as we review below, there is a dearth of measures specifically designed to capture cToM.

Behaviors pertaining to cToM are primarily assessed within the discipline of pragmatic language, which concerns the use of language in social contexts and the ability to infer meanings that depart from the literal meaning of an utterance. Pragmatic impairment has been associated with poor social outcomes (Mok et al., 2014; Whitehouse et al., 2009) and is common in autism (Kalandadze et al., 2018; Ying Sng et al., 2018). However, two characteristics of the pragmatics literature limit its applicability to studying ToM in real-time social interactions. First, the broad category of pragmatics includes a range of heterogeneous skills that vary in the extent to which they depend on ToM. Several authors have proposed a theoretical division between aspects that require taking the perspective of one's conversation partner and those that do not, instead depending solely on structural language (i.e., vocabulary and syntax) and contextual reasoning from the listener's egocentric perspective (Andrés-Roqueta & Katsos, 2017; Kissine, 2012; O'Neill, 2012). However, many existing measures of pragmatic ability conflate these heterogeneous aspects (Matthews et al., 2018). A second limitation of previous pragmatics studies is that, much like ToM studies, they often employ non-interactive measures. Because these measures may simply reflect knowledge of social norms (e.g., Carrow-Woolfolk, 1999), and because pragmatic language use is highly context dependent (Ying Sng et al., 2018), standard laboratory tasks may not adequately capture one's true capacity to use ToM within everyday conversations.

Given the limited ecological validity of task-based measures of pragmatic competence, observational measures of natural language samples are better suited to identifying behaviors that reflect real-world ToM. For example, the Pragmatic Rating Scale (PRS; Landa et al., 1992) has been used to characterize pragmatic difficulties in autism that are observed during interactions with a clinician or experimenter (Greenslade et al., 2019; Lam & Yeung, 2012; Paul et al., 2009) and during free-play among preschoolers (Bauminger-Zviely et al., 2013). However, because the PRS and similar rating scales are designed to quantify only pragmatic deficits, they are ill-suited to measuring positive indicators of cToM such as clarifying ambiguities or explicit perspective-taking. Other observational studies have quantified discrete ToM-relevant features of conversation, such as the number of relevant and appropriately informative statements (Capps et al., 1998; Nadig et al., 2010; Tager-Flusberg & Anderson, 1991). Though focused on limited aspects of discourse, these studies establish the feasibility of analyzing naturalistic conversation for behavioral indicators of cToM.

We argued above that non-interactive ToM tasks are limited in their capacity to measure individuals' ability and propensity to apply ToM during real-world interactions. Nevertheless, understanding how performance on established tasks relates to cToM-related pragmatic abilities could shed light on the cognitive underpinnings of these abilities.

ToM is multifaceted, and different non-interactive tasks likely tap into several distinct underlying abilities (Altschuler et al., 2018; Hayward & Homer, 2017; Schaafsma et al., 2015; Warnell & Redcay, 2019). As measuring every facet of ToM is beyond the scope of a single study, here we focus on two aspects that are likely instrumental to conversation. Face-to-face interactions in particular require sensitivity to multimodal cues from one's interaction partner, including facial expressions conveying affective information relevant to the conversation (e.g., interest or confusion). Tasks measuring *visual-affective ToM* (vaToM), the ability to infer emotional states from facial expressions, are widespread in the general ToM literature and show positive correlations with social functioning in autism (Trevisan & Birmingham, 2016), yet their connection with conversational skills is underexplored. vaToM tasks provide participants with a set of emotion words from which to select the most appropriate label for each facial expression. However, as mentioned above, real-world social interaction lacks such explicit prompts to reason about mental states. In contrast, *spontaneous ToM* (sToM), the propensity to attribute mental states in the absence of explicit prompts<sup>1</sup>, is commonly measured using social animation tasks in which participants are asked to describe scenes depicting interactions between geometric shapes; responses are then coded for the degree of spontaneous mental state attribution (Abell et al., 2000; Castelli et al., 2002; Klin, 2000). sToM may be another key mechanism for cToM, allowing one to actively make inferences or predictions about a conversation partner's mental states to guide and adjust one's own communication.

The present study aims to demonstrate the instrumental role of ToM in social interaction through observational coding of unstructured, face-to-face conversations among autistic and typically developing (TD) children and adolescents. Guided by theoretical arguments (e.g., the "double empathy problem"; Milton, 2012) and empirical work suggesting that dyadic interactions differ depending on the characteristics of each member, particularly autism status (Crompton et al., 2020; Morrison et al., 2019), our study includes dyads that were either matched (both TD or both autistic) or mismatched (one TD, one autistic).

Within this ecologically valid context, we established a novel coding framework for cToM that builds on prior work by focusing on ToM-relevant behaviors rather than general pragmatic competence, and by considering both positive and negative indicators of cToM rather than a deficit-only approach (Clark & Adams, 2020; McCrimmon & Montgomery, 2014). After examining differences between autistic and TD participants on cToM, we examined cToM in relation to the two aforementioned facets of ToM (measured with non-interactive tasks) that are especially relevant to conversation: vaToM and sToM. We hypothesized that both vaToM and sToM would show positive associations with the tendency to apply ToM in conversation (cToM Positive) and negative associations with ToM-related violations of conversational norms (cToM Negative) (Hypothesis 1). Finally, in line with the widespread assumption that ToM plays an instrumental role in social interactions,

---

<sup>1</sup>This distinction between explicit and spontaneous *tasks* should not be confused with the distinction between explicit and implicit *cognitive processes* that is common in the ToM literature (e.g., Frith & Frith, 2008). Throughout this paper, we highlight measures (sToM and cToM) that are spontaneous in that they do not explicitly prompt participants to engage in ToM; we make no claim about whether the cognitive processes participants may spontaneously engage are explicit or implicit in the sense of the two-systems account of ToM (Apperly & Butterfill, 2009).

we hypothesized that cToM Positive would positively predict interaction success, as rated by interaction partners, and cToM Negative would negatively predict success (Hypothesis 2).

## Methods

### Participants

The present study was part of a larger ongoing study approved by the University of Maryland's Institutional Review Board (approval number 733144). Participants were recruited largely based on participation in previous studies in our lab. Additional autistic participants were recruited through the Interactive Autism Network, Simons Foundation Powering Autism Research (SPARK), Facebook advertisements, and flyer at local events. We appreciate obtaining access to recruit participants through SPARK research match on SFARI Base. Additional TD participants were recruited through the University of Maryland's Infant and Child Studies database. Autism diagnoses were confirmed using the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2; Lord et al., 2012) administered by a research-reliable clinician. All participants were native English speakers with a full-scale IQ of at least 80 on the Kaufman Brief Intelligence Test, 2nd edition (KBIT-2; Kaufman & Kaufman, 2004). TD participants had no diagnosis of any neurological or psychiatric disorders, or first-degree relatives with autism or schizophrenia. Additional diagnoses of the autistic participants are reported in the Supplementary Materials.

The sample size was based on a power analysis using the Actor–Partner Interdependence Model (APIM) Power Analysis Shiny Application (Ackerman et al., n.d.), which indicated that a sample of 50 dyads provides 80% power to detect effects with standardized beta weights of at least 0.3. See Supplementary Materials for additional details.

Dyad members were matched on gender and age (within 1 year or grade level) and were arranged into three dyad types: AUT-TD (one autistic, one TD participant), TD-TD, and AUT-AUT. The final sample of 50 dyads included 25 TD-TD (8 female dyads), 18 AUT-TD (4 female dyads), and 7 AUT-AUT (1 female dyad)<sup>2</sup>; across dyad types, the sample comprises 68 (20 female) TD and 32 (6 female) AUT individuals, for a total of 100 individuals. Each participant was included in only one dyad. The average age across the whole sample was 13.34 years, with a range of 8.72–16.91 and a standard deviation of 1.85 years. Participants' gender was assessed via parent report: parents were asked to select between the options of “male” or “female” to indicate their child's gender. Race and ethnicity distributions for the whole sample are as follows: 2% Asian, 11% Black or African American, 65% White, 11% more than one race, 11% not reported; 9% Hispanic or Latino, 78% not Hispanic or Latino, 13% not reported. Additional participant demographics, including indicators of socioeconomic status, are reported in Supplementary Table 1.

<sup>2</sup>We had intended to recruit a balanced number of dyads across the three types; however, data collection was halted indefinitely by the COVID-19 pandemic. While we were sufficiently powered to test the hypotheses presented in this paper, the limited sample—particularly of AUT-AUT dyads—precluded us from testing the effect of (or interactions with) dyad type as a main research question.

## Face-to-face Peer Interaction Task

The unstructured conversations took place within a peer interaction paradigm developed by Usher et al. (2015, 2018). After separately providing informed written consent along with their caregivers, participants were seated across from each other in a behavioral testing room in which three video cameras recorded the interaction from multiple angles. The experimenter told participants they would engage in simple activities together for about 20 minutes, then said, “Before I explain your task, why don’t you get to know each other? I’ll be back in about five minutes.” This five-minute “Get to Know You” period was followed by two semi-structured activities. Given that the present study focused on unstructured conversation, only the Get to Know You portion was coded for cToM. Following these activities, participants were placed in separate rooms, where they individually completed questionnaires and additional behavioral tasks.

## cToM Rating Scale

Conversations were rated along two scales, Positive and Negative, based on whether behaviors reflected the presence (Positive) or absence (Negative) of ToM. Categories of behaviors considered for the cToM rating are described in Supplementary Table 3; where applicable, real examples from the observed sample are provided. Among other categories, the Positive scale included explicit references to the partner’s mental state, whereas the Negative scale included violations of neurotypical conversational norms (e.g., over- or under-informative statements). The Positive and Negative scales were 6 points each (0 to 5), with higher numbers representing higher frequencies of behaviors.

The Get to Know You task was coded for cToM by a team of three trained research assistants supervised by the first author. Coders were masked to participant diagnosis except for three participants who self-disclosed their autism diagnosis and one who self-disclosed his lack of autism diagnosis during the conversation task. Furthermore, due to concern that coders may be biased by participants’ physical appearance or mannerisms such that they could speculate on diagnosis or form a negative impression (Sasson et al., 2017), coding was based on extracted audio files rather than video. Prior to coding, conversations were transcribed from the videos using the Computerized Language Analysis (CLAN) program (MacWhinney, 2000). Transcripts assisted coding by indicating meaningful nonverbal gestures (e.g., nodding the head yes or no, shaking hands, etc.), which provided additional information by which raters could interpret the conversation, and by highlighting mental state words, which were taken into consideration for the explicit ToM category. All transcripts were checked for accuracy by a second research assistant or the first author.

Additional details on the coding system, training and coding procedures, and interrater reliability are reported in the Supplementary Materials. Reliability coefficients for the final dataset were as follows: Positive scale, Krippendorff’s alpha = 0.81; Negative scale, Finn’s  $r = 0.94^3$ . Coefficients of 0.7 or above are considered acceptable for both Krippendorff’s alpha and Finn’s  $r$  (Hayes & Krippendorff, 2007; Heyman et al., 2014).

---

<sup>3</sup>Finn’s  $r$  was used to accommodate the skewed distribution of the Negative scale; see Supplementary Materials.

## Control variables

For both hypotheses, analyses included the following variables (in italics) as covariates of no interest. *Verbal IQ*, consistently linked to performance on ToM tasks (Ronald et al., 2006; Scheeren et al., 2013), was measured via the KBIT-2 Verbal Intelligence score. *Executive functioning* (EF), also known to correlate with ToM (Jones et al., 2018; Lecce et al., 2017) and pragmatic abilities (Matthews et al., 2018), was measured via the Global Executive Composite on the Behavior Rating Inventory of Executive Function, 2nd ed. (Gioia et al., 2015). Higher scores on this measure correspond to greater impairment. *Age, gender, and autism diagnosis* are also associated with variability in ToM skills (Bal et al., 2013; Kirkland et al., 2013). Finally, given that cToM scores were derived from speech, we controlled for the amount of speech produced by each individual to ensure that any effects of interest were specific to cToM and not simply driven by how much an individual talked. *Language productivity* was operationalized as the total number of words per minute (Scott & Windsor, 2000) spoken by each child during the 5-minute interaction, as calculated by CLAN. For Hypothesis 1 only, this total was averaged together with the average word count of responses to the sToM task (described below) to create a language productivity composite.

## Hypothesis 1: Non-interactive ToM predicts cToM

After the peer interaction, participants individually completed two non-interactive ToM tasks: vaToM and sToM.

vaToM was assessed using the Face task from the Cambridge Mindreading Face-Voice Battery for Children (Golan et al., 2015). Children viewed videos of actors expressing complex emotions (emotions involving cognitive states) and chose the appropriate label from among four choices. The task consisted of 27 items representing nine emotional concepts: unfriendly, disappointed, embarrassed, jealous, loving, nervous, bothered, amused, and undecided. The measure of interest was accuracy (% correct). See Supplementary Materials for details on stimulus presentation and task administration.

sToM was assessed using the Frith-Happé Triangles task (Abell et al., 2000). Children viewed four short animations in which two triangles move in a manner that suggests a ToM-related interaction. After viewing each video, children were asked to describe the cartoon while their verbal responses were audio-recorded. Responses were coded for the frequency of unique and appropriate internal state attributions (mental states, intentions, and emotions) following guidelines adapted from Rice and Redcay (2015); see Supplementary Materials. Responses were independently coded by two trained research assistants, with 24% of cases double-coded and discussed with the coding supervisor (a third research assistant) during weekly meetings until consensus was reached. Interrater reliability was excellent (Krippendorff's  $\alpha = 0.85$ ). Scores for each of the three items (excluding a practice item) were averaged for an overall sToM score.

**Analysis strategy**—Due to the dyadic data structure and thus the possibility of non-independence among variables, actor-partner interdependence models (APIM; Kenny et al., 2006) were used to estimate the effects of interest (vaToM and sToM) on cToM. The APIM simultaneously estimates actor effects (e.g., the effect of Person 1's vaToM on Person 1's

cToM) and partner effects (e.g., the effect of Person 1's vaToM on Person 2's cToM); see Figure 1. We ran separate models with each of the cToM scales as outcomes. Each model included actor and partner effects of the non-interactive ToM variables, diagnosis (AUT or TD), verbal IQ, EF, and the language productivity composite; and age (averaged between dyad members) and gender as dyad-level covariates. The hypothesized effects were the actor effects of each of the non-interactive ToM variables. Additionally, we ran reduced versions of these models without partner effects of the individual-level predictors. Results and interpretations for the reduced models were similar to the full models and are reported in the Supplementary Materials, along with details on the modeling approach used for each cToM scale. Because the Negative scale had a highly skewed distribution, leading to violations of the assumptions of linear models, for this model we instead applied generalized estimating equations, which make no distributional assumptions and are recommended for estimating APIMs with non-Gaussian outcomes (Loeys et al., 2014; Loeys & Molenberghs, 2013). All APIM analyses (both hypotheses) were conducted using SPSS Version 26; the data and syntax used for these analyses are available via the [Open Science Framework](#).

### **Hypothesis 2: cToM predicts interaction success**

Following the peer interaction, participants individually completed six items from the Social Interaction Evaluation Measure (Berry et al., 1996) assessing interaction quality, that is, how much the participant enjoyed the interaction and would like to interact again (see Supplementary Materials for specific items). The summed total of responses to these items from each participant's partner served as the operational definition of interaction success. We predicted that each cToM scale would have a significant effect on interaction success (positive effect for Positive, negative effect for Negative).

**Analysis strategy**—For each of the cToM scales, an APIM was estimated using linear mixed effects modeling in SPSS, with actor and partner cToM as the predictors and interaction success as the outcome. Covariates included actor and partner effects of verbal IQ, language productivity, and EF, as well as dyad-level age, gender, and dyad type (TD-TD, AUT-TD, AUT-AUT). We included dyad type rather than individual diagnosis as a covariate based on dyadic research suggesting that perceptions may depend on the match or mismatch between partners' diagnosis rather than an individual's diagnosis alone (Morrison et al., 2019).

We also conducted a planned follow-up analysis examining the effect of vaToM and sToM on interaction success (see Supplementary Materials).

### **Community Involvement Statement**

Community members were not involved in the design, implementation, or interpretation of the study.



## Results

### Preliminary Analyses

Descriptive statistics, group comparisons, and zero-order correlations among all variables are shown in Supplementary Tables 4–7. Notably, Positive and Negative scales were uncorrelated ( $r = .00$ ). Figure 2 depicts the distributions of and group differences in each cToM scale. The Negative scale showed a marked group difference in the expected direction based on previous literature, with autistic participants rated higher (meaning they displayed more frequent negative cToM behaviors;  $t(35.14) = 3.25, p = 0.003$ ). In contrast, TD and autistic participants showed very similar distributions on the Positive scale ( $t(59.18) = 0.02, p = 0.98$ ).

We also examined the frequency of each category of behavior within the cToM coding system. Figure 3 depicts these within-category scores and their averages across the full sample and for each group separately. The most frequent category on the Positive scale was explicit ToM, followed by distinguishing common versus privileged ground. The other Positive categories were relatively infrequent, with only a handful of participants scoring 1 or above. Categories on the Negative scale were more comparable in their frequency, with most participants showing no or very few instances across categories (as reflected in the skewed distribution of the overall Negative score).

Autistic and TD participants showed comparable rates across all Positive categories. On the Negative scale, higher rates within each category were represented almost exclusively by a minority of autistic participants. On average, autistic participants were more likely than TD participants to violate maxims of quantity, relevance, and manner and to interrupt their partners in a manner suggesting they were either not attending to their partner's mental state or were attending but struggled to integrate this information into their responses.

The distributions of the other key variables in Hypotheses 1 and 2 are plotted by diagnostic group in Figure 4. Groups did not significantly differ on sToM ( $t(45.83) = -0.22, p = 0.83$ ), but accuracy on the vaToM task was significantly lower in the autistic group ( $t(42.21) = -3.72, p < 0.001$ ). Groups did not significantly differ on interaction quality (participant's own rating;  $t(50.23) = 0.63, p = 0.53$ ) or interaction success (partner rating;  $t(55.51) = 0.56, p = 0.58$ ). Interaction success did not significantly differ across dyad types ( $F(2,97) = 1.07, p = 0.35$ ; AUT-TD vs. TD-TD:  $t(80.65) = -0.44, p = 0.66$ ; AUT-TD vs. AUT-AUT:  $t(17.19) = 1.17, p = 0.26$ ; AUT-AUT vs. TD-TD:  $t(16.94) = 0.95, p = 0.36$ ). The distributions of all additional variables are plotted by group in Supplementary Figure 1.

### Hypothesis 1 Results

For cToM Negative, there was a significant negative actor effect<sup>4</sup> of vaToM ( $B = -0.24, SE = 0.09, p = 0.01$ ) and a significant negative actor effect of sToM ( $B = -0.30, SE = 0.13, p = 0.02$ ), supporting our hypothesis that non-interactive ToM predicts cToM (Figure 5). However, this hypothesis was not supported for cToM Positive, for which there were no

<sup>4</sup>That is, there was a negative effect of an individual's own vaToM on their own cToM Negative score; see Figure 1.

significant actor effects of either vaToM or sToM. Unstandardized regression coefficients for all effects are reported in Table 1.

## Hypothesis 2 Results

To test the hypothesis that cToM predicts interaction success, we estimated linear mixed effects models with each of the cToM scales as predictors. Contrary to our hypothesis, there were no significant effects of either cToM scale predicting interaction success (Supplementary Table 11).

## Discussion

This study introduced the cToM coding system, a novel approach to measuring the use of ToM in social interaction that capitalizes on properties of naturalistic conversation that inherently involve ToM. We found partial support for our hypotheses in that cToM (Negative) was correlated with non-interactive ToM; however, cToM was not related to interactive success. We also demonstrated dissociations between cToM Positive and Negative scales, and this divergence is relevant to our understanding of cToM in autism.

### Divergence between cToM Positive and Negative scales

The Positive and Negative cToM scales—reflecting the presence or absence, respectively, of ToM-related behaviors—were decidedly uncorrelated ( $r = 0.00$ ). Given that they appear to capture orthogonal dimensions, it is not surprising that we found divergent results for the two scales in Hypothesis 1, as well as in the between-group comparison.

Comparing TD and autistic participants on cToM revealed a few noteworthy patterns, especially considering the prominent mindblindness theory of autism. As expected, at the group level the autistic participants were rated significantly higher than TD participants on the Negative scale (reflecting more frequent negative cToM behaviors). Thus, the Negative scale seems to capture ToM-relevant pragmatic difficulties that have been previously documented in the autism literature, particularly violations of Gricean maxims (Landa, 2000; Paul et al., 2009; Ying Sng et al., 2018). At the same time, autistic and TD participants showed highly similar distributions on the Positive scale, and this was also true at the category level (Figure 3). Consistent with previous findings (reviewed in Bang et al., 2013), autistic participants used mental state language (“explicit ToM” in our categorization) to the same extent as TD participants. Furthermore, while explicit ToM was the most frequent category within the Positive scale, the autistic participants’ scores do not solely reflect superficial knowledge of mental state concepts; they were also comparable to TD participants on the second-most frequent category of distinguishing between common and privileged ground. Behaviors coded in this category include preemptively sharing information that would plausibly be unknown to one’s partner, or asking whether one’s partner shares certain knowledge, both of which reflect an active awareness that others can possess different knowledge states from oneself. Thus, our findings converge with other evidence that autistic individuals’ ToM abilities are often underestimated (Heasman & Gillespie, 2018), and that they are aware of common ground and modify their discourse in response (De Marchena & Eigsti, 2016). Together with the lack of correlation between the

Positive and Negative scales, this pattern of between-group difference and similarity implies even those autistic individuals who violate certain neurotypical conversational norms may be comparable to their TD peers on other aspects of cToM. As such, the abilities captured by the cToM Positive scale could be recognized as an area of strength that could potentially be leveraged in social skills interventions (McCrimmon & Montgomery, 2014).

The discrepancy between the Positive and Negative scales also raises the question of how they may differ in their cognitive demands. Focusing on the dominant category of explicit ToM, simply referencing a partner's mental state does not necessarily require switching between two perspectives (self vs. other) beyond the basic recognition that others have their own mental states. Thus, it is possible that the behaviors captured by the Positive scale do not reflect the degree of perspective-taking necessary for scoring low on the Negative scale.

### **cToM Negative relates to non-interactive ToM**

Non-interactive ToM—specifically, visual-affective (vaToM) and spontaneous (sToM)—was negatively associated with cToM Negative but not Positive. The association between vaToM and cToM Negative suggests that children who struggle to identify complex emotions based on facial expressions also tend to violate neurotypical conversational norms, such as providing too little, too much, or irrelevant information, or expressing themselves in a confusing manner. It is easy to imagine how this association might play out in a conversation. For example, a child who does not pick up on her partner's confused expression would be unaware that her utterances come across as irrelevant or unclear (in the absence of verbal feedback to this effect), and thus would not adjust her behavior. In contrast, explicitly referencing a partner's mental state could reflect a general (rather than partner-specific) understanding of diverse mental states, as argued above, and thus might not necessitate “checking in” with a partner to dynamically integrate feedback from their facial expressions—hence the lack of association between vaToM and cToM Positive.

The negative association between sToM and cToM Negative is consistent with the notion that successfully navigating a conversation, and social interaction more broadly, involves spontaneous mental state attribution. Specifically, participants who described social animations in more mentalistic terms were rated lower on cToM Negative, suggesting that these children spontaneously attributed mental states to their partners during the conversation, and that this allowed them to follow neurotypical conversational norms. The lack of association between sToM and cToM Positive is somewhat surprising given that our sToM measure was the frequency of mental state attributions within a verbal description, and cToM Positive scores were largely driven by explicit mental state language. However, it is important to consider that while the sToM task did not explicitly prompt children to attribute mental states, it did require them to give verbal descriptions of the scenes. In contrast, most mental state attributions made during a naturalistic conversation are almost certainly not vocalized. Whereas cToM Negative may be related to the general tendency to make mental state attributions (which may be unvocalized unless prompted), cToM Positive, specifically the explicit ToM category, indexes the subset of mental state attributions that happen to be vocalized during the conversation.

Altogether, these results add to the relatively sparse literature linking diverse aspects of ToM with conversational skills.

### **cToM does not predict partner-reported interaction success**

Contrary to our hypothesis, neither cToM scale was a significant predictor of interaction success, as measured by partner ratings of interaction quality. Two aspects of our methodology must be considered when interpreting these null findings.

First, this study focused on the specific context of a single interaction with an unfamiliar peer. This setup was meant to represent the type of real-world scenario that, if successful, would lead to further interactions and eventually the formation of a long-term relationship. Autistic individuals often find interacting with unfamiliar peers particularly challenging, as they struggle to apply their social knowledge (including skills learned through explicit training) to novel situations (Bauminger-Zviely et al., 2013; Usher et al., 2015). As such, we expected our paradigm to be more effective at eliciting cToM difficulties than would interacting with familiar individuals. However, it is possible that cToM plays a more important role in the perceived quality of interactions between familiar compared to unfamiliar partners. That is, as two people gain more experience with each other, one might expect the other to make more accurate mental state attributions about oneself, and thus failure to do so would be more striking. More generally, while the present study suggests that cToM is not a significant factor in how individuals judge the quality of a single interaction, the full impact of ToM on relationship quality likely plays out over repeated interactions. Future studies could explore this possibility by following dyads longitudinally and tracking their interaction patterns, including cToM, and relationship outcomes over time.

Second, our choice of partner self-report as the measure of interaction success may have limited our ability to detect a true effect of cToM, as social desirability bias may have led some children to underreport any dissatisfaction with the interaction. We do not believe this was a major concern, as there is a decent spread of scores on this measure (Figure 4). Furthermore, in a post hoc exploration of a subset of participants' future partner preferences (i.e., whether they would prefer the same or a different partner if they were to participate in another session; see Supplementary Materials), which they completed at the end of the session while the experimenter was out of the room, the extent to which participants preferred the same partner correlated positively with their earlier report of interaction quality ( $r = 0.42, p < 0.001$ ). This suggests that our measure of interaction success is a reasonable proxy for the participants' motivation to interact with their partners again, a prerequisite for long-term relationship formation. Nevertheless, future studies could incorporate alternative measures of peer acceptance, such as classroom sociometric ratings, that may be stronger indicators of real-world interaction success.

### **General limitations and future directions**

A few additional caveats apply to our discussion of the overall study. First, while our sample of 50 dyads allowed sufficient power to test our hypotheses, a larger sample would likely produce more robust and reproducible results. Furthermore, the relatively low proportion of autistic participants (32% of the total sample) likely limited the variability we observed

in the cToM Negative scale, on which only a handful of autistic participants received high ratings. This in turn may have limited our ability to detect the predicted effects. The generalizability of our findings to the wider autistic population is also limited, as we excluded individuals with IQs below 80 or who are minimally verbal, and our autistic sample was 84% White. It is also possible that our findings would not generalize to age groups outside of the middle childhood to adolescent range. Additionally, while there is growing recognition that autistic females often differ from their male counterparts in their experience and behavioral presentation of autism, including social behaviors (Dean et al., 2017; Mandy & Lai, 2017; Wood-Downie et al., 2021), our sample was predominantly male. Further research with larger and more balanced samples should assess whether gender moderates the relations under study, and whether gender interacts with autism diagnosis to predict cToM or interaction success. We also recognize the limitation that we assessed gender via parent- and not self-report and provided only binary options. In line with growing recognition of a broader spectrum of gender identities, particularly among the autistic population (Corbett et al., 2022; Strang et al., 2020), future studies should include nonbinary and other options for gender and pose this question to the participants themselves.

Future studies should also further explore AUT-AUT interactions, which constituted only a small portion of our sample. Recent work has suggested that AUT-AUT interactions may qualitatively differ from TD-TD or TD-AUT interactions, including features relevant to cToM such as a generous assumption of common ground (Heasman & Gillespie, 2019), and in ways that impact interaction success (Granieri et al., 2020; Morrison et al., 2019). Such findings support the idea of the double empathy problem (Milton, 2012), in which social interaction difficulties are attributed not solely to the autistic individual, but to a breakdown in communication resulting from the different experiences and expectations of autistic and neurotypical people (Crompton et al., 2021). Thus, rather than being a static trait of an individual, cToM may be intrinsically tied to the dyadic context created between an individual and a particular partner, and an important aspect of this context may be the match or mismatch in autism status. As such, future studies of cToM should be sufficiently powered to investigate the effect of dyad type (i.e., match or mismatch) on both cToM and interaction success.

Other limitations relate to the cToM measure itself. Because coding was based on the audio and transcripts of the conversations, we were unable to capture potentially relevant features like facial expressions and eye gaze; future iterations of the cToM system could be expanded to include the visual modality. Furthermore, any measure based on third-party ratings has the limitation of not directly accessing an individual's subjective experience. We encourage future studies using the cToM coding system to collect self-report measures immediately following the conversation to provide complementary information about how people perceive their own use of ToM in conversation. Ratings about an individual's cToM from their conversation partner would be similarly valuable and may be more likely to uncover associations between cToM and partner-rated interaction success.

Further research is also needed to validate and refine the cToM rating scales. Beyond the above-discussed divergence between the Positive and Negative scales, there may be heterogeneity within each scale in terms of the cognitive processes they reflect. Future

research using a larger sample, and perhaps a slightly more structured interactive task that elicits higher rates of the categories that were infrequent in the present study (e.g., misunderstandings, explicit perspective-taking, and irony comprehension), could employ factor analysis to characterize the pattern of correlations among the categories. Uncovering the latent dimensional structure of cToM could enable more precise assessments of its relations with non-interactive ToM, interaction success, and other constructs.

Finally, our cToM coding system was developed without input from autistic perspectives and thus is biased to reflect the norms of the dominant neurotypical culture; these norms may not be as valued by autistic individuals. Future studies could involve autistic individuals in adapting the cToM system to better capture how autistic individuals consider (or do not consider) each other's perspectives while interacting, and how important this is to their perceptions of interaction quality.

## Conclusion

We introduced the cToM coding system, addressing the need for interactive, ecologically valid measures of ToM. While we did not find evidence that this novel measure improves on standard ToM measures in predicting interaction success, the cToM is a useful framework for characterizing the various ways in which ToM-related difficulties show up in naturalistic conversations between TD and autistic individuals. Crucially, the divergence between the Positive and Negative scales reveals the multidimensionality of ToM in conversation. This finding is valuable in two respects. First, it adds to mounting evidence of divergence between distinct components of ToM in general (Schaafsma et al., 2015; Warnell & Redcay, 2019). Second, it refines our understanding of ToM and pragmatic difficulties among autistic individuals, as even those who struggle with ToM-related violations of neurotypical conversational norms can nevertheless display typical levels of other forms of mental state representation, such as those indexed by the cToM Positive scale. Furthermore, our finding that cToM Negative relates to two forms of non-interactive ToM can inform future studies of the more basic processes that support the application of ToM in social interactions. Altogether, the present study provides a springboard for further investigation into the mechanisms and consequences of ToM-related behavior within naturalistic conversations.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank the members of the cToM coding team: Ryan Regars, Aliceann Trostle, Daniel Friedman-Brown, and Ming Yuan; Katie Beverley, Hema Clarence, Diana Grant, Sarah Gray, Alex Kalomiris, Dahye Kang, and Manasvinee Mayil Vahanan for their assistance with transcription and coding; Jacqueline Thomas, Alexandra Hickey, Tina Nguyen, Micah Plotkin, Kathryn Bouvier-Weinberg, Matthew Kiely, Ryan Stadler, Aranje Sripanjalingam, Dominic Smith-DiLeo, Avi Warshawsky, Maddie Reiter, Bess Bloomer, Miranda Sapoznik, Nicole Chapman, and Aiste Cechaviciute for their assistance with data collection; Dr. Lauren Usher for conceiving of the Get to Know You task; and Drs. Peter Carruthers, Jude Cassidy, and Edward Lemay for their advice on study design, methods, and data analysis. Finally, we thank the participants and their families for making this research possible.

## Funding

Research reported in this publication was supported by the National Institute of Mental Health of the National Institutes of Health under Award number R01MH107441. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

## References

- Abell F, Happé F, & Frith U (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, 15(1), 1–16. 10.1016/S0885-2014(00)00014-9
- Ackerman RA, Ledermann T, & Kenny DA (n.d.). Power analysis for the Actor-Partner Interdependence Model. Unpublished manuscript. <https://robert-ackerman.shinyapps.io/APIMPowerR/>
- Alkire D, Warnell KR, Kirby LA, Moraczewski D, & Redcay E (2021). Explaining Variance in Social Symptoms of Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, 51(4), 1249–1265. 10.1007/s10803-020-04598-x [PubMed: 32676827]
- Altschuler M, Sideridis G, Kala S, Warshawsky M, Gilbert R, Carroll D, ... Faja S (2018). Measuring Individual Differences in Cognitive, Affective, and Spontaneous Theory of Mind Among School-Aged Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, 48(11), 3945–3957. 10.1007/s10803-018-3663-1 [PubMed: 29971661]
- Andrés-Roqueta C, & Katsos N (2017). The Contribution of Grammar, Vocabulary and Theory of Mind in Pragmatic Language Competence in Children with Autistic Spectrum Disorders. *Frontiers in Psychology*, 8(JUN), 996. 10.3389/fpsyg.2017.00996 [PubMed: 28663734]
- Apperly IA, & Butterfill SA (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116(4), 953. Retrieved from <http://psycnet.apa.org/psycinfo/2009-18254-013> [PubMed: 19839692]
- Bal E, Yerys BE, Sokoloff JL, Celano MJ, Kenworthy L, Giedd JN, & Wallace GL (2013). Do social attribution skills improve with age in children with high functioning autism spectrum disorders? *Research in Autism Spectrum Disorders*, 7(1), 9–16. 10.1016/j.rasd.2012.07.004 [PubMed: 23130085]
- Bang J, Burns J, & Nadig A (2013). Brief report: Conveying subjective experience in conversation: Production of mental state terms and personal narratives in individuals with high functioning autism. *Journal of Autism and Developmental Disorders*, 43(7), 1732–1740. 10.1007/s10803-012-1716-4 [PubMed: 23179342]
- Barendse EM, Hendriks MPH, Thoonen G, Aldenkamp AP, & Kessels RPC (2018). Social behaviour and social cognition in high-functioning adolescents with autism spectrum disorder (ASD): two sides of the same coin? *Cognitive Processing*, 19(4), 545–555. 10.1007/s10339-018-0866-5 [PubMed: 29959562]
- Baron-Cohen S, Leslie AM, & Frith U (1985). Does the autistic child have a “theory of mind”? *Cognition*, 21(1), 37–46. 10.1016/0010-0277(85)90022-8 [PubMed: 2934210]
- Baron-Cohen S, Wheelwright S, Spong A, Scahill V, & Lawson J (2001). Are intuitive physics and intuitive psychology independent? A test with children with Asperger Syndrome. *Journal of Developmental and Learning Disorders*, 5(1), 47–78.
- Bauminger-Zviely N, Karin E, Kimhi Y, & Agam-Ben-Artzi G (2013). Spontaneous peer conversation in preschoolers with high-functioning autism spectrum disorder versus typical development. *Wiley Online Library*, 55(4), 363–373. 10.1111/jcpp.12158
- Begeer S, Malle BF, Nieuwland MS, & Keysar B (2010). Using theory of mind to represent and take part in social interactions: Comparing individuals with high-functioning autism and typically developing controls. *European Journal of Developmental Psychology*, 7(1), 104–122. 10.1080/17405620903024263
- Berry DS, Hansen JS, & Berry S (1996). Positive Affect, Negative Affect, and Social Interaction. In *Journal of Personality and Social Psychology* (Vol. 71, Issue 4). <https://psycnet.apa.org/record/1996-06401-014>

- Bosacki S, & Astington JW (1999). Theory of Mind in Preadolescence: Relations Between Social Understanding and Social Competence. *Social Development*, 8(2), 237–255. 10.1111/1467-9507.00093
- Capps L, Kehres J, & Sigman M (1998). Conversational abilities among children with autism and children with developmental delays. *Autism*. 10.1177/1362361398024002
- Caputi M, Pantaleo G, & Scaini S (2017). Do Feelings of Loneliness Mediate the Relationship between Sociocognitive Understanding and Depressive Symptoms During Late Childhood and Early Adolescence? *Journal of Genetic Psychology*, 178(4), 207–216. 10.1080/00221325.2017.1317629 [PubMed: 28682188]
- Carrow-Woolfolk E (1999). CASL: Comprehensive assessment of spoken language.
- Castelli F, Frith C, Happé F, & Frith U (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, 125(8), 1839–1849. 10.1093/brain/awf189 [PubMed: 12135974]
- Clark HH, & Marshall CR (1981). Definite reference and mutual knowledge. *Elements of Discourse Understanding*, 10–63.
- Clark M, & Adams D (2020). The self-identified positive attributes and favourite activities of children on the autism spectrum. *Research in Autism Spectrum Disorders*, 72, 101512. 10.1016/J.RASD.2020.101512
- Corbett BA, Muscatello RA, Klemencic ME, West M, Kim A, & Strang JF (2022). Greater gender diversity among autistic children by self-report and parent-report. *Autism*, 136236132210853. 10.1177/13623613221085337
- Crompton CJ, DeBrabander K, Heasman B, Milton D, & Sasson NJ (2021). Double Empathy: Why Autistic People Are Often Misunderstood. *Frontiers for Young Minds*, 9. 10.3389/FRYM.2021.554875
- Crompton CJ, Ropar D, Evans-Williams CVM, Flynn EG, & Fletcher-Watson S (2020). Autistic peer-to-peer information transfer is highly effective. *Autism*. 10.1177/1362361320919286
- De Marchena A, & Eigsti IM (2016). The art of common ground: emergence of a complex pragmatic language skill in adolescents with autism spectrum disorders. *Journal of Child Language*, 43(1), 43–80. 10.1017/S0305000915000070 [PubMed: 25708810]
- Dean M, Harwood R, & Kasari C (2017). The art of camouflage: Gender differences in the social behaviors of girls and boys with autism spectrum disorder. *Autism*, 21(6), 678–689. 10.1177/1362361316671845 [PubMed: 27899709]
- Devine RT, & Hughes C (2013). Silent Films and Strange Stories: Theory of Mind, Gender, and Social Experiences in Middle Childhood. *Child Development*, 84(3), 989–1003. 10.1111/cdev.12017 [PubMed: 23199139]
- Frith U (1994). Autism and theory of mind in everyday life. *Social Development*, 3(2), 108–124. 10.1111/j.1467-9507.1994.tb00031.x
- Frith CD, & Frith U (2008). Implicit and Explicit Processes in Social Cognition. *Neuron*, 60, 503–510. 10.1016/j.neuron.2008.10.032 [PubMed: 18995826]
- Gioia GA, Isquith PK, Guy SC, & Kenworthy L (2015). BRIEF-2: Behavior Rating Inventory of Executive Function. *Psychological Assessment Resources*.
- Golan O, Sinai-Gavrilov Y, & Baron-Cohen S (2015). The Cambridge Mindreading Face-Voice Battery for Children (CAM-C): complex emotion recognition in children with and without autism spectrum conditions. *Molecular Autism*, 6, 22. 10.1186/s13229-015-0018-z [PubMed: 25932320]
- Granieri JE, McNair ML, Gerber AH, Reifler RF, & Lerner MD (2020). Atypical social communication is associated with positive initial impressions among peers with autism spectrum disorder. *Autism*, 24(7), 1841–1848. 10.1177/1362361320924906 [PubMed: 32498545]
- Greenslade KJ, Utter EA, & Landa RJ (2019). Predictors of Pragmatic Communication in School-Age Siblings of Children with ASD and Low-Risk Controls. *Journal of Autism and Developmental Disorders*, 49(4), 1352–1365. 10.1007/s10803-018-3837-x [PubMed: 30488153]
- Grice HP (1975). Logic and conversation. In Cole P & Morgan J (Eds.), *Syntax and Semantics: Speech Acts* (pp. 41–58). Academic Press.



- Happé FGE (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, 24(2), 129–154. 10.1007/BF02172093 [PubMed: 8040158]
- Happé FGE (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. 10.2307/1131954 [PubMed: 7789204]
- Hayes AF, & Krippendorff K (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures*, 1(1), 77–89. 10.1080/19312450709336664
- Hayward EO, & Homer BD (2017). Reliability and validity of advanced theory-of-mind measures in middle childhood and adolescence. *British Journal of Developmental Psychology*. 10.1111/bjdp.12186
- Heasman B, & Gillespie A (2018). Perspective-taking is two-sided: Misunderstandings between people with Asperger's syndrome and their family members. *Autism*, 22(6), 740–750. 10.1177/1362361317708287 [PubMed: 28683569]
- Heasman B, & Gillespie A (2019). Neurodivergent intersubjectivity: Distinctive features of how autistic people create shared understanding. *Autism*, 23(4), 910–921. 10.1177/1362361318785172 [PubMed: 30073872]
- Hedley D, Uljarević M, Wilmot M, Richdale A, & Dissanayake C (2018). Understanding depression and thoughts of self-harm in autism: A potential mechanism involving loneliness. *Research in Autism Spectrum Disorders*, 46, 1–7. 10.1016/j.rasd.2017.11.003
- Heyman R, Lorber M, Eddy J, & West T (2014). Behavioral observation and coding. In Reis HT & Judd CM (Eds.), *Handbook of research methods in social and personality psychology* (pp. 345–372). Cambridge University Press. <https://psycnet.apa.org/record/2014-12227-014>
- Holt-Lunstad J, Smith TB, & Layton JB (2010). Social Relationships and Mortality Risk: A Meta-analytic Review. *PLOS Medicine*, 7(7), e1000316. 10.1371/journal.pmed.1000316 [PubMed: 20668659]
- Jones CRG, Simonoff E, Baird G, Pickles A, Marsden AJS, Tregay J, ... Charman T (2018). The association between theory of mind, executive function, and the symptoms of autism spectrum disorder. *Autism Research*, 11(1). 10.1002/aur.1873
- Kalandadze T, Norbury C, Nærland T, & Næss KAB (2018). Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism*, 22(2), 99–117. 10.1177/1362361316668652 [PubMed: 27899711]
- Kaufman AS, & Kaufman NL (2004). *Kaufman Brief Intelligence Test, Second Edition*. Pearson, Inc. 10.1002/9780470373699.spedc1189
- Kenny D, Kashy D, & Cook W (2006). *Dyadic Data Analysis*. Guilford.
- Kirkland RA, Peterson E, Baker C, & Pulos S (2013). Meta-analysis reveals adult female superiority in "Reading the Mind in the Eyes Test." Article in *North American Journal of Psychology*. Retrieved from <https://www.researchgate.net/publication/260712981>
- Kissine M (2012). Pragmatics, Cognitive Flexibility and Autism Spectrum Disorders. *Mind & Language*, 27(1), 1–28. 10.1111/j.1468-0017.2011.01433.x
- 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, and Allied Disciplines*, 41(7), 831–846. 10.1111/1469-7610.00671 [PubMed: 11079426]
- Lam YG, & Yeung SSS (2012). Towards a convergent account of pragmatic language deficits in children with high-functioning autism: Depicting the phenotype using the Pragmatic Rating Scale. *Research in Autism Spectrum Disorders*, 6(2), 792–797. 10.1016/J.RASD.2011.08.004
- Landa R (2000). Social language use in Asperger syndrome and high-functioning autism In: Klin F, Volkmar, & Sparrow S.. *Asperger syndrome*.
- Landa R, Piven J, Wzorek M, ... J. G-P, & 1992, U. (1992). Social language use in parents of autistic individuals. *Psychological Medicine*, 22, 245–254. <https://www.cambridge.org/core/journals/psychological-medicine/article/social-language-use-in-parents-of-autistic-individuals/BB9E0F4712E0A6B585E510B628F97E47> [PubMed: 1574562]
- Lecce S, Bianco F, Devine RT, & Hughes C (2017). Relations between theory of mind and executive function in middle childhood: A short-term longitudinal study. *Journal of Experimental Child Psychology*, 163. 10.1016/j.jecp.2017.06.011

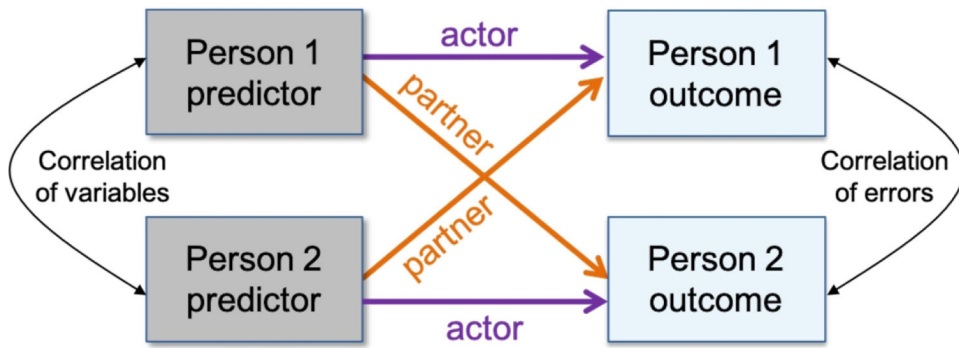
- Loeys T, Cook W, De Smet O, Wietzker A, & Buysse A (2014). The actor-partner interdependence model for categorical dyadic data: A user-friendly guide to GEE. *Personal Relationships*, 21(2), 225–241. 10.1111/pere.12028
- Loeys T, & Molenberghs G (2013). Modeling actor and partner effects in dyadic data when outcomes are categorical. *Psychological Methods*, 18(2), 220–236. 10.1037/a0030640 [PubMed: 23458720]
- Lord C, Rutter M, DiLavore PC, Risi S, Gotham K, & Bishop S (2012). *Autism diagnostic observation schedule* (2nd ed.). Western Psychological Corporation.
- MacWhinney B (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Lawrence Erlbaum Associates.
- Mandy W, & Lai MC (2017). Towards sex- and gender-informed autism research. In *Autism* (Vol. 21, Issue 6, pp. 643–645). SAGE Publications Ltd. 10.1177/1362361317706904 [PubMed: 28749233]
- Matthews D, Biney H, & Abbot-Smith K (2018). Individual Differences in Children’s Pragmatic Ability: A Review of Associations with Formal Language, Social Cognition, and Executive Functions. *Language Learning and Development*, 14(3), 186–223. 10.1080/15475441.2018.1455584
- McCrimmon AW, & Montgomery JM (2014). Resilience-Based Perspectives for Autism Spectrum Disorder. 375–396. 10.1007/978-1-4939-0542-3\_17
- Milton DEM (2012). On the ontological status of autism: The “double empathy problem.” *Disability and Society*, 27(6), 883–887. 10.1080/09687599.2012.710008
- Mok PLH, Pickles A, Durkin K, & Conti-Ramsden G (2014). Longitudinal trajectories of peer relations in children with specific language impairment. *Journal of Child Psychology and Psychiatry*, 55(5), 516–527. 10.1111/jcpp.12190 [PubMed: 24410167]
- Morrison KE, Debrabander KM, Jones DR, Faso DJ, Ackerman RA, & Sasson NJ (2019). Outcomes of real-world social interaction for autistic adults paired with autistic compared to typically developing partners. *Autism*. 10.1177/1362361319892701
- Nadig A, Lee I, Singh L, Bosshart K, & Ozonoff S (2010). How does the topic of conversation affect verbal exchange and eye gaze? A comparison between typical development and high-functioning autism. *Neuropsychologia*, 48(9), 2730–2739. 10.1016/J.NEUROPSYCHOLOGIA.2010.05.020 [PubMed: 20493890]
- O’Neill DK (2012). Components of pragmatic ability and children’s pragmatic language development. In Schmid H (Ed.), *Cognitive Pragmatics* (pp. 261–287). Walter de Gruyter. [https://books.google.com/books?hl=en&lr=&id=thBQGNSQj0C&oi=fnd&pg=PA261&dq=%22Components+of+pragmatic+ability+and+children’s+pragmatic+language+development%22&ots=6TZvVDNi\\_n&sig=SBxUMmnZvh0wGfvYHe3-9uXzRLc](https://books.google.com/books?hl=en&lr=&id=thBQGNSQj0C&oi=fnd&pg=PA261&dq=%22Components+of+pragmatic+ability+and+children’s+pragmatic+language+development%22&ots=6TZvVDNi_n&sig=SBxUMmnZvh0wGfvYHe3-9uXzRLc)
- Paul R, Orlovski SM, Marcinko HC, & Volkmar F (2009). Conversational Behaviors in Youth with High-functioning ASD and Asperger Syndrome. *Journal of Autism and Developmental Disorders*, 39(1), 115–125. 10.1007/s10803-008-0607-1 [PubMed: 18607708]
- Peterson CC, Garnett M, Kelly A, & Attwood T (2009). Everyday social and conversation applications of theory-of-mind understanding by children with autism-spectrum disorders or typical development. *European Child & Adolescent Psychiatry*, 18(2), 105–115. 10.1007/s00787-008-0711-y [PubMed: 18810310]
- Peterson C, Slaughter V, Moore C, & Wellman HM (2016). Peer social skills and theory of mind in children with autism, deafness, or typical development. *Developmental Psychology*, 52(1). 10.1037/a0039833
- Ponnet K, Buysse A, Roeyers H, & De Clercq A (2008). Mind-reading in young adults with ASD: Does structure matter? *Journal of Autism and Developmental Disorders*, 38(5), 905–918. 10.1007/s10803-007-0462-5 [PubMed: 17929156]
- Raczaszek-Leonardi J, Debska A, & Sochanowicz A (2014). Pooling the ground: understanding and coordination in collective sense making. *Frontiers in Psychology*, 5(OCT), 1233. 10.3389/fpsyg.2014.01233 [PubMed: 25426087]
- Raffaelli M, & Duckett E (1989). “We were just talking ...”: Conversations in early adolescence. *Journal of Youth and Adolescence*, 18(6), 567–582. 10.1007/BF02139074 [PubMed: 24272126]

- Rice K, & Redcay E (2015). Spontaneous mentalizing captures variability in the cortical thickness of social brain regions. *Social Cognitive and Affective Neuroscience*, 10(3), 327–334. 10.1093/scan/nsu081 [PubMed: 24847726]
- Roeyers H, Buysse A, Ponnet K, & Pichal B (2001). Advancing Advanced Mind-reading Tests: Empathic Accuracy in Adults with a Pervasive Developmental Disorder. *Journal of Child Psychology and Psychiatry*, 42(2), 271–278. 10.1017/S0021963001006680 [PubMed: 11280423]
- Ronald A, Viding E, Happé F, & Plomin R (2006). Individual differences in theory of mind ability in middle childhood and links with verbal ability and autistic traits: a twin study. *Social Neuroscience*, 1(3–4), 412–425. 10.1080/17470910601068088 [PubMed: 18633802]
- Roth FP, & Spekman NJ (1984). Assessing the pragmatic abilities of children: Part 1. Organizational framework and assessment parameters. In *Journal of Speech and Hearing Disorders* (Vol. 49, Issue 1). 10.1044/jshd.4901.02
- Rubin KH, Bukowski WM, & Parker JG (2007). Peer Interactions, Relationships, and Groups. In Damon W & Lerner RM (Eds.), *Handbook of Child Psychology*. John Wiley & Sons, Inc. <http://doi.wiley.com/10.1002/9780470147658.chpsy0310>
- Sasson NJ, Faso DJ, Nugent J, Lovell S, Kennedy DP, & Grossman RB (2017). Neurotypical Peers are Less Willing to Interact with Those with Autism based on Thin Slice Judgments. *Scientific Reports*, 7(1), 1–10. 10.1038/srep40700 [PubMed: 28127051]
- Schaafsma SM, Pfaff DW, Spunt RP, & Adolphs R (2015). Deconstructing and reconstructing theory of mind. *Trends in Cognitive Sciences*, 19(2), 65–72. 10.1016/j.tics.2014.11.007 [PubMed: 25496670]
- Scheeren AM, De Rosnay M, Koot HM, & Begeer S (2013). Rethinking theory of mind in high-functioning autism spectrum disorder. *Wiley Online Library*, 54(6), 628–635. 10.1111/jcpp.12007
- Schegloff EA (2004). Conversation analysis and socially shared cognition. In *Perspectives on socially shared cognition*. (pp. 150–171). 10.1037/10096-007
- Schilbach L, Timmermans B, Reddy V, Costall A, Bente G, Schlicht T, & Vogeley K (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, 36(04), 393–414. [http://journals.cambridge.org/abstract\\_S0140525X12000660](http://journals.cambridge.org/abstract_S0140525X12000660) [PubMed: 23883742]
- Scott CM, & Windsor J (2000). General Language Performance Measures in Spoken and Written Narrative and Expository Discourse of School-Age Children with Language Learning Disabilities. *Journal of Speech, Language, and Hearing Research*, 43(2), 324–339. 10.1044/jslhr.4302.324
- Sperber D, & Wilson D (1987). *Précis of Relevance: Communication and Cognition*. *Behavioral and Brain Sciences*, 10(04), 697. 10.1017/s0140525x00055345
- Strang JF, van der Miesen AIR, Caplan R, Hughes C, daVanport S, & Lai MC (2020). Both sex- and gender-related factors should be considered in autism research and clinical practice. *Autism*, 24(3), 539–543. 10.1177/1362361320913192 [PubMed: 32299242]
- Tager-Flusberg H, & Anderson M (1991). The Development of Contingent Discourse Ability in Autistic Children. *Journal of Child Psychology and Psychiatry*, 32(7), 1123–1134. 10.1111/j.1469-7610.1991.tb00353.x [PubMed: 1838537]
- Trevisan DA, & 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. 10.1016/j.rasd.2016.08.004
- Usher LV, Burrows CA, Schwartz CB, & Henderson HA (2015). Social competence with an unfamiliar peer in children and adolescents with high functioning autism: Measurement and individual differences. *Research in Autism Spectrum Disorders*, 17, 25–39. 10.1016/j.rasd.2015.05.005 [PubMed: 26161136]
- Usher LV, Burrows CA, Messinger DS, & Henderson HA (2018). Metaperception in Adolescents With and Without Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, 48(2), 533–548. 10.1007/s10803-017-3356-1 [PubMed: 29080926]
- Warnell KR, & Redcay E (2019). Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition*, 191, 103997. 10.1016/J.COGNITION.2019.06.009
- Whitehouse AJO, Watt HJ, Line EA, & Bishop DVM (2009). Adult psychosocial outcomes of children with specific language impairment, pragmatic language impairment and autism. *International*

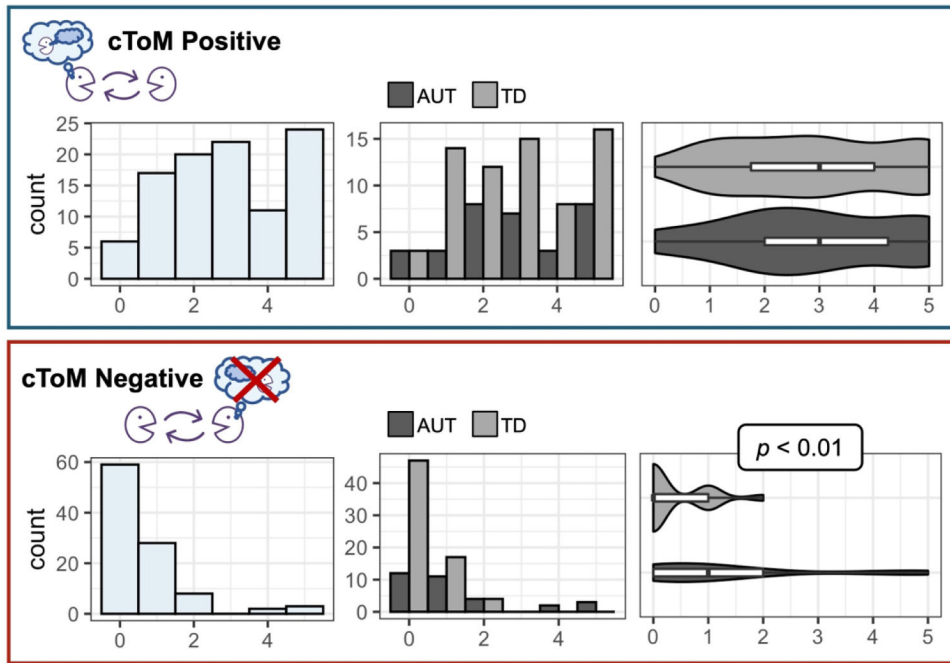
Journal of Language & Communication Disorders, 44(4), 511–528. 10.1080/13682820802708098 [PubMed: 19340628]

Wood-Downie H, Wong B, Kovshoff H, Cortese S, & Hadwin JA (2021). Research Review: A systematic review and meta-analysis of sex/gender differences in social interaction and communication in autistic and nonautistic children and adolescents. *Journal of Child Psychology and Psychiatry*, 62(8), 922–936. 10.1111/JCPP.13337 [PubMed: 33137209]

Ying Sng C, Carter M, & Stephenson J (2018). A systematic review of the comparative pragmatic differences in conversational skills of individuals with autism. *Autism & Developmental Language Impairments*, 3, 239694151880380. 10.1177/2396941518803806



**Figure 1.**  
Schematic of the actor-partner interdependence model.



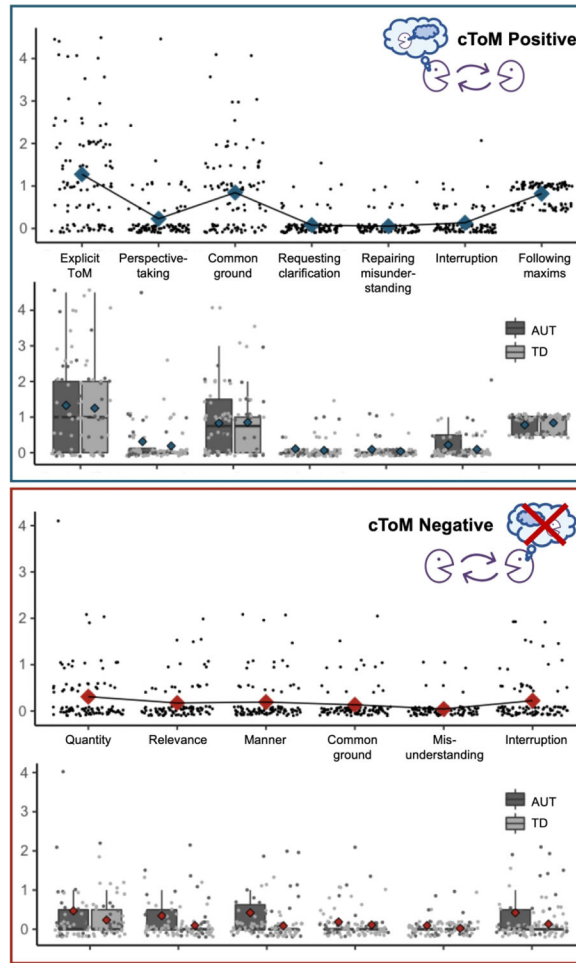
**Figure 2.** Distributions of the cToM scales among the full sample (leftmost graphs) and by diagnostic group. AUT = autistic, TD = typically developing

Author Manuscript

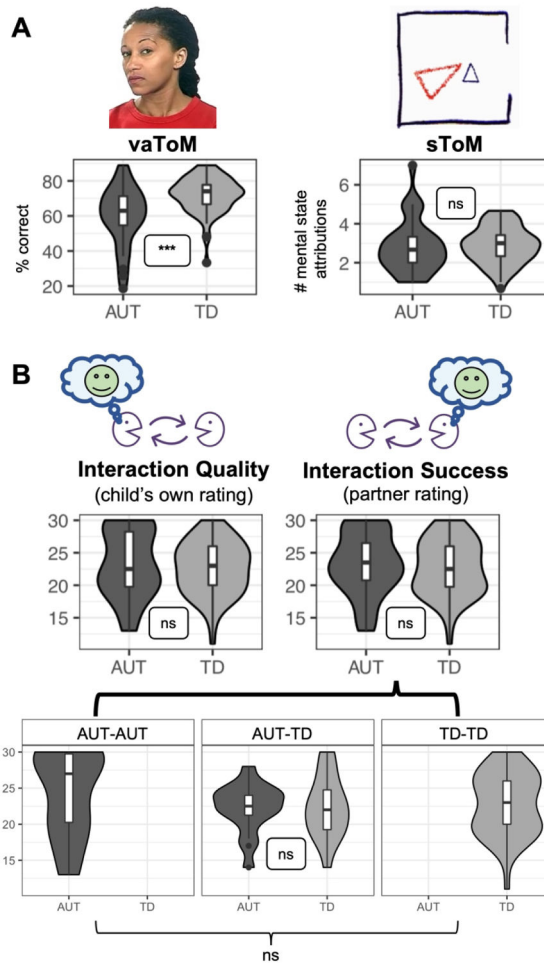
Author Manuscript

Author Manuscript

Author Manuscript

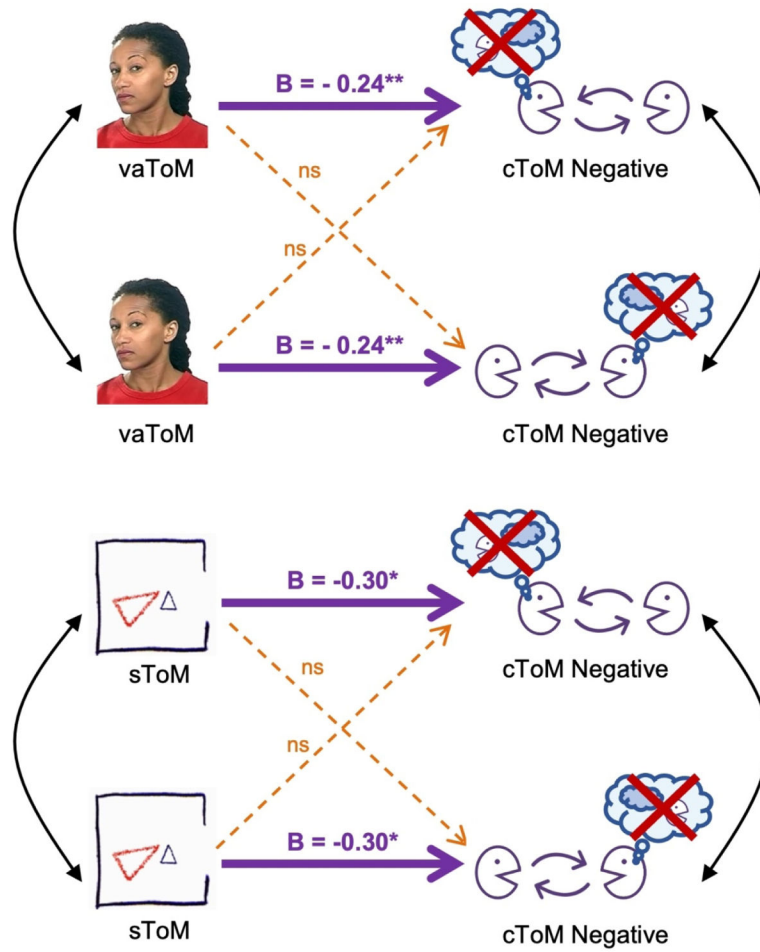


**Figure 3.** Category-level analysis of cToM Positive and Negative scales. Within each scale, sample-wide averages are displayed (with individual data points superimposed) above boxplots comparing the distributions within diagnostic group. On the Positive scale, “Following maxims” refers to consistently following the maxims of quantity, relevance, and manner throughout the conversation; each child was assigned a global score of either 1 (consistent throughout), 0.5 (followed maxims for about half the conversation), or 0 (rarely followed maxims).



**Figure 4.** Distributions plotted by diagnostic group for (A) non-interactive ToM and (B) interaction quality (participant's own rating) and interaction success (partner rating). Interaction success is also plotted by dyad type. vaToM = visual-affective theory of mind. sToM = spontaneous theory of mind. \*\*\*  $p < 0.001$  ns = not significant ( $p > 0.1$ )





**Figure 5.** Hypothesis 1 results for cToM Negative. B = unstandardized beta coefficient. vaToM = visual-affective ToM, sToM = spontaneous ToM. \*  $p < 0.05$  \*\*  $p < 0.01$  ns =  $p > 0.1$

**Table 1.**

Hypothesis 1: Parameter estimates of non-interactive ToM predicting cToM

Effects <sup>a</sup>	B	cToM Negative			cToM Positive				
		SE	95% CI (lower, upper)	P	B	SE	95% CI (lower, upper)	P	
<b>vaToM</b>									
Actor	<b>-0.24**</b>	<b>0.09</b>	<b>-0.42, -0.07</b>	<b>0.01</b>	<b>-0.17</b>	<b>0.15</b>	<b>-0.46, 0.11</b>	<b>0.24</b>	
Partner	0.07	0.12	-0.17, 0.31	0.55	0.03	0.14	-0.25, 0.31	0.84	
<b>sToM</b>									
Actor	<b>-0.30*</b>	<b>0.13</b>	<b>-0.56, -0.04</b>	<b>0.02</b>	<b>-0.01</b>	<b>0.18</b>	<b>-0.35, 0.34</b>	<b>0.98</b>	
Partner	0.02	0.12	-0.27, 0.22	0.84	-0.23	0.18	-0.58, 0.12	0.20	
Verbal IQ									
Actor	-0.01	0.01	-0.03, 0.00	0.13	-0.00	0.01	-0.03, 0.02	0.69	
Partner	-0.01	0.01	-0.03, 0.00	0.10	-0.01	0.01	-0.04, 0.01	0.20	
LangProd									
Actor	<b>0.75**</b>	0.16	0.43, 1.06	0.00	<b>0.64**</b>	0.22	0.20, 1.08	0.00	
Partner	0.03	0.17	-0.31, 0.37	0.86	0.20	0.22	-0.24, 0.64	0.36	
EF									
Actor	0.01	0.01	-0.01, 0.03	0.22	0.00	0.01	-0.01, 0.02	0.61	
Partner	-0.01	0.01	-0.03, 0.01	0.56	0.01	0.01	-0.00, 0.03	0.13	
Dx (AUT)									
Actor	0.28	0.35	-0.40, 0.97	0.42	-0.48	0.54	-1.54, 0.58	0.37	
Partner	0.74	0.55	-0.36, 1.83	0.18	-0.78	0.54	-1.84, 0.28	0.15	
Dyad age	-0.06	0.08	-0.22, 0.10	0.46	<b>0.22#</b>	0.12	-0.02, 0.46	0.07	
Dyad gender (F)	0.38	0.31	-0.22, 0.98	0.22	0.32	0.39	-0.45, 1.09	0.41	

<sup>a</sup>Hypothesized effects in bold\*  $p < 0.05$ \*\*  $p < 0.01$ #  $p < 0.1$ 

B = unstandardized beta coefficient, SE = standard error, CI = confidence interval, LangProd = language productivity composite (see Methods), EF = executive functioning (higher scores correspond to greater impairment), Dx = diagnosis