



Theory of Mind Development in Italian Children with Specific Language Impairment and Autism Spectrum Disorder: Delay, Deficit, or Neither?

Daniela Bulgarelli¹ · Silvia Testa² · Paola Molina³

Accepted: 1 December 2021 / Published online: 11 January 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Some studies report delayed theory of mind (ToM) development in children with specific language impairment (SLI), while others do not. A ToM delay is acknowledged in children with autism spectrum disorder (ASD), while whether these children also display a deficit is still under debate. In the current study, we drew on a developmental trajectory approach to assess whether children with SLI or ASD displayed delays or deficits in their ToM performance. Forty-three children with SLI (age 4–10 years), 44 children with ASD (age 5–12 years), and 227 typically developing children (age 3–11 years) completed the ToM Storybooks. Children with SLI were not found to display either a delay or a deficit in ToM, while children with ASD were found to display a deficit.

Keywords Social cognition · Communication disorders · Developmental speech or language disorders · Autism spectrum disorder · ToM storybooks

Introduction

Theory of mind (ToM) is a well-investigated component of social cognition and consists in attributing mental states such as desires, beliefs, and thoughts to self and others in order to understand, predict, and interpret one's own and others' behavior (Mitchell, 1997). The measure widely used to assess ToM is the first-order false belief (FB) task, originally designed by Wimmer and Perner (1983). Typically developing (TD) 3-year-old children systematically fail the

task, while from 4 years onwards, the probability of solving it correctly goes from below chance to above chance. Several kinds of FB task have been administered to samples of children over the years, and a meta-analysis has suggested that the developmental pattern of TD children's responses does not depend on variations in task type (Wellman, 2018; Wellman et al., 2001).

ToM development has been studied in atypical populations, such as children with specific language impairment (SLI), deafness, or autism spectrum disorder (ASD). The main issue at stake is whether these children display a delay in ToM, i.e., development that is slower but substantially similar to that of typically developing children, or a deficit, i.e., differential patterns of development of this competence.

ASD is characterized by a specific deficit in interaction and communication skills. There is consensus in the literature that a delay is also observed in autistic children's performance on ToM tasks and that the gap between their performance and that of TD children widens with age (Baron-Cohen, 2000; Blijd-Hoogewys et al., 2010; Boucher, 2012; Yirmiya et al., 1998). A longitudinal study showed that the ToM abilities of children with ASD improved during adolescence, but without typical functioning being attained (Ozonoff & McEvoy, 1994). In a recent longitudinal study (Peterson & Wellman, 2019), 3- to 13-year-old children with

✉ Daniela Bulgarelli
daniela.bulgarelli@unito.it

Silvia Testa
s.testa@univda.it

Paola Molina
paola.molina@unito.it

¹ Dipartimento di Psicologia, Università Degli Studi di Torino, via Verdi 10, 10124 Torino, Italy

² Dipartimento di Scienze Umane e Sociali, Università della Valle d'Aosta–Université de la Vallée d'Aoste, Aosta, Italy

³ DIST–Dipartimento Interateneo di Scienze, Progetto e Politiche del Territorio, Università Degli Studi di Torino, Torino, Italy

autism, deafness, or typical development completed the six ToM tasks from the ToM Scale (Peterson et al., 2012) at Time 1 and at Time 2, 18 months later in average. Deaf children and children with ASD obtained lower scores than their TD peers at both Time 1 and Time 2; all three groups displayed steady progress in ToM competence and individual differences among children remained stable over time. Other studies suggest that when the various aspects of the ability to reason about mental states in self and others are compared, children with ASD appear to have greater difficulty attributing beliefs to themselves than to others (for a review, see Williams, 2010). Adults with ASD close this gap and perform similarly to TD adults, although a deficit in false belief processing continues to be observed (Bradford et al., 2018). Furthermore, high-functioning children and adolescents with ASD, aged between 6 and 20 years, can perform as well as TD peers on ToM tasks, but still display limited ToM abilities in everyday social interactions, prompting us to lean towards the conclusion that they ultimately show a deficit (Scheeren et al., 2013).

SLI is characterized by poor performance on tests of language ability, while subjects' non-verbal cognitive functioning is normal and there is no evidence of neurological damage (Leonard, 2014). Given that ToM and language are interdependent competences (Milligan et al., 2007; for a review, see Bulgarelli et al., 2022), several studies have examined how they are related in children with SLI, obtaining mixed results (see Table 1 for a summary of the participants, tasks, and outcomes). More specifically, some studies found a ToM delay in childhood (Andrés-Roqueta et al., 2013; Durrleman et al., 2017; Farrant et al., 2006, 2012; Hanley et al., 2014) and adolescence (Botting & Conti-Ramsden, 2008). Other studies in contrast did not identify any differences between children with and without SLI (Bulgarelli & Molina, 2013; Ziatas et al., 1998). Interestingly, some studies suggest that the development of children with SLI may not actually be delayed; rather, their performance may be hampered by the linguistic nature of ToM tasks: when children with SLI are tested using tasks of low linguistic complexity, they obtain scores similar to those of TD children (Loukusa et al., 2014; Miller, 2001, 2004; Schaeffer et al., 2018; van Buijsen et al., 2011). Furthermore, inter-individual differences may depend on the severity of the disorder: children with Phonological Language Impairment (who display difficulty in properly forming the sounds of words) seem to be more likely to perform similarly to TD children, while children with Expressive Language Impairment (difficulty understanding and producing words and sentences) or Pragmatic Language Impairment (difficulty making appropriate use of language in social situations) seem more likely to perform poorly. Nevertheless, given that it is difficult to recruit large groups of children with SLI to participate in research, only a couple of studies to date have

been sufficiently well powered to compare children with different degrees of severity of SLI (Bulgarelli & Molina, 2013; Shields et al., 1996). A meta-analysis examining 17 studies published between 1998 and 2014 found that children with SLI display a delay in ToM compared with aged-matched TD children, and that age and gender do not moderate the difference in performance between the two groups (Nilsson & de Lopez, 2016). This meta-analysis does not yet provide sufficient evidence to fully reject the hypothesis that the linguistic complexity of ToM tasks may affect children's performance, given that it did not include studies that matched participants on language competence.

Notably, the studies summarized in Table 1 cannot shed light on the question of whether children with SLI show a ToM deficit, because they do not test the correlation of ToM performance with age across the groups of children (with and without language impairment), with the sole exception of van Buijsen et al. (2011). The last-mentioned study found that the pattern of correlations between chronological age and ToM tasks was quite similar in children with SLI and TD children, offering initial support for the hypothesis that children with language impairment do not have a ToM deficit.

The aim of this study was to contribute to the debate about whether children with SLI and ASD display a ToM delay or a ToM deficit. Our first research hypothesis was that children with SLI would not display a deficit or a delay in ToM competence, given that, on controlling for verbal competence, we expected both their performance (i.e., mean ToM scores) and their developmental patterns (i.e., correlation between ToM scores and chronological age) to be similar to those of TD children. Our second hypothesis was that children with ASD would display a delay in ToM, but we did not formulate a hypothesis about whether or not they would display a deficit, given that the literature reports mixed findings in this regard, while recent longitudinal studies have reported substantially similar developmental processes in ASD and TD children. The present study represents an innovative contribution to the literature because it drew on a comprehensive ToM test, the ToM Storybooks, which enables a composite and more precise measure of Theory of Mind competence than FB tasks alone. Furthermore, this instrument enables comparison of children across a wide age range, given that it meaningfully differentiates between the performances of TD children between 3 and 8 years.

Method

Participants

Three different groups of Italian children participated in the study. Children with SLI (N = 43, 13 girls, age range:

Table 1 Summary of the literature about theory of mind in children with specific language impairment

Authors	Year	Country	Age of children with SLI (range and average age) ^a	N SLI	N ASD	N TD	ToM tasks	Presence of delay in SLI compared to TD
Andrés-Roqueta et al	2013	Spain	R: 3;5–7;5 years M = 5;4, SD = 15.17 months (SLI) M = 5;4, SD = 12.85 months (PLI)	19 + 12	0	31 + 31	False beliefs: CL, UC	Yes, about 1-year delay No differences between SLI and PLI
Bulgarelli & Molina	2013	Italy	R: 51–98 months M = 76.4, SD = 10.8 months (PhLI) M = 69.1, SD = 14.0 months (ELI) M = 69.4, SD = 9.4 months (PLI)	10 + 13 + 6	0	29	Multiple ToM + false beliefs: CL	No PhLI performed better than ELI and PLI
Durrleman et al	2017	France	R: 6.5–11.7 years, M = 9.2	20	34	30	False beliefs: CL (low-verbal)	Yes, about 1-year delay
Farrant et al	2006	Australia	R: 58–68 months M = 62.9 months, SD = 3.2	20	0	20	Multiple ToM, including CL	Yes
Farrant et al	2012	Australia	R: 48–74 months M = 63.0 months, SD = 9.0	30	0	91	False beliefs: UC	Yes
Guiberson & Rodriguez	2013	USA (Mexican)	R: 3;0–5;11 years M = 4;3 years, SD = 8 months	22	0	24	False beliefs: UC	No in 3- and 4-year olds Yes in 5-year olds
Hanley et al	2014	Ireland	R: 8;1–11;1 years M = 115.2 months, SD = 10	14	17	16	False beliefs: BD, CL, UC	Yes
Loukusa et al	2014	Finland	R: 5;0–7;7 years M = 6;2 years, SD = 0;9	18	14	25	Multiple ToM	No in contextual tasks Yes in verbal tasks
Miller	2001	USA	R: 4;5–7;1 years M = 5;6 years	10	0	10 + 9	False beliefs: CL	No in low linguistic complexity tasks Yes in high linguistic complexity tasks
Miller	2004	USA	R: 47–71 months M = 59 months, SD = 7.3	15	0	15 + 15	False beliefs: CL	No in low linguistic complexity tasks Yes in high linguistic complexity tasks
Perner et al	1989	UK	R: 6;11–9;11 years M = 8;8	12	26	0	False beliefs: UC	No comparison between SLI and TD SLI performed better than ASD
Schaeffer et al	2018	The Netherlands	R: 6–14 years M = 9;8, SD = 2;2	27	27	27	False beliefs: CL	No in non verbal task
Shields et al	1996	UK	R: 7–11 years M = 105.9, SD = 19.78 (PhLI) M = 104.4, SD = 19.44 (PLI)	9 + 9	8	8	False beliefs: CL, UC	No in PhLI Yes in PLI
van Buijsen et al	2011	The Netherlands	R: 4–7 years M = 6.7 years	27	27	27	False beliefs: BD, CL, UC	No in line-drawing tasks Yes in verbal tasks
Ziatas et al	1998	Australia	M = 6;11, SD = 2;2 M = 8;2, SD = 2;7	12 + 12	12 + 12	12 + 12	False beliefs: CL	No

CL change of location, UC unexpected content, MA misleading appearance, BD belief desire, SLI specific language impairment, PhLI phonological language impairment, ELI expressive language impairment, PLI pragmatic language impairment, ASD autism spectrum disorder

^aInformation missing in the table is not reported in the original papers

4–9 years) attending the public child neuropsychology service: they had received this diagnosis during routine health care checks, based on their parents' accounts and tests of their linguistic production and understanding; 29 of these children also participated in a study by Bulgarelli and Molina (2013). Children with ASD (N = 47, 8 girls, age range: 5–12 years) attending two centers run by the national health system and specialized in the treatment of ASD; this group had been diagnosed directly at the centers. TD children (N = 227, 106 girls, age range: 3–10 years) were recruited at mainstream kindergartens and primary schools across several districts of a large Northern Italian city. Teachers and/or parents did not report any psychological or developmental problems that could hamper these participants' performance on the study measures. The age and gender distribution of the groups is reported in Table 2.

Measures and Procedures

Three measures were collected: ToM was evaluated through the Italian version of the ToM Storybooks (Blijd-Hoogewys et al., 2008; Bulgarelli et al., 2015; Molina & Bulgarelli, 2012); receptive vocabulary was assessed through the Italian version of the PPVT-R (Stella et al., 2000), and non-verbal cognitive ability was measured with the Leiter-R Brief IQ (Roid & Miller, 2011).

The ToM Storybooks is a comprehensive test that evaluates five ToM abilities drawn from Wellman's theory of mental state understanding (1990): recognizing emotions, making a distinction between physical and mental entities, appreciating that perception leads to knowledge, understanding how desires affect behavior, and understanding how beliefs affect behavior. The test is based on six books with full color illustrations, which present tasks in the context of a story about a character called Sam. Some tasks are repeated several times in different contexts throughout the stories, with a view to obtaining a more reliable measure. The test

comprises 95 items. The *Quantitative* score ranges from 0 to 77 and is based on 77 closed-ended questions, some of which require non-verbal responses (e.g., pointing at images). The *Qualitative* score varies from 0 to 36 assigned based on the child's response to 18 open-ended questions, which investigate whether the child spontaneously attributes mental states to the story characters (e.g., "Sam looks for the skates in the box because he thinks they are there": 2 points), only invokes situational aspects (e.g., "Sam looks for the skates in the box because he put them there": 1 point), or provides wholly inconsistent or incorrect explanations (zero points). Notably, the scores for the qualitative items partly depend on the quantitative items preceding them: if a closed-ended question is answered incorrectly, then a score of zero is automatically assigned to the following open-ended item. The *Total* score is obtained by summing the *Quantitative* and *Qualitative* scores.

The ToM Storybooks have been standardized for use with Dutch-speaking (Blijd-Hoogewys et al., 2008) and Italian-speaking (Bulgarelli et al., 2015; Molina & Bulgarelli, 2012) populations. The instrument offers good internal consistency (Cronbach's alpha = .90), test-retest reliability ($r = .86$, $p < .001$), and inter-rater reliability (range of Cohen's Kappa values among coders = .81–.97). It also displays discriminant validity (differentiating children with ASD from TD children), and divergent and convergent validity (Blijd-Hoogewys et al., 2008, 2010; Molina et al., 2020). Concerning content validity, a Principal Component Analysis of the data in the Dutch validation study showed that a five-component model (belief action, emotion recognition, mental physical, belief emotion, and desire emotion) offered the best theoretical interpretation. This solution accounted for 53.8% of variance. A confirmatory factor analysis applied to the data from the Italian study also supported a five-component solution (emotion, desire, mental-physical, belief, and perception knowledge). For a discussion of these two factorial structures, see Bulgarelli et al. (2015). The ToM Storybooks

Table 2 Total sample by age and gender

Years	TD			ASD			SLI		
	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total
3	2	7	9	–	–	–	–	–	–
4	22	25	47	–	–	–	3	1	4
5	32	24	56	2	1	3	10	5	15
6	24	26	50	9	2	11	9	4	13
7	17	9	26	5	1	6	3	1	4
8	11	5	16	10	0	10	4	2	6
9	6	2	8	8	2	10	1	0	1
10	7	8	15	0	2	2	–	–	–
11	–	–	–	2	0	2	–	–	–
12	–	–	–	3	0	3	–	–	–
Total	121	106	227	39	8	47	30	13	43

have been translated into multiple languages (English, Finnish, French, Italian, and Spanish), and an adapted Italian version of the test has been developed for blind children (Bartoli et al., 2019).

The children were individually assessed in a quiet room at the healthcare center they attended, or at their kindergarten or primary school. The children agreed to take the tests and both parents provided their written informed consent for the administration of the tests.

Data Analysis

To assess whether children with SLI displayed delayed or atypical development in terms of their ToM performance and whether their pattern of ToM development differed from that of children with ASD, we adopted a developmental trajectory approach (Thomas et al., 2009). When subjects of a wide range of ages are available to form both clinical and typical development samples, the developmental trajectory approach outperforms the matching approach because it permits discrimination between different forms of developmental delay (delayed onset, slowed rate, and delayed onset + slowed rate) as well as different forms of atypical development (non-linear trend, premature asymptote) as compared to the linear pattern usually observed in the TD group.

First, we examined the relationship between the three scores for ToM ability (*Quantitative*, *Qualitative*, and *Total* ToM Storybooks scores) and chronological age (CA) by calculating Pearson's correlation coefficients and creating a scatter plot for each of the three groups of participants. To control for the effects of differences in verbal and general cognitive ability, these descriptive analyses were conducted partialling out the influence of receptive vocabulary scores (VQ) and non-verbal cognitive ability scores (IQ) from the ToM Storybooks scores.

Second, we tested three regression models in which the different ToM Storybooks scores (*Quantitative*, *Qualitative*, and *Total* scores) were the respective dependent variables, and gender, CA, IQ, VQ, and two between-group factors ASD and SLI (TD was the reference group) were the independent variables. To test whether the developmental trajectory of the children with ASD and SLI differed from that of the TD children, we included the two interactions between group factors and CA in the regression models (ASD \times CA and SLI \times CA) and conducted simple slope analyses. We entered CA, IQ, and VQ in the models as mean-centered variables. The regression coefficients of the between-group factors (ASD and SLI) express the mean difference between these clinical groups and the TD group at the mean value of CA, allowing us assess for delays in the development of ToM competence as evaluated by the ToM Storybooks. More specifically, a ToM average score that was lower than

that obtained by the group of TD children was considered to show a delay when it occurred in combination with a similar regression slope for CA (i.e., a non-statistically significant coefficient for the interaction); a lower ToM average score combined with a lower regression slope compared to that of the TD group was considered to show a deficit.

We evaluated the overall goodness of fit of the regression models by calculating R^2 and used the variance inflation factor (VIF) to assess collinearity among the independent variables (values greater than 2.5 were taken to indicate collinearity). We measured the effect size of the interactions by calculating the difference in R^2 between a model that omitted the interaction term and one including the interaction term. We used IBM SPSS Statistics 26 to perform these analyses and PROCESS v3 macro (Hayes, 2012–2020) for the simple slope analyses.

Results

Descriptive statistics for the three groups are reported in Table 3.

In terms of the associations between age and the other variables, we found a significant correlation between age and IQ and VQ in the TD and SLI children, whereas the ASD children displayed a different pattern: their correlation coefficients for age were lower in general, and the correlation between age and *Qualitative* score, representing the ability to explain behavior in term of mental states, was particularly weak (see Table 4 and supplementary materials).

Moving on to the results of the regression analyses (Table 5), overall goodness of fit was satisfactory for all three models estimated; explained variance was 58% for *Quantitative* scores, 61% for *Qualitative* scores, and 63% for *Total* scores.

All VIF values were lower than 2.5, indicating that collinearity was not an issue. With regard to the control variables, the coefficient for gender was not statistically significant in any of the regression models, while IQ and VQ scores both exerted a positive and statistically significant influence on

Table 3 Sample characteristics by groups

	TD	ASD	SLI
N (girls)	227 (106)	47 (8)	43 (13)
Age in months (SD)	71.11 (21.43)	93.91 (22.25)	71.68 (14.04)
IQ (SD)	99.13 (13.11)	91.15 (17.42)	96.84 (13.05)
VQ (SD)	83.54 (14.13)	76.57 (14.57)	78.74 (12.20)
ToM Storybooks scores			
Quantitative (SD)	53.97 (12.04)	49.30 (13.09)	54.88 (8.08)
Qualitative (SD)	9.19 (6.83)	5.15 (5.09)	8.14 (5.38)
Total (SD)	63.16 (18.23)	54.45 (17.34)	63.02 (12.87)

Table 4 Partial correlation with age, controlling for IQ and VQ in TD, ASD, and SLI children

	TD		ASD		SLI	
	<i>r</i>	<i>p</i> *	<i>r</i>	<i>p</i> *	<i>r</i>	<i>p</i> *
Quantitative score	.685	<.001	.421 ^a	<.005	.570	<.001
Qualitative score	.696	<.001	.121 ^a	.43	.520	<.001
Total score	.725	<.001	.361 ^a	<.05	.582	<.001

*Pearson partial correlation, two-tailed; the superscript (a) means that the correlation is significantly different from the correlation in the TD group at $p < .05$

Table 5 Regression results

	Quantitative ToM score		Qualitative ToM score		Total ToM score	
	B	beta	B	beta	B	Beta
(Costant)	55.02***		9.87***		64.89***	
Gender	−0.72	−0.03	−0.57	−0.04	−1.29	−0.04
VQ	0.20***	0.24	0.12***	0.26	0.33***	0.26
IQ	0.13***	0.16	0.06**	0.12	0.19***	0.15
CA	0.36***	0.67	0.21***	0.69	0.57***	0.71
SLI_dummy	1.79	0.05	−0.61	−0.03	1.18	0.02
ASD_dummy	−7.90***	−0.24	−4.12***	−0.22	−12.02***	−0.24
SLI×CA	−0.02	−0.01	−0.02	−0.02	−0.05	−0.01
ASD×CA	−0.14*	−0.13	−0.18***	−0.30	−0.32***	−0.20
R ² full model	.58		.61		.63	

VQ verbal quotient, IQ non-verbal intelligence quotient, SLI specific language impairment, ASD autism spectrum disorder, CA chronological age

* $p < .05$

** $p < .01$

*** $p < .001$

the *Qualitative*, *Quantitative*, and *Total* scores. With regard to the variables CA and group membership and their interactions, the coefficient for CA—which measured the influence of chronological age among TD children (the reference group)—was positive and statistically significant for all three models. Its standardized value ranged from .67 to .71, reproducing the pattern of associations identified in the correlational analysis (Table 4, first column). The coefficient for the SLI group factor was not statistically significant in any of the models, meaning that the mean ToM Storybooks scores of the SLI group did not significantly differ from those of the TD group, once the influence of the control variables was held constant and the mean difference was evaluated at the mean chronological age. A different outcome was identified for the ASD group condition factor: in this case, the coefficient was statistically significant in all the models and the relationship was negative and strong. Specifically, the standardized regression coefficient (see the column headed ‘Beta’ in Table 5), ranged from −.22 to −.24, meaning that the mean ToM Storybooks scores of the ASD group were

around 1/5 or 1/4 of a standard deviation lower than the mean ToM Storybooks scores obtained by the TD group. The mean differences between the TD group and the ASD and SLI groups, respectively, are represented in Fig. 1, along with their 95% confidence interval values.

When we added the interaction between CA and group membership, this led to a statistically significant increase in R² in all the models. The greatest improvement was observed in relation to the *Qualitative* score outcome variable ($\Delta R^2 = 0.04$, $p < .0001$), followed by the *Total* score outcome variable ($\Delta R^2 = 0.02$, $p < .0005$), and lastly the *Quantitative* score outcome variable ($\Delta R^2 = 0.01$, $p < .05$). In all three models, only the coefficient for the ASD × CA interaction was statistically significant. As illustrated in Figs. 2, 3 and 4, simple slope analyses revealed that, compared to the slope in the TD group, the slope of the ASD group was not significantly different to zero for the *Qualitative* scores model and weaker than in the TD group for the other two. As may be observed in the figures, the simple slopes estimated for the SLI group overlapped with those of the TD group,

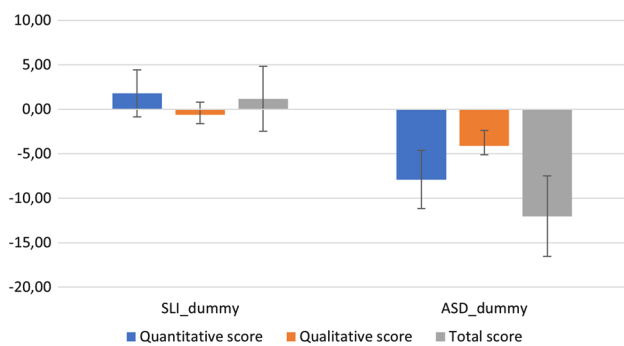
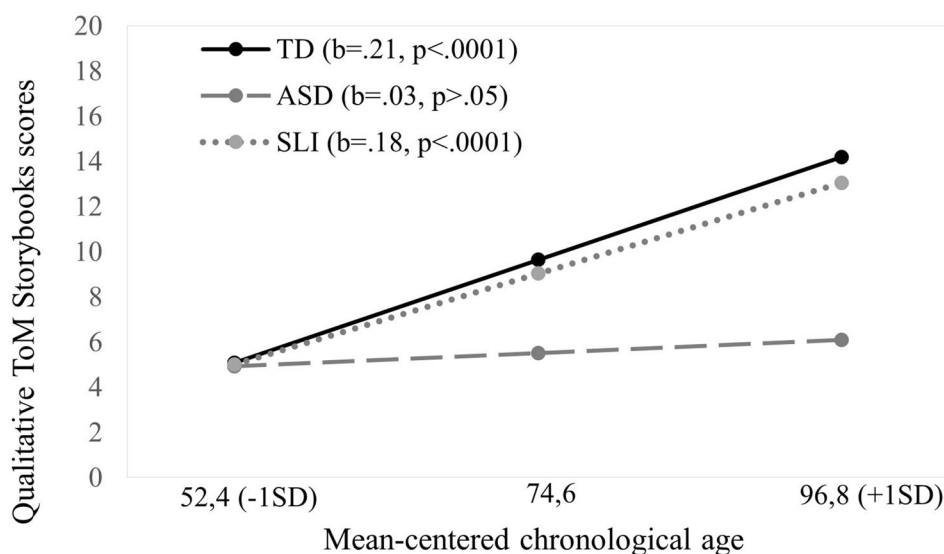


Fig. 1 Mean differences and 95% confidence intervals for SLI and ASD scores compared to TD scores (regression estimates)

Fig. 2 Qualitative ToM score: simple slope tests



reflecting the fact that the interaction term was not statistically significant.

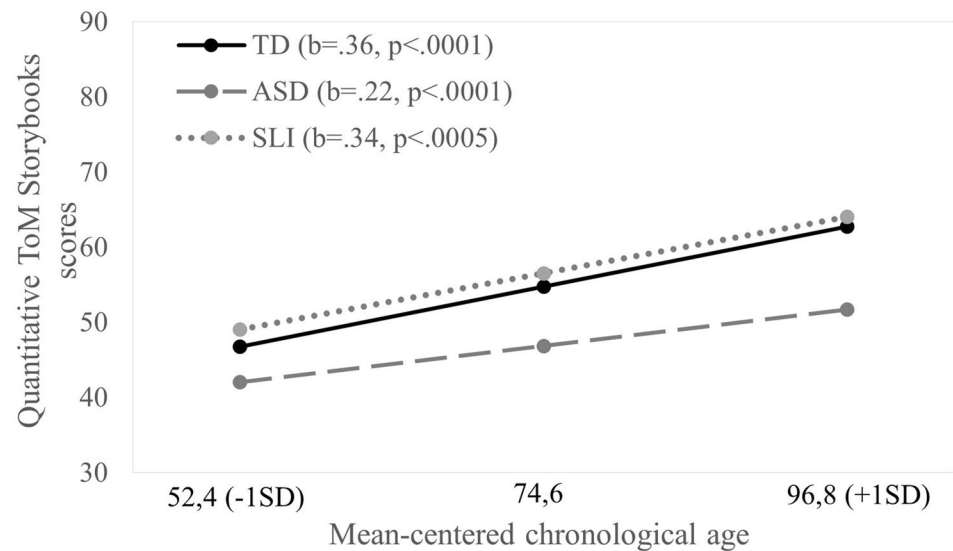
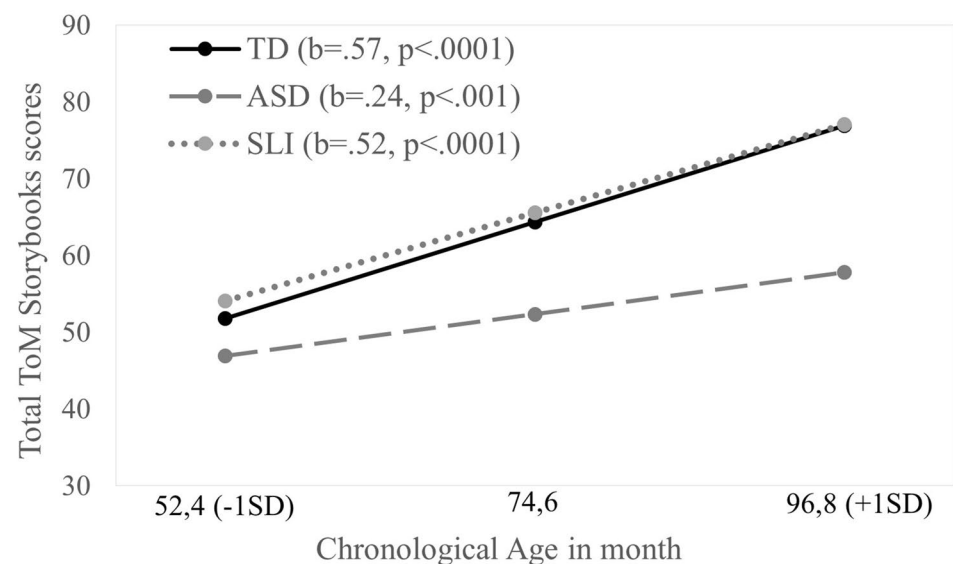
Discussion

The literature offers mixed evidence concerning ToM development in children with SLI, insofar as some studies have found that these children are delayed in their development (Andrés-Roqueta et al., 2013; Durrleman et al., 2017; Farrant et al., 2006, 2012; Hanley et al., 2014; Nilsson & de Lopez, 2016), while others have not (Bulgarelli & Molina, 2013; Ziatas et al., 1998), and there is only preliminary evidence suggesting that they may not show a ToM deficit (van Buijsen et al., 2011). The findings of this study contribute to the literature by confirming our first research hypothesis: the children with SLI in our sample displayed neither a delay nor a deficit in ToM competence,

once their receptive vocabulary had been controlled for; indeed, their mean ToM scores did not significantly differ from those of the TD children and the patterns of correlation between age and ToM scores were substantially similar in the two groups. Furthermore, in the regression analysis, neither membership of the SLI group nor the interaction between SLI status and age yielded a statistically significant effect.

With regard to the second hypothesis, the children with ASD displayed a deficit in ToM, not only a delay. Indeed, compared to the TD children, the children with ASD obtained lower mean ToM scores, and lower correlations between age and ToM scores. Furthermore, in the regression analysis, age, language, non-verbal cognitive functioning, and ASD were the factors that explained variance in the

ToM scores. Notably, the *Quantitative* ToM scores of children with ASD improved with age, albeit at a slightly lower rate than those of the TD children and children with SLI, while their *Qualitative* score was similar to that of the other two groups in the younger cohort (at the age of 52 months) and only differed in the older ones. This outcome reflects the specificity of the *Qualitative* score, which measures the ability to explicitly explain responses to the ToM tasks in term of mental states; in other words, in the qualitative items, children are not simply asked to recognize the effects of mental states on behaviors and/or emotions, but rather to refer to these mental states of their own initiative. Hence, a growing gap between children with and without ASD is to be expected from the age of 3 years, when TD children usually have not yet acquired the ability to spontaneously invoke mental states as explanations, to 12 years of age, when spontaneous references to mental states are extremely common in TD children but still rare in children with ASD.

Fig. 3 Quantitative ToM score: simple slope tests**Fig. 4** Total ToM score: simple slope tests

These findings partly differ from those of the two longitudinal studies reported in the literature that appeared to exclude a specific deficit in ToM. The first of these studies, by Blijd-Hoogewys et al. (2010), found that children with ASD displayed nonlinearities in their ToM development just as TD children do; the overall developmental pattern was similar in both groups, with children on the autism spectrum only displaying a delay. Blijd-Hoogewys and colleagues only examined children's development over a limited time period and from a micro-genetic perspective, while our own investigation was conducted with subjects of a broad range of ages in a cross-sectional sample. These two different levels of developmental analysis may not reveal the same pattern, and thus the question of whether children with ASD display a ToM deficit or not remains open. The limitation of

Blijd-Hoogewys et al.'s work (2010) is the restricted period of time investigated, whereas the shortcoming of our own work was that we could not track developmental patterns over time, because we lacked a sufficient number of cases and longitudinal data. The second longitudinal study, by Peterson and Wellman (2019), based on the 6-step ToM Scale (including measures of diverse desires, diverse beliefs, knowledge access, false belief, hidden emotion, sarcasm), found that children with ASD displayed steady progress in ToM competence over time as did TD children. This outcome appears to suggest a ToM delay in the former group. However, Peterson and Wellman (2019) reported that 65% of the children with ASD who passed the Hiding Emotion Task failed the False Belief task, while all TD children who passed the Hiding Emotion Tasks also passed the False

Belief task. This inversion in the scaling of the tasks in the two populations is a sign of qualitatively different development in children with ASD as compared to TD children, thus suggesting a ToM deficit. This is in line with our own observations of a differential pattern of ToM scores across groups and a different pattern of correlation between ToM scores and age. In Peterson and Wellman's article, further possible signs of qualitatively different development in children with and without ASD might be detected, in support of the deficit hypothesis: specifically, the participants were divided into two groups as a function of their performance at Time 1: children with *low* ToM scores and children with *high* ToM scores. In the ASD sample, the children in the *low* scoring group displayed a bigger increase in their scores from Time 1 to Time 2 than did the children in the *high* scoring group. In the TD sample, the opposite pattern was observed, with the children in the *high* scoring group displaying stronger gains at Time 2 than their peers in the *low* scoring group.

At present, our results offer no clear contribution to the debate about the impact of the linguistic complexity of the ToM tasks used to assess children with SLI. Some studies have shown that these children are more likely to obtain similar scores to TD children when tested using tasks of low linguistic complexity (Loukusa et al., 2014; Miller, 2001, 2004; Schaeffer et al., 2018; van Buijsen et al., 2011). Overall, the ToM Storybooks administered in our study is not an instrument of low linguistic complexity, although the *Quantitative* items only involve receptive language (requiring the child to answer yes/no or point to images), while the *Qualitative* questions also involve expressive language and the use of mental state lexicon. Neither *Quantitative* or *Qualitative* ToM scores differed significantly across the SLI and TD groups, suggesting that children with SLI could succeed on the tasks irrespective of the type of language involved. One possible explanation for this outcome, which suggests that children with SLI develop typically in terms of their ToM abilities, is that we administered all six books in the ToM Storybooks instrument, thus maximizing the number of similar ToM tasks that were repeated, and arguably thus obtaining a more reliable measure of the children's competence than if we had only used classical FB tasks. Further research is required to test the linguistic complexity hypothesis. It should also be noted that the impact of task linguistic complexity may be mediated by severity of linguistic impairment, as already discussed in Bulgarelli and Molina (2013). To test this hypothesis, future studies should compare children with phonological, expressive, and pragmatic language impairment.

With respect to the limitations of the current study, first, the number of participants in each age group was generally low and some age groups were not represented at all. This is usually addressed as the "common support problems", that may increase the root mean square error

of the regression estimator (Lechner & Strittmatter, 2019). A second limitation is that the children with SLI and with ASD were not divided on the basis of the severity of their condition: this information was available for the children with SLI, but the small number of participants did not allow statistical comparisons to be conducted; the equivalent information was not available for the children with ASD. A third limitation concerns the use of the PPVT-R, which only evaluates general language competence in terms of receptive language and vocabulary size. However, receptive language is the skill that children require to take the ToM test, so it seems reasonable to view it as a sufficient measure of their linguistic competence for the purposes of this study. Another limitation is the low mean score for receptive vocabulary obtained by the TD participants. Nevertheless, the children were recruited at mainstream schools and neither the participants' parents nor their teachers reported developmental issues of any kind. It is possible that this limitation was related to the wider range of socio-economic backgrounds represented in our sample compared to the PPVT-R normative sample. Nevertheless, we cannot test this hypothesis, given that the standardization study for the Italian version of the PPVT-R did not report information about the socio-economic background of the children who took part in it. Moreover, to control for this issue, we conducted independent samples t-tests, which showed that the mean PPVT-R scores of the TD children differed significantly from the average PPVT-R scores of the children with SLI and with ASD, suggesting that the instrument was able to discriminate adequately between the groups. Finally, we should also acknowledge the wider call in the literature to revisit research perspectives on autism (e.g., Woods et al., 2018): notably, the Double Empathy Problem suggests that the limited ToM abilities observed in persons with ASD are actually the outcome of barriers created when neurotypical interlocutors misunderstand and misperceive their communications. Given that little research on such reciprocal dynamics has been conducted to date (Mitchell et al., 2021), our study cannot contribute to this debate.

Future research about ToM development in children with SLI and with ASD should be focused on obtaining longitudinal data to validate the developmental patterns observed in our cross-sectional data and on observing whether delays or deficits in these groups persist beyond childhood. In addition, further research on different types of SLI and different degrees of severity in ASD is needed, given the crucial clinical value of being able to discriminate amongst children with ASD and children with SLI (and specifically those with Pragmatic Language Impairment). Comprehensive tools such as the ToM Storybooks could then be used in support of more reliable diagnostic processes, across a broad range of ages during childhood.

The novel contribution of the present study to the literature on ToM development in atypical populations lies in our use of a comprehensive test, the ToM Storybooks, which yields a composite and more reliable measure of theory of mind ability than FB tasks used in isolation. Furthermore, we did not find any ceiling effect across the age range under study, an issue that arises when relying on classical FB tasks (see also Thomas et al., 2009). The impact of this type of research on clinical practice is potentially crucial, given that ToM tasks and tests can be used to better describe children's mental state understanding and related abilities and to plan more tailored interventions accordingly. This seems particularly important for children with SLI who mainly need support in relation to their development of language and communication skills, given that their ToM ability is mainly preserved.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10803-021-05389-8>.

Author Contributions PM conceived and coordinated the Study, DB and ST participated in its design and coordination, DB coordinated the data collection, DB performed the literature review, ST and PM performed its measurement and the statistical analysis. All authors wrote, read and approved the final manuscript.

Declarations

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Ethical Approval This study was carried out in accordance with the ethical recommendations of the Ethical Code of the Italian Psychologists Association. All subjects gave their written informed consent in accordance to the Declaration of Helsinki.

References

- Andrés-Roqueta, C., Adrian, J. E., Clemente, R. A., & Katsos, N. (2013). Which are the best predictors of theory of mind delay in children with specific language impairment? *International Journal of Language & Communication Disorders*, 48(6), 726–737. <https://doi.org/10.1111/1460-6984.12045>
- Baron-Cohen, S. (2000). Theory of mind and autism: A fifteen year review. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (pp. 3–20). Oxford University Press.
- Bartoli, G., Bulgarelli, D., & Molina, P. (2019). Theory of mind development in children with visual impairment: The contribution of the adapted comprehensive test ToM Storybooks. *Journal of Autism and Developmental Disorders*, 49(9), 3494–3503. [https://doi.org/10.1016/0010-0277\(85\)90022-8](https://doi.org/10.1016/0010-0277(85)90022-8)
- Blijd-Hoogewys, E. M. A., van Geert, P. L. C., Serra, M., & Minderaa, R. B. (2008). Measuring theory of mind in children: Psychometric properties of the ToM storybooks. *Journal of Autism and Developmental Disorders*, 38(10), 1907–1930. <https://doi.org/10.1007/s10803-008-0585-3>
- Blijd-Hoogewys, E. M. A., Van Geert, P. L. C., Serra, M., & Minderaa, R. B. (2010). Development of theory-of-mind and the theory-of-mind storybooks research in typically developing children and children with autism spectrum disorders. *European Psychiatric Review*, 3(2), 34–38.
- Botting, N., & Conti-Ramsden, G. (2008). The role of language, social cognition, and social skill in the functional social outcomes of young adolescents with and without a history of SLI. *British Journal of Developmental Psychology*, 26(2), 281–300. <https://doi.org/10.1348/026151007X235891>
- Boucher, J. (2012). Putting theory of mind in its place: Psychological explanations of the socio-emotional-communicative impairments in autistic spectrum disorder. *Autism*, 16(3), 226–246. <https://doi.org/10.1177/2F1362361311430403>
- Bradford, E. E., Hukker, V., Smith, L., & Ferguson, H. J. (2018). Belief-attribution in adults with and without autistic spectrum disorders. *Autism Research*, 11(11), 1542–1553. <https://doi.org/10.1002/aur.2032>
- Bulgarelli, D., & Molina, P. (2013). Teoria della Mente nei bambini con Disturbo Specifico di Linguaggio: una questione di competenza o di performance? [Theory of Mind in children with Specific Language Impairment: A matter of competence or performance?]. *Giornale Italiano di Psicologia*, 60(4), 761–784. <https://doi.org/10.1421/76946>
- Bulgarelli, D., Henning, A., & Bertin, E. (2022). 33. Social cognition and its main correlates in childhood. In P. K. Smith & C. Hart (Eds.), *Wiley-Blackwell Handbook of Childhood Social Development*, 3rd Ed. Wiley-Blackwell.
- Bulgarelli, D., Testa, S., & Molina, P. (2015). Factorial structure of the 'ToM Storybooks': A test evaluating multiple components of theory of mind. *British Journal of Developmental Psychology*, 33(2), 187–202. <https://doi.org/10.1111/bjdp.12062>
- Durrleman, S., Burnel, M., & Reboul, A. (2017). Theory of mind in SLI revisited: Links with syntax, comparisons with ASD. *International Journal of Language & Communication Disorders*, 52(6), 816–830. <https://doi.org/10.1111/1460-6984.12317>
- Farrant, B. M., Fletcher, J., & Maybery, M. T. (2006). Specific language impairment, theory of mind, and visual perspective taking: Evidence for simulation theory and developmental role of language. *Child Development*, 77(6), 1842–1853. <https://doi.org/10.1111/j.1467-8624.2006.00977.x>
- Farrant, B. M., Maybery, M. T., & Fletcher, J. (2012). Language, cognitive flexibility, and explicit false belief understanding: Longitudinal analysis in typical development and specific language impairment. *Child Development*, 83(1), 223–235. <https://doi.org/10.1111/j.1467-8624.2011.01681.x>
- Guiberson, M., & Rodriguez, B. L. (2013). False belief understanding in language impaired and typically developing Spanish-speaking preschoolers from lower income backgrounds. *Early Education & Development*, 24(4), 517–535. <https://doi.org/10.1080/10409289.2012.693429>
- Hanley, M., Riby, D. M., McCormack, T., Carty, C., Coyle, L., Crozier, N., Robinson, J., & McPhillips, M. (2014). Attention during social interaction in children with autism: Comparison to specific language impairment, typical development, and links to social cognition. *Research in Autism Spectrum Disorders*, 8(7), 908–924. <https://doi.org/10.1016/j.rasd.2014.03.020>
- Hayes A. F. (2012–2020). *The PROCESS macro for SPSS, SAS, and R*. Retrieved on: <https://www.processmacro.org>
- Lechner, M., & Strittmatter, A. (2019). Practical procedures to deal with common support problems in matching estimation. *Econometric Reviews*, 38(2), 193–207. <https://doi.org/10.1080/07474938.2017.1318509>

- Leonard, L. B. (2014). *Children with specific language impairment*. MIT Press.
- Loukusa, S., Mäkinen, L., Kuusikko-Gauffin, S., Ebeling, H., & Moilanen, I. (2014). Theory of mind and emotion recognition skills in children with specific language impairment, autism spectrum disorder and typical development: Group differences and connection to knowledge of grammatical morphology, word-finding abilities and verbal working memory. *International Journal of Language & Communication Disorders*, *49*(4), 498–507. <https://doi.org/10.1111/1460-6984.12091>
- Miller, C. A. (2001). False belief understanding in children with specific language impairment. *Journal of Communication Disorders*, *34*(1–2), 73–86. [https://doi.org/10.1016/S0021-9924\(00\)00042-3](https://doi.org/10.1016/S0021-9924(00)00042-3)
- Miller, C. A. (2004). False belief and sentence complement performance in children with specific language impairment. *International Journal of Language and Communication Disorders*, *39*(2), 191–213.
- Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and Theory of Mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, *78*(2), 622–646. <https://doi.org/10.1111/j.1467-8624.2007.01018.x>
- Mitchell, P. (1997). *Introduction to theory of mind children, autism and apes*. London: Arnold.
- Mitchell, P., Sheppard, E., & Cassidy, S. (2021). Autism and the double empathy problem: Implications for development and mental health. *British Journal of Developmental Psychology*, *39*(1), 1–18. <https://doi.org/10.1111/bjdp.12350>
- Molina, P., & Bulgarelli, D. (2012). La standardizzazione italiana del test ToM Storybooks: Dati preliminary [The Italian standardization of the ToM Storybooks test: Preliminary data]. *Giornale Italiano di Psicologia*, *39*(4), 863–880. <https://doi.org/10.1421/73146>
- Molina, P., Testa, S., & Bulgarelli, D. (2020). Évaluer la compréhension des émotions et des états mentaux dans le développement typique et atypique: Résultats de deux nouveaux outils [Evaluating emotion and mental state understanding in typical and atypical development: Results from two new tools]. In Youssef, T. (Ed.), *Actes du colloque de la 13^e édition du RIPSYPDEVE 2020 «La psychologie du développement et de l'éducation pour le 21^e siècle: Nouveaux objets, espaces et temporalités»* (pp. 166–170). 2LPN - Université de Lorraine. <https://hal.univ-lorraine.fr/hal-03168825>
- Nilsson, K. K., & de Lope, K. J. (2016). Theory of mind in children with specific language impairment: A systematic review and meta-analysis. *Child Development*, *87*(1), 143–153. <https://doi.org/10.1111/cdev.12462>
- Ozonoff, S., & McEvoy, R. E. (1994). A longitudinal study of executive function and theory of mind development in autism. *Development and Psychopathology*, *6*(3), 415–431. <https://doi.org/10.1017/S0954579400006027>
- Perner, J., Frith, U., Leslie, A., & Leekam, S. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. *Child Development*, *60*, 689–700. <https://doi.org/10.2307/1130734>
- Peterson, C. C., & Wellman, H. M. (2019). Longitudinal theory of mind (ToM) development from preschool to adolescence with and without ToM delay. *Child Development*, *90*(6), 1917–1934. <https://doi.org/10.1111/cdev.13064>
- Peterson, C. C., Wellman, H. M., & Slaughter, V. (2012). The mind behind the message: Advancing theory of mind scales for typically developing children, and those with deafness, autism, or Asperger syndrome. *Child Development*, *83*(2), 469–485. <https://doi.org/10.1111/j.1467-8624.2011.01728.x>
- Roid, G. H., & Miller, L. J. (2011). *leiter international performance scale-revised (Leiter-R)*. Pymtec.
- Schaeffer, J., Van Witteloostuijn, M., & Creemers, A. (2018). Article choice, theory of mind, and memory in children with high-functioning autism and children with specific language impairment. *Applied Psycholinguistics*, *39*(1), 89–115. <https://doi.org/10.1017/S0142716417000492>
- Scheeren, A. M., de Rosnay, M., Koot, H. M., & Begeer, S. (2013). Rethinking theory of mind in high-functioning autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *54*(6), 628–635. <https://doi.org/10.1111/jcpp.12007>
- Shields, J., Varley, R., Broks, P., & Simpson, A. (1996). Social cognition in developmental language disorders and high-level autism. *Developmental Medicine and Child Neurology*, *38*, 487–495. <https://doi.org/10.1111/j.1469-8749.1996.tb12109.x>
- Stella, G., Pizzoli, C., & Tressoldi, P. E. (2000). *PPVT-R, peabody picture vocabulary test-revised*. Omega Edizioni.
- Thomas, M. S. C., Annaz, D., Ansari, D., Scerif, G., Jarrold, C., & Karmiloff-Smith, A. (2009). Using developmental trajectories to understand developmental disorders. *Journal of Speech, Language, and Hearing Research*, *52*(2), 336–358. [https://doi.org/10.1044/1092-4388\(2009/07-0144\)](https://doi.org/10.1044/1092-4388(2009/07-0144))
- van Buijsen, M., Hendriks, A., Ketelaars, M., & Verhoeven, L. (2011). Assessment of theory of mind in children with communication disorders: Role of presentation mode. *Research in Developmental Disabilities*, *32*(3), 1038–1045. <https://doi.org/10.1016/j.ridd.2011.01.036>
- Wellman, H. M. (1990). *The child's theory of mind*. MIT Press.
- Wellman, H. M. (2018). Theory of mind: The state of the art. *European Journal of Developmental Psychology*, *15*(6), 728–755. <https://doi.org/10.1080/17405629.2018.1435413>
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655–684. <https://doi.org/10.1111/1467-8624.00304>
- Williams, D. (2010). Theory of own mind in autism: Evidence of a specific deficit in self-awareness? *Autism*, *14*(5), 474–494. <https://doi.org/10.1177/2F1362361310366314>
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding and deception. *Cognition*, *13*(1), 103–128. [https://doi.org/10.1016/0010-0277\(83\)90004-5](https://doi.org/10.1016/0010-0277(83)90004-5)
- Woods, R., Milton, D., Arnold, L., & Graby, S. (2018). Redefining critical autism studies: A more inclusive interpretation. *Disability & Society*, *33*(6), 974–979. <https://doi.org/10.1080/09687599.2018.1454380>
- Yirmiya, N., Erel, O., Shaked, M., & Solomonica-Levi, D. (1998). Meta-analyses comparing theory of mind abilities of individuals with autism, individuals with mental retardation, and normally developing individuals. *Psychological Bulletin*, *124*(3), 283–307. <https://doi.org/10.1037/0033-2909.124.3.283>
- Ziatas, K., Durkin, K., & Pratt, C. (1998). Belief term development in children with autism, Asperger-syndrome, specific language impairment and normal development: Links to theory of mind development. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *39*, 755–763. <https://doi.org/10.1111/1469-7610.00374>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.