Journal of Autism and Developmental Disorders
https://doi.org/10.1007/s10803-021-05277-1

ORIGINAL PAPER

Second-Order False Beliefs and Linguistic Recursion in Autism Spectrum Disorder

Irina Polyanskaya[1] · Inge-Marie Eigsti[2] · Torben Brauner[1] · Patrick Blackburn[3]

Accepted: 4 September 2021

Abstract

This study investigates the role of recursive language and working memory (WM) in second-order false belief skills in Danish-speaking children with autism spectrum disorder (ASD; $n=62$; 8 females) and typical development ($n=41$; 15 females), ages 6–16. Second-order false belief skills correlated with receptive grammar, vocabulary, and age; sentential complement production predicted second-order false beliefs, controlling for age, receptive grammar and WM. Regressions showed that second-order false belief was associated with age across groups, but with sentential complements in the ASD group only. Second-order false belief skills improved in children who received training in *either* recursive phrases ($d=0.21$) or WM ($d=0.74$), compared to an active control group. Results suggest that false belief skills are entwined with both linguistic and executive functions.

Keywords Verbal mediation · Sentential complementation · Compositional semantics · Theory of mind · Second-order false belief

First-Order and Second-Order Mental State Understanding

Theory of Mind (ToM) involves understanding that others have mental, epistemic, and emotional states, understanding the content of these states, and predicting behavior based on this knowledge. The development of ToM knowledge is robustly associated with language and executive abilities, as well as social communicative differences, in autism spectrum disorder (ASD). One *specific* component of ToM, second-order mental state reasoning, is thought to require knowledge of a specific component of linguistic structure: sentential complements. The present study uses correlational analyses and a training approach to test the roles of recursive sentential complementation and working memory in the acquisition of second-order ToM.

First-order mental state reasoning requires understanding that other individuals have their own distinct beliefs, or representations, of the world: *Molly believes that it is raining; Desmond knows that the rain has stopped*. This form of reasoning is critical for multiple aspects of social communication, such as deception (Bosco & Gabbatore, 2017; Sodian & Frith, 1992; Sodian et al., 1991), recognizing surprise (Baron-Cohen et al., 1993), pretend play (Leslie, 1987), and protodeclarative pointing (Baron-Cohen, 1989b). Second-order mental state reasoning skills emerge later in development, and require understanding that one person's belief can be about another person's belief: *Desmond thinks that Molly believes that it is still raining*. Such beliefs are embedded one inside another; that is, they are recursive (Braüner et al., 2020). Second-order mental state reasoning skills seem to be critical for complex pragmatic language skills such as understanding idioms (Caillies & Le Sourn-Bissaoui, 2013), irony, sarcasm, and metaphor (Bosco &

✉ Inge-Marie Eigsti
inge-marie.eigsti@uconn.edu

Irina Polyanskaya
irinapol.translations@gmail.com

Torben Brauner
torben@ruc.dk

Patrick Blackburn
patrickb@ruc.dk

1 Department of People and Technology, Roskilde University, Building 08.2, P.O. Box 260, 4000 Roskilde, Denmark

2 Department of Psychological Sciences, University of Connecticut, 146 Babbidge Road, U-1020, Storrs, CT 06269, USA

3 Department of Communication and Arts, Roskilde University, Roskilde, Denmark

Gabbatore, 2017; Filippova & Astington, 2008), embarrassing statements (Hillier & Allinson, 2002), coordinating with peers (Grueneisen et al., 2015), reasoning about evidence (Astington et al., 2002), maintaining consistency between a lie and subsequent statements (Talwar et al., 2007), and moral judgement (Fu et al., 2014).

While we can understand another person's representation of the world when it is either true or false, *measuring* these representations often requires situations that are inconsistent with reality. For example, if it is raining, and Molly believes that it is raining, and Desmond reports that *Molly believes that it is raining*, this is an ambiguous demonstration of Desmond's ability to represent Molly's beliefs; he may simply be describing reality. Accordingly, most ToM research relies on false belief (FB) paradigms, in which the participant must demonstrate her knowledge that a second person holds a belief that is inconsistent with reality. Reporting that *Molly believes that it is raining* on a sunny day demonstrates the speaker's grasp of the distinction between representation and reality. In *unexpected transfer* paradigms (e.g., Edwards & Low, 2017), an object is moved from an initial location (known to the agent), unbeknownst to that agent; the child's task is to describe the agent's belief about the location of the object. In *unexpected contents* paradigms (e.g., Rubio-Fernandez, 2019), a container that typically holds a particular type of content (e.g., Smarties candies in a Smarties Tube) is revealed to contain something different (e.g., paper clips). After seeing the contents, the child is asked what someone else (who has *not* seen the unexpected contents) will believe it contains. Both tasks require the participant to understand and report someone else's belief when it differs from the child's own knowledge. This ability emerges by around age 4 years (Wellman et al., 2001).

Second-order reasoning requires more complex paradigms. To investigate this form of false belief understanding, paradigms typically present brief anecdotes featuring characters who hold conflicting beliefs, and ask the child to report that Character A holds a false belief *about* Character B's belief by answering questions about those beliefs and providing an explanation of the responses. For example, in the "Ice Cream" story (Perner & Wimmer, 1985; see also Online Appendix A), Frederik, Katrine and an ice cream seller are in the park. Frederik goes home to get money to buy ice cream. After he leaves, the ice cream seller tells Katrine that he is going to the city center. En route, the ice-cream man encounters Frederik and tells him the same thing. Later, Katrine sees Frederik's mother, who tells her that Frederik has gone out to buy ice-cream; Katrine goes to look for Frederik. Participants are asked 'Where does Katrine think that Frederik has gone?' and are asked to justify their answer.

Mental State Understanding in ASD

Theory of Mind research has been shaped by studies of autism spectrum disorder (ASD). A number of studies indicate that children with ASD pass first-order false belief tasks several years later than neurotypical children, and that there are meaningful links between ToM abilities and the broader symptoms of ASD (Baron-Cohen, 2000; Tager-Flusberg, 2001). The original ("mindblindness") version of this theory, suggesting that ToM impairments are universal and specific to ASD, has been discounted on several grounds (Brynskov et al., 2015); they are not specific to ASD, as Deaf children with delayed access to language show similar impairments(Schick et al., 2007); ToM cannot account for other characteristics of ASD (Happé, 2003; Peterson et al., 2005); symptoms of ASD emerge prior to ToM delays (Tiede & Walton, 2020); and some autistic individuals display intact ToM abilities (Tager-Flusberg, 2001). Between 15 and 60% of children with ASD are found to pass first-order FB tasks at the typical developmental age (Happé, 1995). Second-order false belief tasks present a more mixed picture; nonetheless, in essentially every study we could find, either the entire sample or a subsample of children with ASD displayed *intact* second-order false belief skills (Baron-Cohen, 1989a) (the sole exception was Baron-Cohen, 1989a, which reported impairments in all participants).

There are several potential explanations for the finding that some (but not all) individuals with ASD show delayed or impaired ToM. First, it has been proposed that ToM tasks can be solved via two distinct methods: a social-perceptual (implicit) method or a social-cognitive (explicit) method. Children with ASD may struggle with the former and relying more than their neurotypical peers on the latter (Tager-Flusberg, 2001; Tager-Flusberg & Joseph, 2005). The social-perceptual/implicit route involves online understanding of mental states based on social stimuli such as faces, body, and voices, while the social-cognitive/explicit route involves information processing and deliberate, conscious reasoning skills. Although mental-state inference also depends on perceptual cues, its development starts when children are beginning to talk and reason about epistemic states (Tager-Flusberg & Joseph, 2005). Of course, there is infant research (using looking time and violation of expectation paradigms) demonstrating false belief understanding as early as 15 months (Buttelmann et al., 2009; Scott & Baillargeon, 2009; Southgate et al., 2010; Surian et al., 2007) and even at age 7 months (Kovács et al., 2010). The distinction between implicit and explicit false belief understanding implies that children younger than 4 years of age rely on *implicit* false belief understanding, and that improvements in language, executive and processing skills at around age four help make such false belief understanding explicit, enabling children to succeed in verbal elicited-response tasks (Low, 2010; Low & Perner, 2012).

A second explanation suggests that general cognitive skills and experience may allow children to learn a restricted set of solutions to a given problem, to compensate for difficulty solving such problems via implicit intuition. Examples of such compensatory cognitive skills include logical reasoning (Frith, 2004), application of rule-bound, cognitively acquired heuristics (Williams & Happé, 2009), and syntactic skills (Lind & Bowler, 2009; Tager-Flusberg & Sullivan, 1994b). A study of Danish-speaking schoolchildren with and without ASD found significant correlations between verbal IQ and social-cognitive (explicit) mental state inference tasks, but no correlations with the Eyes in the Mind task, thought to tap social-perceptual (implicit) processes (Kaland et al., 2008). For the typically developing group, verbal IQ and ToM performance were uncorrelated. Such findings suggest that individuals with ASD may rely more on general logical reasoning and linguistic skills to solve advanced ToM tasks. As such, ToM performance may be differently intertwined with cognitive and language abilities in individuals with ASD.

The current study provides a strong test of this second explanation. We investigate second-order false belief reasoning abilities and their association with relevant features of linguistic structure (namely, syntactic recursion and semantic compositionality) and with domain-general (working memory) processes; we also test whether training on these critical processes shapes second-order false belief processing. As a novel contribution to the literature, we hypothesize that the mastery of *recursive sentential complementation* is a compensatory skill that helps children with ASD to solve theory of mind problems.

Associations Between Mental State Understanding and Language Acquisition

Developmental research has established strong links between development of first-order false beliefs and language skills development in both ASD and neurotypical development. Both longitudinal and training studies have shown a causal relationship between first-order false beliefs and semantic and syntactic skills (Astington & Baird, 2005; de Villiers, 2007; Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003; Miller, 2006; Milligan et al., 2007). "Semantics" here encompasses words such as thinking, hoping, or wishing, which describe unobservable mental acts. More frequent or developmentally earlier production of mental states terms is associated with better ToM performance (Farrar & Maag, 2002; Guajardo & Watson, 2002; Peskin & Astington, 2004).

A syntactic construction called "sentential complementation," central to this paper, has long been linked to first-order false belief understanding; indeed, the theory of "linguistic determinism" originally proposed that the comprehension of sentential complementation is required for success on first-order false belief tasks (de Villiers & Pyers, 2002). In sentential complementation, one of the arguments of a verb is a complement clause: In "Molly believes that it is raining," "that it is raining" is a complement clause in a singly embedded sentential complement sentence. Reflecting the focus on second-order false belief, this paper focuses on *doubly embedded* sentential complements such as "Desmond says that Molly believes that it is raining." Here one sentential complement clause (about Molly) is embedded in another (about Desmond); this embedding is recursive, in that one could embed a theoretically infinite number of clauses: "Molly says that Desmond believes that Jack wishes that Keisha said that it is raining." While the strong version of linguistic determinism is likely incorrect, sentential complementation may provide a well-defined index of multiple aspects of development, including working memory and syntactic knowledge.

The correlation of sentential complement comprehension and false belief understanding has been tested in multiple languages, including English (de Villiers & Pyers, 2002), American Sign Language (Schick et al., 2007), and Danish (Boeg Thomsen, 2016). These studies use both unexpected transfer and unexpected contents tasks to assess first-order false belief understanding. Mastery of complementation predicted *explicit* but not *implicit* FB understanding in three and four year-olds (Low, 2010). Further, intervention studies with preschoolers reported that training in sentential complementation contributed to improvements in first-order false belief understanding (Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003). One training study tested the impact of a four-week training on sentential complements on false belief reasoning in children with ASD, developmental language disorder, or typical development, ages 2–11 (Durrleman et al., 2019). Results clearly indicated a benefit of complement training on both the production of sentential complements, and on ToM reasoning.

The importance of syntactic knowledge, and specifically sentential complementation, for FB reasoning has been supported by studies of ASD (Tager-Flusberg & Sullivan, 1994a, 1994b). In the absence of intuitive or implicit understanding, children with ASD may use complementation to reason (or "hack") solutions to first-order false belief tasks (Durrleman & Franck, 2015; Durrleman et al., 2016; Lind & Bowler, 2009; Tager-Flusberg & Joseph, 2005).

Second-Order False Belief and Language

While many studies document the critical role of sentential complementation in first-order false belief reasoning, there is limited empirical work investigating the role of language development in second-order false belief (SOFB) reasoning. Evidence against the central role of language was reported in a study of Dutch-speaking children showing that double-embedded sentences (*X says/thinks that Y says/thinks that P*) were spontaneously produced at all levels of false belief mastery, including by children who failed second-order FB questions (Bogaerds-Hazenberg & Hendriks, 2016). Furthermore, children who had mastered SOFB produced both single and mixed embeddings, suggesting that recursive language was not a prerequisite for recursive FB reasoning, contesting the notion of linguistic determinism. In contrast, Lockl and Schneider (Lockl & Schneider, 2007) suggested that general language abilities were fundamental in both first-order and SOFB, for different reasons; in the first-order case, language supports comprehension of belief representation; in the second-order case, language support *expression* of belief representation, reflecting increased processing demands.

In several studies, overall linguistic competence (including vocabulary) predicted SOFB abilities (Astington et al., 2002; Filippova & Astington, 2008). General semantic, syntactic abilities and complex working memory were *each* found to be significant contributors to SOFB performance in a study of 60 preschoolers ages 38–71 months, using the birthday puppy stories (Hasselhorn et al., 2005). Vocabulary was the most important limiting factor, and findings suggested that vocabulary and working memory were each independent contributors to ToM development. Arslan and colleagues reported that the main predictor of SOFB reasoning in Turkish-speaking children was complex working memory, assessed using a listening span task (Arslan et al., 2017). Findings were explained in terms of a serial processing bottleneck, in which SOFB performance reflects the demands of holding multiple representations in working memory while also serially processing them.

Linguistic Recursion

The current study focuses on the shift from mastery of single to double embedded sentential complements. Linguistic structures can be described as *constituents*, or groups of words (such as clauses) that function as a single unit. Constituents can contain a constituent of the same type (Jackendoff, 1990), as in *Linda likes that Henry learned that Susan hopes that Mary says that John thinks that the car is cool*, which has five levels of embedding, and could be diagrammed as follows: *(^{5}Linda likes that (^{4}Henry learned that (^{3}Susan hopes that (^{2}Mary says that (^{1}John thinks that (^{0}the car is cool)))))).* Embeddings can be viewed either *structurally* (as above), or *algorithmically*, as in: Add a prefix of the form Proper-Name + Verb + “that” to a zero-order sentence.

While we often think of recursive beliefs in structural terms, where one belief can be nested inside another, the algorithmic perspective is informative about how one accesses the information embedded inside these structures. The sentences we use to talk about beliefs are often embedded sentential complements with a recursive structure: the zero-order competency reflects a basic representation of the world, which is altered by the awareness that others may hold different (possibly false) representations: *Emma believes that the ball is in the basket.* This leads to the realization that there can be beliefs about beliefs. Here, we probe whether second-order false belief reasoning depends primarily on linguistic or on domain-general processes.

Working Memory

Working memory involves holding information in mind while performing one or more mental operations. For example, we use working memory to re-order items while holding them in mind and seeing how they relate to one another; this is critical for reasoning and problem solving. Working memory is conceptualized as an executive function (Shah & Miyake, 1999), and has been found to correlate with performance on false belief tasks (Arslan et al., 2017; Gordon & Olson, 1998; Hughes, 1998; Hughes & Ensor, 2007; Keenan et al., 1998). Some studies suggest that executive functions precede Theory of Mind development (Carlson et al., 2004; Perner & Lang, 1999); the ability to monitor one’s actions may precede Theory of Mind understanding (Russell, 1997).

The Present Study

As reviewed here, language acquisition, and specifically syntax, plays an important role in false belief reasoning for children with ASD. Previous studies have examined first-order false belief understanding; the current study focused on the role of syntax in *second-order* false belief development, testing the hypothesis that children with ASD depend on mastery of recursive sentential complementation (and linguistic recursion more generally) in order to perform second-order mental reasoning. The work reported here utilizes two approaches. Study 1 investigates associations among second-order false beliefs, linguistic recursion, and working memory in children with and without ASD. We collected data on language skills and working memory to identify profiles of linguistic, second-order false belief, and working memory skills. This study also provided Time 1 data for Study 2, in which children with ASD were randomly assigned to one of three interventions: (1) recursive language training, (2) working memory training, and (3)

Table 1 Participant information: Study 1

Measures	ASD group	NT group	F/χ^2	*P*	*Cohen's d*
Total n; male/female	62; 54/8	39; 26/15	28.48	< 0.001	
Chronological age	11.8 (2.5); 6.1–16.9	8.4 (1.9); 4.5–11.6	54.64	< 0.001	1.53
Verbal comprehension—WISC	109 (12.6); 87–138 ($n=57$)	N/A			
Working memory—WISC (RS)	30.9 (4.6); 16–40 ($n=61$)	28.5 (6.7); 13–42 ($n=35$)	4.25	0.04	0.42
TROG grammar comprehension	105 (7); 89–118 ($n=59$)	N/A			
Social Responsiveness Scale (T-score)	61 (12); 38–93 ($n=53$)	46 (6); 39–55 ($n=12$)	19.44	< 0.001	1.58

Data are presented as M(SD); range. Results for the WISC and TROG measures are shown as standard scores, which have a mean of 100 and SD of 10; the SRS is reported as T-scores, with $M(SD)=50(10)$, and where higher scores suggest greater severity of symptoms

an interaction-only (control) condition. This training study investigated causal relationships between recursive structures, working memory, and SOFB reasoning. This is the first study to our knowledge in which children with ASD were trained on recursive linguistic structures, and the first to juxtapose linguistic and working memory training.

Study 1: Associations of Linguistic Recursion, Working Memory, and Second-Order False Belief Reasoning

The goal of Study 1 was to test linguistic recursion as a predictor of SOFB mastery, controlling for age, working memory and receptive grammar, in children with and without ASD.

Methods

Procedures

Participants with ASD aged 6–16 years were recruited from six schools for children with special educational needs in the Sjælland region of Denmark (which includes Copenhagen) and by word of mouth. Inclusion criteria were: formal diagnosis of ASD; Danish as first language; and the ability to sit through an assessment session (as evaluated by their teachers) and a training program (as evaluated by the psychologist who administrated the assessment). The only exclusion criterion was psychosis. All parents gave informed written consent; all children gave verbal assent prior to each session. The study was reviewed by the National Committee on Health Research Ethics, which determined that it met ethical standards. There was no financial compensation for participation.

Initial Assessment

Children completed the Danish version of the Wechsler Intelligent Scale for Children (WISC-IV) Verbal Comprehension Index (including subtests measuring vocabulary and verbal concept formation, knowledge of concepts, and the ability to apply reasoning skills to verbally-presented information) and Working Memory Index (including subtests measuring the ability to memorize new information, to hold it in short-term memory, and to concentrate and manipulate that information to produce some result) (Wechsler, 2003). They also completed the Danish version of the Test for Reception of Grammar—2nd Ed (TROG; Bishop, 2009), as an assessment of grammar comprehension.

These tests provided quantitative data for the secondary inclusion criteria. Participants were included if they had standard scores of 80 or higher on WISC Verbal Comprehension and Working Memory indices and on the TROG. A total of 62 children met these criteria and were included in Study 1. These participants completed a set of four SOFB stories and a set of five recursive embedding stories (described below), in random order. Teachers completed the Danish version of the Social Responsiveness Scale (SRS) questionnaire, to provide a quantitative assessment of autistic social impairments (Constantino, 2003). Testing was carried out at school or home, in a private room with a table. All children were tested individually over two or three sessions of between 45 and 60 min per session, depending on the child.

Neurotypical (NT) children were recruited from local schools ($n=25$) and via word of mouth ($n=16$). Given the likely emergence of first-order false belief skills at age 4 years (as reviewed above), children as young as 4;0 were recruited; the age range was limited at 12 years to facilitate age-matching with the ASD sample. Inclusion criteria were: Danish as first language, and a history of typical development per teacher and parent report. Due to time constraints at school, children in the NT group did not complete the WISC Verbal Comprehension or TROG standardized assessments, but completed all other assessments. All children were tested individually in a private room during a single 30–45 min session.

Demographic characteristics are shown in Table 1. Missing data were entered as missing values, and treated pairwise. Missing scores included the WISC, due to concerns about re-testing ($n = 2$); and other missing scores due to poor compliance with testing (WISC, $n = 3$), TROG ($n = 3$) and working memory ($n = 1$). The Teacher SRS was missing due to time constraints ($n = 27$).

Measures

Second-Order False Belief Tasks

The four second-order false belief tasks were Danish translations of four second-order false belief stories used in prior research: Ice-Cream (Perner & Wimmer, 1985), Birthday Puppy (Sullivan et al., 1994), Sally Anne (Baron-Cohen et al., 1999), and Bake-Sale (Hollebrandse et al., 2014). The Ice-Cream and Puppy stories were taken from the Danish version of the validated Dutch ToM Storybook (Blijd-Hoogewys et al., 2008), translated and validated for Danish (Clemmensen et al., 2016). The Sally Anne and Bake Sale stories were translated by the first author; English translations of accompanying pictures for all tasks are shown in Online Appendix A, and Danish versions are available upon request. Each story was read aloud, with accompanying pictures; participants then answered multiple questions including one or two memory questions, one or two first-order false belief questions, one or two second-order false belief questions, and one justification question. The SOFB and justification questions required a child to understand what one protagonist believed about another protagonist's understanding of reality, when the first protagonist's beliefs were false. To minimize the use of sentential complements, direct speech was used whenever possible, and the test questions contained no double embedded constructions. This ensured that participants did not need to process any *linguistic* double embedded structures; of course, correct responding required doubly-embedded (second-order) false belief *reasoning*. For example, questions for the Ice-Cream story (described above, and see Online Appendix A) were:

> Memory questions: *Has Katrine heard what the ice-cream man said to Frederik? Has Frederik heard what the ice-cream man was saying to Katrine?*
> SOFB questions: *Where does Frederik think Katrine has gone? Why does Frederik think that?*
> Justification question: *Why does he think that?*
> Control question 3 (reality): *Where did Katrine really go?*

SOFB responses were scored only if participants answered the control questions correctly. The justification questions were scored as in previous studies (Arslan et al., 2015): Incorrect explanation = 0; correct reasoning with no reference to mental states ("Because that is what the ice-cream man told him") = 1; correct reasoning including one mental state verb ("Because he does not know that she saw the ice-cream man") = 2; correct reasoning using two or more mental verbs ("Because he thinks that she does not know that the ice-cream man left the park") = 3. Scores for individuals were correlated in a pairwise fashion across the four stories (all p's < 0.05); as such, response scores were collapsed across the four stories, for a maximum of 18 SOFB points.

Recursive Embedding Task

We could not find any validated Danish test of the comprehension or production of recursive syntactic constructions. Thus, we developed the Recursive Embedding Task (RET), piloted with a sample of 240 typically developing Danish-speaking children; the final version was validated with 70 typically developing Danish-speaking children (36 girls and 34 boys) and 15 adults. Further description of the RET design and validation can be found elsewhere (Polyanskaya, 2019).

The RET consists of five short stories involving both a single-embedded and a double-embedded question, each with a colorful illustration (Hollebrandse & van Hout, 2015; Hollebrandse et al., 2008); Fig. 1 presents a sample item. Each story included truth-value contrasts between clauses (de Villiers et al., 2014). Note that a single complement answer to Q2 (*Mom says that the rabbit has run away*) would contradict the fact that Mom can see the rabbit in Sofie's arms (the truth value contrast). Participants received 1 point for a correct answer to the double embedding question. The English translation of the RET test is shown in Online Appendix B; the maximum score was 5.

Statistical Analyses

All study variables were inspected to test whether they met standard statistical assumptions of normality and heteroscedasticity. Several did not, even after transformations; this included the RET and SOFB measures. For analyses including these variables, nonparametric Kendall's tau-b (τ_b) tests were used. The interrelationships among study variables were tested using bivariate correlational analyses. Given the large age differences between groups, we created subgroups of ASD and NT participants, matched for age, by excluding the oldest ASD and youngest NT participants until both subgroups (ASD $n = 30$; NT $n = 30$) were normally distributed for age ($p > 0.10$) and group differences in age were not statistically significant ($p > 0.08$). Subgroup differences were evaluated using t-tests for normally-distributed measures (WM and age) and Mann–Whitney tests for non-normal measures (SOFB, RET). Given the relatively limited literature on SOFB development, we also present performance data for the full unmatched samples. Regression analyses were used to test the role of WM, and RET in

Fig. 1 Sample item, recursive embedding test (RET)

Dad is calling home from work and says to Mom: "*I couldn't find our rabbit. It has run away*." Sofie is standing next to Mom and holds the rabbit in her arms. Mom tells Sofie that Dad says that the rabbit has run away.
Q1. What did Dad say?
Q2. What did Mom say to Sofie?

Table 2 Theory of mind/second-order false belief (SOFB), recursive embedding test (RET) and working memory scores for the entire sample and for age-matched subgroups

	ASD, *n* = 62	NT, *n* = 39	*z/F*	*p*	*Cohen's d*
SOFB	8.8 (4.3); 1–17	6.3 (4.5); 0–17	7.65	0.007	0.57
RET	3.1 (1.6); 0–5	2.7 (1.9); 0–5	1.18	0.28	0.23
WISC WM (RS)	30.9 (4.6); 16–40	28.5 (6.7); 13–42	4.25	0.04	1.32
	ASD, *n* = 30	NT, *n* = 30			
Age	9.8 (1.3), 6.1–11.7	9.2 (1.4), 6.8–11.6	3.42	0.07	0.44
SOFB	7.3 (4.5), 1–17	7.7 (4.1), 0–17	0.18	0.68	0.68
RET	2.9 (1.5), 1–5	3.1 (1.9), 0–5	0.38	0.54	0.54
WISC WM (RS)	29 (5), 16–40	30 (6), 17–42	0.25	0.62	0.18

SOFB abilities; regressions were conducted separately for ASD versus NT age-matched groups. Because age might play a significant role in performance, and because groups were not well-matched for age, this variable was included in all regressions.

Results

First, an analysis of the entire sample of ASD/NT participants indicated no group differences in sentential complements (RET), but group differences in second-order false belief task performance (SOFB) and working memory (WM) abilities; see Table 2. Results indicated that the ASD group, which was significantly older, had significantly higher SOFB and working memory scores compared to their non-autistic peers. Comparing the age-matched subgroups, there were *no differences* on second-order false belief task performance (SOFB), sentential complements (RET), or working memory (WM).

Next, we probed for correlations among second-order false belief, recursive complements, grammar comprehension (TROG), WISC verbal comprehension, WISC working memory, age, and autistic severity (SRS). These correlations are presented in Table 3, with the ASD group above and the NT group below the diagonal. Second-order false belief was significantly correlated with all language measures (recursive complements, WISC verbal comprehension and TROG grammar comprehension; the latter two measures were available in the ASD group only) and with age and working memory, in both groups. Autism severity (SRS) did not correlate with other variables.

We used hierarchical regression analysis to investigate the contribution of recursive complements mastery (RET

Table 3 Correlations among Study 1 measures

	Age	SRS	SOFB	RET	TROG	VC	WM
Age	–	− 0.16	0.44***	0.13	**0.30****	**0.56*****	**0.30*****
SRS		–	− 0.16	0.11	0	**− 0.24***	0.25
SOFB	*0.54****	*− 0.23*	–	**0.30****	**0.27****	**0.43*****	**0.36****
RET	*0.46****	*0.20*	*0.38***	–	0.18	**0.23***	0.18
TROG					–	**0.46*****	**0.30****
VC						–	**0.45*****
WM	*0.50****	*0.00*	*0.58****	*0.44****			–

SRS autism severity Social Responsiveness Scale, *SOFB* second-order false belief, *RET* recursive embedding task with sentential complements, *TROG* Test of the Reception of Grammar, *VC* WISC verbal comprehension index, raw score, *WM* WISC working memory raw score

*$p<0.05$; **$p<0.01$; ***$p<0.001$; correlations calculated using Kendall's tau. Correlations for the full ASD group are shown above the diagonal; the full NT group is shown below the diagonal, in *italic font*

Table 4 Full ASD sample: second-order false belief (SOFB) abilities as a function of age, working memory (WM), receptive syntax (TROG scores) and recursive complement knowledge (RET)

Variable	Step 1			Step 2			Step 3		
	B	*SE*	*P*	*B*	*SE*	*p*	*B*	*SE*	*p*
Constant	− 0.94	2.57	0.71	− 10.37	5.68	0.07	− 9.68	5.53	0.08
Age	**0.81****	**0.21**	**< 0.001**	0.43	0.25	0.09	0.46	0.24	0.07
WM				0.26	0.13	0.05	0.22	0.13	0.09
TROG				0.33	0.36	0.36	0.23	0.35	0.5
RET							**0.63***	**0.31**	**0.04**
R^2	0.21			0.29			0.34		
F	**14.93**		**< 0.001**	**7.31**		**< 0.001**	**6.89**		< 0.001
ΔR^2	0.21			0.07			0.05		
ΔF	**14.93**		**< 0.001**	2.97		0.06	**4.25**		**0.04**

Significant variables are highlighted in bold font

B unstandardized regression coefficient, *SE* standard error of the coefficient. Data represent mean raw scores

*$p < 0.05$; **$p < 0.001$

score) to second-order false belief (composite SOFB score), controlling for individual differences in age, general grammar, and working memory, within the full ASD sample. The assumption of normal distribution of residuals was met, as assessed by P–P Plot, and there was no evidence of multicollinearity. Independent variables were added in three steps: age was added at Step 1; WISC working memory and TROG raw scores were added in the second step; and RET score was added at the third step. WISC Verbal Comprehension scores were removed because the participants as a group had very strong verbal skills, with mean scores in the high end of the average range (and nearly a third with above-average range scores). As such, including verbal IQ in the analysis would likely swamp the effect of a more specific language variable such as recursive complements (see, e.g., Dennis et al., 2009). Model results are shown in Table 4.

This regression analysis indicates that, in the ASD sample, recursive complements (RET) accounted for significant variance in SOFB reasoning, controlling for age, working memory and receptive syntax. With age alone, the model explained 21% of the SOFB variance, and age was a significant predictor. With age, WM and TROG, the model explained a further 7% of the variance in SOFB understanding. Individual beta weights suggested that WM and age were each marginally significant, and TROG was not significant. With RET scores, the model explained an additional 5% of the variance with a large effect size (Cohen's $f^2 = 0.51$). The individual beta weights suggested that, in the full model, recursive complement comprehension was significant, age and WM were marginally significant, and TROG was not significant.

Regression with ASD and NT Subsamples

To contrast the distinct contributions of age, RET and WM to SOFB in children with and without ASD, the same multiple regression analyses were conducted in the age-matched ASD and NT subgroups. The models were significant predictors of SOFB scores for both the ASD group, $F(3,25) = 6.82$, $p = 0.002$, $R^2 = 0.45$, with a large effect

Table 5 Predictors of second-order false belief in age-matched subgroups

ASD group ($n = 30$)				NT group ($n = 30$)		
Variable	B	SE	p	B	SE	P
Constant	− 12.61	5.24	0.02	− 12.53	4.05	0.005
Age	1.84	0.59	**0.004**	1.41	0.52	**0.01**
RET	1.11	0.47	**0.03**	− 0.01	0.38	0.97
WM	− 0.05	0.16	0.77	0.25	0.12	**0.04**

Significant variables are highlighted in bold font

$n = 30$ per group

B unstandardized regression coefficient, *SE* standard error of the coefficient

(Cohen's $f^2 = 0.81$), and for the NT group, $F(3,26) = 11.28$, $p < 0.001$, $R^2 = 0.57$, again with a large effect (Cohen's $f^2 = 1.32$); see Table 5. Age was a significant predictor in both groups; further, the regression coefficients for age were similar, suggesting similar developmental trajectories in the two samples. In contrast, recursive embedding complements accounted for significant variance in false belief *only* in the ASD subgroup. Working memory was a significant predictor *only* in the NT subgroup. When working memory was further split into simple (short-term) and complex (the ability to manipulate information kept in short-term memory) functions, complex working memory was the significant predictor (see Polyanskaya, 2019 for further details).

Interim Discussion

Study 1 examined working memory, age, and syntactic abilities as predictors of second-order theory of mind processing in a large cohort of children with ASD, and in age-matched subgroups of children with ASD and typical development. Results indicated similar abilities on all variables, for the age-matched groups. This failure to find group differences in performance of false belief ToM tasks in children with ASD with a mean age of 11 years (range, 6–16 years) is consistent with numerous studies, reviewed above. Furthermore, regression analyses suggested that age, recursive knowledge, and working memory together contributed to variance in second-order theory of mind processing. One intriguing group difference was that recursive embedding complements abilities were important in ToM processing *only* in the ASD subgroup. This result is consistent with the hypothesis that comprehension of recursive complements scaffolds second-order theory of mind development in children with ASD, in contrast to typically developing children, for whom these processes are relatively independent. Further, the finding that working memory was critical for ToM processing in the NT subgroup, is consistent with other studies as reviewed in the Introduction. It is possible that limits in power and variability precluded us from capturing the role played by other predictor variables.

Study 2: The Impact of Training in Recursive Embedding and Working Memory on Second-Order False Belief Processing

Study 2 provided a stringent experimental test of the relationships among recursive embedding, working memory and second-order false belief reasoning abilities in autism.

Methods

Participants

Study 1 participants with ASD were included in the training study if they scored below chance (nine or fewer points out of 18) in the SOFB tasks; children who already displayed adequate SOFB reasoning skills were unlikely to show measurable benefits from training. A total of 27 children met this criterion and were randomly assigned to one of three training conditions: recursive embedding training, working memory training, or interaction only (control condition). The exception to random assignment was three children who were allocated to the working memory group at the request of their classroom teachers. Age, verbal comprehension, grammar comprehension, working memory, SRS, SOFB and RET baseline scores did not differ across training conditions. Table 6 shows participant characteristics.

An ANOVA analysis tested whether participants differed in demographic characteristics shown in Table 6 as a function of school district; there was one significant difference, in age, $F(5,26) = 2.987$, $p = 0.03$, with children from one school being older. The districts did not differ in other characteristics, and as a result, schools were collapsed in subsequent analyses.

Overview of Training Procedures The training comprised five individual 30-min sessions, generally scheduled across

Table 6 Group characteristics for each of the three training conditions

	Recursive embedding, n = 9	Working memory, n = 10	Interaction only, n = 8	*H*	*P*
Age (years)	10.65 (2.26), 8.17–15.75	10.87(1.56), 8.83–13.92	11.02 (1.78), 8.42–13.58	1.03	0.6
SRS	65.4 (10.6), 52–80	60.7 (5.8), 53–70	62.3 (10.0), 48–75	0.83	0.66
TROG	103 (6), 96–113	107 (9), 91–118	102 (5), 96–111	2.52	0.28
WISC verbal comprehension	102 (11), 85–116	108 (12), 93–128	109 (13), 91–132	1.01	0.61
WISC working memory	94 (11), 80–110	91 (7), 83–101	94 (10), 83–110	0.56	0.76
SOFB, pre-training score	4.9 (1.9), 1–8	6.2 (2.2), 2–10	5.4 (3.1), 1–10	1.71	0.43
RET, pre-training score	3.0 (1.7), 1–5	3.1 (1.6), 1–5	2.5 (1.5), 1–5	0.75	0.69

TROG, WISC and SRS are standard scores. All data are shown as *M*(*SD*), range

five consecutive days except when school activities or sickness intervened; in all cases, the training was completed within two weeks. One examiner (the first author) conducted all of the training sessions across conditions; this approach led to enhanced treatment fidelity, as did the examiner's reliance on manualized scripts (as described below). Within 3 days after the end of training, participants once again completed the SOFB, RET and working memory tasks. The pre-and post-training tasks were identical in structure, but incorporated different content, to reduce the impact of practice and familiarity. After children completed all study measures, teachers and parents received detailed reports describing each child's results.

Across training conditions, the trainer presented the concept of "the brain as a muscle" that can be strengthened; this short discussion was motivating for many participants. During all sessions, children were given a red "I need a break" card that they could utilize as needed. Several children made use of this option, later returning to the activity. Across the three conditions, the participants seemed similarly engaged and interested in participating.

Recursive Embedding Training The recursive embedding program was developed de novo for this study, utilizing ideas from *compositional semantics*, which investigates how the meaning of a whole is built out of the meaning of its parts, in concert with syntactic structure (Janssen, 2020, Fall; Szabó, 2020, Fall). Specifically, syntactic structure provides a 'frame,' and the semantics of the top-level constituent are constructed by combining the meanings of components using the pattern provided by the syntactic frame. The training was designed to convey four ideas concerning compositionality, namely: (1) *Multiple linguistic constituents of the same type may be combined;* (2) *Constituents may be embedded inside each other;* (3) *Changing the order of embedded components changes the meaning of the whole;* and (4) *The number of embedded constituents is theoretically infinite.* These concepts were communicated in terminology appropriate for children; a translation is included in Online Appendix B.

The aim of the training was to expose children to a large number of sentences with embedded constituents that do *not* involve mental state terms, to get children to produce such sentences, and to provide the children with explicit awareness of the four principles. During training, children heard and produced multiple examples of recursively embedded possessive noun phrases and sentential complements (none of which involved mental vocabulary). The examples were presented in concert with, and mentioned, images of familiar characters (e.g., Harry Potter) and scenes from popular books and magazines. The participants were asked to produce sentences to describe the pictures, using the rules that had been discussed. On the final day, children were asked to repeat all four rules aloud. Each day's session began with a recap of the previous day's activities and a preview of the day's session.

Working Memory Training Training utilized an off-the-shelf computerized adaptive working memory game, "Jungle Memory" (Memosyne Ltd., 2008), designed for children ages 7–16 years. The program included three games with multiple difficulty levels and activities, presented with motivational features and regular feedback: memorizing and use of word endings, mental rotation of letters, and sequential memory of mathematical solutions. The WM training was completed online, with minimal verbal input from the examiner; however, the examiner and participant sat side by side, to enable the examiner to address questions and to monitor engagement.

Interaction-Only Condition The interaction-only program was designed to function as an active control condition. It mirrored the RET program in terms of materials and length, but without discussion of recursion. Children heard and produced sentences based on presented images. The central differences between RET and IOC were that (1) there was no mention of compositionality, recursion, or embedding, and (2) only singly-embedded sentences (e.g., *the girl's dog*) were included, with no use of double embeddings (e.g., *the girl's dog's tail*).

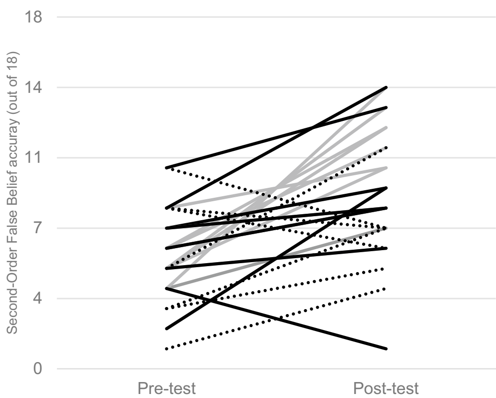

Fig. 2 Pre and post-training accuracy, second-order false belief test. Note: Recursive embedding training, grey lines; Working memory training, solid black lines; Interactive control, dotted lines

Results

To test whether the training conditions led to a change in second-order false belief processing, pre- and post-training scores were calculated; see Table 6. All variables met standard statistical assumptions of normality, skewness, and heterogeneity. A repeated-measures ANOVA was conducted with training condition (Recursive embedding, Working memory, or Interactive control) as a between-subjects factor and pre-training vs. post-training second-order false belief score as a within-subjects factor. The results showed a significant main effect of training condition, $F(1,24) = 29.526$, $p < 0.001$, with a large effect size (partial $\eta^2 = 0.55$). Second-order false belief scores increased from pre- to post-training timepoints in all conditions. This significant main effect was moderated by a significant interaction of training condition and time, $F(2, 24) = 5.311$, $p = 0.01$, partial $\eta^2 = 0.30$, suggesting that the change in second-order false belief scores differed as a function of training type; see Fig. 2. Paired sample t-tests showed a significant increase in false belief ability following Recursive embedding, $t(8) = 6.27$, $p < 0.001$, Cohen's $d = 2.1$, and Working memory, $t(9) = 2.5$, $p = 0.04$, $d = 0.74$, trainings, both with large effect sizes. The Interactive control group scores showed no significant change, $t(7) = 1.14$, $p = 0.29$, $d = 0.48$. Because participants were older in one school, age was included as a covariate in a second ANOVA. The interaction between age and time was not significant, $F(1, 23) = 1.14$, $p = 0.29$, partial $\eta^2 = 0.04$, and the pattern of results was unchanged.[1]

[1] To test whether differences in teachers, classroom structure, geographical location, etc., might influence results, school district was added as a covariate; the interaction of district and time was not significant, $F(1, 23) = 0.20$, $p = 0.88$, partial $\eta^2 = 0.001$, and the pattern of results was unchanged.

These results confirmed that there was a significant improvement from pre- to post-test in both the Recursive embedding training and Working memory training condition, but not in the Interaction-only control condition. The improvement in the working memory training group of a smaller effect size than in the Recursive embedding condition; this is relatively unsurprising, given that the latter involves a much more distant transfer of training knowledge.

The exciting results of Study 2 show that the capacity to reason through second-order false belief tasks by children with ASD is impacted by the development of recursive complementation and working memory skills; it is both novel and exciting to find that this sophisticated form of theory of mind reasoning can be shaped by intervention. Even a relatively abbreviated five-day training course led to significant changes in reasoning performance immediately following the training. This work lays a foundation for a large-scale intervention study that includes a long-term follow-up.

General Discussion

Neurotypical children are thought to understand and reason through theory of mind problems using relatively intuitive, perceptually-driven, and implicit skills and knowledge. The development of theory of mind processing has been proposed to follow a distinct developmental pathway in individuals with ASD, with performance more directly tied to cognitive skills and experience. Here we examine a proposal that a late-developing aspect of theory of mind, second-order false belief skills, specifically depends on the ability to perform algorithmic recursive operations on linguistic materials.

The present study tested associations among second-order theory of mind, language, and working memory, in children with ASD aged 6–16 and typically developing children aged 4 to 11 years, using two approaches. Importantly, the ASD group showed no impairments in second-order theory of mind performance, compared to an age-matched group of neurotypical children. Cross-sectional correlational analyses examined links among performance on theory of mind. The results provided support for the associations in the ASD group among semantic knowledge (as measured by the WISC Verbal Comprehension index), grammar comprehension (as measured by the TROG), recursive sentential complement skills (as measured by the recursive embedding test), working memory (as measured by the WISC Working Memory index), and false belief reasoning skills. Crucially, our results showed that recursive sentential complementation skills predicted second-order false belief reasoning skills (in a test battery including all four types of second-order reasoning task), supporting previous findings about such a link in ASD (Boeg Thomsen, 2016; Lind & Bowler, 2009; Tager-Flusberg & Joseph, 2005). Furthermore, results showed that second-order false belief skills were predicted by recursive sentential complementation skills for children with ASD, but not for neurotypical children; in contrast, second-order false belief skills were predicted by working memory skills for neurotypical children, but not for the ASD group. This suggests that children with ASD and NT children arrive at the correct solutions differently, consistent with the hypothesis that children with ASD rely to a greater extent on language, compared to neurotypical children.

Correlational results cannot determine whether relationships are causal, nor can they address the role one capacity plays in facilitating the development of another. Study 2 addressed this issue, by providing five sessions of training in linguistic recursion or working memory, as compared to an active control training condition, for the children with ASD who had the most impaired second-order theory of mind skills at the outset. Results of this intervention indicated that *both* linguistic recursion and working memory skills training boosted the ability to make correct predictions in second-order false belief reasoning. These results are nicely consistent with effects of a different form of linguistic training on first-order ToM reasoning (Durrleman et al., 2019).

Both training conditions involved far transfer. The far-transfer effect in the recursion training group (from the linguistic use of recursive embeddings, to the use of recursive mental state reasoning) suggests that some children with ASD benefit from enhanced language skills in solving false belief reasoning tasks; this is corroborated by the fact that references to mental states were carefully avoided in the training part and in general communication with the children who took part in the study. These results further support the proposal that sentential complements are an effective tool for 'hacking out' solutions to false belief tasks; that is, a deeper grasp of compositional semantics influences second-order false belief reasoning because this tool helps build a foundation of competency in recursive structures in language, to support the processing of recursive false belief knowledge.

In order to understand *how* such a recursive tool may function and affect SOFB reasoning, it seems useful to draw a distinction between recursion in structural and algorithmic terms. To understand recursion in structural terms is to think of it in terms of its subparts: for example, we can think of a sentence as consisting of a number of embedded sentences, or think of a belief as consisting of a number of embedded beliefs. Thinking structurally about the development of mental state understanding, we might say that at the first-order stage, the child learns that beliefs are 'mental objects,' and that people can have false beliefs; at the second-order stage, the child learns that these mental objects can be about other beliefs, that is, that they have recursive structure. Viewing the current study from this structural perspective, our results suggest that learning recursive structures in language enhanced or strengthened children's conceptual understanding of recursion and led to improvement in applying this concept to beliefs.

On the other hand, recursion can be described in algorithmic terms, as a procedure to process several levels of information in a uniform fashion. Under this reading, recursion becomes a solution for dealing with growing complexity, which may be applicable to both language and false belief reasoning domains. In the current study, results suggest that learning the procedure of embedding complements one inside another led to improvements in reasoning about beliefs: learning how they are built helps children to produce them.

Whether from a structural or an algorithmic perspective, development from first-order to second-order false belief reasoning requires additional processing to handle the increased complexity, requiring increased working memory capacity. The far-transfer effects in the working memory group cannot serve as a stringent evidence of the working memory's direct impact on the false belief reasoning, because there was no direct control for this condition (such as a computer game that trains other, non-executive skills). Nonetheless, children in the working memory group were not exposed to any language training, and the tasks were not even verbally mediated; this argues against the possibility that WM training lead to enhanced second-order false belief processing by affecting language. Rather, the far-transfer effects of WM training support the proposal that language and working memory each have an independent impact on SOFB development (Hasselhorn et al., 2005).

The interplay of mental state understanding, working memory, and two perspectives on recursion, corresponds with explanations of the development from the first-order to the second-order false belief stage: the *conceptual change* and the *complexity-only* theories (Perner, 1988; Tager-Flusberg & Sullivan, 1994b). According to the conceptual change theory, SOFB competency involves a qualitative transformation of the underlying thought system, requiring new conceptual resources. According to the complexity-only position, SOFB development requires higher information processing skills rather than conceptual change.

Understanding recursion in structural terms fits well with the conceptual change position: the new conceptual resources that are needed for this development may include understanding recursion—learning that representations (words, utterances, thoughts, figures, mental states) can be meaningfully embedded into each other, and the order of embedding is relevant for meaning. Acquiring this concept could impact the cognitive, perceptual, and linguistic domains. Along with increasing working memory, a child who can explicitly understand the concept of recursion across different contexts will do better in recursive mental state reasoning tasks.

Understanding recursion in algorithmic terms, on the other hand, seems to correspond to the complexity-only position: by learning the procedure of embedding, a child learns how to handle complex information, and coupled with growing working memory, is better equipped to process recursive false belief reasoning. This procedural knowledge could be formulated as *the skill needed to embed an unlimited number of objects inside one another.* The training study involved learning about four simple principles of compositional semantics (i.e., procedural knowledge) for recursion, and helping children to comprehend and produce a large number of sentences with embedded constituents of the same type. This provided hands-on experience with the role played by the principles (i.e., conveyed knowledge of structural recursion).

Limitations

This study provides new data on the links among domain-general cognitive skills, linguistic skills, and mental state reasoning, in children with ASD and with typical development, and demonstrates the effectiveness of a relatively brief training on such reasoning skills. However, results are necessarily limited in several dimensions. It is impossible to pinpoint the precise mechanism of improvement in the recursive embedding training condition. Learning the principles of recursive embedding may have helped to highlight the procedural and structural aspects of recursion; this training also included concentrated exposure to multiple examples of recursion in language, and the opportunity to produce recursive structures. Any of these, or other factors, may have been the "active ingredient." A second important limitation is the lack of a follow-up assessment months after the training and post-test, which would help to establish the depth of knowledge and skill-building that occurred as a result of training. A third and important limitation is the relatively small size of the three training groups. Furthermore, for pragmatic reasons, the TD comparison group did not complete the training, nor the entire set of characterizing measures, which limits our ability to compare predictors, among other analyses. Finally, the same examiner administered the pre- and post-tests, and the training activities in all three conditions. This consistency likely contributed to participant engagement, and to training consistency within conditions. However, the fact that the examiner was not naïve to study hypotheses could have led to subtle examiner effects. While the results are not conclusive, given these limitations, they provide motivation for pursuing a larger study.

This study breaks new ground in training children with ASD to use recursive linguistic structures, and in comparing the effects of linguistic and working memory training on second-order false belief reasoning skills. It suggests that sentential complementation is important in the acquisition of the more complex aspect of theory of mind processing. Furthermore, our training results suggest that mastery of linguistic recursion is among the compensatory skills that lets children with ASD succeed in second-order mental state reasoning.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10803-021-05277-1.

Acknowledgments Many thanks to children, parents and teachers that participated in the study. We also thank Lars Clemmensen for sharing the Danish version of the Dutch ToM Storybook. We acknowledge funding from NIMH R01MH112687-01A1 to I.M.E. This work formed the basis of Irina Polyanskaya's Ph.D. thesis.

Author Contributions IP: Conceptualization Original draft preparation, Methodology, Formal analysis, Investigation, Writing - Original Draft. IME: Conceptualization, Methodology, Formal analysis, Data Curation, Writing - Review & Editing, Visualization. TB: Conceptualization, Methodology, Writing - Review & Editing. PB: Conceptualization, Methodology, Writing - Review & Editing.

Data Availability Data from this study are available on the Open Science Foundation website at: https://osf.io/qf6ch/?view_only=abc784e9ffc345508cc1a6e03a4354cd.

Declarations

Conflict of interest The authors report no conflict of interest.

References

Arslan, B., Hohenberger, A., & Verbrugge, R. (2017). Syntactic recursion facilitates and working memory predicts recursive theory of mind. *PLoS ONE, 12*(1), e0169510. https://doi.org/10.1371/journal.pone.0169510

Arslan, B., Verbrugge, R., Taatgen, N., & Hollebrandse, B. (2015). Teaching children to attribute second-order false beliefs: A training study with feedback. In *Proceedings of the 37th annual meeting of the Cognitive Science Society* (pp. 108–113).

Astington, J. W., & Baird, J. A. (2005). Why language matters for theory of mind. *Oxford University Press*. https://doi.org/10.1093/acprof:oso/9780195159912.001.0001

Astington, J. W., Pelletier, J., & Homer, B. (2002). Theory of mind and epistemological development: The relation between children's second-order false-belief understanding and their ability to reason about evidence. *New Ideas in Psychology, 20*(2–3), 131–144.

Baron-Cohen, S. (1989a). Perceptual role taking and protodeclarative pointing in autism. *British Journal of Developmental Psychology, 7*, 113–127. https://doi.org/10.1111/j.2044-835x.1989.tb00793.x

Baron-Cohen, S. (1989a). The autistic child's theory of mind: a case of specific developmental delay. *Journal of Child Psychology and Psychiatry, 30*(2), 285–297. https://doi.org/10.1111/j.1469-7610.1989.tb00241.x

Baron-Cohen, S. (2000). Theory of mind and autism: A fifteen year review. In *Understanding other minds: Perspectives from* *developmental cognitive neuroscience* (2nd ed., pp. 3–20). Oxford University Press.

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders, 29*(5), 407–418. https://doi.org/10.1023/a:1023035012436

Baron-Cohen, S., Spitz, A., & Cross, P. (1993). Do children with autism recognise surprise? A research note. *Cognition and Emotion, 7*(6), 507–516. https://doi.org/10.1080/02699939308409202

Bishop, D. (2009). *Test for Reception of Grammar—Version 2, Dansk Version*. Pearson.

Blijd-Hoogewys, E. M., van Geert, P. L., Serra, M., & Minderaa, R. B. (2008). Measuring theory of mind in children. Psychometric properties of the ToM storybooks. *The Journal of Autism and Developmental Disorders, 38*(10), 1907–1930. https://doi.org/10.1007/s10803-008-0585-3

Boeg-Thomsen, D. (2016). *Linguistic perspective marking and mental-state reasoning in children with autism: A training study with complement clauses*. University of Copenhagen.

Bogaerds-Hazenberg, S. T. M., & Hendriks, P. (2016). Complex language, complex thought? *Linguistics in the Netherlands, 33*(1), 28–40.

Bosco, F. M., & Gabbatore, I. (2017). Sincere, deceitful, and ironic communicative acts and the role of the theory of mind in childhood. *Frontiers in Psychology, 8*, 21. https://doi.org/10.3389/fpsyg.2017.00021

Braüner, T., Blackburn, P., & Polyanskaya, I. (2020). Being deceived: Information Asymmetry in second-order false belief tasks. *Topics in Cognitive Science, 12*(2), 504–534. https://doi.org/10.1111/tops.12422

Brynskov, C., Krøjgaard, P., & Eigsti, I. M. (2015). Language and communication in children with autism: A critical look at three assumptions and their clinical implications. *Nordic Psychology, 68*(2), 100–113. https://doi.org/10.1080/19012276.2015.1071199

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition, 112*(2), 337–342. https://doi.org/10.1016/j.cognition.2009.05.006

Caillies, S., & Le Sourn-Bissaoui, S. (2013). Nondecomposable idiom understanding in children: recursive theory of mind and working memory. *Canadian Journal of Experimental Psychology, 67*(2), 108–116. https://doi.org/10.1037/a0028606

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology, 87*(4), 299–319.

Clemmensen, L., Bartels-Velthuis, A. A., Jespersen, R. A. F., van Os, J., Blijd-Hoogewys, E. M. A., Ankerstrøm, L., Væver, M., Daniel, P. F., Drukker, M., Jeppesen, P., & Jepsen, J. R. M. (2016). A psychometric evaluation of the Danish version of the theory of mind storybook for 8–14 year-old children. *Frontiers in Psychology, 7*, 330–330. https://doi.org/10.3389/fpsyg.2016.00330

Constantino, J. N. (2003). *Social Responsiveness Scale (SRS)*. Western Psychological Services.

de Villiers, J. G. (2007). The interface of language and theory of mind. *Lingua, 117*(11), 1858–1878. https://doi.org/10.1016/j.lingua.2006.11.006

de Villiers, J., Hobbs, K., & Hollebrandse, B. (2014). Recursive Complements and Propositional Attitudes. In T. Roeper & M. Speas (Eds.), *Recursion: Complexity in Cognition* (pp. 221–242). Dordrecht: Springer. https://doi.org/10.1007/978-3-319-05086-7_10

de Villiers, J. G., & Pyers, J. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development, 17*, 1037–1060.

Dennis, M., Francis, D. J., Cirino, P. T., Schachar, R., Barnes, M. A., & Fletcher, J. M. (2009). Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. *Journal of the International Neuropsychological Society, 15*, 1–13.

Durrleman, S., Burnel, M., De Villiers, J. G., Thommen, E., Yan, R., & Delage, H. (2019). The impact of grammar on mentalizing: A training study including children with autism spectrum disorder and developmental language disorder. *Frontiers in Psychology, 10*, 2478. https://doi.org/10.3389/fpsyg.2019.02478

Durrleman, S., Burnel, M., Thommen, E., Foudon, N., Sonié, S., Reboul, A., & Fourneret, P. (2016). The language cognition interface in ASD: Complement sentences and false belief reasoning. *Research in Autism Spectrum Disorders, 21*, 109–120.

Durrleman, S., & Franck, J. (2015). Exploring links between language and cognition in autism spectrum disorders: Complement sentences, false belief, and executive functioning. *Journal of Communication Disorders, 54*, 15–31. https://doi.org/10.1016/j.jcomdis.2017.03.005

Edwards, K., & Low, J. (2017). Reaction time profiles of adults' action prediction reveal two mindreading systems. *Cognition, 160*, 1–16. https://doi.org/10.1016/j.cognition.2016.12.004

Farrar, M. J., & Maag, L. (2002). Early language development and the emergence of a theory of mind. *First Language, 22*(2), 197–213.

Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development, 79*(1), 126–138. https://doi.org/10.1111/j.1467-8624.2007.01115.x

Frith, U. (2004). Emanuel Miller lecture: Confusions and controversies about Asperger syndrome. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 45*(4), 672–686. https://doi.org/10.1111/j.1469-7610.2004.00262.x

Fu, G., Xiao, W. S., Killen, M., & Lee, K. (2014). Moral judgment and its relation to second-order theory of mind. *Developmental Psychology, 50*(8), 2085–2092. https://doi.org/10.1037/a0037077

Gordon, A. C., & Olson, D. R. (1998). The relation between acquisition of a theory of mind and the capacity to hold in mind. *Journal of Experimental Child Psychology, 68*(1), 70–83. https://doi.org/10.1006/jecp.1997.2423

Grueneisen, S., Wyman, E., & Tomasello, M. (2015). "I know you don't know I know…" children use second-order false-belief reasoning for peer coordination. *Child Development, 86*(1), 287–293. https://doi.org/10.1111/cdev.12264

Guajardo, N. R., & Watson, A. C. (2002). Narrative discourse and theory of mind development. *Journal of Genetic Psychology, 163*(3), 305–325. https://doi.org/10.1080/00221320209598686

Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science, 6*(3), 346–359. https://doi.org/10.1111/1467-7687.00289

Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66*(3), 843–855.

Happé, F. (2003). Theory of mind and the self. *Annals of the New York Academy of Sciences, 1001*, 134–144.

Hasselhorn, M., Mähler, C., & Grube, D. (2005). Theory of mind, working memory, and verbal ability in preschool children: The proposal of a relay race model of the developmental dependencies. In *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind* (pp. 219–237). Lawrence Erlbaum Associates Publishers.

Hillier, A., & Allinson, L. (2002). Beyond expectations: Autism, understanding embarrassment, and the relationship with theory of mind. *Autism, 6*(3), 299–314. https://doi.org/10.1177/1362361302006003007

Hollebrandse, B., Hobbs, K., de Villiers, J., & Roeper, T. (2008). Second order embedding and second order false belief. In A. Gavarró & M. J. Freitas (Eds.), *Language acquisition and development: Proceedings of GALA 2007*. Cambridge Scholars Publishing.

Hollebrandse, B., & van Hout, A. (2015). Comprehension and production of double-embedded structures: A kindergarden path in long-distance dependencies in Dutch school-aged children? In C. Hamann & E. Ruigendijk (Eds.), *Proceedings of GALA 2013* (pp. 190–204). Cambridge Scholars Publishing.

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*(3), 321–333. https://doi.org/10.1007/s11229-012-0169-9

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16*(2), 233–253.

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology, 43*(6), 1447–1459.

Jackendoff, R. (1990). *Semantic structures*. MIT Press.

Janssen, T. M. V. (2020, Fall). Montague semantics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2020 ed.). https://plato.stanford.edu/archives/spr2020/entries/montague-semantics.

Kaland, N., Callesen, K., Moller-Nielsen, A., Mortensen, E. L., & Smith, L. (2008). Performance of children and adolescents with Asperger syndrome or high-functioning autism on advanced theory of mind tasks. *The Journal of Autism and Developmental Disorders, 38*(6), 1112–1123. https://doi.org/10.1007/s10803-007-0496-8

Keenan, T., Olson, D. R., & Marini, Z. (1998). Working memory and children's developing understanding of mind. *Australian Journal of Psychology, 50*(2), 76–82.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010, Dec 24). The social sense: susceptibility to others' beliefs in human infants and adults. *Science, 330*(6012), 1830–1834. https://doi.org/10.1126/science.1190792

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind." *Psychological Review, 94*(4), 412–426. https://doi.org/10.1037/0033-295X.94.4.412

Lind, S. E., & Bowler, D. M. (2009). Language and theory of mind in autism spectrum disorder: The relationship between complement syntax and false belief task performance. *Journal of Autism and Developmental Disorders, 39*(6), 929–937. https://doi.org/10.1007/s10803-009-0702-y

Lockl, K., & Schneider, W. (2007). Knowledge about the mind: Links between theory of mind and later metamemory. *Child Development, 78*(1), 148–167. https://doi.org/10.1111/j.1467-8624.2007.00990.x

Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development, 74*(4), 1130–1144. https://doi.org/10.1111/1467-8624.00597

Low, J. (2010). Preschoolers' implicit and explicit false-belief understanding: Relations with complex syntactical mastery. *Child Development, 81*(2), 597–615. https://doi.org/10.1111/j.1467-8624.2009.01418.x

Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: state of the art. *British Journal of Developmental Psychology, 30*(Pt 1), 1–13. https://doi.org/10.1111/j.2044-835X.2011.02074.x

Memosyne Ltd. (2008). *Jungle Memory*. Memosyne Ltd.

Miller, C. A. (2006). Developmental relationships between language and theory of mind. *American Journal of Speech and Language Pathology, 15*(2), 142–154. https://doi.org/10.1044/1058-0360(2006/014)

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

Perner, J. (1988). Higher-order beliefs and intentions in children's understanding of social interaction. In *Developing theories of mind* (pp. 271–294). Cambridge University Press.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences, 3*(9), 337–344. https://doi.org/10.1016/s1364-6613(99)01362-5

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that…": Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology, 39*(3), 437–471. https://doi.org/10.1016/0022-0965(85)90051-7

Peskin, J., & Astington, J. W. (2004). The effects of adding metacognitive language to story texts. *Cognitive Development, 19*(2), 253–273. https://doi.org/10.1016/j.cogdev.2004.01.003

Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in theory-of-mind development for children with deafness or autism. *Child Development, 76*(2), 502–517. https://doi.org/10.1111/j.1467-8624.2005.00859.x

Polyanskaya, I. (2019). *Second-order false belief reasoning by children with autism: A correlation and training study*. Roskilde University.

Rubio-Fernandez, P. (2019). Jan). Memory and inferential processes in false-belief tasks: An investigation of the unexpected-contents paradigm. *Journal of Experimental Child Psychology, 177*, 297–312. https://doi.org/10.1016/j.jecp.2018.08.011

Russell, J. (1997). Autism as an Executive Disorder. In J. Russell (Ed.), *How executive disorders can bring about an adequate 'theory of mind'* (pp. 256–304). Oxford University Press.

Schick, B., de Villiers, P., de Villiers, J., & Hoffmeister, R. (2007). Language and theory of mind: A study of deaf children. *Child Development, 78*(2), 376–396. https://doi.org/10.1111/j.1467-8624.2007.01004.x

Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development, 80*(4), 1172–1196. https://doi.org/10.1111/j.1467-8624.2009.01324.x

Shah, P., & Miyake, A. (1999). Models of working memory: An introduction. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 1–27). Cambridge University Press.

Sodian, B., & Frith, U. (1992). Deception and sabotage in autistic, retarded and normal children. *Journal of Child Psychology and Psychiatry, 33*(3), 591–605. https://doi.org/10.1111/j.1469-7610.1992.tb00893.x

Sodian, B., Taylor, C., Harris, P. L., & Perner, J. (1991). Early deception and the child's theory of mind: False trails and genuine markers. *Child Development, 62*(3), 468–483. https://doi.org/10.2307/1131124

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science, 13*(6), 907–912. https://doi.org/10.1111/j.1467-7687.2009.00946.x

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994) Preschoolers can attribute second-order beliefs. *Developmental Psychology, 30*, 395–402. https://doi.org/10.1037/0012-1649.30.3.395

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18*(7), 580–586. https://doi.org/10.1111/j.1467-9280.2007.01943.x

Szabó, Z. G. (2020). Compositionality. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2020 ed.). https://plato.stanford.edu/archives/fall2020/entries/compositionality.

Tager-Flusberg, H. (2001). A re-examination of the Theory of Mind hypothesis of autism. In J. Burack, T. Charman, N. Yirmiya, & P. R. Zelazo (Eds.), *Development in autism: Perspectives from theory and research* (pp. 173–193). Erlbaum.

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

Tager-Flusberg, H., & Sullivan, K. (1994a). Predicting and explaining behavior: A comparison of autistic, mentally retarded and normal children. *Journal of Child Psychology and Psychiatry, 35*(6), 1059–1075. https://doi.org/10.1111/j.1469-7610.1994.tb01809.x

Tager-Flusberg, H., & Sullivan, K. (1994b). A second look at second-order belief attribution in autism. *Journal of Autism and Developmental Disorders, 24*(5), 577–586. https://doi.org/10.1007/bf02172139

Talwar, V., Gordon, H. M., & Lee, K. (2007). Lying in the elementary school years: Verbal deception and its relation to second-order belief understanding. *Developmental Psychology, 43*(3), 804–810. https://doi.org/10.1037/0012-1649.43.3.804

Tiede, G. M., & Walton, K. M. (2020). Social endophenotypes in autism spectrum disorder: A scoping review. *Developmental Psychology*. https://doi.org/10.1017/s0954579420000577

Wechsler, D. (2003). *Wechsler intelligence scales for children* (4th ed.). Psychological Corporation.

Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development, 72*, 655–684.

Williams, D. M., & Happé, F. (2009). What did I say? Versus what did I think? Attributing false beliefs to self amongst children with and without autism. *Journal of Autism and Developmental Disorders, 39*(6), 865–873. https://doi.org/10.1007/s10803-009-0695-6

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