

From Understanding to Mindreading: The Role of Scenario Comprehension and Verbal Demand on Theory of Mind

Teresa FACCHETTI¹, Gianna COCCHINI¹ & Evelyne MERCURE¹

Affiliations

¹Department of Psychology, Goldsmiths University of London, London (UK)

Corresponding author:

Teresa Facchetti, BSc, MSc

Goldsmiths, University of London

Department of Psychology

Email: tf2604@cumc.columbia.edu

Phone number: +1 917 526 7186

Keywords: Language, Theory of Mind, Scenario Comprehension, Verbal Demand

Acknowledgements

The authors would like to thank MSc students Sabereen Munye and Jimena Larrea Mijares for their support in collecting the data for this study. We would also like to thank Cecilia Mijares for the development of the PTOMs illustrations.

Competing interest Declaration

Competing interests: The authors declare none.

Abstract

While a role of language in the development of Theory of Mind (ToM) is well established, the interplay with a child's ability to understand structured scenarios remains unclear. A new scale (PTOMs), assessing true and false belief comprehension at different levels of linguistic complexity, was used to explore language effects on ToM while accounting for scenario comprehension. 39 children (aged 4-6 years; 53.8% female) participated in this study. Results showed that 46.8% of 4-6-year-olds can understand false beliefs from picture-based scenarios with limited language output. Both language and scenario comprehension contributed to ToM in first-order false beliefs, whereas only scenario comprehension predicted true beliefs. In contrast, only language predicted second-order false beliefs, highlighting their different roles in ToM development.

1.Introduction

The ability to attribute mental states to oneself and others, a concept known as Theory of Mind (ToM), is a crucial mechanism for our social cognition (Mar, 2011). Most children typically develop this ability by age 4 (Wellman & Lagattuta, 2000), when they begin to predict and understand behaviour by recognizing that the mind is a representational system rather than a mere reflection of reality (Hale & Tager-Flusberg, 2003). Early in development, children first grasp the concept of true beliefs - mental representations that correspond to reality. As their cognitive abilities develop, they begin to understand first-order false beliefs (1FB), recognising that others may hold beliefs that differ from reality. More sophisticated ToM skills, such as second-order false beliefs (2FB; beliefs about other people's beliefs), develop around 6-7 years of age (Perner & Wimmer, 1985; Miller, 2013). Language plays a critical role in the development of ToM, serving as both a facilitator and a complex interdependent factor in its development (Astington & Jenkins, 1999). Previous studies have shown that the link between ToM and language remains important throughout child development (Hughes, 2011). This relation has been observed across typical (Milligan et al., 2007), and diverse/atypical developmental trajectories, including studies of children with autism (Tager-Flusberg & Joseph, 2005; Happé, 1995), developmental language disorder (DLD; Nillsson & de López, 2016), and deaf children (Peterson & Siegal, 1995; Lundy, 2002). The relation between language skills and success in ToM tasks could be explained in two alternative ways: (1) language could be essential for the development of ToM skills, or (2) language could be essential for understanding traditional ToM scenarios that rely heavily on language.

Over the years, several studies have highlighted the essential role that language plays in the development of ToM skills. Woolfe et al. (2002) developed a task with a reduced verbal

component to assess ToM in deaf children and found that those with early access to sign language, typically from deaf parents, demonstrated superior ToM skills compared to their peers with reduced language access. Thus, it appears that language not only enables children to fully participate in cultural and social activities that promote the development of ToM (Nelson, 2005), but also provides essential representational resources for understanding and managing false beliefs (Astington & Jenkins, 1999; De Villiers, 2005). Studies of second-order false belief comprehension have shown that successful completion of these tasks relies heavily on children's ability to manipulate complex linguistic constructions and to understand narratives that involve recursive reasoning about others' beliefs about beliefs (Perner et al., 1987). This complexity suggests that while rudimentary language skills may be sufficient for understanding first-order false beliefs, advanced language skills, including the understanding of more complex syntactic structures, are crucial for understanding second-order false beliefs. Therefore, it seems that the development of ToM is closely linked to the development of other cognitive skills that pave the way (De Villiers & De Villiers, 2000). Although significant milestones in ToM are typically reached by 4 years of age, when children can pass conventional false belief tasks (Wellman et al., 2001), some authors have argued that the foundations of ToM can be traced back to the early years of life, even before language acquisition begins (Bloom & German, 2000; Wellman & Lagattuta, 2000). For example, eye-tracking studies have shown that infants as young as 15 months can successfully perform non-verbal false belief tasks (Onishi & Baillargeon, 2005). This suggests that delays in passing false belief tasks may be due to the verbal and complexity demands of the tasks rather than a lack of ToM skills (Lewis & Mitchell, 1994; Siegel, 1999), supporting the hypothesis that language is critical for understanding traditional ToM scenarios that rely heavily on language. Alternatively, Wiesmann and Southgate (2021) argued that early signs of ToM in infancy may not

reflect genuine understanding or mental states or the ability to represent one's own and other's perspectives. Rather, they proposed that these apparent ToM skills may stem from infants' heightened attention to events observed with others rather than those experienced alone. According to this view, early ToM-like behaviours may stem from shared attention or social referencing mechanisms rather than true perspective-taking.

Research with neurotypical adults suggests that language plays a critical role in mediating ToM performance. Indeed, individuals tend to perform more successfully on tasks that include verbal elements than on those that are purely non-verbal (Marinis et al., 2023). In developmental psychology, ToM tasks with minimal verbal components have often been used with clinical populations. For example, studies of ToM in deaf children without access to fluent sign language have shown below average performance on both verbal and non-verbal tasks for their age group (Peterson & Siegal, 1995), suggesting that the difficulties with these tasks are not solely due to linguistic demands (Gale et al., 1996; De Villiers & De Villiers, 2000). Similarly, Peristeri (2021) found that children with autism performed worse than their neurotypical peers on low-verbal false belief tasks, despite performing well on control questions. Although evidence from typical development remains limited (Marinis et al., 2023), a study by Hollebrandse et al. (2014) found that children aged 6-7 years performed equally well on both verbal and non-verbal first-order false belief tasks. However, in the case of second-order false belief comprehension, children were successful on verbal tasks by age 7, but had difficulties with similarly complex non-verbal tasks until age 8 or 9, highlighting the supportive role of language in advanced ToM reasoning.

If the relation between ToM and language goes beyond mere task demands, it becomes crucial to investigate the role of different aspects, such as the child's ability to comprehend scenarios within structured narratives (Mar, 2011). Indeed, the nature of false belief tasks, which are often based

on narratives, raises the concern that the link between ToM and language may not be exclusive to false belief comprehension, but may rather be indicative of the child's ability to understand complex story plots. The underlying concept is that the cognitive processes used to interpret the mental states of fictional characters, such as those in novels or movies, are comparable to those used to understand the mental states of real people (Gerrig, 1993; Oatley, 1999). Neuroimaging evidence supports this overlap, showing that certain neural networks involved in ToM processes are associated with regions activated during narrative comprehension (Mar, 2011). While these studies often focus on the broader process of narrative understanding, they also suggest that children's ability to track key events and character intentions within structured scenarios - such as those presented in false belief tasks—relies on similar cognitive mechanisms. Effective storytelling, which requires an understanding of multiple perspectives and intentions, reflects the core cognitive skills essential for ToM. Studies of children with Social Communication Disorder have shown that difficulties in generating coherent narratives are related to broader ToM challenges (Bishop & Adams, 1992). These children often produce disjointed narratives that seem irrelevant, reflecting their struggle to integrate and articulate different mental states, a critical component of both narrative and ToM skills. Thus, the interdependence between narrative comprehension and ToM implies that the influence of language on ToM goes beyond simply understanding false beliefs. Given the complex nature of these processes, it is important to understand whether ToM assessments target underlying ToM skills and not merely the child's understanding of the narrative line. To mitigate this problem, researchers often use control questions designed to test scenario comprehension, thereby excluding data from children who do not understand the task correctly. Because ToM and narrative competence may be closely linked, it is important to account for scenario comprehension (Mar, 2018). Therefore, it may be

informative to consider performance on standard check questions as a proxy for scenario comprehension, given their established role in verifying understanding of task content. To address these issues, this study introduces a novel task, the Pictorial Theory of Mind Scale (PTOMs), which has been designed to assess a spectrum of ToM abilities, including both true and false beliefs, across different levels of linguistic demand. Our study aims to investigate: i) whether 4-to 6-year-olds understand false beliefs in a task with limited language input, thereby examining the role of the language content in the task as a facilitating factor; ii) the extent to which language skills predict performance on both first- and second-order false belief tasks when measured with limited language input; and iii) whether language skills contribute specifically to ToM or more broadly to the child's ability to understand structured scenarios within a task.

2. Methods

2.1. Participants

A sample of 39 developing children aged 4 to 6 years ($M = 5.36$, $SD = 0.60$, range = 4.10 - 6.60, 53.8% female) was randomly recruited from three schools in the United Kingdom. Children were enrolled in the British educational system at Reception (ages 4 – 5) or Year 1 (ages 5 – 6). The sample included 16 White, 18 Asian, and 5 Black children. Children were eligible to participate in the study if they were English speakers and had parental consent.

2.2. Procedures

Language was assessed by a subset of tasks from the Clinical Evaluation of Language Fundamentals (CELF-5; Wiig et al., 2013), including sentence comprehension, formulated

sentences, and word classes. Raven's Progressive Matrices-Short digital version (Arthur & Day, 1994) was employed as a measure of non-verbal reasoning. CELF-5 and Raven were administered via Q-Global, an online assessment platform from Pearson Clinical. First- and second-order false belief understanding, and true belief understanding were measured through the novel PTOMs Scale, whilst scenario comprehension was extrapolated from the check questions of the PTOMs (see description below). Ethical approval for the study was granted by the Goldsmiths, University of London Ethics Committee in accordance with the Declaration of Helsinki. Written informed consent was obtained from one of the parents before the study began.

2.3. Measures

The Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5; Wiig et al., 2013) is a battery of tasks that measures language skills during development (from 5 to 21 years). Three of the 16 subtests were selected to measure receptive and expressive skills in our study: Word Classes, Sentence Comprehension and Formulated Sentences. Sentence Comprehension consists of 26 items testing children's ability to comprehend increasingly complex spoken sentences through picture selection. The Word Classes subtest consists of progressively challenging items that assess the child's understanding of word relations, including semantic, functional, location, and temporal aspects. Formulated Sentences includes 24 items that assess the production of complex sentences within grammatical constraints and reflect the child's attention, comprehension, and analytical thinking. The composite score for this study was derived by calculating the average of the percentile scores from the three subscales (0–100).

Raven's Progressive Matrices – short digital form (Arthur & Day, 1994) assesses a wide range of cognitive functions, including the ability to generate novel ideas in response to new information, to interpret ambiguous or unclear contexts and to perform systematic logical analysis. Key cognitive processes assessed by the matrices are inductive reasoning, categorization, spatial intelligence, concurrent processing, detailed visual perception, and working memory capacity. Percentile scores were used for the analysis in our study (0–100).

The Pictorial Theory of Mind Scale (PTOMs) was developed for this study, building on the work of Woolfe et al. (2002), to assess true and false belief understanding in children in children aged 4 to 7 years by manipulating the verbal component of the task. The scale includes four items assessing first-order false belief with minimal verbal component and three complex narratives with increased verbal component assessing true beliefs and second-order false beliefs (**Table 1**). The first section of the scale is designed to assess first- order false beliefs using a visual "*thought bubble*" technique, inspired by Woolfe et al.'s (2002) work. This section contains four test items and one practice item, each representing a simple narrative of a character in everyday scenarios who holds a false belief (i.e., a belief that does not correspond to current reality, **Figure 1**). The first drawing shows the character with an obstructed view of an object, leading to a false belief. For example, in the first item, a girl fishing believes she has caught a fish when her view of the actual catch – a boot – is obscured by seaweed (**Figure 1.1**). The next illustration (**Figure 1.2**) reveals the nature of the object, thereby clarifying the false belief (e.g., the catch is a boot, not a fish). Children are presented with the illustrations one at a time to help them construct a coherent narrative. Next, children see an illustration of the character surrounded by four thought bubbles, each representing a possible response. To assess their understanding of the character's false belief,

children are asked: "*What does the girl think she has caught?*" (**Figure 1.3**). To assess their understanding of the scenario, another drawing integrates the four options directly into the scene and children are asked: "*What did the girl really catch?*" (**Figure 1.4**). Children are asked to point to the answer or give a verbal response. If no response is given within 30 seconds, the experimenter repeats the options. Children score 1 point if they answer the first-order false belief question correctly (e.g., fish), with a maximum score of 4 across the first-order false belief items. If a prompt is required, the score is reduced by 0.5 points. Children receive 0 points for an incorrect answer. Check questions were not part of the ToM score.

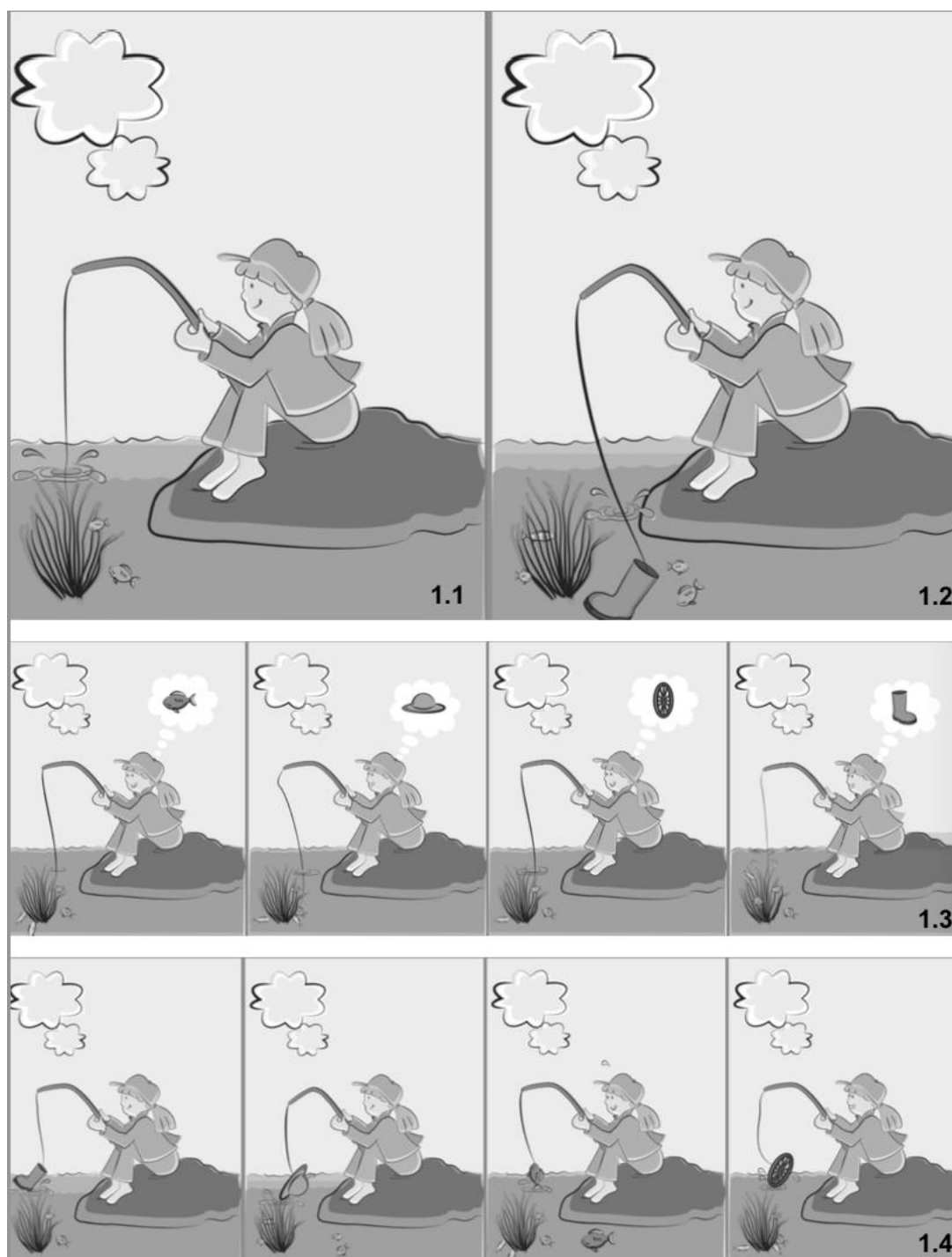


Figure 1. Example of a simple narrative item assessing first order false belief in the Pictorial ToM scale; **1.1**): A girl is fishing, the view of the object is covered by seaweed; **1.2**): The object is revealed to be a boot, challenging the girl's initial belief that she had caught a fish; **1.3**): first order

false belief question: "What does the girl think she caught? (Correct response: fish; Incorrect responses: hat, wheel, boot; **1.4**): Check question: "What did the girl really catch?", (Correct response: boot; Incorrect responses: hat, fish, wheel)

The second part of the assessment consists of three complex narratives, each designed to assess true beliefs and second-order false belief understanding (**Table 1**). The narratives are read by the experimenter and supported by sequential visual illustrations to facilitate comprehension (**Figure 2**). For example, in the first item, named '*Surprise*', children were presented with the following story: "Mum puts the phone on the table (**Figure 2.1**), and she leaves the room (**Figure 2.2**). Mark wants to surprise her, so he moves the phone from the table to her purse (**Figure 2.3**). His mom is looking through the door, but Mark does not know that (**Figure 2.4**). Each segment of the story is clearly illustrated to help children integrate the story elements. Children are then presented with a second-order false belief question, represented by an illustration of the main character surrounded by four thought bubbles, consistent with the format of the initial task. For example, in the previous item children are asked: "*Where does Mark think his mum will look for the phone?*" (**Figure 2.5**). Answers are indicated by pointing or selecting one of the four options. This is followed by a true belief question, e.g., "*Where will Mom look for the phone?*" (**Figure 2.6**), followed by a check question, e.g., "*Where is the phone?*" (**Figure 2.7**). Children are asked to point at the answer or give a verbal response. If no response is given within 30 seconds, the experimenter repeats the options. Children score 2 points if they respond correctly to all the questions, 1 point if they respond correctly to the true belief question, with a maximum score of 6 across the 3 scenarios. If a prompt is given, 0.5 points are deducted from the final score. Check questions were not part of the ToM score.

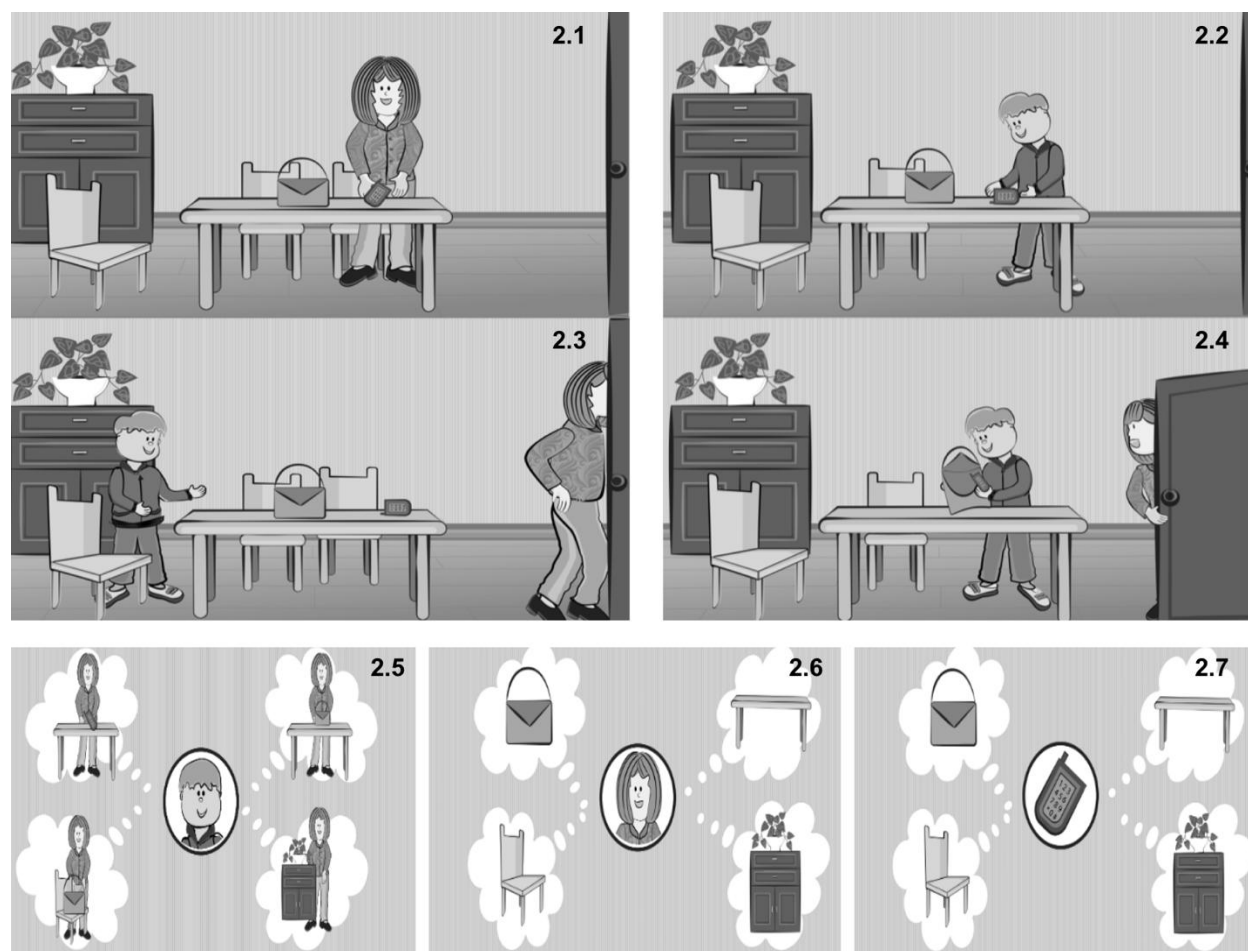


Figure 2: Complex narratives, first item- “Surprise!”. **1)** Mom leaves her phone on the table and; **2)** **she** leaves the room; **3)** Mark wants to surprise her and moves the phone from the table to the purse; **4)** His mum is looking through the door, but Mark does not know that. **5)** Second-order false belief question: Where does Mark think his mom will look for the phone? (possible responses: on the table, in the purse, on the chair, in the drawer); **6).** True belief question: Where will Mom look for the phone? (possible responses: in the purse, on the chair, in the drawer, on the table); **7)** Check question: Where is the phone? (Possible responses: in the purse, on the chair, in the drawer, on the table).

Table 1. Summary of the items in the PTOMs. The labels represent the central theme of each item's story.

Item	Verbal Demand	1 st order FB	2 nd order FB	True Belief	Check question
Fish	Simple Narratives Minimal verbal demand	<i>What does the girl think she caught?</i> (Boot = 0, Fish =1, Hat=0, Wheel=0)			<i>What did the girl really catch?</i> (Boot=1, Fish=0, Hat=0, Wheel=0)
Cat	Simple Narratives Minimal verbal demand	<i>What does the girl think is on the chair?</i> (Cat=1, Sweater=0, Pillow=0, Teddy bear=0)			<i>What was really on the chair?</i> (Cat=0, Sweater=1, Pillow=0, Teddy bear=0)
Glass	Simple Narratives Minimal verbal demand	<i>What does the man think he is taking from the cupboard?</i> (Glass=1, Apple=0, Fork=0, Pan=0)			<i>What did he really take from the cupboard?</i> (Glass=0, Apple=1, Fork=0, Pan=0)
Mermaid	Simple Narratives Minimal verbal demand	<i>What does the boy think he is seeing?</i> (Mermaid=0, Can=0, Boot=0, Fish=1)			<i>What did he really see?</i> (Mermaid=1, Can=0, Boot=0, Fish=-0)
Surprise	Complex Narratives Increased verbal demand		<i>Where does Mark think his mum will look for the phone?</i> (Bag=0, Table=1, Chair=0, Cupboard=0)	<i>Where will the mum look for the phone?</i> (Bag=1, Table=0, Chair=0, Cupboard=0)	<i>Where is the phone?</i> (Bag=1, Table=0, Chair=0, Cupboard=0)
Chocolate	Complex Narratives Increased verbal demand		<i>Where does Luke think Annie will look for the chocolate?</i> (Table=1, Chair=0, Drawer=0, Window=0)	<i>Where will Annie look for the chocolate?</i> (Table=0, Chair=0, Drawer=1, Window=0)	<i>Where is the chocolate?</i> (Table=0, Chair=0, Drawer=1, Window=0)
Cake	Complex Narratives Increased verbal demand		<i>Where does Sarah think David will look for the cake?</i> (Table=0, Oven=1, Fridge=0, Chair=0)	<i>Where will David look for the cake?</i> (Table=0, Oven=0, Fridge=1, Chair=0)	<i>Where is the cake?</i> (Table=0, Oven=0, Fridge=1, Chair=0)

2.4. Statistical analyses

IBM SPSS Statistics Version 27 was used for all analyses. Descriptive statistics were computed for PTOMs, CELF, and Raven scores. A principal component analysis (PCA) was conducted on the 7-item PTOMs scale to explore its factor structure and alignment with theoretical constructs. A priori power analysis using G*Power indicated that a larger sample (~70 participants) would be required to detect medium effects ($f^2 = 0.15$, 80% power). However, recruitment limitations prevented achieving this target. Sample adequacy was confirmed, with a Kaiser–Meyer–Olkin (KMO) value of 0.72 and a significant Bartlett’s test of sphericity, $\chi^2(45) = 180.00$, $p < .001$, supporting the suitability of the data for factor analysis. Given the limited sample size ($N = 39$ children), this study should be considered a preliminary exploration rather than a formal psychometric validation of the PTOMs scale. Spearman correlations were employed to assess relationships between variables due to the non-parametric nature of the data (Shapiro Wilk, $p < .005$). Hierarchical regression analyses were conducted to examine the relationship between language, scenario comprehension, and false belief understanding, while controlling for age and non-verbal reasoning, given their established role in theory of mind development.

3. Results

3.1. CELF and RAVEN

Table 2 presents descriptive statistics for our sample, and the children's scores on Raven and CELF composite score, which combines receptive and expressive language skills into a global language score (see Tomblin & Zhang, 2006). The table also includes scores from the PTOMs, describing the performance on first and second order false belief. Scenario comprehension was inferred from the check questions in the ToM task.

Table 2. Descriptive Statistics for the Study Sample.

Variable	%,(n) or Mean (SD); Range
Age	5.36 (0.60); 4.10–6.60
Year group	
Reception	48.7% (n = 19)
Year 1	51.3% (n = 20)
Ethnicity	
White	41.0% (n = 16)
Asian	46.2% (n = 18)
Black	12.9% (n = 5)
RAVEN*	51.9 (28.3); 0.3–93
CELF*	37.9 (25.8); 0.7–98
PTOMs	
Simple Narratives	
First-order FB	1.82, (1.61); 0–4
Scenario Comprehension	3.05, (1.10); 1–4
Complex Narratives	
True beliefs	2.41, (0.88); 0–3
Second-order FB	0.87, (1.24); 0–3
Scenario comprehension	2.74, (0.54); 1–3

Note: * Scores represent percentile ranks

3.2. PTOMs: Factorial Structure and Internal Consistency of the Scale

To examine the factorial structure of the PTOMs and to verify the components assessed by each part of the scale, an exploratory principal component analysis (PCA) with oblique rotation (Oblimin) was employed in line with our sample size. Children's scores in all items, excluding check questions, were included in the analysis. The Kaiser-Meyer-Olkin (KMO) measure validated the adequacy of the sample for this analysis, with a KMO value of 0.72. Bartlett's sphericity test was significant, $\chi^2(45) = 180$, $p < .001$, confirming the fitness of the correlation matrix for factor

analysis. Using maximum likelihood extraction and a factor loading threshold of .30, along with Kaiser's criterion of retaining factors with eigenvalues greater than 1, a three-factor structure appeared as the most appropriate, explaining 70.9% of the total variance. These factors included: a) first-order false belief items with minimal verbal component (i.e., simple narratives), which accounted for 29% of the variance; b) true belief questions with increased verbal component (i.e., complex narratives), which accounted for 16.0% of the variance and c) second-order false belief items (i.e., complex narratives), which explained 24.9% of the variance. During the validation of this ToM instrument, cross-loadings were found in two items, which were retained in the final model following precedents in the field (Lee et al., 2023), see **Table 3**. This structure is consistent with the original design of the scale and the verbal component of the items. Internal consistency was assessed using Cronbach's alpha. Overall internal consistency was high ($\alpha = .82$). Reliability was high for simple narratives ($\alpha = .83$) and second-order false belief items ($\alpha = .88$). In contrast, true belief items showed lower reliability ($\alpha = .58$)

Table 3. Factor Loadings and Uniqueness of PTOMs items in our tool. The labels represent the central theme of each item's story.

	1	2	3	Uniqueness	Dimension
Fish	0.69			0.37	Simple Narratives 1 st order FB
Cat	0.82			0.34	
Glass	0.67			0.34	
Mermaid	0.92			0.16	
Surprise		0.81		0.35	Complex Narratives True Beliefs
Chocolate		0.64	0.31	0.43	
Cake		0.70		0.43	
Surprise	0.46		0.57	0.24	Complex Narratives 2 nd order FB
Chocolate			0.94	0.10	
Cake			0.94	0.10	

3.3. Performance in the PTOMs

On first-order false belief tasks with minimized verbal demands, children's mean score was 1.85 ($SD = 1.62$, range: 0–4; **Figure 3**). The average pass rate across all items was 46.8%, with 25.6% of the children obtaining the maximum score in all the items. Our study included children who failed check questions. Excluding these children adjusts the pass rate across items to 60.5%.

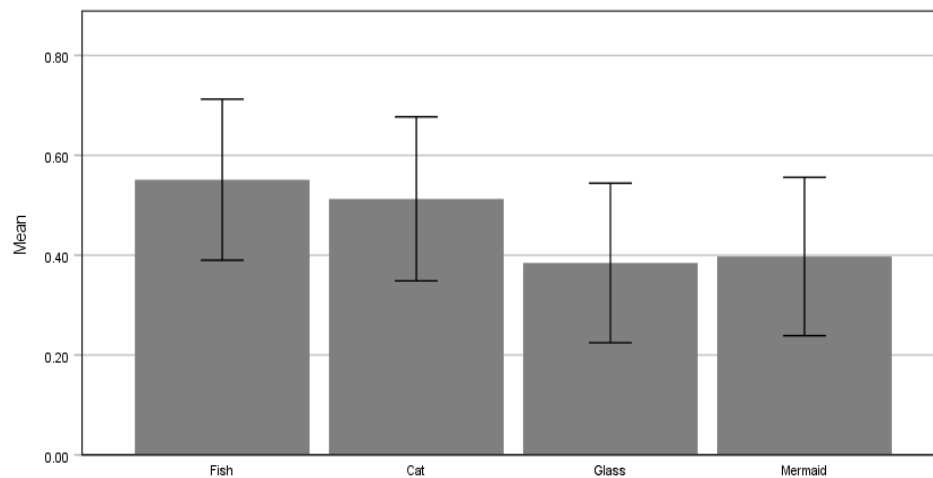


Figure 3. Mean scores and SE (confidence level 95%) for first-order false belief items with reduced verbal component in our sample.

Complex narratives involving true beliefs, and second order understanding of false beliefs were assessed, with 64% of the children successfully completing all items related to true beliefs ($M = 2.41$, $SD = 3.00$; range = 0–3). Performance on second-order false belief items was significantly lower with 20.5% of children successfully completing all items ($M = 0.87$, $SD = 1.24$; range = 0–3; see **Figure 4**).

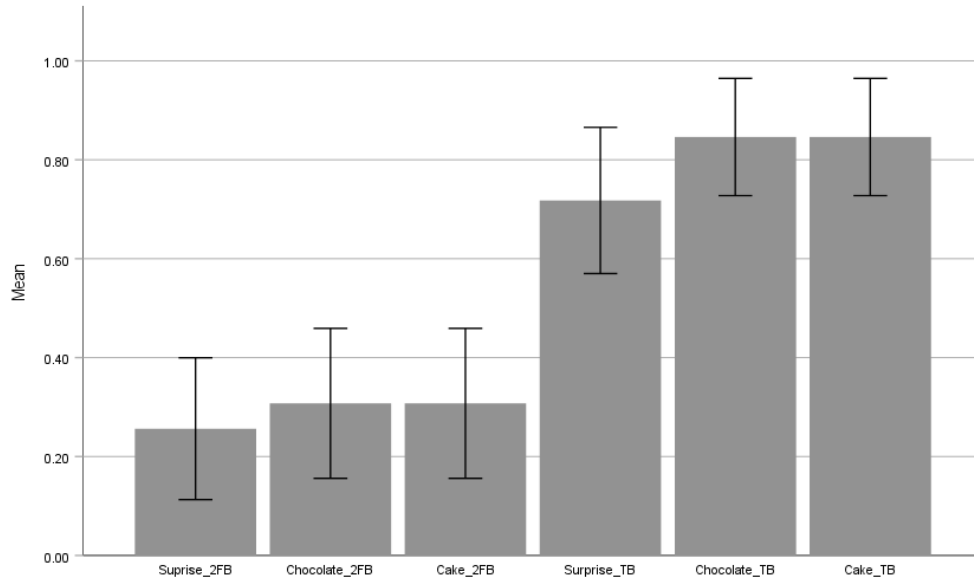


Figure 4. Mean scores and SE (confidence level 95%) for second order false belief and true belief items in our scale (complex narratives).

3.4. Is language linked to Theory of Mind?

Spearman correlation analysis showed that first-order false belief performance in simple narratives was associated with language skills ($r = .57, p < .001$). No significant correlation was found with language skills on true belief performance ($r = .10, p = .52$). However, these results might be explained by a potential ceiling effect on this task. Second-order false belief performance was significantly correlated with language skills ($r = .63, p < .001$). Scenario Comprehension was related to first-order false belief ($r = .66, p < .001$), true belief ($r = .62, p < .001$) and second-order false belief understanding ($r = .38, p = .01$) but not to language scores (see **Table 4**).

Table 4. Strength of Spearman's correlations between CELF, first-order false belief (1FB), scenario comprehension (SC; simple narratives), true beliefs (TB), second order false-belief (2FB), scenario comprehension (SC; complex narratives).

	CELF	RAVEN	1FB	SC (Single Narratives)	TB	2FB	SC (Complex Narratives)
CELF	-						
RAVEN	.35* 37 .025	-					
1FB	.57*** 37 <.001	.06 37 .68	-				
SC (Single Narratives)	.28 37 .07	.18 37 .26	.66** * 37 <.001	-			
TB	.10 37 .52	-.05 37 .72	.27 37 .09	.42** 37 .007	-		
2FB	.63*** 37 <.001	.15 37 .35	.49** 37 .001	.17 37 .27	.35* 37 .02	-	
SC (Complex Narratives)	.38* 37 .01	-.11 37 .47	.34* 37 .03	.28 37 .08	.62*** 37 <.001	.38* 37 .01	-

Note. Each cell presents Spearman's rho, degrees of freedom ($df = 37$), and p value. $p < .05^*$, $p < .01^*$, $^{**}p < .001$.

3.5. The relation between language, ToM and Scenario Comprehension

Hierarchical regression analysis was employed to assess the effect of language on first-order false belief performance, controlling for variance explained by scenario comprehension, age, and Raven's scores. In Model 1, neither age nor Raven's scores emerged as significant predictors (see

Table 5). In Model 2, scenario comprehension emerged as a significant predictor ($\beta = .79, p < .001$), underscoring its critical role in children's ability to navigate first-order false belief tasks. This model accounts for 39.82% of the variance ($\Delta R^2 = 0.379$). The inclusion of language abilities in Model 3 increased the variance explained by an additional 19.68% ($\Delta R^2 = 0.197$). Language skills emerged as a critical predictor ($\beta = .01, p = .001$) with scenario comprehension maintaining its significance in predicting first-order false belief performance ($\beta = .87, p < .001$). AIC and BIC show consistent improvements across models, decreasing from 154 and 161 in Model 1 to 122 and 134 in Model 3, respectively, indicating superior model fit. Examining the interaction effect between the two variables in predicting first-order false belief performance, it appears that scenario comprehension does not moderate the relation between ToM and language ($\beta = .002, p = .70$). Instead, these variables independently predict first-order false belief comprehension. Another hierarchical regression analysis was employed to analyse the relation between language and true beliefs (**Table 5**). In Model 1, neither Raven scores nor age were significant. Interestingly, whilst Model 2 reveals that scenario comprehension is a significant predictor of true beliefs performance, increasing the explained variance by 33.06% ($\Delta R^2 = 0.33$), language does not appear as a significant predictor in our model. Hierarchical regression analyses were conducted to examine the link between second-order false belief and language proficiency, controlling for age, Raven's progressive matrices, and scenario comprehension. In Model 2, scenario comprehension significantly predicted second-order false belief performance ($\beta = .84, p = .022$). However, when language skills were introduced in Model 3, scenario comprehension lost its significance and language proficiency became the only significant predictor of second-order false belief performance ($\beta = .02, p < .001$), see **Table 5**).

Table 5. Standardized regression coefficients (β), standard error and p-value for each predictor for the hierarchical regression analyses in our study, with first-order false beliefs, true beliefs and second order false beliefs items as dependent variables; and age, raven, scenario comprehension and language skills as covariates.

Measure	Mode l	R^2	Adj R^2	AIC	BIC	F (df1 - df2)	P- value	Significant predictors
First-order FB (Simple narratives)	1	.01	-.03	154	161	0.34 (2, 36)	.71	None
	2	.63	.34	137	146	7.71 (3, 35)	<.001	Scenario comprehension ($\beta = .79, p < .001$)
	3	.63	.57	122	133	11.4 (5, 33)	<.001	Scenario Comprehension ($\beta = .87, p < .001$), CELF ($\beta = .01, p = .001$)
True beliefs (Complex Narratives)	1	.009	-.04	107	114	0.17 (2, 36)	.84	None
	2	.34	.28	93.5	102	6.01 (3, 35)	.002	Scenario Comprehension ($\beta = 0.9, p < .001$)
	3	.37	.30	93.3	103	5.13 (4, 34)	.002	Scenario Comprehension ($\beta = 1.09, p < .001$)
Second- order FB (Complex Narratives)	1	.03	-.01	133	140	0.71 (2, 36)	.49	None
	2	.17	.10	129	137	2.43 (3, 35)	.08	Scenario Comprehension ($\beta = .84, p = .02$)
	3	.42	.35	117	127	6.22 (4, 34)	<.001	CELF ($\beta = .02, p < .001$)

Note. Model 1 includes age and Raven scores. Model 2 adds scenario comprehension. Model 3 includes language skills (CELF scores). β = standardized beta coefficient. p = significance level. FB = False belief. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion.

4. Discussion

The interface between ToM and language remains a central area of debate. The current study aims to address some of the open questions on the link between language and ToM using the PTOMs,

a pictorial ToM measure that assesses both first- and second-order false and true beliefs across different levels of verbal demand.

4.1. Can 4–6-year-olds understand false beliefs in a task with limited language input?

Building on Woolfe et al. (2002), our study aimed to assess first-order false belief performance by reducing verbal task complexity using the PTOMs. Prior research (Onishi & Baillargeon, 2005) indicates that nonverbal false belief reasoning can emerge as early as 15 months, suggesting that language might not be crucial at early stages. While our study did not directly compare different task modalities, our success rate (46.8%) was lower compared to traditional verbal false belief tasks, where reported success rates range from 74.6% (Wellman et al., Watson, 2001) to 85% (Baron-Cohen et al., 1985). Excluding children who did not pass the check questions increased the success rate to 60.5%, though still below standard verbal tasks. This discrepancy raises an important question. Does reducing verbal input affect or facilitate false belief reasoning? One possibility is that task modality differences affects cognitive load. Indeed, the PTOMs task allowed children to shift from visual processing to verbal questioning and, in some cases, to provide a verbal response. Although this shift was not required, it may have increased cognitive demands for those who chose to respond verbally (Dantzig et al., 2008). Another explanation is that false belief comprehension performance relies on language, suggesting that the verbal component of the task itself inherently facilitates task comprehension and false belief reasoning. Hollebrandse et al. (2014) found that 6- to 7-year-olds who had already developed first-order false beliefs understanding performed similarly across verbal and non-verbal first-order false belief tasks. However, they performed better on verbal than nonverbal second-order false-belief tasks, suggesting that the verbal component of the task itself inherently facilitates task comprehension

and false belief reasoning (Hollebrandse et al., 2014). More broadly, these findings are consistent with evidence that verbal scaffolding enhances ToM performance. False belief tasks vary in the extent to which verbal cues are provided, influencing children's success rates. Comparisons between Perner's et al. (1989) highly verbal tasks and later adaptations with reduced linguistic complexity suggest that verbal input facilitates ToM reasoning. Our findings indicate that 46.8% of 4–6-year-olds successfully engaged in false-belief reasoning using picture-based scenarios with limited language output. This underscores the facilitating role of language in supporting false belief understanding, even in contexts designed to minimize verbal demands. Future research should consider that our sample may not be directly comparable to previous studies, as variability in language exposure and task structure may influence success rates. Differences in the amount of verbal scaffolding provided across studies may account for performance discrepancies (Perner et al., 1989). Moreover, our study included children from mainstream classrooms with a wide range of verbal and non-verbal abilities. Direct comparisons between studies must therefore take these methodological differences into account to accurately interpret differences in false-belief reasoning.

4.2. Does a child's language skills associate with their performance on picture-based ToM tasks?

The second aim of our study was to examine whether language remains a significant predictor of ToM when verbal task demand is minimised. Regression analyses controlling for age and non-verbal reasoning confirmed that language skills were significantly associated with both first- and second-order false belief performance (Tager-Flusberg & Joseph, 2005), across different levels of language demands. This indicates that language skills are essential for interpreting and

understanding mental states (Gavilán & García-Albea, 2011), even when tasks are designed with reduced verbal demands. In line with these findings, Schick et al. (2007) conducted a study in which they assessed the ToM abilities of 4- to 6- year-olds deaf children by employing both verbal and low-verbal tasks. Notably, even the low-verbal tasks, which minimized linguistic demands, revealed that deaf children with hearing parents, who are likely to experience reduced language access, performed significantly worse compared to hearing children or deaf children with deaf parents who use sign language. This suggests that early and accessible language exposure is crucial for ToM development, even in tasks with reduced verbal requirements. Most importantly, this link is valid for both first and second order false belief items, with successful negotiation of second-order false belief tasks relying heavily on children's ability to manipulate complex linguistic constructions (Perner et al., 1987). Taken together, these observations suggest that the impact of language on ToM goes beyond simple linguistic barriers, such as the semantic and syntactic challenges posed by traditional assessment methods. The reduced performance on ToM tasks with a minimized verbal component, together with the correlation between language proficiency and ToM performance even under reduced language conditions, may suggest an inherent function of language in fostering our ability to attribute mental states to others (De Villiers, 2005).

4.3. How do language skills and narrative competence contribute to ToM?

The third aim was to determine whether the relationship between language and ToM is specific to false beliefs or whether it also relates to children's ability to understand structured scenarios. While previous studies have highlighted a strong link between ToM and narrative competence (Capps et al., 1998; Mar & Oatley, 2008), narrative competence encompasses broader storytelling and discourse skills. However, our study focused specifically on scenario comprehension, assessing

children's ability to follow key events and character actions within a structured context. Rather than assessing storytelling or story retelling (Roch et al., 2016), we inferred scenario comprehension through check questions and included it as a covariate in our analyses to examine its role alongside language in ToM reasoning (Mar, 2018). Unlike previous research, which often excluded data from children who misunderstood the story, our method allowed us to retain participants who may show a dissociation between ToM and scenario comprehension, and to advance our understanding of the interplay between these variables. Overall, our findings suggest a complex interplay among language, false beliefs, and scenario comprehension. Indeed, when considering first-order false beliefs, we found that both verbal and scenario comprehension skills were significant predictors of ToM performance, even after controlling for age and nonverbal reasoning. However, there was no significant interaction effect between them, suggesting that these domains may influence ToM at different levels, with language ability playing a critical role in effectively interpreting mental states (De Villiers & Pyers, 1997) whereas scenario comprehension aids in the structuring and interpretation of narratives (Mar, 2011). This distinction becomes particularly significant considering the design of our study, in which first-order false belief items were presented with minimal linguistic input. Interestingly, our results indicate that performance on true belief tasks was significantly associated with scenario comprehension, but not with language. This suggests that success in true belief reasoning may depend more on the ability to understand narrative structure and event sequences than on broader linguistic skills. Conversely, the fact that the association with language was specific to false belief performance underscores the potential role of language in providing a cognitive framework that supports the processing of mental states (De Villiers, 2005), challenging the notion that the link between ToM and language merely refers to the broader ability to understand stories. Furthermore, considering

that true belief tasks are simpler and less demanding, they may not require the same level of linguistic articulation and mental state processing as false belief tasks (Dennet, 1978). CELF-5, as a broad language measure, may not directly capture the skills most relevant to true belief reasoning. Because it assesses multiple language domains, its predictive power for true belief performance could be limited. In contrast, scenario comprehension may be more closely aligned with the demands of the task, as it reflects a child's ability to process structured events, which is essential for understanding true beliefs. Additionally, it is also necessary to consider that a potential ceiling effect in our data (where 64% of children achieved maximum scores on true belief tasks) might have influenced these results. In second-order false belief tasks, our analysis revealed that only language emerged as a significant predictor, while narrative competence did not contribute to task performance. This contrasts with first-order false belief tasks, where both language and scenario comprehension were significant predictors. These results may underscore the increasing reliance on advanced linguistic processing for interpreting and predicting others' beliefs in higher-order ToM tasks (Hughes, 2011). These findings prompt further investigation of how these skills contribute differently at different stages of ToM development. A possible explanation is that, as ToM reasoning progresses, reliance on direct scenario comprehension may diminish and be replaced by more complex language functions that provide greater explanatory power for understanding the mental states of others.

Overall, our findings emphasize the importance of considering both language skills and scenario comprehension in ToM assessments. Our study highlights the role of check questions, traditionally employed to ensure participants' understanding of task scenarios. Often, failure on these questions leads to participant exclusion or the reiteration of the story. This latter practice may obscure genuine comprehension difficulties, potentially inflating ToM performance by providing

additional scaffolding. In light of these considerations, it may be worthwhile to revisit how control questions are used in ToM studies. Rather than serving solely as screening tools, responses to these questions should be systematically recorded and analysed as indicators of scenario comprehension. This approach would allow researchers to discern whether ToM task performance reflects true mental state reasoning or is influenced by scenario comprehension. Reanalysing existing datasets with this perspective could yield valuable insights into the developmental trajectory of ToM and the foundational role of scenario comprehension. Furthermore, for future research a longitudinal approach is recommended to further investigate the development of narrative and language skills in ToM across different age groups. While the diverse demographics of our sample enhance generalisability, the relatively small sample size may limit broader applicability. Larger and more culturally diverse samples, clinical populations, and psychometric validation of the PTOMs scale are needed. Future studies should also include more specific mental state language assessments to refine our understanding of ToM development.

In conclusion, our study reaffirms the critical role of language in ToM and highlights the unique contribution of scenario comprehension. Both language and scenario comprehension contributed to ToM in first-order false beliefs, whereas only scenario comprehension predicted true beliefs. In contrast, second-order false belief reasoning was predicted only by language, highlighting their different roles in the development of ToM.

Acknowledgements

The authors would like to thank MSc students Sabereen Munye and Jimena Larrea Mijares for their support in collecting the data for this study. We would also like to thank Cecilia Mijares for the development of the PTOMs illustrations.

Fundings Statement

The authors have not received any fundings for the conduct of this study.

Data Statement

The PTOMs scale is publicly available at the following URL: <https://giannacocchini.wixsite.com/gcpage/neuropsychological-tests>. Data are available from the first author upon reasonable request.

References

- Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relationship between language and theory of mind development. *Developmental Psychology*, 35(5), 1311-1320. doi.org/10.1037/0012-1649.35.5.1311
- Baron-Cohen, S. (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21(1), 37-46. doi.org/10.1016/0010-0277(85)90022-8
- Bishop, D. V., & Adams, C. (1991). What do referential communication tasks measure? A study of children with specific language impairment. *Applied Psycholinguistics*, 12(2), 199-215. doi.org/10.1017/S0142716400009140
- Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, 77(1), B25-B31. doi.org/10.1016/S0010-0277(00)00096-2

- Capps, L., Kehres, J., & Sigman, M. (1998). Conversational abilities among children with autism and children with developmental delays. *Autism*, 2(4), 325-344. doi.org/10.1177/1362361398024002
- Dantzig, S., Pecher, D., Zeelenberg, R., & Barsalou, L. (2008). Perceptual processing affects conceptual processing. *Cognitive Science*, 32, 579–590. doi.org/10.1080/03640210802035365
- De Villiers, P. A. (2005). The role of language in the development of theories of mind: What deaf children tell us. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 266-297). Oxford University Press.
- De Villiers, J. G., & De Villiers, P. A. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 191-228). Psychology Press/Taylor & Francis.
- De Villiers, J., & Pyers, J. (1997). Complementing cognition: The relation between language and theory of mind. In Proceedings of the 21st Annual Boston University Conference on Language Development, (pp. 136-147). Cascadilla Press.
- Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, 1, 568-570. doi.org/10.1017/S0140525X00076664
- Gale, E., de Villiers, P., de Villiers, J., & Pyers, J. (1996). Language and Theory of Mind in Orally Deaf Children. In A. Stringfellow, D. Cahama-Amitay, E. Hughes, & A. Zukowski (Eds.), *Proceedings of the 20th annual Boston University Conference on Language Development* (Vol. 1, pp.423-448). Somerville, MA: Cascadilla Press.

- Gavilán, J. M., & García-Albea, J. E. (2011). Theory of mind and language comprehension in schizophrenia: Poor mindreading impairs figurative language comprehension beyond intelligence deficits. *Journal of Neurolinguistics*, 24(1), 54-69. doi.org/10.1016/j.jneuroling.2010.07.006
- Gerrig, R. J. (1993). *Experiencing narrative worlds: On the psychological activities of reading*. Yale University Press.
- Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science*, 6(3), 346-359. doi.org/10.1111/1467-7687.00289
- Happé, F. G. E. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, 66(3), 843–855. doi.org/10.2307/1131954
- Hollebrandse, B., Van Hout, A., & Hendriks, P. (2014). Children's first- and second-order false belief reasoning in a verbal and a low-verbal task. *Synthese*, 191, 321-333. doi.org/10.1007/s11229-012-0169-9
- Hughes, C. (2011). *Social understanding and social life: From toddlerhood to the transition to school*. Psychology Press.
- Lewis, C., & Mitchell, P. (Eds.). (1994). *Children's early understanding of mind: Origins and Development*. Lawrence Erlbaum Associates, Inc.
- Lundy, J. E. B. (2002). Age and language skills of deaf children in relation to theory of mind development. *Journal of Deaf Studies and Deaf Education*, 7(1), 41-56. doi.org/10.1093/deafed/7.1.41

- Nelson, K. (2005). Language pathways to the community of mind. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 26-49). Oxford University Press. doi.org/10.1093/acprof:oso/9780195159912.003.0002
- Nilsson K.K., de López K.J. (2016) Theory of mind in children with specific language impairment: a systematic review and meta-analysis. *Child Development*, 87(1), 143–153. doi.10.1111/cdev.12462
- Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology*, 62, 103-134. doi.org/10.1146/annurev-psych-120709-145406
- Mar, R. A. (2018). Evaluating whether stories can promote social cognition: Introducing the Social Processes and Content Entrained by Narrative (SPaCEN) framework. *Discourse Processes*, 55 (5-6), 454-479. doi.org/10.1080/0163853X.2018.1448209
- Mar, R. A., & Oatley, K. (2008). The function of fiction is to abstract and simulate social experience. *Perspectives in Psychological Science*, 3(3), 173-192. doi.org/10.1111/j.1745-6924.2008.00073.x
- Marinis, T., Andreou, M., Bagioka, D. V., Baumeister, F., Bongartz, C., Czipionka, A., Golegos, A., Peristeri, E., Skrimpa, V., Durrleman, S., & Terzi, A. (2023). Development and validation of a verbal and nonverbal first- and second-order theory of mind task battery. *Frontiers in Language Sciences*, 1, 1052095. doi.org/10.3389/flang.2022.1052095
- Miller, S. A. (2013). Children's understanding of second-order false beliefs: Content and assessment method comparisons. *Infant and Child Development*, 22, 649-658. doi.org/10.1002/icd.1810

Miller, C. A., Kail, R., Leonard, L. B., & Tomblin, J. B. (2001). Speed of Processing in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 44(2), 416-433. doi.org/10.1044/1092-4388(2001/034)

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: A meta-analysis of the relationship between language ability and false belief comprehension. *Child Development*, 78(2), 622-646. doi.org/10.1111/j.1467-8624.2007.01018.x

Oatley, K. (1999). Why fiction may be twice as true as fact: Fiction as cognitive and emotional simulation. *Review of General Psychology*, 3(2), 101–117. doi/10.1037/1089-2680.3.2.101

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308(5719), 255-258. doi.org/10.1126/science.1107621

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false beliefs: The case for a conceptual deficit. *British Journal of Developmental Psychology*, 5(2), 125-137. doi.org/10.1111/j.2044-835X.1987.tb01048.x

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that...": Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, 39(3), 437-457. doi.org/10.1016/0022-0965(85)90051-7

Peristeri, E., Baldimtsi, E., Vogelzang, M., Tsimpli, I. M., & Durrleman, S. (2021). The cognitive benefits of bilingualism in autism spectrum disorder: Is theory of mind enhanced and by what underlying factors? *Autism Research*, 14, 1695-1709. doi.org/10.1002/aur.2542

- Peterson, C. C., & Siegal, M. (1995). Deafness, conversation, and theory of mind. *Journal of Child Psychology and Psychiatry*, 36(3), 459-474. doi.org/10.1111/j.1469-7610.1995.tb01303.x
- Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian-English bilingual children between the ages of 5 and 7. *Applied Psycholinguistics*, 37(1), 49-67. doi.org/10.1017/S0142716415000417
- Wellman, H. M., Cross, D., & Watson, J. (2001). A meta-analysis of the development of theories of mind: The truth about false beliefs. *Child Development*, 72(3), 655-684. doi.org/10.1111/1467-8624.00304
- Wellman, H. M., & Lagattuta, K. H. (2000). Developing an understanding of mind. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed., pp. 21-49). Oxford University Press.
- Wiesmann, C. G., & Southgate, V. (2021). Early theory of mind development: Are infants innately altercentric? In M. Gilead & K. N. Ochsner (Eds.), *The neural basis of mentalizing* (pp. 1-10). Springer.
- Coret, M. C., & McCrimmon, A. W. (2015). Test Review: Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals—Fifth Edition (CELF-5)*. *Journal of Psychoeducational Assessment*, 33(5), 495-500. doi.org/10.1177/0734282914557616
- Woolfe, T., Want, S. C., & Siegal, M. (2002). Signposts to development: Theory of mind in deaf children. *Child Development*, 73(3), 768-778. doi.org/10.1111/1467-8624.00437
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of false beliefs in young children's understanding of deception. *Cognition*, 13(1), 103-128. doi.org/10.1016/0010-0277(83)90004-5

Schick B., De Villiers J., Hoffmeister R. (2007). Language and theory of mind: a study of deaf children. *Child Development*. 78 (2), 376-396. doi.org/10.1111/j.1467-8624.20007.01004.x.

Siegel, D. J. (1999). *The developing mind: Toward a neurobiology of interpersonal experience*. Guilford Press.

Tager-Flusberg, H., & Joseph, R. M. (2005). How language facilitates the acquisition of false belief understanding in children with autism. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 298-318). Oxford University Press. doi.org/10.1093/acprof:oso/9780195159912.003.0014

Tomblin J.B., Zhang X. (2006). The dimensionality of language ability in school-age children. *Journal of Speech, Language, and Hearing Research*, 49(6), 1193-1208. doi.org/ 10.1044/1092-4388(2006/086).