LANGUAGE ACQUISITION
https://doi.org/10.1080/10489223.2021.1877705

Routledge
Taylor & Francis Group

Check for updates

Verbal mediation of theory of mind in verbal adolescents with autism spectrum disorder

Inge-Marie Eigsti and Christina A. Irvine

Department of Psychological Sciences, University of Connecticut

ABSTRACT

This study tests the role of verbal mediation during theory of mind processing in autism spectrum disorder (ASD). Adolescents with ASD or typical development completed a false belief task while simultaneously performing a verbal or nonverbal load task. There was no group difference in false belief *accuracy*; however, under verbal load, the ASD group was relatively less *efficient*, with slower reaction times, in false belief compared to true belief trials. Faster false belief task performance under verbal but not nonverbal load was associated with pragmatic language ability for the ASD group only. Results were consistent with the theory that there are two (implicit, nonverbal and explicit, verbal) processes that support cognitive reasoning about other people's minds and that people with ASD rely more on the explicit system. Verbal mediation may be critical for false belief understanding in individuals with ASD but not typical development.

ARTICLE HISTORY
Received 7 October 2018
Accepted 12 January 2021

1. Introduction

Inner speech—or the silent, internal, unvoiced expression of thoughts—is said to be the primary medium for representing and relating abstract concepts; it provides a foundation for flexible cognition and self-regulation (Vygotsky 1987). Extensive evidence indicates that inner speech is important in executive control (Barkley 1997); inner speech, and language more generally, provides a critical medium for inhibition, planning, and working memory (Carruthers 2002), all of which are higher-order cognitive functions. Higher-order social functioning, including the ability to attribute intentions and mental representations to others, may similarly rely on inner speech. However, some evidence suggests that this sort of reasoning about others, called theory of mind (ToM) reasoning, may be an intuitive, implicit process that bypasses language. The current study examines the roles of inner speech (verbal mediation) in ToM reasoning in autism spectrum disorder (ASD), a neurodevelopmental condition characterized by early-emerging social and communication deficits (American Psychiatric Association 2013) and often including difficulties in ToM (Baron-Cohen, Leslie & Frith 1985).

Four-year-old children display a remarkable and sophisticated ensemble of linguistic, cognitive, and social skills. They learn new concepts quickly and can communicate and comprehend a wide variety of complex propositions such as "if you are gentle with the kitty, it will come sit in your lap." Alongside this sophistication, however, children at this age struggle to process seemingly straightforward statements such as "Annie thinks that the apple is in the box"; this struggle is most apparent when those statements conflict with the actual state of affairs (e.g., if the apple is not in the box but rather in the basket). False belief tasks require children to contrast another individual's false belief (the apple is in the box) with their own (accurate) belief (it is in the basket). In a typical false belief task, a child hears a story about (for example) a character named Annie who places her apple in one location (e.g., a basket); after Annie leaves the room, a second character moves the apple (e.g., under the bed).

CONTACT Inge-Marie Eigsti inge-marie.eigsti@uconn.edu Department of Psychology, University of Connecticut, 406 Babbidge Road U-1020, Storrs, CT 06250.

Upon Annie's return, the child is asked to indicate where Annie will look for the apple. A report about Annie's *belief* about the apple's location conflicts with the apple's *actual* location; as such, a child's responses provide a clear indication of whether the child understands and is able to report on other people's mental representations.

2. Language and theory of mind

Language has long been proposed as a critical ingredient of ToM and false belief reasoning (Wellman 1990). Deaf children exposed early in life to sign language perform better on false belief tasks than late language learners (de Villiers 2005; Woolfe, Want & Siegal 2002), as do those children whose parents have ASL language skills that are accessible to the children (Schick et al. 2007). The theory of linguistic determinism proposes that the ability to comprehend and produce sentential complements (e.g., "Mom knows that Dad is going to the market") is critical for belief reasoning (de Villiers 2007; Durrleman & Franck 2015). Supporting at least some aspects of this theory, a meta-analysis found a significant relationship between language skills and false belief performance, independent of age, in children younger than age 7, with language ability accounting for a significant 18% of the variance (Milligan, Astington & Dack 2007).

In contrast, several lines of research suggest that language is *not* essential for false belief reasoning. Preverbal infants appear to be influenced by their perception of the goals, perceptions, and beliefs of other individuals in tasks where an actor looks for a hidden toy (Onishi & Baillargeon 2005). Results of a study using a dual-task method (described further in the following) suggest that belief attribution and verbal reasoning are separable in adults (Forgeot d'Arc & Ramus 2011). A study of 30 children with Williams syndrome who were presented with multiple false belief tasks found that performance was predicted by cognitive demands of the tasks rather than by receptive vocabulary or grammatical ability (Van Herwegen, Dimitriou & Rundblad 2013). Children with developmental language disorder show difficulties on false belief tasks requiring an explicit verbal response, suggesting the involvement of an "internal monologue" in belief reasoning (Durrleman, Burnel & Reboul 2017). People with aphasia, who struggle with language production and comprehension, nevertheless show intact ToM (Varley, Siegal & Want 2001; Willems et al. 2011); of course, their language difficulties are, by definition, acquired late in life.

These results, and others, thus suggest either that language *is* or *is not* critical for false belief and ToM reasoning. One possible resolution of these conflicting results comes from a proposal of two distinct systems supporting ToM task performance (Apperly & Butterfill 2009). A fast, inflexible, intuitive, automatic system might be available early in development, permitting people to engage in social-perceptual processing. In contrast, a verbally mediated, deliberate, effortful, and slower system is thought to develop later in life and to mediate social-cognitive processing (Hale & Tager-Flusberg 2003; Tager-Flusberg 2001). Individuals with no access to language (e.g., infants or nonhuman primates) could presumably solve some ToM tasks by utilizing the former system.

It is important to explicitly note the proposed role of language. Language ability, broadly construed via scores on standardized or other assessments, could undergird theory of mind reasoning because knowledge of a particular construct (e.g., sentential complements) provides a scaffold by which individuals develop a way of contrasting a true versus a represented state of reality; in essence, this is the argument of linguistic determinism. Verbal mediation is a related but distinct construct; it can be defined as subvocal or internal speech (in essence, our internal "voice"). While verbal mediation is likely highly dependent on and correlated with linguistic knowledge, the degree to which verbal mediation is required for various kinds of problem solving appears to differ strikingly across individuals and groups. For example, studies of set-shifting (Whitehouse, Maybery & Durkin 2006) and planning (Wallace et al. 2009; Williams, Bowler & Jarrold 2008) tasks show that individuals with ASD appear to rely *less* on verbal mediation compared to typically developing peers. In contrast, other studies indicated ASD-specific differences and increased reliance on verbal mediation (Joseph et al. 2005; Williams, Bowler & Jarrold 2008); these results are revisited in the Discussion. In general, structural language abilities are necessary for verbal mediation, but the latter construct involves additional processes.

3. Language and autism spectrum disorder

Studies of ASD are relevant to our understanding of the centrality of language in ToM reasoning. Communication skills are highly variable in ASD, with approximately 30% of individuals with the diagnosis attaining single-word or two-word-utterance language productions and a productive vocabulary of fewer than 50 words (Tager-Flusberg & Kasari 2013). Even those who develop fully fluent language abilities display significant delays in development and often have lifelong difficulties with aspects of discourse and pragmatic language (Eigsti et al. 2011). As such, ASD studies can provide a means of testing the role of linguistic knowledge in other developmental processes, such as ToM. Given the remarkable heterogeneity of language skills, it is useful to establish the structural and pragmatic language abilities of research participants to provide a meaningful comparison group.

Happé and colleagues reported that whereas typically developing children begin to pass belief-reasoning tasks between ages 3;05 and 4;00 (Wellman 1990), children with ASD typically pass significantly later, at a verbal mental age of 9 years (Happé 1995); this study also indicates that ToM performance is highly correlated with language ability, but not chronological age, in ASD; in contrast, chronological age is a strong predictor in children with typical development or with global developmental delay. Many children with ASD do ultimately solve belief-reasoning tasks (Channon et al. 2001; Kaland et al. 2008). Happé (1995) and Tager-Flusberg et al. (2001) proposed that autistic people rely more on the aforementioned latter processes—the social-cognitive route—solving false belief problems in a "verbally mediated" manner. Multiple studies to date indicate that autistic people rely more heavily on verbal processes during belief reasoning (Durrleman & Franck 2015; Fisher, Happé & Dunn 2005; Hale & Tager-Flusberg 2003; Happé 1995; Lind & Bowler 2009). That is, verbal individuals with ASD may utilize compensatory verbal skills (the aforementioned social-cognitive processes) to represent others' minds (Fisher, Happé & Dunn 2005; Happé 1995; Tager-Flusberg 2001). An illustrative study tested whether language and ToM were more tightly linked in ASD than in individuals with typical development; results indicated that performance on sentential complements explained a significant 30% of the variance in nonverbal ToM task performance in children with ASD with age-appropriate cognitive abilities (Durrleman et al. 2016). Language abilities were measured via scores on experimental and standardized assessments.

These correlational results are suggestive; studies in which the role of language is directly manipulated, however, are needed to directly test the role of language, and specifically online verbal mediation, in false believe reasoning. Intervention studies provide one avenue for investigation (e.g., Durrleman & Delage 2020). A direct test of the role of verbal mediation comes from dual task paradigms, in which the verbal channel is occupied by the articulation of a repeated word or phrase, and leading to reduced processing of verbal information (Murray 1967). Verbal shadowing tasks that utilize overt speech to interfere with covert inner speech have a long history (Sokolov 1972).

In a verbal shadowing study, Newton & de Villiers (2007) used a dual task paradigm to probe whether adults actively rely on language, and specifically on inner speech, while solving false belief problems. Eighty-one participants saw brief movie clips depicting a mouse hiding a piece of cheese; a second character entered the scene and moved the cheese to a new location. The movie clips then provided two possible endings: the mouse searching in (i) the original location and (ii) the new location. Subjects saw one true belief trial, in which the mouse observed the cheese being moved, and one false belief trial, in which the mouse exited the scene before the cheese was moved. While watching the movie clips, participants were asked to perform a secondary (dual) task. In the verbal load condition, they had to verbally shadow a complex narrative, and in the nonverbal load condition, they had to tap on a pair of wooden blocks using a specified rhythm; conditions differed in a between-participant manner. Results indicated that accuracy was high for the true belief condition in both verbal and nonverbal load groups and for the nonverbal load group in the false belief condition; in contrast, participants in the verbal load group were at chance in the false belief condition, suggesting that inner speech is critical for ToM. There were no significant effects of trial order (e.g., whether the false or true belief trial came first). Importantly, the verbal and nonverbal load tasks were not equated for difficulty, meaning that performance differences related to load were confounded by difficulty.

Following up on this result, Forgeot d'Arc & Ramus (2011) employed a similar method but with a larger number of trials (38) and utilizing a within-subjects design. A total of 58 adults with an unremarkable developmental history solved nonverbal false or true belief attribution tasks while performing verbal shadowing or in silence. Results suggested no impact of the verbal load task on either false or true belief attribution. One important constraint on the interpretation of such dual-load paradigms is that they rely on an indirect comparison of verbal versus nonverbal tasks; it is critical to establish that these tasks are comparable in the burden they place on processing, imposing a similar load. To address such questions, it is critical to provide a direct comparison of load tasks using a second task that does *not* rely heavily on inner speech.

4. The current study

Given conflicting results about the role of language in false belief reasoning, from studies utilizing correlational and experimental methods, the current study was designed to test the impact of a verbal dual task on false belief and true belief reasoning in verbally adept individuals with ASD and typical development. We contrasted the impact of a verbal load task, which suppressed inner speech and therefore was expected to impede reliance on a verbal route for solving ToM problems, with the impact of a nonverbal load task of equivalent difficulty, designed to leave inner speech intact. True belief trials were expected to be relatively straightforward for all participants and therefore provided a test of task comprehension and a performance baseline for false belief trials. A greater performance decrement under verbal load as compared to nonverbal load would serve as evidence of verbal mediation for the belief task. If language is critical in processing false belief, and there is only a single "system" (e.g., under the strong version of linguistic determinism), participants across groups would be expected to show a performance decrement in the verbal compared to nonverbal trials. If, in contrast, there are two systems subserving false belief reasoning (the fast, automatic, social-perceptual system, and the slower, more deliberate, social-cognitive system), and if participants with autism rely more on the second system, we would predict a significantly greater performance decrement for the ASD group when solving false belief problems under verbal load.

Executive functioning is frequently impaired in ASD (Eigsti 2011), and dual task paradigms are intrinsically heavily loaded for executive processes (Strobach 2020). Furthermore, ToM processing likely also involves executive processes (e.g., Russell, Saltmarsh & Hill 1999). As such, a measure of executive functioning (specifically, inhibition) was included to enable us to test whether executive processes interacted differently with language and ToM.

5. Study 1: Belief reasoning in a dual-task paradigm

5.1. *Methods*

5.1.1. *Participants*

Participants were 31 adolescents between the ages of 12 and 18 years old, with autism spectrum disorder (ASD; $n = 15$) or typical development (TD; $n = 16$). All participants had full-scale IQ scores falling within or above 1.5 *SD* of the mean (e.g., SS > 85) as measured by the *Stanford-Binet Abbreviated Intelligence Scales—Fifth Edition* (Roid 2003); scores are shown in Table 1. The Stanford-Binet provides an estimate of verbal IQ (VIQ) and nonverbal IQ (NVIQ). Groups differed in VIQ (TD higher than ASD) and NVIQ (ASD higher than TD). Structural language ability was measured using the *Clinical Evaluation of Language Fundamentals* (CELF; Semel, Wiig & Secord 2003) Core Language composite. The groups did not differ significantly on gender, age, or general language ability (CELF score). Participants also completed the

Table 1. Characteristics of participants with autism spectrum disorder (ASD) and typical development (TD).

	ASD	TD	F/χ^2	*p*	Cohen's *d*
n (M:F)	13:2	13:3	0.17	0.68	
Age (y)	14.4 (1.9) 12–17	14.6 (1.7) 12–17	0.35	0.56	0.11
FSIQ	102.4 (9.7) 85–121	102.8 (8.8) 88–124	0.16	.90	0.04
Verbal IQ*	9.7 (2.7) 5–13	11.5 (2.0) 9–16	4.89	0.04	.76
Nonverbal IQ*	11.2 (2.4) 7–15	9.5 (1.9) 6–13	4.68	0.04	.79
Social Communication Questionnaire***	27.2 (13.4) 16–59	2.3 (2.2) 0–7	53.45	0.001	2.64
ADOS Total (Module 3)	10.1 (2.1) 7–14	n/a			
CELF Core Language SS	109 (10) 93–123	114 (9) 93–124	2.35	0.14	0.42
TLC Ambiguous sentences SS	10.1 (3.3) 4–14	10.9 (2.6) 6–16	0.56	0.46	0.28
D-KEFS Color-Word Inhibition SS	10.3 (1.4) 8–13	11.5 (2.5) 5–16	2.87	0.10	0.59

Note. Data are shown as *M*(*SD*), range. SS = Standard Score, *M(SD)* = 100(15); SS = Subtest Score, *M(SD)* = 10(3); FSIQ = Stanford-Binet Abbreviated IQ; ADOS = Autism Diagnostic Observation Schedule, for which higher scores suggest more symptomatology; CELF = Clinical Evaluation of Language Fundamentals; TLC = Test of Language Competence; D-KEFS = Delis-Kaplan Executive Function Scale.

Color-Word Interference subtest from the *Delis-Kaplan Executive Function System* (D-KEFS; Delis, Kaplan & Kramer 2001) as a measure of executive inhibitory control, and the *Test of Language Competence* (Wiig & Secord 1989) *Ambiguous Sentences* subtest as a measure of pragmatic language. In this subtest, which evaluates the ability to comprehend lexically or syntactically ambiguous sentences, examinees hear a spoken sentence with two possible interpretations and must choose an appropriate pictorial representation.

5.1.1.1. ***ASD group.*** Participants met diagnostic criteria for ASD using DSM-IV-TR criteria (American Psychiatric Association 2000), according to scores on Module 3 of the *Autism Diagnostic Observation Schedule* (ADOS; Lord et al. 2002), the parent-report *Social Communication Questionnaire—Lifespan version* (SCQ; Rutter, Bailey & Lord 2003), which measures autism symptomatology across development and clinical judgment. All participants had an ADOS composite score at or above the clinical cutoff of 7 and an SCQ score at or above the clinical cutoff of 15.

5.1.1.2. ***TD group.*** Participants in the TD group did *not* meet criteria for ASD based on clinical judgment or the SCQ and had no first-degree relatives with ASD. To avoid a hypernormative group, TD children were not excluded for other learning or psychiatric disorders (though no participants carried a current diagnosis).

All procedures were approved by University of Connecticut Institutional Review Board. Parents provided written assent, and participants provided written assent for study participation; participants received financial remuneration.

5.1.2. *Low-verbal false belief task*

The belief-reasoning task was designed to test false belief performance under cognitive load using a dual task design. The primary belief reasoning task comprised eight silent 1-minute live-action video trials. In each trial, "Annie" placed an apple in Location A; the "Custodian" entered the room and then moved the apple to Location B. In *true belief* trials, Annie was present and watched while the apple was moved; in *false belief* trials, Annie left the room and did not see the final location of the apple. Each trial ended with two still images, showing (i) Annie searching in Location A, and (ii) Annie searching Location B. For true belief trials, Location B was correct; for false belief trials, Location A was correct. The scene (video setting) changed across trials such that Locations A and B were always trial-unique; however, the characters (Annie and Custodian) were consistent. Participants responded by pressing one of two keys (O and P) with their right middle or index fingers to select the image on the left or right. Colored tactile stickers were placed on the O and P keys to facilitate accurate responding. Participants were told to respond as quickly as possible without making mistakes. Response side was counterbalanced by trial. All videos were trial-unique, with different settings and A/B locations. The task was programmed using SuperLab 5 software on a MacBook Pro. Responses were measured via button press, which recorded both accuracy and RT. Given the simplicity of the videos and the age of the participants, RT was expected to serve as the most sensitive metric of individual differences.

Each participant completed two trials per condition (false belief/verbal load; false belief/nonverbal load; true belief/verbal load; true belief/nonverbal load), for a total of eight trials. Participants were randomly assigned to one of four pseudorandom trial lists with the first trial type balanced across groups.

While performing the primary belief reasoning task, participants simultaneously performed one of two secondary load tasks in time to a metronome ticking at 1-second intervals. In the *verbal load* condition, participants recited the days of the week (Sunday, Monday, Tuesday . . .). In the *nonverbal load* condition, participants used their left index finger to tap a sequence of blocks in a “Z” configuration. The four 1.5” cubes used for the tapping task were affixed to a 4 x 6” foam board in a rectangle. All participants used the finger on the left hand, regardless of handedness, to maintain consistency across subjects, as this was not a skill- or speed-based exercise. Both the verbal and nonverbal tasks were designed to be easy—that is, not imposing a significant cognitive demand. Task instructions for both the belief reasoning task and load tasks were delivered verbally by the experimenter and were also printed on the computer screen.

5.1.3. *Procedure*

Testing was conducted at the participant’s home or at the University of Connecticut during two 3-hour sessions. Participants completed several additional standardized and experimental measures not relevant to the current study.

Prior to the belief task, participants completed task training. First, they received instructions on the load tasks and practiced them in time to the metronome until they were comfortable. Participants also practiced selecting images on the screen using the keyboard response buttons. No participant had difficulty understanding or carrying out these tasks.

To become familiar with the belief task procedures, participants watched a video of a causal event (a person’s hand spilling a glass of milk), followed by an image of a full glass and an image of an empty glass. They responded under verbal load (days of the week) and nonverbal load (block tapping), one practice trial per condition. All participants showed excellent comprehension and were able to respond correctly while performing the load tasks. They then received training on the belief-reasoning task before proceeding, via the following script. All participants expressed comprehension.

> This story is about a girl named Annie. This is Annie. [*Cartoon picture of Annie.*] Today, Annie brought her favorite snack to school—a green apple. [*Picture of apple.*] Annie is looking for a safe place to hide her apple until lunch. But there’s a problem: Whenever Annie hides her apple, the Custodian moves it! [*Cartoon picture of Custodian.*] Sometimes, Annie is in the room when the Custodian moves her apple, but sometimes she’s not. You will see two photos on the screen that show two possible endings to the story. Your job is to pick the ending that shows where Annie *thinks* her apple is.

6. Results

6.1. *Preliminary analyses*

Between-subjects variables were group (ASD, TD) and Trial List (A, B, C, D). Within-subjects conditions were belief (true, false) and load (verbal, spatial). Dependent variables were accuracy (number of correct trials) and RT (ms). Trials with RT of less than 100 ms (Whelan 2010) or greater

Table 2. Performance on the nonverbal belief task as a function of group.

		ASD *M(SD)*	TD *M(SD)*	*F*	*p*	*Cohen's d*
Accuracy (proportion correct)	True belief, verbal load	.74 (.37)	.69 (.44)	.097	.76	.10
	False belief, verbal load	.77 (.32)	.91 (.27)	1.72	.20	.47
	True belief, nonverbal load	.77 (.32)	.85 (.35)	.405	.53	.24
	False belief, nonverbal load	.84 (.31)	.94 (.17)	1.38	.25	.4
RT (ms)	True belief, verbal load	1757 (639)	1365 (415)	3.69	.06	.73
	False belief, verbal load	2224 (1268)	1271 (357)	8.82	.006	1.02
	True belief, nonverbal load	2243 (1030)	1360 (443)	10.12	.003	1.11
	False belief, nonverbal load	2008 (707)	1268 (350)	15.51	<.001	1.33

Note. Data are presented as nontransformed (raw) mean scores.

than 2.5 *SD* above the mean (Balota & Spieler 1999) were removed (nine trials total). There were no main effects of trial list (Lists A, B, C, and D represented pseudorandomly ordered presentations of items) for either RT or accuracy,[1] so these groups were collapsed in subsequent analyses. Accuracy data met the assumption of homogeneity of variance but were negatively skewed and kurtotic (Shapiro-Wilk *p*s < .05); these data could not be normalized using standard transformations. Generalized estimating equation (GEE) analyses, which are robust to unmeasured dependence between outcomes, were thus used to probe effects of group, load, and belief with VIQ and NVIQ scores as covariates. The RT data were normally distributed but did not show homogeneity of variance between groups (Levene statistic $p < .05$); a square-root transformation was applied, after which the assumption was met. Repeated-measure ANOVAs were used to test for group differences in RT, calculated separately for Verbal and Nonverbal conditions. Effect sizes were calculated with Cohen's *d* (Cohen, 1988). To examine within-group correlations of RT with standardized clinical assessments, Pearson correlations were used (conventions for effect size, *r*: small = .10; medium = .30; large = .50).

6.2. *Accuracy*

We examined effects of group, load, and belief in a first-order autoregressive (AR1) correlation matrix using GEE. Task accuracy was high overall: *M* (*SD*) = 6.48 (1.48) out of 8, with a range of 4–8 (chance = 4). Analyses revealed no main effects of group, belief, or load and no two-way or three-way interactions (all *p*'s >.50); see Table 2. These findings suggest that neither belief nor load conditions impacted the ability of participants in either group to accurately choose the correct location and that participants across groups were similarly attentive and engaged with the task. Accuracy data are shown in Table 2.

6.3. *Reaction time*

Performance was measured via RT, a metric with a "rich tradition" (see also Farmer, Misyak & Christiansen 2012; Martin & McElree 2018:770) of being used to examine group differences in psycholinguistic tasks. In the Verbal load condition results showed a significant main effect of group, $F(1,29) = 7.85$, $p = .009$, $d = .91$, such that ASD participants were significantly slower than TD participants with a large effect size. There was no main effect of belief condition, $F(1,29) = 1.28$, $p = .27$, $d = .16$. There was a trend for a significant interaction of group X belief condition, $F(1,29) = 4.34$, $p = .05$, $d = .82$, such that the ASD group was slower in False belief compared to True belief trials, whereas the TD group showed the opposite tendency, performing faster in the False belief compared to the True belief trials. Paired-sample *t*-tests provided weak support for this finding; in the

[1]There was a significant three-way interaction of trial list x group x load, $\chi^2(1) = 7.03$, $p < .01$. Post hoc analyses revealed that for List A the difference between spatial and verbal load trials in the TD group was significantly different than the difference between spatial and verbal load for ASD participants, and for List D the TD group had greater accuracy than the ASD group, $F(1,12) = 9.55$, $p < .05$, $\eta_p^2 = .66$. Because effects were nonsystematic across orders, these results were thought to reflect idiosyncratic effects of small trial numbers, and order was collapsed for subsequent analyses; more information is available on request.

ASD group, RT for the False and True belief trials in the Verbal condition differed at the trend level, $t(14) = 1.717$, $p = .099$; RT for the False and True belief trials in the Nonverbal condition did not differ, $t(14) = .865$, $p = .402$. In the TD group, RT did not differ in either set of conditions, Verbal: $t(14) = 1.036$, $p = .32$, Nonverbal: $t(14) = 0.738$, $p = .47$.

In the Nonverbal load condition, results showed again a significant main effect of group, $F(1,29) = 16.27$, $p < .001$, $d = 2.44$, such that ASD participants were significantly slower than TD participants with a large effect size. There was no main effect of belief condition, $F(1,29) = 1.31$, $p = .26$, $d = .10$, and no interaction of group X belief condition, $F(1,29) = 0.097$, $p = .76$, $d = .10$. These findings are depicted in Figure 1. This finding is consistent with the proposal that individuals with ASD rely more on social-cognitive verbal processes to solve false belief ToM problems, as results showed a slowing in false belief trials compared to true belief trials only for the ASD group and only when the verbal channel was occupied with a concurrent verbal task.

The repetition across trials (the presence of Annie and the Custodian, the search for an apple) may have detracted from the salience of ToM in solving the task, as participants may have been able to "get into set" after one or two belief task trials. Thus, as a check on the primary analyses, we examined performance on the first trial alone (recall that the condition of the first trial was balanced across four orders). Note that each cell in the analysis contained only four participants; results should therefore be interpreted with caution due to the risk of Type II error. Between-subjects ANOVAs were used to examine Trial 1 Accuracy and RT as a function of trial condition (belief/load) and group. There were no main or interaction effects, all *p*s > .10.

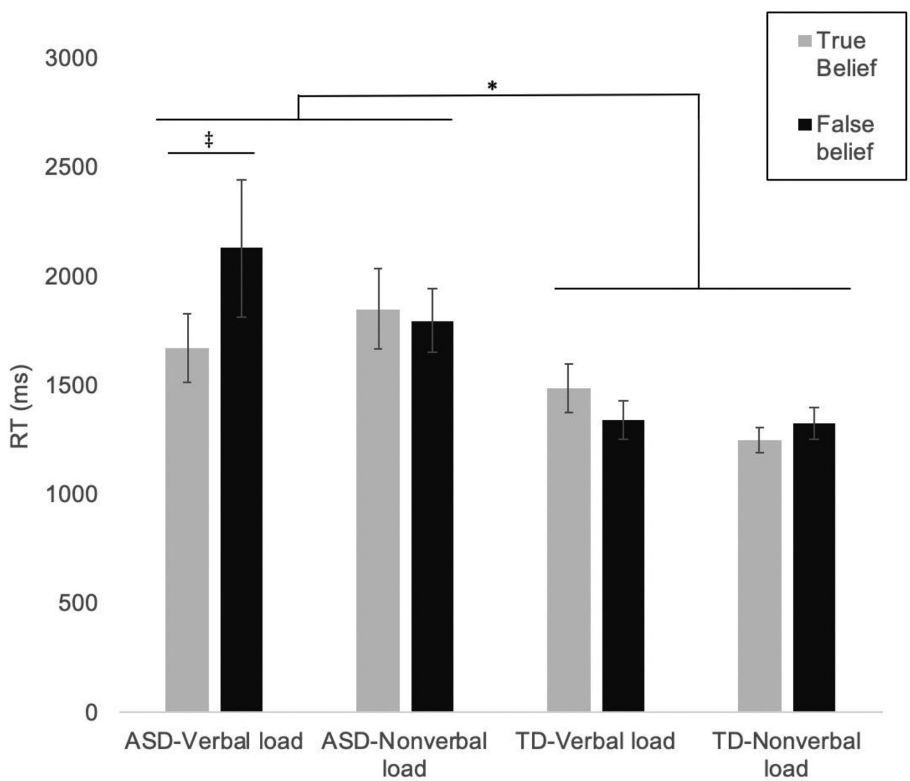

Figure 1. RT by group as a function of belief and load conditions.

6.4. *Correlational analyses: Structural and pragmatic language and inhibitory control*

A final analysis examined the correlates of false belief task performance under verbal and load. Planned within-group analyses tested CELF Core Language (as a measure of structural language knowledge), TLC Ambiguous Sentences (as a measure of pragmatic language knowledge), and Color-Word Inhibition (as a measure of inhibitory executive control) as correlates of true and false belief RT. We also examined whether RT correlated with ASD symptom severity in either group, indexed using total ADOS scores in the ASD group and SCQ scores in the TD group. Results are shown in Table 3. In the ASD group, False belief, Verbal load was correlated with TLC Ambiguous Sentences, $r(14) = -.542$, $p = .045$; higher pragmatic scores were associated with faster RT. The correlation of TLC Ambiguous Sentences with False belief, Nonverbal load missed significance, $r(14) = -.463$, $p = .09$, as did the correlation with CELF Core Language, $r(14) = -.470$, $p = .09$. False belief RT in the Verbal load condition was correlated with False belief RT in the Nonverbal load condition for both groups; ASD, $r(14) = .597$, $p = .02$; TD, $r(14) = .635$, $p = .007$. In the TD group, there were no other significant associations, all *p*s > .25. These findings suggest that within the ASD group, faster, more efficient false belief task performance under verbal but not nonverbal load was associated with pragmatic language skills.

6.5. *Interim summary*

Results of a low-verbal false belief dual task paradigm revealed no *general* false belief deficit in teens with ASD with cognitive abilities in the average range compared to their typically developing peers. The lack of differences in accuracy is unsurprising given the straightforward nature of the task. RT measures were expected to provide a more sensitive index of individual differences in online belief reasoning. The ASD group was generally slower across all trials; however, under verbal load, participants with ASD were significantly slower in the false belief condition.

These results are consistent with two possibilities. First, individuals with ASD may typically rely more on explicit, verbal, social-cognitive processes when solving false belief problems (e.g., when beliefs conflict with reality) such that performance is impeded when that verbal channel is "occupied." In contrast, their TD peers may utilize fast, implicit, nonverbal social-perceptual processes such that even a verbal task load imposes no particular constraint on performance. Correlational findings were consistent with this hypothesis; performance on false belief trials under verbal load was associated with pragmatic language skills in the ASD group only (it was also weakly associated with structural

Table 3. False belief RT under verbal and nonverbal load: Correlations with structural and pragmatic language measures, and inhibitory control.

	False belief RT, Verbal load	False belief RT, Nonverbal load	TLC ambiguous sentences	CELF core language	D-KEFS Stroop-like inhibition	ADOS (ASD)
False belief RT, Verbal load		**.597***	**−.542***	**−0.470±**	−0.161	0.095
False belief RT, Nonverbal load	**.635****		**−0.463±**	−0.442	−0.157	0.237
TLC ambiguous sentences	−0.166	0.292		**.577***	−0.152	−0.033
CELF core language	−0.197	−0.037	0.205		−0.026	**−.610***
D-KEFS Stroop-like inhibition	−0.042	−0.16	−0.194	−0.299		−0.225
SCQ (TD)	0.026	−0.122	−0.299	−0.179	−0.382	

*$p < .05$ (two-tailed); **$p < .01$ (two-tailed); ± <.10.

Correlations are presented *above* the diagonal for the ASD group and *below* for the TD group. Standardized measures are presented as standard scores. TLC = Test of Language Competence; CELF = Clinical Evaluation of Language Fundamentals; D-KEFS = Delis-Kaplan Executive Function Scale; ADOS = Autism Diagnostic Observation Schedule (administered in the ASD Group); SCQ = Social Communication Questionnaire (administered in the TD group).

language skills); false belief performance under nonverbal load was also weakly associated with pragmatic language skills.

One potential confound in interpreting these results is that the verbal and nonverbal load tasks were differentially demanding. The verbal task may have imposed a greater processing demand; if so, and if the ASD group were slightly less efficient in solving false belief problems in general, then one would predict a specific decrement in that condition. Study 2 was designed to compare the difficulty of the verbal and nonverbal load tasks that were utilized in Study 1.

7. Study 2: Comparison of verbal and nonverbal load tasks

7.1. *Participants*

Thirty-two undergraduates at the University of Connecticut participated for course credit. All were native English speakers without significant visual or auditory impairment.

7.2. *Perceptual matching task*

To compare the verbal and nonverbal load, we utilized a dual-task methodology (Hegarty, Shah & Miyake 2000) in which participants solved a series of visuospatial puzzles while also performing the secondary load tasks utilized in Study 1: reciting the days of the week (verbal load) or tapping blocks in an ordered pattern (nonverbal load). All participants completed both load conditions.

The primary task was adapted from the Identical Pictures test in the *Kit of Factor-Referenced Cognitive Tests* (Ekstrom et al. 1976). Participants were asked to match a target figure (presented on the left side) to one of four figures on the right as quickly and as accurately as possible by pressing one of four buttons corresponding with the four fingers of the right hand. The task was programmed in SuperLab, and there were 50 trials total. This task is thought to require minimal verbal mediation and executive function (Hegarty, Shah & Miyake 2000), but it was challenging and required cognitive effort. Participants simultaneously performed the secondary verbal or visuospatial task. Unlike in Study 1, participants performed the dual tasks continuously rather than starting and stopping for each trial.

Trials were blocked by secondary task (Verbal or Nonverbal), with 25 trials per block and a short break between blocks. Participants were randomly assigned to one of two counterbalanced orders (Verbal-Nonverbal; Nonverbal-Verbal). Responses were made on a Cedrus RB-740 response pad, which has four horizontally arrayed response buttons, similar to the array of presented multiple choice items. The experiment was conducted in a quiet room and lasted 20 minutes. Participants completed a brief training to familiarize them with the verbal and nonverbal load tasks, which were practiced in time to a metronome until they were comfortable. They also practiced pressing the response buttons. Before starting the experiment, the participants completed several practice trials of the perceptual matching task, first without a secondary load task and then with each load task. They were instructed to respond to the perceptual matching task as quickly as possible without making mistakes.

8. Results

8.1. *Preliminary analyses*

Dependent variables (accuracy and RT) were checked for outliers and missing data. As in Study 1, outliers, defined as RT of 2.5 *SD* from the mean and RTs faster than 100 ms, were removed from the data. The accuracy data were nonnormally distributed with negative skew (Shapiro Wilk $p < .05$); a reflected logarithmic transformation was applied. RT data were skewed and kurtotic (Shapiro-Wilk $ps < .05$); a logarithmic transformation was applied. After transformations, both measures met statistical assumptions. Two-way repeated-measures mixed-model ANOVAs (load x order) were

Table 4. Perceptual matching task performance as a function of secondary load condition (verbal versus nonverbal).

	Verbal block	Nonverbal block
Overall accuracy (proportion)	.88 (.07)	.84 (.12)
Accuracy: Verbal first, Nonverbal second	.86 (.08)	.84 (.15)
Accuracy: Nonverbal first, Verbal second	.91 (.06)	.85 (.09)
Overall RT (ms)*	1794 (357)	1926 (440)
RT: Verbal first, Nonverbal second	1771 (300)	1738 (317)
RT: Nonverbal first, Verbal second*	1820 (421)	2140 (471)

Note. Data are presented as nontransformed (raw) mean scores.

used to test for main and interaction effects of accuracy and RT. Data are presented as geometric means (GM), calculated as the back-transformed value of the log, using the product of their values (Table 4).

8.2. *Accuracy*

Mean task accuracy was high (.90, with a range of .68 to 1.00). Analyses revealed no main effect of load, $F(1,30) = 2.59$, $p = .19$, $d = .42$, or order, $F(1,30) = 1.84$, $p = .19$, $d = .49$, and no load X order interaction, $F(1,30) = .78$, $p = .38$, $d = .63$, suggesting that neither the load task nor the task order affected accuracy on the figure matching task.

8.3. *Reaction time*

We next examined effects of load and order on RT, again using two-way repeated-measures ANOVAs. Analyses revealed a main effect of load, $F(1,30) = 6.48$, $p = .04$, $d = .31$, with slower RT under nonverbal load ($GM = 1880$; 95% CI [1734, 2038]) versus verbal load ($GM = 1759$; 95% CI [1635, 1892]). This main effect was qualified by a load x order interaction, $F(1,30) = 10.92$, $p = .005$, such that participants in the Nonverbal-Verbal order were slower under nonverbal load as compared to verbal load, $t(31) = -4.23$, $p < .001$, $d = .72$. There was no load difference in the Verbal-Nonverbal order, $t(30) = .21$, $p = .84$, $d = .12$. These results suggest that participants in the Verbal-Nonverbal order may have benefitted from practice, which in turn suggests that the verbal load task was not more difficult; if anything, it may have been *less* demanding of cognitive and attention resources than the nonverbal load task.

9. Discussion

The current study examined the role of inner speech on false belief performance in 12–17-year-old adolescents with high-functioning ASD and TD. Participants responded to nonverbal true and false belief videos while performing either a verbal task designed to suppress inner speech or a nonverbal tapping task. A second study compared the cognitive load imposed by the verbal and nonverbal tasks.

9.1. *Belief reasoning results*

Our findings revealed no group differences in accuracy for a theory of mind task. Specifically, there was no impairment in the ability to predict where someone might believe an object was, even when that belief was incorrect, in adolescents with ASD with cognitive abilities in the average range. This finding is consistent with a number of other studies suggesting that when individuals with ASD reach a verbal mental age of approximately 9 years, they show mastery of false belief reasoning (Happé 1995).

Reaction time analyses were expected to provide a more sensitive index of individual differences in online task performance, under the assumption that slower responding indicates more effortful execution. These analyses revealed that overall, participants with ASD were significantly slower than the TD comparison group. This suggests that although participants with ASD generally chose the correct ending to the silent videos, they were slower to do so across conditions. This could reflect broad motor slowing; it could indicate that this ToM task was generally more effortful; finally, it could also indicate that performing a secondary load task, regardless of modality, imposed a greater relative burden on the ASD group. We are inclined to find the third explanation most plausible, given the literature on dual task performance in ASD (e.g., Garcia-Villamisar & Della Sala 2002), but this question cannot be addressed within the current study.

Analyses of RT further indicated a group by belief condition interaction in the Verbal load condition only; participants with ASD were slower for false belief relative to true belief trials, whereas participants with TD showed the opposite pattern. The within-group comparison of RTs across belief conditions provided some support for this analysis, showing a trend toward faster RT in the true belief condition in the ASD group only, although this trend did not reach significance. Overall, this result suggested that participants in the ASD group struggled more with false belief reasoning compared to true belief reasoning.

A correlational RT analysis for both Verbal and Nonverbal load conditions indicated, first, that RT across those conditions was correlated; this was fully expected given that the structure of the task was identical across conditions. Analyses were conducted separately for ASD and TD groups. Results indicated that more efficient (faster RT) on False belief, Verbal load trials was associated with better pragmatic language abilities. Associations with a structural language measure, CELF Core Language, approached but did not reach significance. In addition, pragmatic language abilities were associated with False belief, Nonverbal load RT at a trend level. Associations between belief reasoning RT and external measures of language ability in the TD group were not significant; in this group, RT for Verbal and Nonverbal conditions was the sole significant correlation. This result suggests that language abilities, assessed via external standardized measures, were correlated with online RT measures of belief reasoning in youth with ASD but not TD and that this association was most apparent for false belief reasoning when the verbal channel was otherwise occupied (e.g., under verbal load conditions).

9.2. *Comparing load task difficulty*

Given that group differences interacted with belief and load, it was important to test whether the load tasks were similarly difficult. Study 2 compared the Verbal and Nonverbal load tasks using a dual-task methodology in a sample of 32 college undergraduates. Findings indicated generally slower RT for the Nonverbal load condition; further examination of this difference indicated that it held only for the group that completed the Nonverbal trials first with *no* Verbal/Nonverbal RT difference for those participants who completed the Verbal trials first. This pattern of results suggests that the Verbal and Nonverbal conditions exerted roughly comparable demands on cognitive resources, at least in the absence of learning or practice effects. If there was a difference in difficulty, it appears that it was the Nonverbal load task that was more demanding than the Verbal task.

The current study is the first to examine true and false belief performance under verbal and nonverbal load conditions in ASD. The findings are consistent with correlational reports indicating that performance in multiple ToM tasks is associated with language abilities in ASD (Durrleman et al. 2016; Durrleman & Franck 2015) and that performance of false belief tasks specifically is correlated with sentential complements specifically (Durrleman, Hinzen & Franck 2018). These results are consistent with the hypothesis that language processing and TOM are more closely linked in ASD compared to typical development.

9.3. *True versus false belief task performance*

ToM is often assessed using false belief paradigms; only false belief presents a contrast between the (incorrect) belief and the true state of affairs such that false belief tasks present the optimal means of evaluating belief reasoning. Though false belief trials are seen as the "true test" of ToM, as they do not confound reality with a person's internal representations, true belief trials also require ToM (see Forgeot d'Arc & Ramus 2011 for a cogent discussion). This framework is consistent with findings from a verbal shadowing task, reported by Forgeot d'Arc & Ramus (2011). As discussed, their results indicated that verbal interference impacted performance, with no differences in true and false belief conditions; this is also the finding in the current study. These findings are inconsistent with those reported by Newton & de Villiers (2007), who found that verbal load specifically disrupted false belief performance (42% correct) while sparing true belief performance (91% correct) in typical adults. That said, Newton & de Villiers's results were confounded by lack of clear comparability in task difficulty (the verbal task may have been more difficult than the nonverbal task).

9.4. *Executive demands of false belief reasoning*

False belief reasoning likely exerts a slightly greater demand on executive processing because one must presumably inhibit one's knowledge of the true state of affairs in acknowledging the incorrect belief state. As such, the current results could speak to the involvement of executive control processes in balancing conflicting ideas as much as to differences in ToM in autism per se. The current study does not resolve this possibility; belief task RT was uncorrelated with performance on one executive measure (of inhibition) in either group, consistent with some prior research (Durrleman & Franck 2015). However, the sample size was small, and we did not comprehensively assess multiple types of executive processes; for example, the belief reasoning task (tracking the location of Annie and the apple, maintaining dual task activity) may have loaded more onto working memory than on inhibitory control. Additional research is needed to more fully examine the relationships among language, EF, and belief reasoning in ASD.

More broadly, the present findings may highlight difficulties in the *application* as opposed to the *representation* of false belief. As argued by Bloom & German (2000), selecting the correct response on a false belief task requires an individual to override useful heuristics about the world. Successful false belief task performance may be related to the ability to inhibit a "default" (true) response in favor of the correct (false) response (Leslie, Friedman & German 2004). In Newton & de Villiers (2007), participants completed a single trial per false/true belief condition. The current study included two true belief and two false belief trials, providing a more robust estimate of individual differences in performance (though the difference between one and two trials is likely small). Perhaps more importantly, participants completed multiple belief trials in succession in a within-subjects design. Though each video was trial-unique, each required the participant to attend to whether Annie was present when the Custodian moved the apple. Debriefing responses suggested that many participants recognized this consistency and used it to facilitate their response. That is, participants across groups reported that, after exposure to multiple trials, they learned to memorize the original location of the apple when Annie left the room (the signature of the false belief trials), knowing that it would be the correct answer. This heuristic may have reduced the impact of verbal interference on task performance for those participants who drew on this heuristic as a "short cut" to belief reasoning.

The verbal and nonverbal load tasks also differed across dual load studies. In the current experiment, participants said a sequence of words or tapped a sequence of blocks in time to a metronome. In contrast, in Newton & de Villiers (2007) participants either shadowed a complex verbal narrative or heard a tapping sequence that they repeated by hitting a block with a mallet. Pilot testing with these load tasks (generously shared with us by the original authors) indicated that participants were simply unable to perform the verbal shadowing task while attending to the false belief videos; they reported (and their behavior indicated) that verbal shadowing task was subjectively significantly more difficult. Apparently, the cognitive demand of the load tasks differed between conditions.

9.5. *Explicit versus implicit ToM representations*

Both observational and experimental ToM work suggests that although children younger than age 4 consistently fail traditional false belief tasks, they *do* have the capacity to represent the minds of others. Young children are capable of attributing goals to other agents (Gergely et al. 1995), engaging in pretend play and understanding pretense (Leslie 1994), and imitating the intended (but not accidental) actions of others (Meltzoff 1995). O'Neill (1996) found that 2-year-old children were more likely to reference a toy and point to its location when their parent was aware of the presence of the toy (i.e., witnessed the toy being placed on the shelf) than when the parent was unaware. These toddlers altered their behavior according to the knowledge states of others. Studies of ToM in 13-month (Surian, Caldi & Sperber 2007) and 15-month (Onishi & Baillargeon, 2005) infants report that they looked significantly longer on trials in which an agent searched for an object in a location that was correct but inconsistent with the agent's knowledge (Surian, Caldi & Sperber 2007). Thus, infants as young as 13 months account for others' beliefs when predicting behavior, suggesting that infants represent at least some aspects of ToM almost *two years* before passing traditional false belief tasks.

The results reported here support the theory of parallel ToM systems: an "implicit" system, which is rapid and efficient but inflexible (present in toddlers), and an "explicit" system, which is flexible but slow, effortful, and mediated by language (Apperly & Butterfill 2009). Infants and toddlers, and potentially even Deaf adults with limited access to a shared linguistic system (Gagne & Coppola 2017) may rely on the prelinguistic implicit ToM system, not making use of the explicit system until the executive (Bialystok & Viswanathan 2009; Leslie, Friedman & German 2004) and language functions that support it are in place. Tager-Flusberg (2001) proposed a similar contrast between social-perceptive (i.e., implicit) ToM and social-cognitive (i.e., explicit) ToM. In this model, the explicit system requires deliberate and effortful engagement, whereas the implicit system is activated automatically.

9.6. *Does language mediate theory of mind in ASD?*

In the current study, the participants with ASD responded more slowly to false belief trials when concurrently engaged in a verbal task. Furthermore, in that group, there was an association between false belief RT and a pragmatic language measure; faster RTs were associated with greater language skill. These findings suggest that stronger language skills in ASD were associated with more efficient false belief reasoning, a relationship that was not evident in the TD group. RT correlations with another measure of language ability—CELF Core Language—missed significance (at $p = .08$), likely due to limited range and the CELF's limited sensitivity to individual differences in cognitively able individuals. In the current study, the lowest CELF standard score was 93 (well within the average range), with a standard deviation of just 10 points (i.e., within the test's normative *SD* of 15). The restricted range in the CELF may have prevented the detection of relationships; this is particularly likely given the magnitude of the correlations ($r = .54$ for Verbal load, $r = .46$ for Nonverbal load); these are moderate (Nonverbal) to large (Verbal) effect sizes.

Taken together, the RT and correlational results suggest that language is relevant for mediating ToM in ASD, consistent with studies showing the use of inner speech to mediate some (but not all) cognitive tasks in people with ASD with adequate language (Williams, Happé & Jarrold 2008). The current results suggest that linguistic competence has a unique influence such that people with ASD rely *more* heavily on language to reason about beliefs, consistent with findings of a stronger relationship between language level and false belief in ASD as compared to other populations (Durrleman & Franck 2015; Fisher, Happé & Dunn 2005; Hale & Tager-Flusberg 2003; Happé 1995; Lind & Bowler 2009). Whereas typically developing individuals may be able to rely more upon implicit "social-perceptive" ToM when solving false belief tasks, people with ASD may reason their way through false belief tasks, using deliberate, effortful, explicit "social-cognitive" ToM (Frith 2004; Tager-Flusberg

2001). If so, people with ASD may be more dependent upon language to bootstrap false belief task performance (Joseph, McGrath & Tager-Flusberg 2005; Tager-Flusberg 2001). In the absence of intuitive social-perceptive ToM, complement syntax affords a scaffold for solving false belief tasks (de Villiers & Pyers 2002; Lind & Bowler 2009). The explicit ToM system is posited to be a cognitively taxing process and to draw more heavily on language. Consistent with this suggestion, language skills in this study were correlated with false belief processing among ASD participants. The ASD group was, on average, almost a full second slower than the TD group on the belief-reasoning task. While generally consistent with other research showing slower processing speed (Mayes & Calhoun 2007) and slower reaction time (Schmitz, Daly & Murphy 2007) in ASD, this result is also consistent with a slower, explicit, language-driven approach to belief reasoning.

Data from an elegant study examining comprehension of the mental state verb *think* in 4-year-old children suggest that ToM difficulties arise at least in part because of children's difficulty in understanding the pragmatics of sentences; that is, they fail to understand that the focus of a statement is on a mental state rather than the reality (Lewis, Hacquard & Lidz 2017). This study was not designed to address this interesting possibility, which remains open for further research.

9.7. *Reliance on verbal mediation in domains other than ToM*

Current findings are inconsistent with several studies showing reduced reliance on inner speech during task performance in adults with ASD. For example, a verbal load task caused *less* performance decrement during set-shifting (Whitehouse, Maybery & Durkin 2006) and planning (Wallace et al. 2009; Williams, Bowler & Jarrold 2008) in ASD compared to TD groups. However, this result is consistent with a finding that children with ASD showed impairments in a verbal, but not a nonverbal, self-ordered pointing task that seems to rely on verbal mediation (Joseph et al. 2005) and with a finding that verbal load led to a decrement in performance of a short-term memory task in ASD (Williams, Bowler & Jarrold 2008). Fernyhough (2008) proposed a possible account of this discrepancy: *Monologic* inner speech is critical to linear processes, such as working memory, whereas *dialogic* inner speech (e.g., an internal dialogue between different perspectives held by oneself) is critical to the executive functions of flexibility and inhibition. Because ToM involves the simultaneous representation of multiple perspectives, one would hypothesize that ToM involves dialogic inner speech. This may be particularly relevant in the case of ASD, which is characterized by more limited engagement in dialogue as well as by delays in the development of ToM. Of course, the current study does not directly address this intriguing hypothesis.

9.8. *Limitations and future directions*

This study excluded individuals with low-average cognitive and language abilities. Given the complexity of this particular method, this population provided a logical starting point, but necessary next steps are to examine ToM processing in individuals with more limited abilities, in part by relying on online and implicit response measures. Furthermore, given the small sample size of the current study, the current results require replication.

A second limitation concerns the nature of the false belief task, which involves multiple task demands including sustained attention to the movies and motivation to participate and to be accurate; as such, the current conclusions are necessarily tentative. The repeated-measures experimental design presents a related methodological limitation; while the task permitted the comparison of load and belief conditions in a within-subjects design, the repetition of multiple belief trials may have impacted the task's sensitivity to individual differences in ToM processing. It is possible that some participants utilized a heuristic of memorizing the initial location of the apple when Annie remained in the room, which would mean that the task draws on working memory as much or more than belief reasoning. The analysis of the first trial results partially addressed this possibility but did not resolve it.

A third limitation concerns the secondary load tasks. Performance accuracy was not recorded for these tasks; it is possible that there were trade-offs between load and belief-reasoning task performance. In fact, subjective experimenter impressions indicated that all participants reliably and consistently performed the load tasks, but some participants may have sacrificed speed in favor of belief task performance in a manner that was not recorded here. Furthermore, while Study 2 suggested that the load tasks were of equivalent difficulty, or that the nonverbal load task imposed a slightly greater demand, this study was conducted with college students with an unremarkable developmental history; it is possible that had we normed the load tasks within an ASD sample the results would differ.

Finally, the correlational findings suggested that pragmatic language abilities were more strongly associated with efficiency of false belief reasoning in the ASD group compared to structural language measures; however, the CELF may be less useful in indexing individual differences in structural language in adolescents with age-appropriate cognitive abilities. Future studies may consider using more comprehensive measures of VIQ or more sensitive online language processing measures (see Eigsti & Schuh 2016 for further discussion).

10. Conclusions

This study directly examined the relationship between inner speech and ToM in individuals with ASD, using a belief-reasoning task depicting human actors in realistic settings. Our finding of a decrement in performance for false belief trials under verbal but not nonverbal load, along with an association between language abilities and false belief performance in the ASD group only, is consistent with the suggestion that ToM involves two parallel systems: an automatic, intuitive, social-perceptive ToM and an effortful social-cognitive ToM. Depending on the context, people with ASD may rely more heavily on language to bootstrap ToM. The current results suggest that autistic people may rely more on the verbal system rather than the intuitive social-perceptive ToM system. This choice does not likely reflect a conscious choice; rather, we suggest that for some people, the social-perceptive system is less automatic and less accessible; the pathways to this system may be less accessible, whereas access to the verbal social-cognitive system may reflect more experience and use (especially if this system was utilized more throughout development). This might be akin to what happens when one moves to a new home; although one knows the route from work to the new house, when cognitive resources are taxed (e.g., because of fatigue or multitasking), one might inadvertently navigate to the old, familiar home location.

If supported by additional research, this finding would inform our conceptualization of ToM deficits in ASD. Interventions could explicitly teach a compensatory strategy of relying on inner speech to provide a narrative of ToM representations, for example. Results highlight the importance of fostering early language skills among children with ASD, which might facilitate the development of earlier or more efficient reasoning about other people's minds.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Inge-Marie Eigsti http://orcid.org/0000-0001-7898-1898

References

American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders IV-TR*. IV-TR ed. Washington, DC: American Psychiatric Association.

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Washington, DC: American Psychiatric Association.

Apperly, I. A., and S. A. Butterfill. 2009. “Do Humans Have Two Systems to Track Beliefs and Belief-like States?” *Psychological Review* 116 (4):953–70. doi:10.1037/a0016923.
Balota, D. A., and D. H. Spieler. 1999. “Word Frequency, Repetition, and Lexicality Effects in Word Recognition Tasks: Beyond Measures of Central Tendency.” *Journal of Experimental Psychology: General* 128 (1):32. http://psycnet.apa.org/journals/xge/128/1/32/.
Barkley, R. A. 1997. “Behavioral Inhibition, Sustained Attention, and Executive Functions: Constructing a Unifying Theory of ADHD.” *Psychological Bulletin* 121 (1):65–94. doi:10.1037/0033-2909.121.1.65
Baron-Cohen, S., A. Leslie, and U. Frith. 1985. “Does the Autistic Child Have a “Theory of Mind”?” *Cognition* 21:37–36. doi:10.1016/0010-0277(85)90022-8.
Bialystok, E., and M. Viswanathan. 2009. “Components of Executive Control with Advantages for Bilingual Children in Two Cultures.” *Cognition* 112 (3):494–500. S0010-0277(09)00157-7 [pii]. doi:10.1016/j.cognition.2009.06.014.
Bloom, P., and T. P. German. 2000. “Two Reasons to Abandon the False Belief Task as a Test of Theory of Mind.” *Cognition* 77 (1):B25–31. http://www.ncbi.nlm.nih.gov/pubmed/10980256.
Carruthers, P. 2002. “The Cognitive Functions of Language.” *Behavioral and Brain Sciences* 25 (6):657–674; discussion 674–725. http://www.ncbi.nlm.nih.gov/pubmed/14598623.
Channon, S., T. Charman, J. Heap, S. Crawford, and P. Rios. 2001. “Real-life-type Problem-solving in Asperger’s Syndrome.” *Journal of Autism and Developmental Disorders* 31 (5):461–69. https://www.ncbi.nlm.nih.gov/pubmed/11794411.
Cohen, J. (1988). Statistical power analysis for the behavioral sciences, 2nd Ed.. Hillsdale, NJ: Erlbaum
de Villiers, J. G. 2005. “Can Language Acquisition Give Children a Point of View?” In *Why Language Matters for Theory of Mind*, edited by J. Astington and J. Baird, 186–219. New York: Oxford University Press.
de Villiers, J. G. 2007. “The Interface of Language and Theory of Mind.” *Lingua* 117 (11):1858–78. doi:10.1016/j.lingua.2006.11.006.
de Villiers, J. G., and J. Pyers. 2002. “Complements to Cognition: A Longitudinal Study of the Relationship between Complex Syntax and False-belief-understanding.” *Cognitive Development* 17:1037–60. doi:10.1016/S0885-2014(02)00073-4.
Delis, D. C., E. Kaplan, and J. H. Kramer. 2001. *Delis-Kaplan Executive Function System (D-KEFS) Examiner’s Manual.* San Antonio, TX: The Psychological Corporation.
Durrleman, S., and H. Delage. 2020. “Training Complements for Belief Reasoning in Developmental Language Disorder.” *Journal of Speech, Language, and Hearing Research* 63 (6):1861–77. doi:10.1044/2020_jslhr-19-00075.
Durrleman, S., and J. Franck. 2015. “Exploring Links between Language and Cognition in Autism Spectrum Disorders: Complement Sentences, False Belief, and Executive Functioning.” *Journal of Communication Disorders* 54:15–31. doi:10.1016/j.jcomdis.2017.03.005
Durrleman, S., M. Burnel, and A. Reboul. 2017. “Theory of Mind in SLI Revisited: Links with Syntax, Comparisons with ASD.” *International Journal of Language & Communication Disorders* 52 (6):816–30. doi:10.1111/1460-6984.12317.
Durrleman, S., M. Burnel, E. Thommen, N. Foudon, S. Sonie, A. Reboul, and P. Fourneret. 2016. “The Language Cognition Interface in ASD: Complement Sentences and False Belief Reasoning.” *Research in Autism Spectrum Disorders* 21:109–20. doi:10.1016/j.rasd.2015.10.003.
Durrleman, S., W. Hinzen, and J. Franck. 2018. “False Belief and Relative Clauses in Autism Spectrum Disorders.” *Journal of Communication Disorders* 74:35–44. doi:10.1016/j.jcomdis.2018.04.001
Eigsti, I. M. 2011. “Executive Functions.” In *Neuropsychology of Autism*, edited by D. Fein, 185–204. New York: Oxford University Press.
Eigsti, I. M., A. B. de Marchena, J. M. Schuh, and E. Kelley. 2011. “Language Acquisition in Autism Spectrum Disorders: A Developmental Review.” *Research in Autism Spectrum Disorders* 5:681–91. doi:10.1016/j.rasd.2010.09.001
Eigsti, I. M., and J. M. Schuh. 2016. “Language Acquisition in Autism Spectrum Disorders: Beyond Standardized Language Measures.” In *Innovative Investigations of Language in Autism Spectrum Disorder*, edited by L. Naigles, 183–200. Berlin, Germany: APA/Walter de Gruyter.
Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Kit of factor-referenced cognitive tests* Princeton, NJ: Educational Testing Service.
Farmer, T., J. Misyak, and M. H. Christiansen. 2012. “Individual Differences in Sentence Processing.” In *Cambridge Handbook of Psycholinguistics*, edited by M. Spivey, M. F. Joanisse and K. McRae, 353–63. Cambridge, UK: Cambridge University Press.
Fernyhough, C. 2008. “Getting Vygotskian about Theory of Mind: Mediation, Dialogue, and the Development of Social Understanding.” *Developmental Review* 28 (2):225–62. doi:10.1016/j.dr.2007.03.001
Fisher, N., F. Happé, and J. Dunn. 2005. “The Relationship between Vocabulary, Grammar, and False Belief Task Performance in Children with Autistic Spectrum Disorders and Children with Moderate Learning Difficulties.” *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 46 (4):409–19. JCPP371 [pii]. doi:10.1111/j.1469-7610.2004.00371.x.
Forgeot d’Arc, B., and F. Ramus. 2011. “Belief Attribution despite Verbal Interference.” *Quarterly Journal of Experimental Psychology* 64 (5):975–90. doi:10.1080/17470218.2010.524413.

Frith, U. 2004. “Emanuel Miller Lecture: Confusions and Controversies about Asperger Syndrome.” *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 45 (4):672–86. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15056300.
Gagne, D. L., and M. Coppola. 2017. “Visible Social Interactions Do Not Support the Development of False Belief Understanding in the Absence of Linguistic Input: Evidence from Deaf Adult Homesigners.” *Frontiers in Psychology* 8. doi:10.3389/fpsyg.2017.00837.
Garcia-Villamisar, D., and S. Della Sala. 2002. “Dual-task Performance in Adults with Autism.” *Cognitive Neuropsychiatry* 7 (1):63–74. doi:10.1080/13546800143000140.
Gergely, G., Z. Nádasdy, G. Csibra, and S. Biro. 1995. “Taking the Intentional Stance at 12 Months of Age.” *Cognition* 56 (2):165–93. http://www.sciencedirect.com/science/article/pii/001002779500661H.
Hale, C. M., and H. Tager-Flusberg. 2003. “The Influence of Language on Theory of Mind: A Training Study.” *Developmental Science* 6 (3):346–59. doi:10.1111/1467-7687.00289.
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):843–55. https://pubmed.ncbi.nlm.nih.gov/7789204/ http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7789204.
Hegarty, M., P. Shah, and A. Miyake. 2000. “Constraints on Using the Dual-task Methodology to Specify the Degree of Central Executive Involvement in Cognitive Tasks.” *Memory & Cognition* 28 (3):376–85. http://www.ncbi.nlm.nih.gov/pubmed/10881555.
Joseph, R. M., L. M. McGrath, and H. Tager-Flusberg. 2005. “Executive Dysfunction and Its Relation to Language Ability in Verbal School-age Children with Autism.” *Developmental Neuropsychology* 27 (3):361–78. doi:10.1207/s15326942dn2703_4
Joseph, R. M., S. Steele, E. Meyer, and H. Tager-Flusberg. 2005. “Self-ordered Pointing in Children with Autism: Failure to Use Verbal Mediation in the Service of Working Memory?” *Neuropsychologia* 43 (10):1400–11. doi:10.1016/j.neuropsychologia.2005.01.010
Kaland, N., K. Callesen, A. Moller-Nielsen, E. L. Mortensen, and L. Smith. 2008. “Performance of Children and Adolescents with Asperger Syndrome or High-functioning Autism on Advanced Theory of Mind Tasks.” *Journal of Autism and Developmental Disorders* 38 (6):1112–23. doi:10.1007/s10803-007-0496-8.
Leslie, A. M. 1994. “Pretending and Believing: Issues in the Theory of ToMM.” *Cognition* 50 (1):211–38. doi:10.1016/0010-0277(94)90029-9
Leslie, A. M., O. Friedman, and T. P. German. 2004. “Core Mechanisms in ‘Theory of Mind’.” *Trends in Cognitive Sciences* 8 (12):528–33. https://pubmed.ncbi.nlm.nih.gov/15556021/
Lewis, S., V. Hacquard, and J. Lidz. 2017. “‘Think’ Pragmatically: Children’s Interpretation of Belief Reports.” *Language Learning and Development* 13 (395–417):119. doi:10.1080/15475441.2017.1296768
Lind, S. E., and D. M. Bowler. 2009. “Language and Theory of Mind in Autism Spectrum Disorder: The Relationship between Complement Syntax and False Belief Task Performance.” *Journal of Autism and Developmental Disorders* 39 (6):929–37. doi:10.1007/s10803-009-0702-y.
Lord, C., M. Rutter, P. C. DiLavore, and S. Risi. 2002. *Autism Diagnostic Observation Schedule (ADOS)*. Los Angeles: Western Psychological Services.
Martin, A. E., and B. McElree. 2018. “Retrieval Cues and Syntactic Ambiguity Resolution: Speed-accuracy Tradeoff Evidence.” *Language, Cognition and Neuroscience* 33 (6):769–83. Epub 23272018 Jan 23273722. doi:10.1080/23273798.23272018.21427877.
Mayes, S. D., and S. L. Calhoun. 2007. “Learning, Attention, Writing, and Processing Speed in Typical Children and Children with ADHD, Autism, Anxiety, Depression, and Oppositional-defiant Disorder.” *Child Neuropsychology* 13 (6):469–93. doi:10.1080/09297040601112773
Meltzoff, A. N. 1995. “Understanding the Intentions of Others: Re-enactment of Intended Acts by 18-month-old Children.” *Developmental Psychology* 31 (5):838. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4137788/pdf/nihms459908.pdf.
Milligan, K., J. W. Astington, and L. A. Dack. 2007. “Language and Theory of Mind: Meta-analysis of the Relation between Language Ability and False-belief Understanding.” *Child Development* 78 (2):622–46. doi:10.1111/j.1467-8624.2007.01018.x.
Murray, D. J. 1967. “The Role of Speech Responses in Short-term Memory.” *Canadian Journal of Psychology/Revue canadienne de psychologie* 21 (3):263. doi:10.1037/h0082978
Newton, A. M., and J. G. de Villiers. 2007. “Thinking while Talking: Adults Fail Nonverbal False-belief Reasoning.” *Psychological Science* 18 (7):57. PMID: 17614864. doi:10.1111/j.1467-9280.2007.01942.x.
O’Neill, D. K. 1996. “Two-year-old Children’s Sensitivity to a Parent’s Knowledge State When Making Requests.” *Child Development* 67 (2):659–77. http://ezproxy.lib.uconn.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1996-01743-028&site=ehost-live.
Onishi, K. H., and R. Baillargeon. 2005. “Do 15-month-old Infants Understand False Beliefs?” *Science* 308 (5719):255–58. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3357322/pdf/nihms375148.pdf.
Roid, G. 2003. *Stanford-Binet Intelligence Scales*. 5th ed. Itasca, IL: Riverside Publishing.

Russell, J., R. Saltmarsh, and E. Hill. 1999. “What Do Executive Factors Contribute to the Failure on False Belief Tasks by Children with Autism?” *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 40 (6):859–68. doi:10.1111/1469-7610.00504

Rutter, M., A. Bailey, and C. Lord. 2003. *The Social Communication Questionnaire (SCQ)*. Los Angeles: Western Psychological Services.

Schick, B., P. de Villiers, J. de Villiers, and R. Hoffmeister. 2007. “Language and Theory of Mind: A Study of Deaf Children.” *Child Development* 78 (2):376–96. doi:10.1111/j.1467-8624.2007.01004.x.

Schmitz, N., E. Daly, and D. Murphy. 2007. “Frontal Anatomy and Reaction Time in Autism.” *Neuroscience Letters* 412 (1):12–17. http://www.sciencedirect.com/science/article/pii/S0304394006007178.

Semel, E., E. H. Wiig, and W. A. Secord. 2003. *Clinical Evaluation of Language Fundamentals*. 4th ed. San Antonio, TX: Harcourt Assessment, Inc.

Sokolov, A. N. 1972. *Inner Speech and Thought*. New York, NY: Plenum Press.

Strobach, T. 2020. “The Dual-task Practice Advantage: Empirical Evidence and Cognitive Mechanisms.” *Psychonomic Bulletin & Review* 27 (1):3–14. doi:10.3758/s13423-019-01619-4.

Surian, L., S. Caldi, and D. Sperber. 2007. “Attribution of Beliefs by 13-month-old Infants.” *Psychological Science* 18 (7):580–86. http://pss.sagepub.com/content/18/7/580.long.

Tager-Flusberg, H. (2001). „A re-examination of the theory of mind hypothesis of autism.„ In Development in autism: Perspectives from theory and research, edited by J. Burack, T. Charman, N. Yirmiya, and P. R. Zelazo, 173–193. Hillsdale, NJ: Erlbaum

Tager-Flusberg, H. 2001. “A Re-examination of the Theory of Mind Hypothesis of Autism.” In *Development in Autism: Perspectives from Theory and Research*, edited by J. Burack, T. Charman, N. Yirmiya and P. R. Zelazo, 173–93. Hillsdale, NJ: Erlbaum.

Tager-Flusberg, H., and C. Kasari. 2013. “Minimally Verbal School-aged Children with Autism Spectrum Disorder: The Neglected End of the Spectrum.” *Autism Research* 6 (6):468–78. Epub 2013 Oct 1007. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3869868/pdf/nihms520261.pdf.

Van Herwegen, J., D. Dimitriou, and G. Rundblad. 2013. “Performance on Verbal and Low-verbal False Belief Tasks: Evidence from Children with Williams Syndrome.” *Journal of Communication Disorders* 46 (5–6):440–48. doi:10.1016/j.jcomdis.2013.10.002.

Varley, R., M. Siegal, and S. C. Want. 2001. “Severe Impairment in Grammar Does Not Preclude Theory of Mind.” *Neurocase* 7:489–93. doi:10.1093/neucas/7.6.489.

Vygotsky, L., Rieber, R., & Carton, A. (1988). The collected works of L.S. Vygotsky: Problems of General Psychology, Including the Volume *Thinking and Speech*. New York, NY: Springer.

Wallace, G. L., J. A. Silvers, A. Martin, and L. E. Kenworthy. 2009. “Brief Report: Further Evidence for Inner Speech Deficits in Autism Spectrum Disorders.” *Journal of Autism and Developmental Disorders* 39 (12):1735–39. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2974759/pdf/nihms-223889.pdf.

Wellman, H. M. 1990. *The Child’s Theory of Mind*. Cambridge, MA: Bradford Books/MIT Press.

Whelan, R. 2010. “Effective Analysis of Reaction Time Data.” *The Psychological Record* 58 (3):9.

Whitehouse, A. J., M. T. Maybery, and K. Durkin. 2006. “Inner Speech Impairments in Autism.” *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 47 (8):857–65. JCPP1624 [pii]. doi:10.1111/j.1469-7610.2006.01624.x.

Wiig, E. H., and W. Secord. 1989. *Test of Language Competence (TLC)*. San Antonio, TX: The Psychological Corporation.

Willems, R., Y. Benn, P. Hagoort, I. Toni, and R. Varley. 2011. “Communicating without a Functioning Language System: Implications for the Role of Language in Mentalizing.” *Neuropsychologia* 49:3130–35. doi:10.1016/j.neuropsychologia.2011.07.023

Williams, D., F. Happé, and C. Jarrold. 2008. “Intact Inner Speech Use in Autism Spectrum Disorder: Evidence from a Short-term Memory Task.” *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 49 (1):51–58. JCPP1836 [pii]. doi:10.1111/j.1469-7610.2007.01836.x.

Williams, D. M., D. M. Bowler, and C. Jarrold. 2008. Intact inner speech use in autism spectrum disorder: evidence from a short-term memory task. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 49(1), 51-58. doi:10.1111/j.1469-7610.2007.01836.x.

Woolfe, T., S. C. Want, and M. Siegal. 2002. “Signposts to Development: Theory of Mind in Deaf Children.” *Child Development* 73 (3):768–78. http://www.ncbi.nlm.nih.gov/pubmed/12038550.