Journal of Experimental Child Psychology 191 (2020) 104739

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Theory of mind, language, and reading: Developmental relations from early childhood to early adolescence

Susanne Ebert *

University of Bamberg, 96047 Bamberg, Germany
Norwegian University of Science and Technology, 7491 Trondheim, Norway

ARTICLE INFO

Article history:
Received 8 April 2019
Revised 25 October 2019
Available online 9 December 2019

Keywords:
Theory of mind
Language
Longitudinal study
Listening comprehension
Reading comprehension
Preschool age
Early adolescence

ABSTRACT

This study longitudinally investigated the relation between theory of mind (ToM) and verbal language skills in 231 children from preschool to early adolescence. Further, links to reading comprehension of texts at age 13;7 (years;months) were examined. To assess ToM, children completed false belief tasks at 5;6 and the Strange Stories at 12;8. To assess language, children completed a receptive grammar/sentence comprehension test at 3;6 and 5;6, a receptive vocabulary test at 3;6, 5;6 and 12;8, as well as a test of listening comprehension of texts at 13;7. A bidirectional relation between early and advanced measures of children's language skills and ToM was found: Changes in ToM were predicted by language skills, especially by receptive grammar/sentence comprehension; changes in children's receptive vocabulary were predicted by early ToM. However, early ToM had no direct or indirect effect on later listening comprehension or reading comprehension after controlling for early language skills. Only children's advanced ToM had a small indirect effect on reading comprehension, via listening comprehension. The results are discussed in light of ToM stability over time, and theories on how language and ToM development are intertwined.

* Corresponding author at: Norwegian University of Science and Technology, 7491 Trondheim, Norway.
E-mail address: susanne.ebert@ntnu.no

https://doi.org/10.1016/j.jecp.2019.104739
0022-0965/

Introduction

Between 3 and 5 years of age, children gain increasingly greater knowledge of mental states and processes; that is, they have notable advancement in their theory of mind (ToM) development. This is reflected in their mastery of explicit false-belief tasks in which children are asked to predict how a protagonist will act or what a protagonist thinks based on a mistaken belief. This understanding of false beliefs has been shown to be closely connected to children's language skills (Milligan, Astington, & Dack, 2007). Moreover, longitudinal data and training studies have revealed that in preschool age language skills are more predictive of ToM than vice versa (de Villiers, 2005; Ebert, 2015; Hale & Tager-Flusberg, 2003). However, less is known about how the two domains are related longitudinally beyond the preschool years. For instance, little is known about whether early language skills are related to advanced ToM and whether early explicit false-belief understanding is related to children's further development in ToM, language, and other language-related or cognitive domains (Apperly, Samson, & Humphreys, 2009; Hughes, 2016; Lockl, Ebert, & Weinert, 2017). Thus, my main aim in this study was to longitudinally investigate the relation between language and ToM from preschool to early adolescence. In addition, I asked how both are connected to children's later reading comprehension.

The role of language in children's early ToM

Various theoretical accounts have explained why children's language skills might be important for the emergence of ToM in preschool years. On the one hand, children's language skills are a means of *communication* that enable them to take part in and make sense of verbal communication, especially when people talk about nonvisible mental states and processes (e.g., Harris, 2005; Nelson, 2005; Wellman & Peterson, 2013). Training and longitudinal studies show that verbal communication, especially talking about and elaborating on mental entities, promotes children's ToM development (e.g., Ebert, Peterson, Slaughter, & Weinert, 2017; Lohmann & Tomasello, 2003). On the other hand, language is an important means for *representing* mental states and separating them from reality. For example, performance in false-belief tasks was found to improve after the experimenter provided labels for the different locations of hidden objects (Low & Simpson, 2012). Labeling may support children's representation of nonobservable objects. In addition, the ability to use specific mental terms and mastery of syntax, especially structures with propositions embedded in clauses (e.g., "Hannes thinks that Joshua is climbing outdoors"), may help children to represent mental states and multiple propositions simultaneously (see also Astington & Baird, 2005a).

Some language components seem more theoretically relevant for ToM development than others. Whereas pragmatic features of language and ToM are related by definition (Astington & Jenkins, 1999), there has been some deeper discussion of whether semantics or syntax are more important for keeping track of and representing false-beliefs (Astington & Jenkins, 1999; Harris, de Rosnay, & Pons, 2005; Slade & Ruffman, 2005). A meta-analysis by Milligan et al. (2007) reported a larger effect size for the longitudinal correlation between ToM and more general language measures than for receptive vocabulary but reported no other differences in correlations with various language measures. This suggests that in children's preschool years various facets of language support their understanding of mental states and no single or specific language component is of primary importance (see also Harris et al., 2005; Miller, 2016). However, although this conclusion can be drawn with respect to children's ToM development in preschool, the situation may be different beyond children's preschool years.

ToM beyond the preschool years

At about 5 years of age, when children explicitly understand false beliefs, they are said to have developed a metarepresentational understanding of the mind (Perner, 1991; Wellman, 2014). However, their understanding of mental states and processes continues to develop beyond this point.

In particular, children learn that people differ in their opinions and interpretations of the same entity and gain a better understanding of mental states and processes as well as nonliteral meanings of people's utterances in more complex social situations (e.g., Carpendale & Chandler, 1996; Weimer, Dowds, Fabricius, Schwanenflugel, & Suh, 2017). Accordingly, advanced ToM tests such as the Strange Stories task often incorporate social scenarios (White, Hill, Happe, & Frith, 2009). In these tasks, children need to refer to a character's mental state to explain her or his actions correctly.

Many researchers believe that advanced ToM includes no further changes in children's conceptual ToM understanding (Apperly et al., 2009; Devine, White, Ensor, & Hughes, 2016; Keenan, 2003; Lecce, Bianco, Devine, & Hughes, 2017). Nevertheless, it is assumed that individual differences in the use of mental concepts in social situations exist. The "genuine variation" account states that although all typically developing children develop a metarepresentational understanding eventually, individual differences in the time point at which children acquire this understanding reflect "differences in the ease or fluency with which children or adults use their theory of mind to attribute mental states to others" (Hughes & Devine, 2015, p. 151). Thus, individual differences in early ToM should be related to individual differences in advanced ToM.

The few studies that have measured ToM longitudinally in preschool and middle childhood report medium correlations, supporting this account (Devine et al., 2016; Ensor, Devine, Marks, & Hughes, 2014; Lecce, Caputi, & Pagnin, 2014). These studies also suggest that individual differences in language skills are related to advanced ToM, similarly as for early ToM.

Relations between early language and advanced ToM

Under the assumption that no further conceptual change occurs after having acquired false-belief understanding, language might play a different role for developing an advanced ToM than for developing a metarepresentational ToM in preschool. For example, it is possible that they correlate less or that correlations are rooted in different underlying mechanisms. For example, language might be correlated only with advanced ToM due to shared task demands or because it becomes a necessary component of ToM (Apperly et al., 2009; Miller, 2009).

Recent studies show that pragmatic, semantic, and syntactic language measures are significantly related to ToM in middle childhood but more weakly in early adolescence (Banerjee, Watling, & Caputi, 2011; Devine & Hughes, 2013; Im-Bolter, Agostino, & Owens-Jaffray, 2016; Lecce, Ronchi, Del Sette, Bischetti, & Bambini, 2019). However, from a developmental and theoretical point of view, cross-sectional associations say nothing about developmental relations. To my knowledge, three longitudinal studies have assessed language measures as control variables and provide hints on whether early language skills scaffold ToM beyond the preschool years. Lecce et al. (2014) reported medium correlations between early receptive vocabulary at about 6 years of age and advanced ToM (social scenarios) at 10 years in a group of 49 children. Devine et al. (2016) similarly found relations between receptive vocabulary at 6 years of age and ToM at 10 years (various measures) in a group of 137 children even after controlling for earlier ToM, executive functions, and socioeconomic status (SES). Ensor et al. (2014) found an association between more general early verbal abilities at 3 years of age and advanced ToM (social scenarios) at 10 years.

Although these three studies suggest a longitudinal relation between early language and advanced ToM, they all included only one language measure at a single time point. In addition, the studies differed in whether they considered early vocabulary or a more general language comprehension measure. Moreover, there were differences in the time point when the language measures were assessed; Lecce et al. (2014) and Devine et al. (2016) measured receptive vocabulary at about 5 or 6 years of age, whereas Ensor et al. (2014) assessed a more comprehensive language measure at 3 years. They also reported that the impact of early language on advanced ToM at 10 years of age was mediated by false-belief understanding at 6 years. Thus, it is not clear whether these various language measures actually were differentially related to advanced ToM.

I extended these previous studies while simultaneously including two different language measures, vocabulary and grammar (sentence comprehension), at 3 *and* 5 years of age and comparing their developmental relations with ToM. Moreover, because early ToM was measured at 5 years of age, I was also able to specify direct and indirect effects of language at 3 years via early ToM on advanced ToM.

Relations between early ToM and later language

The relation between ToM and language skills might change as children get older and start school. Now, more complex language comprehension skills, such as those necessary in text comprehension tasks, become more central. ToM skills might support these language skills. For example, Pelletier and Astington (2004) assumed that children with more advanced understanding of mental entities are better able to connect settings, events, and actions described in a story with the characters' thoughts, motives, and emotions. Thus, based on Bruner (1986) idea, children with more advanced ToM may integrate the landscape of action with the landscape of consciousness more easily when listening to or reading a story. In addition, an advanced ability to represent mental states and processes also promotes metacognitive knowledge and skills, which may further support text comprehension. Moreover, ToM may support understanding the author's intentions, that is, understanding the aims of the text, the ability to connect information in the text with earlier knowledge, and the construction of a mental situation model (see Atkinson, Slade, Powell, & Levy, 2017; Guajardo & Cartwright, 2016; Kim, 2015).

Indeed, some recent longitudinal studies have reported associations between ToM in preschool and children's later text comprehension even after controlling for early language (Atkinson et al., 2017; Kim, 2015, 2016). Another study points to a longitudinal association between early ToM and later receptive vocabulary (Lecce et al., 2014). However, to my knowledge, no existing studies have examined the longitudinal association between early ToM in preschool and various advanced language skills systematically.

Relations among ToM, language, and reading

Early ToM may be predictive of not only advanced language skills but also later reading comprehension, which is closely related to children's language skills. Indeed, according to the simple view of reading, reading comprehension is a product of listening comprehension and decoding skills (Hoover & Gough, 1990). Moreover, there is good evidence that early language skills predict later reading comprehension alongside early phonological information processing and decoding skills (e.g., Dickinson, McCabe, Anastasopoulos, Peisner-Feinberg, & Poe, 2003; Ebert & Weinert, 2013; Storch & Whitehurst, 2002). However, other higher-order cognitive skills, including ToM, are also discussed as further potential contributors. Theoretical explanations for the relation between ToM and reading comprehension of texts are like those for the relation between ToM and listening comprehension of texts (Atkinson et al., 2017; Guajardo & Cartwright, 2016; Kim, 2015, 2016). Thus, ToM may predict reading comprehension indirectly via listening comprehension, i.e. verbal language skills.

The results of the few studies that have investigated the impact of early ToM on later reading comprehension are mixed. Whereas some studies found no direct link between ToM and reading comprehension at all or only via language skills (Guajardo & Cartwright, 2016; Kim, 2015, 2016; Lockl et al., 2017), others found direct relations even after controlling for early language skills (Atkinson et al., 2017; Boerma, Mol, & Jolles, 2017).

The lack of association between early ToM and later reading comprehension reported in some studies might be explained by the fact that these studies measured reading comprehension at the beginning of learning to read, and tests at this stage may be so easy that higher-order cognitive processes are not relevant. However, Atkinson et al. (2017) found a direct link between early ToM and reading comprehension, even in 6-year-old children, although they did not directly model the role of early language skills in this relation. In contrast, Boerma et al. (2017), who also found a relation between ToM and reading comprehension, investigated only the relations between reading comprehension and advanced measures of ToM and language. Thus, it is not clear how early and advanced ToM as well as language skills together contribute to children's reading comprehension in older children.

The current study

The main aim of the current study was to provide a more systematic investigation of the longitudinal relation between language and ToM from preschool to early adolescence. If language is associated only with advanced ToM due to shared task demands or because language is a necessary component of advanced ToM, only concurrent but no longitudinal associations should be detected. I assumed that language supports children in understanding and keeping track of what is going on in social situations and, thus, in gaining better insights into people's mental states and processes, as has been found for developing a metarepresentational ToM in preschool. Furthermore, like in preschool, children's language skills enable them to take part in verbal exchanges and communicate and learn about others' mental states and processes while inferring the fine-grained meaning of language in more complex social situations. Thus, I expected that early language skills also scaffold the development of advanced ToM both directly and indirectly via early ToM.

The current study replicates and extends previous studies in various ways. First, the developmental period up to 13 years of age was investigated, extending the developmental periods observed in previous studies. Second, not only one early indicator for language at each time point was assessed; rather, two were investigated: one for vocabulary and one for grammar (sentence comprehension). This opens the opportunity to analyze whether different language indicators play a differential role in children's ToM development between preschool and early adolescence. Third, ToM *and* language measures were assessed in early childhood *and* early adolescence; thus, the reciprocal relation between ToM and language beyond preschool can be investigated in more detail. Given that knowledge about mental states should support children's development of metacognitive knowledge and inference-making skills, I assumed that early ToM is a stronger predictor of later listening comprehension of texts as an advanced language measure, than it is for later vocabulary.

Furthermore, I investigated how the developmental relation between ToM and language skills is linked to reading comprehension. ToM and language skills are interrelated over time, and both are considered to predict later reading comprehension. However, to my knowledge, how early and advanced ToM and language skills together are related to reading comprehension in early adolescence has never before been investigated. From a theoretical point of view, I expected only indirect effects of early ToM and language measures on reading comprehension via advanced ToM and advanced language measures.

A subordinate aim of the study was to investigate whether individual differences in early ToM are connected to children's advanced ToM longitudinally by covering the age range from 5 to 12 years. In preschool false-belief tasks were used to assess ToM, and in early adolescence social scenario tasks (i.e., a measure of children's use of mental states in social situations) were administered. According to previous studies and in line with the "genuine variation" account, I expected to find substantial correlations even after controlling for other cognitive and language measures. I assumed that children who understand mental concepts earlier than their peers are those who have learned to pay more attention to mental states and use them more easily and flexibly in early adolescence (see also Devine et al., 2016).

Because nonverbal cognitive abilities, working memory, and family SES are related to both language and ToM, I also controlled for these variables in all analyses.

Method

Study design

The current study was based on a subsample of a German longitudinal study that included children from diverse socioeconomic backgrounds as well as rural and urban areas. This subsample was tested for ToM at Wave 5 of the more comprehensive study. According to the study aims, various additional measurement points and measures were included (see Fig. 1 for an overview). For easier comprehension, I renumbered the waves of the more comprehensive longitudinal study and refer to Wave 1 and Wave 5 in preschool as Time 1 and Time 2 and to Wave 11 and Wave 12 in early adolescence as Time 3 and Time 4, respectively. Measures of Wave 1/Time 1 took place in 2005 and were included as baseline measures.

Fig. 1. Overview of study design. ToM, theory of mind. Ages are in years;months.

Ethical approval for the comprehensive study was given by the university, and compliance with ethical standards was confirmed by the German Research Foundation (DFG), which funded the study. Informed consent to the children's participation was obtained from their parents, and all information was provided voluntarily. The children were tested individually by trained research assistants in their preschools. In early adolescence, they were tested at home. The children had the opportunity to withdraw from testing at any point and received a small gift (e.g., sticker) after each testing session. The parents were also rewarded with a small gift for their participation in interviews and questionnaires.

Participants

The current study's participants were from a subsample of 267 children from a more comprehensive longitudinal study. These 267 children should have received ToM measures at Time 2. However, 47 children could not be reached at this time point due to dropout or absence during the testing session. Of these 47 children, 11 were reached at Time 3 when advanced ToM was measured. Thus, those 231 children (125 boys) were included in the current sample.

At Time 1, the children had a mean age of 3;6 (years;months) (M = 41.62 months, SD = 3.95). At the other measurement points included in this study, the children had mean ages of 5;6 (M = 63.62 months, SD = 3.95), 12;8 (M = 151.70 months, SD = 3.98), and 13;7 (M = 162.86 months, SD = 3.72).

All children were born in Germany, and most (n = 213, 92.2%) had at least one parent with German as her or his native tongue. The educational and socioeconomic backgrounds of the sample were diverse. To measure SES, I refer to the family's Highest International Socioeconomic Index (HISEI; Ganzeboom, De Graaf, Treiman, & De Leeuw, 1992), an international index of occupational status. The mean HISEI in the sample was 52.30 (SD = 15.93) on a scale ranging from 16 (e.g., cleaner, unskilled farm worker) to 90 (e.g., judge in a court of law). An example occupation with an ISEI of 52 is an administrator for electronic data processing.

Measures

Theory of mind: False belief

At Time 2, children completed one first-order unexpected content false-belief task (based on Perner, Leekam, & Wimmer, 1987) and one second-order false-belief task (Sullivan, Zaitchik, & Tager-Flusberg, 1994). Both tasks were acted out with small figures.

First-order task. After demonstrating that a peanut box unexpectedly contains a ball instead of peanuts, a naive protagonist (P1) arrived and children were asked the false-belief question ("What does P1 think is in the box?") and a control question ("Did P1 look inside the box?"). Children needed to answer the control question correctly to be given 1 point on the false-belief question. Children were also given a second test question about their own belief ("Before you had a look inside the box, what did you think was inside?"). Total first-order false-belief scores ranged from 0 to 2 (M = 1.26, SD = 0.73).

Second-order task. Children were told a story about Peter, a boy who found his actual birthday present (a puppy) unbeknownst to his mother (Mum) who had told him that he would receive a different present (a toy). When Peter was absent, Peter's Mum talked to Grandma about Peter's present.

Children needed to answer two control questions ("What has Mum really got Peter for his birthday?" and "What did Peter's Mum say to him that he got for his birthday?") and three test questions. These were one first-order knowledge access question ("Does Mum know that Peter saw the dog?") and one second-order knowledge access question ("What does Mum answer to Grandma's question: Does Peter know what you got him for his birthday?") as well as one second-order false-belief question ("What does Mum answer to Grandma's question: What does Peter think you got him for his birthday?"). In total, children could earn up to 3 points for the second-order ToM task (M = 1.71, SD = 1.12). If children responded incorrectly to one of the control questions, they received a score of 0 for the test questions.

Scores on the first-order and second-order tasks were correlated, $r(220)$ = .40 and, thus, were summed to form a comprehensive ToM score.

Theory of mind: Strange Stories

At Time 3, we assessed ToM using two stories of deception, two stories of misunderstanding, and two stories of double bluff. One of the double-bluff stories was rewritten based on a story by White et al. (2009); all other stories were taken from a German translation of the Strange Stories (Rakoczy, Harder-Kasten, & Sturm, 2012). Children listened twice to each story and a subsequent open-ended question about one of the actors' behavior from an MP3 player via loudspeaker. They answered the questions verbally, and their responses were recorded, transcribed, and coded according to White et al. (2009). Children received 1 point for a partially correct response and 2 points for a fully correct response. About 25% of the transcripts were coded by a second rater. Interrater reliability was good to excellent (intraclass correlation coefficient for absolute agreement between .78 and .89; Cohen's kappa between .76 and .86). The scores for all items were summed to form a total score. Cronbach's alpha was .56.

Language: Receptive vocabulary

At Time 1, Time 2, and Time 3, children's receptive vocabulary was measured using a German research version of the Peabody Picture Vocabulary Test–Revised (PPVT; Dunn & Dunn, 1981). This test contained 175 items ordered in sets of 12 (with 7 items in the last set). The maximum possible score was 175.

Language: Receptive grammar

At Time 1, the sentence comprehension subtest of the German Language Development Test for 3- to 5-year-old children was administered (SETK 3–5; Grimm, 2001). In this test, children were given sentences varying in grammatical complexity. In the first section (9 items), they needed to determine which one of four pictures corresponded to the sentence they had just heard. In the second section (10 items), they needed to act out the content of the given sentence (e.g., "Put the blue pen under the bag"). The maximum possible score was 19.

At Time 2, children's sentence comprehension using a research version of the German adaptation of the Test for the Reception of Grammar (TROG; Bishop, 1983/1989; German version: TROG-D; Fox, 2006) was assessed. This adaptation contained 48 items and required children to select which one of four pictures corresponded to a verbally presented sentence. The research version included all the grammatical structures of the original test; the only difference was that the first three sets had 2 items rather than 4 items. The maximum possible score was 48.

Language: Text comprehension

At Time 4, six stories (each with approximately 100–150 words) from a paper-and-pencil language comprehension test for adolescents adopted from the DELKO project (Marx & Stanat, 2009) were administered. The stories varied in the complexity of vocabulary and syntax and were set in everyday contexts (e.g., conversation in a supermarket) or were more informational (e.g., text about a rare animal). Children listened to each story twice and were then asked three to five questions (multiple-choice and open-ended questions; 25 questions in total). For example, children were asked to recall or compare information or make inferences. Partially correct answers were given 0.5 points. The maximum possible score was 25. About 22% of the answers were coded by a second rater, and interrater reliability was good to excellent (intraclass correlation coefficient [absolute agreement] between .90 and .98; Cohen's kappa between .76 and .96). The scores for all items were summed to form a total score. Cronbach's alpha was .64.

For some analyses, the two language measures conducted at the same time point were z-standardized, summed, and averaged. Correlations between the vocabulary and grammar measures were $r(225) = .70$ at Time 1 and $r(211) = .57$ at Time 2.

Reading: Text comprehension

At Time 4, a paper-and-pencil test developed as part of the German National Educational Panel Study (NEPS; Gehrer, Zimmermann, Artelt, & Weinert, 2012) was administered. The test version used was originally developed for ninth graders and encompassed five texts of different types (informational, commentary, literary, instructional, and advertising). Each text contained about 230 words and was followed by five to seven questions, mostly multiple-choice questions with one correct option out of four options. The other questions took the form of matching and decision-making tasks. All items required children to complete tasks such as extracting information and making inferences based on the text. Children had 28 min to complete the entire test. For some matching and decision-making tasks, partially correct answers were possible. The maximum possible score was 33.

Nonverbal cognitive abilities

At Time 1, the Analogies and Categories subtests from the Snijders–Oomen Nonverbal Intelligence Test for 2½- to 7-year-old children (SON-R 2½-7; Tellegen, Winkel, Wijnberg-Williams, & Laros, 2005) were administered as an indicator for children's nonverbal reasoning skills. This test asks children to infer sorting and classification principles and sort, classify, or categorize either abstract materials of various shapes and colors or picture cards. The maximum possible score was 17 for Analogies and 15 for Categories. Scores for the two subtests were standardized, summed, and averaged. The correlation between the subtests was $r(224) = .48$.

Working memory

At Time 1 two subtests from the German version of the Kaufman Assessment Battery for Children (K-ABC; Melchers & Preuß, 2003) were administered. In the Digit Span subtest, children answered 12 items clustered into sets of 3 items of identical length (2–5 digits). They earned 1 point for each correctly repeated item. In the Hand Movements subtest, children needed to repeat a sequence of taps on the table performed by the research assistant with their fist, palm, or side of their hand. The test consisted of 12 items differing in length (2–4 hand movements each) grouped into sets of 4 items. Children earned 1 point for each correctly repeated sequence. The maximum possible scores were 12. Test scores for the two subtests were standardized, summed, and averaged. The correlation between the subtests was $r(218) = .51$.

Family background

Parents' education and SES were assessed via a computer-assisted personal interview (CAPI) with the main caregiver in each child's home at Time 1.

Results

Missing data and analytical strategy

Table 1 presents descriptive statistics for the cognitive and language measures included in the current study. For various reasons (no further consent to participate was given, the family moved, illness, technical problems, child refusal, etc.), I did not have valid data for all children at all measurement points. Due to the large developmental period under investigation, the dropout rate between the

Table 1
Descriptive statistics of child variables.

	N	*M*	*SD*	Min	Max
Theory of mind					
False belief (age 5;6)		2.98	1.56	0	5
Strange stories (age 12;8)	112	7.70	2.07	2	12
Language					
Receptive vocabulary (PPVT) (age 3;6)	225	29.54	14.99	0	82
Receptive vocabulary (PPVT) (age 5;6)	211	75.66	21.37	13	139
Receptive vocabulary (PPVT) (age 12;8)	114	147.02	10.30	98	167
Receptive grammar (SC) (age 3;6)	224	11.06	4.28	0	19
Receptive grammar (TROG) (age 5;6)	220	36.32	4.99	17	45
Text comprehension (DELKO) (age 13;7)	110	19.62	3.05	9	24.50
Reading					
Text comprehension (NEPS) (age 13;7)	111	21.39	6.03	8.50	31.75
Nonverbal reasoning					
Analogies (SON) (age 3;6)	224	6.10	2.32	0	12
Categories (SON) (age 3;6)	214	5.68	1.95	0	10
Working memory					
Digit span (K-ABC) (age 3;6)	217	3.15	2.58	0	9
Hand movements (K-ABC) (age 3;6)	218	3.15	2.14	0	10

Note. PPVT, Peabody Picture Vocabulary Test; SC, test for sentence comprehension; TROG, Test for the Reception of Grammar; ToM, theory of mind; DELKO, test for listening comprehension; NEPS, test for reading comprehension; SON, Snijders–Oomen Nonverbal Intelligence Test; K-ABC, Kaufman Assessment Battery for Children. Ages are in years;months.

measurement points in early childhood and early adolescence was particularly high. However, there are good statistical solutions for missing data, especially in longitudinal research, that do not lead to biased estimates, although missingness mechanisms need to be considered appropriately (Graham, 2009).

I did not assume that the data were missing completely at random (MCAR). Moreover, Little's MCAR test was significant, $\chi^2(308) = 517.39$, $p < .00$. However, there is no reason to believe that advanced ToM itself predicts whether a participant has missing data on advanced ToM; rather, children from lower SES backgrounds are more likely to drop out of the study, for example, due to more frequent moves or stress. Moreover, it is known that SES is associated with early cognitive abilities. Indeed, the 112 children from whom we had a valid measure in advanced ToM differed in neither age, $t(223) = -0.37$, $p = .71$, nor language background from the children who left the study, $\chi^2(2) = 1.85$, $p = .40$, but were advanced in cognitive and language skills, $F(9, 185) = 2.26$, $p = .02$, and their family's HISEI was higher, $t(229) = -2.12$, $p = .04$. Thus, I assumed that the data are missing at random. This means that missing values are systematically related to other observed variables (Enders, 2013) and that after controlling for "all the variables one has, any remaining missingness is completely at random" (Graham, 2009, p. 552).

I used a full information maximum likelihood (FIML) approach to account for the missing data and included early cognitive and language variables as well as background variables for which missingness was rare as control variables in the model. FIML including control variables is a good means of handling data that are missing at random, especially incomplete outcome variables, and results in less biased parameter estimates than older methods (Enders, 2013; Graham, 2003). Thus, FIML is superior to listwise deletion, pairwise deletion, and similar response imputation, especially in small sample sizes (Enders & Bandalos, 2001).

To analyze the longitudinal association between language and ToM as well as between these variables and reading comprehension in more detail, I specified path models using Mplus 7 (Muthén & Muthén, 2012). I refrained from estimating latent variables to keep the structural equation simple and the sample size-to-parameter ratio low. This increases the likelihood that the statistical requirements will be met even though our sample size is not huge and missing data are estimated (Kline, 2016).

I controlled for autoregressive effects in all models, which enables determining whether the relations are bidirectional or solely in one direction. Only if a relation between early and later measures is observed after controlling for autoregressive effects can the earlier measure be said to predict developmental change and might be causally related to the later measure (see Ruffman, Slade, & Crowe, 2002). I also considered direct and indirect links between language and ToM over time and allowed the predictor variables assessed at one time point to correlate.

Moreover, I included HISEI and the composite scores for nonverbal cognitive abilities and working memory at Wave 1 as control variables in all models. Hence, I specified paths between these control variables and all outcome measures at all measurement points. I also controlled for age by specifying paths between concurrent age and the cognitive and language measures. I report all significant and nonsignificant paths that were specified among ToM, language, and reading measures in the figures. For the other variables, I report only those paths and standardized beta weights that approached significance at $p < .10$ for easier readability.

The criteria for model fit were as follows: a root mean square error of approximation (RMSEA) < .08, a comparative fit index (CFI) > .95, and a nonsignificant chi-square test of model fit (Brown, 2006; Hu & Bentler, 1999).

Relations between ToM and language over time

As shown in Table 2, longitudinal and concurrent correlations between the aggregate language measures and ToM were moderate to high at all measurement points. Table 2 also shows that the association between vocabulary and ToM seems to change across waves and varies for early versus advanced ToM. Thus, PPVT at age 3;6 correlates stronger with ToM at age 5;6 than with ToM at age 12;8, whereas PPVT at age 5;6 correlates stronger with ToM at age 5;6 than with ToM at age 12;8. In contrast, the association between grammar and ToM seems to be more stable over time. Thus, the grammar measure correlates similarly with early ToM and advanced ToM at both measurement points in preschool.

To further investigate the reciprocal relations between different facets of language and ToM, I specified three equivalent path models including different language measures for preschool (see Fig. 2) to predict advanced ToM and language. I used the vocabulary measure in Model 1, the grammar measure in Model 2, and the aggregate language score consisting of both vocabulary and grammar in Model 3.

All models fit the data very well (see Fig. 2). However, the models revealed both similarities and differences in the longitudinal relation between language and ToM. In all models, there was a significant direct effect of language at Time 1 on ToM in preschool and adolescence. However, the beta weight between vocabulary at age 3;6 and preschool ToM (β = .24) was much smaller than the one between grammar at age 3;6 and preschool ToM (β = .51). In addition, the relation between language measures at age 5;6 and advanced ToM differed across the three language measures; even after controlling for language and various other control variables 2 years earlier, grammar had a direct effect on the change in ToM between preschool and early adolescence, whereas vocabulary and the aggregate score for language did not. Furthermore, ToM significantly predicted changes in vocabulary between ages 5;6 and 12;8, even when controlling for earlier vocabulary and other control variables, but had no direct effect on listening comprehension.

Indirect effects of early language on advanced ToM via preschool ToM were found only for vocabulary at age 3;6 (β = .07, $p < .05$). In addition, there was an indirect effect of early vocabulary on vocabulary at age 12;8 via preschool ToM (β = .06, $p < .05$).

Relations among ToM, language, and reading

To investigate whether ToM is related to later reading comprehension after controlling for language skills, I specified path models that, besides the models in Fig. 2, included a measure of reading comprehension (see Fig. 3).

First, I specified a model that was similar to Model 3 (see Fig. 2) but included only preschool measures and reading comprehension in early adolescence (Model a in Fig. 3). This was done to test for

Table 2
Concurrent and longitudinal correlations between language and cognitive measures.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Language (3;6 years)	–	.92**	.92**	.49**	.42**	.67**	.60**	.60**	.51**	.58**	.53**	.50**	.48**
2. PPVT (3;6 years)	.91**	–	.70**	.45**	.40**	.61**	.58**	.53**	.39**	.52**	.50**	.45**	.39**
3. SC (3;6 years)	.91**	.67**	–	.46**	.39**	.62**	.54**	.58**	.56**	.53**	.47**	.47**	.48**
4. WM (3;6 years)	.43**	.39**	.40**	–	.37**	.41**	.35**	.37**	.35**	.31**	.28**	.30**	.34**
5. NON (3;6 years)	.35**	.32**	.32**	.30**	–	.46**	.41**	.41**	.26**	.33**	.25**	.27**	.31**
6. Language (5;6 years)	.66**	.59**	.61**	.38**	.43**	–	.89**	.89**	.55**	.39**	.64**	.52**	.40**
7. PPVT (5;6 years)	.56**	.53**	.50**	.31**	.36**	.88**	–	.57**	.51**	.23*	.49**	.45**	.26**
8. TROG (5;6 years)	.61**	.53**	.58**	.36**	.39**	.89**	.56**	–	.48**	.46**	.63**	.46**	.45**
9. ToM (5;6 years)	.51**	.37**	.55**	.33**	.24**	.54**	.50**	.48**	–	.39**	.48**	.35**	.24*
10. ToM (12;8 years)	.56**	.50**	.51**	.27**	.29**	.35**	.17+	.44**	.38**	–	.49**	.47**	.45**
11. PPVT (12;8 years)	.56**	.52**	.49**	.29**	.27**	.66**	.52**	.64**	.48**	.49**	–	.61**	.56**
12. DELKO (13;7 years)	.52**	.47**	.48**	.31**	.28**	.54**	.47**	.47**	.35**	.47**	.61**	–	.64**
13. NEPS (13;7 years)	.49**	.40**	.49**	.34**	.32**	.40**	.26**	.45**	.24*	.45**	.56**	.64**	–
Gender	.03	−.02	.08	.00	.17**	.00	−.01	.06	.09	.07	.00	−.07	.15
HISEI	.36**	.30**	.37**	.07	.18**	.37**	.34**	.33**	.21**	.22**	.34**	.35**	.26**

Note. Language, aggregate score of language measures at that wave; PPVT, Peabody Picture Vocabulary Test; SC, test for sentence comprehension; WM, working memory; NON, nonverbal abilities; TROG, Test for the Reception of Grammar; ToM, theory of mind; DELKO, test for listening comprehension; NEPS, test for reading comprehension; HISEI, Highest International Socioeconomic Index. Partial correlations controlling for age are presented under the diagonal. Concurrent correlations control for concurrent age, whereas longitudinal correlations control for age at the earlier time point.
*p < .05., **p < .01.

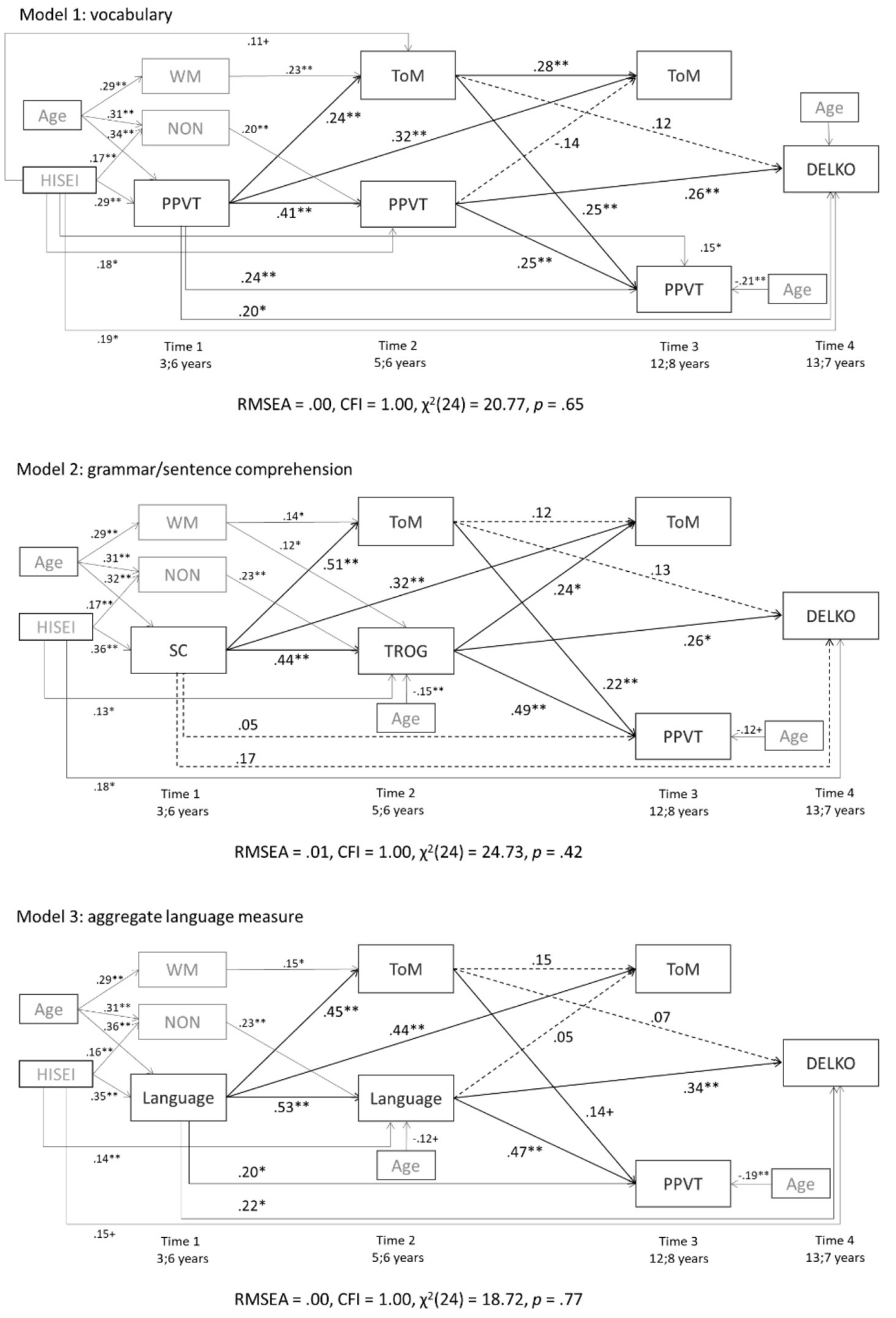

Fig. 2. Path models depicting longitudinal relations (standardized beta weights) between language and theory of mind. HISEI, Highest International Socioeconomic Index; WM, working memory; NON, nonverbal cognitive abilities; ToM, theory of mind; PPVT, Peabody Picture Vocabulary Test; SC, test for sentence comprehension; TROG, Test for the Reception of Grammar; Language, aggregate score of vocabulary and grammar measures; DELKO, test for listening comprehension; RMSEA, Root Mean Square Error of Approximation; CFI, Comparative Fit Index. +$p < .10$; *$p < .05$; **$p < .01$.

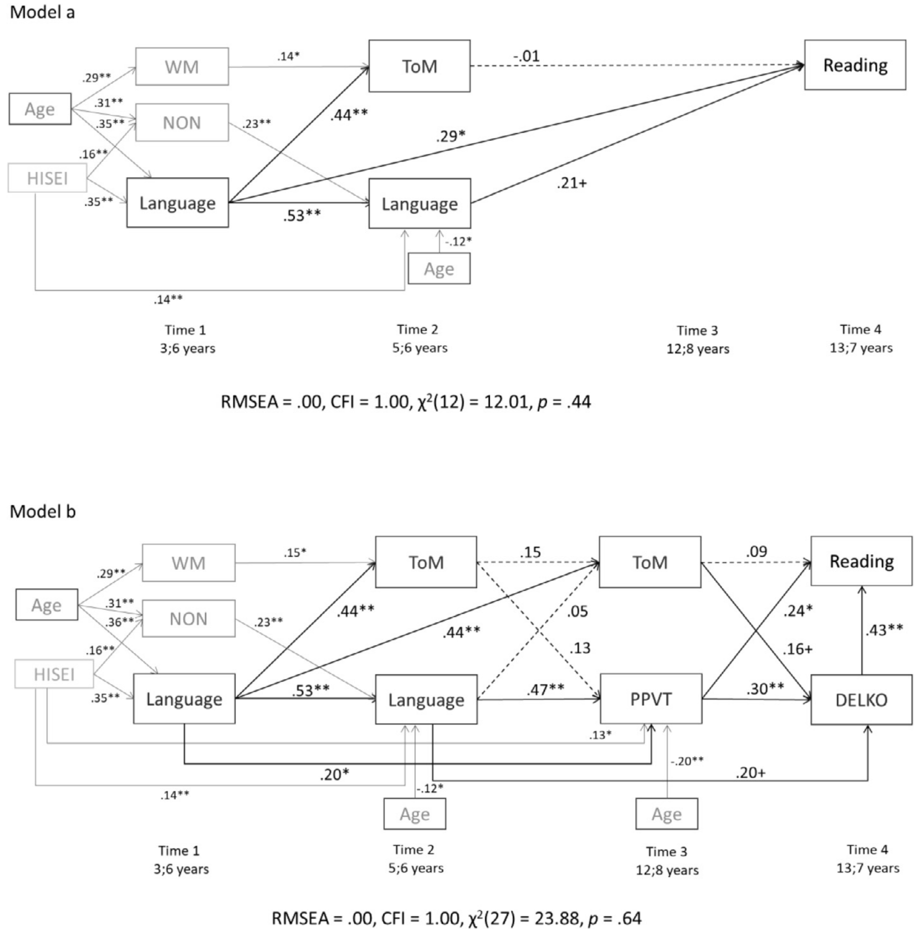

Fig. 3. Path diagrams depicting longitudinal relations (standardized beta weights) among theory of mind, language, and reading. HISEI, Highest International Socioeconomic Index; WM, working memory; NON, nonverbal abilities; Language, aggregate score of vocabulary and sentence comprehension measures; ToM, theory of mind; PPVT, Peabody Picture Vocabulary Test; SC, test for sentence comprehension; TROG, Test for the Reception of Grammar; DELKO, test for listening comprehension; RMSEA, Root Mean Square Error of Approximation; CFI, Comparative Fit Index. $+p < .10$; $^{*}p < .05$; $^{**}p < .01$.

direct effects of early language and ToM on later reading comprehension. In a second model, I also included advanced ToM and language measures (Modell b in Fig. 3).

Model a in Fig. 3 shows that early ToM has no direct effect on later reading comprehension. There is only a direct effect of early language measures on later reading comprehension. There were also no significant indirect effects of early ToM and later reading comprehension via language and ToM measures in early adolescence (Model b in Fig. 3). However, a small indirect effect of advanced ToM via listening comprehension on reading comprehension that approaches significance was found ($\beta = .07$, $p < .10$). In contrast, both language measures in early preschool showed indirect effects on later reading comprehension. These effects were mainly mediated by language measures in early adolescence.

Relations between early and advanced ToM

As can be seen in Table 1, early ToM at age 5;6 and advanced ToM at age 12;8 were moderately correlated, $r(101) = .39$, $p < .01$. This correlation hardly changed after controlling for age ($r_p = .38$,

$p < .01$) and remained substantial even after additionally accounting for early nonverbal abilities ($r_p = .32$, $p < .01$). However, Fig. 2 shows that after also controlling for early sentence comprehension ($\beta = .12$; see Model 2), the correlation between early and advanced ToM was substantially reduced, whereas it remained significant when controlling for early vocabulary ($\beta = .28$; see Model 1).

Given that the advanced ToM measure required a large amount of listening comprehension, I tested whether a specific correlation between early ToM and advanced ToM existed beyond the effect of general listening comprehension needed for the Strange Stories. That is, I controlled for children's general listening comprehension in adolescence. Therefore, I regressed the advanced ToM score on the listening comprehension score (DELKO). The correlation between early ToM and the residual of advanced ToM was substantial, $r(93) = .22$, $p < .05$.

Discussion

The current study replicates and extends former research by showing that individual differences in language measures—especially in early sentence comprehension—predict changes in ToM from preschool until early adolescence. This is a time interval of more than 7 years. The study also revealed a reciprocal relation between language and ToM in this developmental period and showed that early ToM predicts children's vocabulary. However, early ToM had neither direct nor indirect effects on reading comprehension at 13;7 years of age, and only advanced ToM exerted small indirect effects on reading comprehension via listening comprehension. Moreover, the results confirmed and extended prior studies by demonstrating that individual differences in ToM exhibit a moderate level of stability across a period of 7 years. These results are discussed in more detail in the following sections.

Relations between early language and ToM

Consistent with earlier research, the study clearly demonstrated that children's language skills are strongly correlated with ToM in their preschool years. In addition, language skills measured as early as age 3;6 predicted the change in ToM between preschool and early adolescence. Even though there were no indirect effects of early language skills via ToM or later language skills on children's advanced ToM, the findings underscore the importance of early language skills, which provide children with a strong start and continue to promote their social understanding until early adolescence at least.

In addition, the study provides evidence that there are differential effects of language measures on ToM development and that these differential effects might change as children grow older. In early preschool, individual differences in vocabulary and grammar both predicted children's advanced ToM after controlling for early ToM and language; two years later, only children's grammar predicted their further ToM development.

There are at least two possible explanations for why language and especially receptive grammar/sentence comprehension may be important for advanced ToM. First, advanced ToM tasks such as the Strange Stories are linguistic tasks. Children need to comprehend and verbally respond to stories that are not acted out with puppets (in contrast to early ToM tests). One way to investigate whether language is associated only due to linguistic requirements with advanced ToM would be to measure children's advanced ToM using tasks that are less verbal or even nonverbal. It has been shown in numerous ways that language is associated with early ToM over and above the two constructs' shared linguistic nature (Astington & Baird, 2005b). With regard to advanced ToM, Devine et al. (2016) study may provide some insight. The authors of that study administered various advanced ToM tasks that differed in their verbal requirements. Unfortunately, they did not report whether differences were found in the associations between the respective tasks and language measures. Thus, it is not known whether the relation between early language and advanced ToM varies with verbal task requirements. However, I assume that the Strange Stories are not simply a test of text comprehension. This assumption is supported by the finding that the advanced ToM measure and the listening comprehension measure were predicted differently by our three language measures. In addition, and as expected, reading comprehension was more strongly correlated with listening comprehension than with advanced ToM. If the advanced ToM measure were simply a measure of listening comprehension of texts, listening comprehension of texts and advanced ToM should show stronger similarities in correlational patterns. Moreover, several prior studies support my assumption by showing that comprehension skills for texts including mental content (like the Strange Stories) are differently correlated with other ToM measures and executive functions than comprehension skills for texts including physical content (Lecce et al., 2019; Rakoczy et al., 2012).

Another possible explanation for the relation between language and advanced ToM is that advanced ToM is typically applied in social situations based on language. Misunderstandings, double bluffs, and persuasion are impossible without interpersonal and primarily verbal communication. Thus, the question arises whether advanced ToM occurs only in linguistic situations. I assume that even if advanced ToM means in particular paying more attention to mental states in social situations and, thus, mentally interpreting social situations more easily and flexibly, this does not necessarily mean that the social situation must be a verbal one. Hence, alternative measures of ToM suggest nonverbal forms of advanced ToM. For example, in the "Reading the Mind in the Eyes Test" (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001), people must infer a mental state from photographs of the eye region. Nevertheless, language might have supported the development of one's ability to pay more attention to people's mental states. Indeed, this is exactly what the current study demonstrates, namely that it is particularly children's early language that predicts children's advanced ToM. Language and in particular syntactical skills allow children to represent mental phenomena, take part in verbal interactions, and provide opportunities to learn about mental phenomena as well as about people and their mental states. As children grow older, the more "functional properties of language as a vehicle for communication and social exchange with others" (Hughes, 2005, p. 321) may become more important for understanding others.

In contrast to syntactical skills, a rich vocabulary was less related to advanced ToM. This finding seems to contradict Lecce et al. (2014) and Devine et al. (2016), both of whom found effects of vocabulary at 5 or 6 years of age on advanced ToM. However, in contrast to the current study, they did not consider language skills at an earlier time point. Nevertheless, even when running a model that includes only vocabulary at age 5;6 and no controls for language and cognition at age 3;6, no effect of vocabulary at age 5;6 on advanced ToM was found.

Admittedly, I cannot exclude the possibility that differences in the predictability of various language measures at different measurement points might have emerged by chance, and I agree with Slade and Ruffman (2005) that children's general language skills are more important than any single component when it comes to taking part in the community of minds. However, the current study showed that language measures that are strongly correlated may nevertheless differ in their predictive utility for ToM. Indeed, receptive vocabulary and receptive grammar/sentence comprehension were very strongly correlated at Time 1, meaning that it was not methodologically possible to include them as separate measures in a single analysis without suppression effects occurring. However, the predictive value of the language measured differed despite the fact that the same sample was considered. Thus, whenever effects of language skills are discussed, it seems important to consider which facets of language one is talking about.

The current study's result that vocabulary also has an indirect effect via early ToM on advanced ToM was consistent with Ensor et al. (2014), who found an indirect effect of general language skills via early ToM on advanced ToM. However, in the current study, no indirect effects of the grammar measure or the aggregate language measure via early ToM on advanced ToM were found. In addition, there were also direct effects of all three language measures on advanced ToM. This suggests that the link between early language and advanced ToM is not mainly due to the fact that language supports a metarepresentational understanding, and this in turn helps to build an advanced mental understanding, but rather that advanced ToM emerges directly due to the support of early language skills. Most important is that language skills predicted even changes in ToM between preschool and early adolescence.

Apperly et al. (2009) used a metaphor of cement and scaffolding to describe the roles a variable might play in development. If a variable is like cement, it becomes part of the construct; if it is like a scaffold, it is no longer important after a certain point in development. Based on neuropsychological data, those authors proposed that in adulthood grammar is independent of ToM and, thus, not part of the construct and not cement, although it might have had a scaffolding role earlier in development. The current study suggests that for ToM in early adolescence, language has both roles; the high concurrent correlations between language and ToM in early adolescence propose that language is part of the construct, and the longitudinal relations suggest that language might have scaffolded the development of advanced ToM. Thus, ToM in early adolescence might be an intermediate state before children achieve an adult ToM. In addition, the results support the idea that grammar and vocabulary may play different roles in ToM development and that ToM is particularly related to more complex language skills needed for communication and not simply to a richer vocabulary.

Relations between early ToM and later language

It was hypothesized that children's early ToM supports their metacognitive knowledge and inference-making skills and, thus, is particularly predictive for listening comprehension of texts but less for vocabulary. However, the opposite was found to be the case; early ToM was predictive of changes in children's vocabulary between ages 5;6 and 12;8, but it was not predictive of text comprehension, after controlling for earlier language skills. Three possible explanations for these results are proposed. First, the vocabulary measure used at older ages may have included more abstract and mental words and, thus, may have indirectly tested children's advanced ToM. However, the more complex terms in the PPVT are not specifically mental. Another possibility is that the effect of ToM on vocabulary actually reflects the effect of a general text comprehension measure and not a specific effect of ToM. However, there is a unique effect of ToM on vocabulary at age 12;8 even after controlling for early sentence comprehension instead of early vocabulary (see Model 2 in Fig. 2). A third explanation is that as children grow older a receptive vocabulary test such as the PPVT not only might measure children's vocabulary size but also, and more strongly, might reflect children's verbal intelligence. If this is the case, the current study's results are even more intriguing. Early ToM would then be an early predictor of children's later verbal intelligence, which is important for so many aspects of their lives. That vocabulary in early adolescence might be more a measure of verbal knowledge and intelligence may also be reflected in the fact that early receptive grammar/sentence comprehension is more strongly correlated with advanced ToM than early receptive vocabulary. However, because I did not expect a relation between early ToM and later vocabulary, the result will need to be replicated and investigated in more detail in subsequent studies.

Relation among ToM, language, and reading comprehension

In the current study, a small indirect effect of advanced ToM on children's reading comprehension via listening comprehension of texts was observed. This finding is consistent with the theoretical assumption that ToM facilitates the development of higher-order comprehension skills in children (Atkinson et al., 2017; Guajardo & Cartwright, 2016; Kim, 2015). However, the effect was small and found only for advanced ToM, not for early ToM. Moreover, the advanced ToM measure did not mask the effects of early ToM. This might be explained by the fact that, in contrast to other studies, I controlled for early vocabulary and grammar and also included a measure of early and advanced ToM. Indeed, many prior studies that investigated the link between ToM and reading comprehension also found no effect or only indirect effects of early ToM on later reading comprehension via language skills (Guajardo & Cartwright, 2016; Kim, 2015, 2016; Lockl et al., 2017). However, in the current study, reading comprehension was assessed in early adolescence when children are advanced in basic reading skill and, thus, higher-order cognitive skills should be more strongly correlated with reading comprehension than in these previous studies. Nevertheless, our reading comprehension test may have been too demanding for children of this age given that it was originally designed for older children; thus, the demands made on basic reading processes may have masked the effects of higher-order inference-making skills, meaning that even advanced ToM was of little help. Contrary to this assumption, early and advanced language measures both were found to be important predictors for reading comprehension. Thus, there seems to be no specific effect of early ToM on children's reading comprehension after considering language skills.

Stability of individual differences in ToM

The results demonstrate that young adolescents differ in *how* they answer the Strange Stories and that these individual differences are related to differences in children's understanding of basic ToM concepts 7 years earlier. Hence, children who developed a more basic understanding of mental states earlier than their peers during preschool also exhibit better advanced ToM in early adolescence. However, the study's correlational findings are silent with respect to whether the Strange Stories indeed assess the application of already understood mental concepts or reflect another step in children's conceptual ToM development (see Peterson & Wellman, 2019). The moderate internal consistency of the Strange Stories suggests that different aspects of advanced ToM may be related to early ToM to differing degrees. Indeed, when differentiating between the mental concepts measured (misunderstanding, double bluff, and deception), the highest correlations between false beliefs in preschool and advanced ToM were found for double bluff stories (r = .50). Because there were only two tasks per concept included, there is a need for further research including more tasks per mental concept in order to examine whether early ToM is associated with better social understanding in general or only with specific mental concepts. There is also recent evidence that ToM shows a diverse structure, at least beyond the preschool years (Warnell & Redcay, 2019).

In confirmation of other studies, the current study supports the assumption of stable individual differences in ToM over time (Ensor et al., 2014; Devine et al., 2016), although children were followed for a longer time period, that is, up to 12 years of age. One implication of this is that children's school environment changed fundamentally during this period: Children left preschool about 6 years of age, went to elementary school for another 4 years, and eventually entered secondary school. In Germany, children enter different school tracks after Grade 4 (about 10 years of age) depending on their performance in elementary school. Hence, children had very different educational environments at 12 years of age (i.e., when advanced ToM was assessed), and they had undergone numerous experiences between assessments. Given that I did not find lower stability in ToM than Ensor et al. (2014) and Devine et al. (2016), an "environmental account of stability" (Devine et al., 2016, p. 769) seems unlikely. However, the factors that support ToM development might not vary greatly across different school environments and, thus, might not affect stability. Moreover, this does not mean that environmental factors have no influence on the development of ToM after preschool. Stability was moderate at best, and there is much room for other factors to affect development as, for example, was shown for language skills. Thus, the study shows that the correlation between early and advanced ToM was dramatically reduced when it was controlled for early grammar, but not when it was controlled for early vocabulary. This indicates that early ToM also reflects language skills.

Limitations and conclusion

The current study adds to previous research on ToM with a more thorough analysis of how ToM and language skills are linked over an extended time period of 7 years. Nevertheless, some limitations need consideration. First, internal consistency of advanced ToM was rather low, and scores were close to ceiling. In addition, both reading and listening comprehension tests were newly designed measures, which may implicate reduced reliability. Thus, the internal consistency of the listening comprehension test was rather low, and a number of children had problems in completing the reading comprehension test in time. Thus, the reading comprehension test might cover not only reading comprehension but, to an extent, also reading speed.

Another challenge of the current study was the high dropout rate of about 50%. Nevertheless, given that the study covers a period of more than 10 years in childhood including multiple measurement points, the sample size was still substantial and included families of a wide variety of SES backgrounds from urban and rural areas. I am convinced that even if attrition is high, the use of longitudinal data is critical for gathering information about relations between variables over time and development, pending good methods to deal with missing data. In addition, given that previous studies on ToM often refer to SES homogeneous samples, despite our study's high attrition, the study makes an important contribution.

Unfortunately, it was not possible to use the same measures of ToM and language (except for vocabulary) in preschool and adolescence. Thus, it was not possible to deliver evidence of how individual differences in early ToM and language contribute to "real" growth in these concepts. In addition, in the current study, only a verbal measure of children's advanced ToM was included. This made it somewhat difficult to investigate the "functional" link between early language skills and advanced ToM. However, it must be taken into account that social understanding and language may be inextricably intertwined and that early ToM develops out of language skills (e.g., Astington & Jenkins, 1999). Thus, ToM is interwoven with language. This is also reflected in the results demonstrating that language and ToM are closely reciprocally intertwined over time; language skills drive developmental changes in ToM, and ToM also drives changes in language development.

Acknowledgments

This study was part of the research group BiKS (Bildungsprozesse, Kompetenzentwicklung und Selektionsentscheidungen im Vorschul- und Grundschulalter; English: Educational Processes, Competence Development and Selection Decisions in Preschool and School-Age Children) and the follow-up project BiKSplus$_{3-13}$ (Long-Term Effects of Early Global and Domain-Specific Educational Experiences and Developments) at the University of Bamberg, funded by grants from the German Research Foundation (DFG). For the most part, data were collected within the developmental subproject (principal investigator: Sabine Weinert; Grants WE 1478/4-1, 4-2, 4-3, and 8-1). I thank Sabine Weinert for discussion and useful comments as well as Katharina Mursin for proofreading earlier drafts of the manuscript and for her work during data collection. I also thank all other colleagues involved in data collection for the BiKS study. Finally, I thank all participating children, their parents, and their (preschool) teachers, as well as all students engaged in data collection for their most active cooperation.

References

Apperly, I. A., Samson, D., & Humphreys, G. W. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental Psychology, 45*, 190–201.

Astington, J. W., & Baird, J. A. (2005a). Introduction: Why language matters. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 3–25). New York: Oxford University Press.

Astington, J. W., & Baird, J. A. (2005b). *Why language matters for theory of mind*. New York: Oxford University Press.

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology, 35*, 1311–1320.

Atkinson, L., Slade, L., Powell, D., & Levy, J. P. (2017). Theory of mind in emerging reading comprehension: A longitudinal study of early indirect and direct effects. *Journal of Experimental Child Psychology, 164*, 225–238.

Banerjee, R., Watling, D., & Caputi, M. (2011). Peer relations and the understanding of faux pas: Longitudinal evidence for bidirectional associations. *Child Development, 82*, 1887–1905.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42*, 241–251.

Bishop, D. V. M. (1989). TROG—Test for the Reception of Grammar. Manchester, UK: University of Manchester. (Original work published 1983).

Boerma, I. E., Mol, S. E., & Jolles, J. (2017). The role of home literacy environment, mentalizing, expressive verbal ability, and print exposure in third and fourth graders' reading comprehension. *Scientific Studies of Reading, 21*, 179–193.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.

Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.

Carpendale, J., & Chandler, M. J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development, 67*, 1686–1704.

de Villiers, J. G. (2005). Can language acquisition give children a point of view? In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 187–219). New York: Oxford University Press.

Devine, R. T., & Hughes, C. (2013). Silent Films and Strange Stories: Theory of mind, gender, and social experiences in middle childhood. *Child Development, 84*, 989–1003.

Devine, R. T., White, N., Ensor, R., & Hughes, C. (2016). Theory of mind in middle childhood: Longitudinal associations with executive function and social competence. *Developmental Psychology, 52*, 758–771.

Dickinson, D. K., McCabe, A., Anastasopoulos, L., Peisner-Feinberg, E. S., & Poe, M. D. (2003). The comprehensive language approach to early literacy: The interrelationships among vocabulary, phonological sensitivity, and print knowledge among preschool-aged children. *Journal of Educational Psychology, 95*, 465–481.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test-Revised (PPVT-R)*. Circle Pines, MN: American Guidance Service.

Ebert, S. (2015). Longitudinal relations between theory of mind and metacognition and the impact of language. *Journal of Cognition and Development, 16*, 559–586.

Ebert, S., Peterson, C., Slaughter, V., & Weinert, S. (2017). Links among parents' mental state language, family socioeconomic status, and preschoolers' theory of mind development. *Cognitive Development, 44*, 32–48.
Ebert, S., & Weinert, S. (2013). Predicting reading literacy in primary school: The contribution of various language indicators in preschool. In P. Maximilian, C. Artelt, & S. Weinert (Eds.), *The development of reading literacy from early childhood to adolescence—Empirical findings from the Bamberg BiKS longitudinal studies* (pp. 93–149). Bamberg, Germany: University of Bamberg Press.
Enders, C. K. (2013). Dealing with missing data in developmental research. *Child Development Perspectives, 7*, 27–31.
Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*, 430–457.
Ensor, R., Devine, R. T., Marks, A., & Hughes, C. (2014). Mothers' cognitive references to 2-year-olds predict theory of mind at ages 6 and 10. *Child Development, 85*, 1222–1235.
Fox, A. (2006). *TROG-D: Test zur Überprüfung des Grammatikverständnisses [Test for the Reception of Grammar]*. Idstein, Germany: Schulz-Kirchner.
Ganzeboom, H. B. G., De Graaf, P. M., Treiman, D. J., & De Leeuw, J. (1992). A standard international socioeconomic index of occupational-status. *Social Science Research, 21*, 1–56.
Gehrer, K., Zimmermann, S., Artelt, C., & Weinert, S. (2012). *The assessment of reading competence (including sample items for Grade 5 and 9): Scientific Use File 12, Version 1.0.0*. Bamberg, Germany: University of Bamberg, National Educational Panel Study.
Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling, 10*, 80–100.
Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549–576.
Grimm, H. (2001). *Sprachentwicklungstest für drei- bis fünfjährige Kinder (SETK 3–5) [Language development test for three- to five-year-old children]*. Göttingen, Germany: Hogrefe.
Guajardo, N. R., & Cartwright, K. B. (2016). The contribution of theory of mind, counterfactual reasoning, and executive function to pre-readers' language comprehension and later reading awareness and comprehension in elementary school. *Journal of Experimental Child Psychology, 144*, 27–45.
Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science, 6*, 346–359.
Harris, P. L. (2005). Conversation, pretense, and theory of mind. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 70–83). New York: Oxford University Press.
Harris, P. L., de Rosnay, M., & Pons, F. (2005). Language and children's understanding of mental states. *Current Directions in Psychological Science, 14*, 69–73.
Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*, 127–160.
Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.
Hughes, C. (2005). Genetic and environmental influences on individual differences in language and theory of mind: Common or distinct? In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 319–339). New York, NY: Oxford University Press.
Hughes, C. (2016). Theory of mind grows up: Reflections on new research on theory of mind in middle childhood and adolescence. *Journal of Experimental Child Psychology, 149*, 1–5.
Hughes, C., & Devine, R. T. (2015). Individual differences in theory of mind from preschool to adolescence: Achievements and directions. *Child Development Perspectives, 9*, 149–153.
Im-Bolter, N., Agostino, A., & Owens-Jaffray, K. (2016). Theory of mind in middle childhood and early adolescence: Different from before? *Journal of Experimental Child Psychology, 149*, 98–115.
Keenan, T. (2003). Individual differences in theory of mind: The preschool years and beyond. In B. Repacholi & V. Slaughter (Eds.), *Individual differences in theory of mind: Implications for typical and atypical development* (pp. 121–142). New York: Psychology Press.
Kim, Y.-S. (2015). Language and cognitive predictors of text comprehension: Evidence from multivariate analysis. *Child Development, 86*, 128–144.
Kim, Y.-S. (2016). Direct and mediated effects of language and cognitive skills on comprehension of oral narrative texts (listening comprehension) for children. *Journal of Experimental Child Psychology, 141*, 101–120.
Kline, R. B. (2016). *Principles and practice of structural equation modeling* (fourth ed.). New York: Guilford.
Lecce, S., Bianco, F., Devine, R. T., & Hughes, C. (2017). Relations between theory of mind and executive function in middle childhood: A short-term longitudinal study. *Journal of Experimental Child Psychology, 163*, 69–86.
Lecce, S., Caputi, M., & Pagnin, A. (2014). Long-term effect of theory of mind on school achievement: The role of sensitivity to criticism. *European Journal of Developmental Psychology, 11*, 305–318.
Lecce, S., Ronchi, L., Del Sette, P., Bischetti, L., & Bambini, V. (2019). Interpreting physical and mental metaphors: Is theory of mind associated with pragmatics in middle childhood? *Journal of Child Language, 46*, 393–407.
Lockl, K., Ebert, S., & Weinert, S. (2017). Predicting school achievement from early theory of mind: Differential effects on achievement tests and teacher ratings. *Learning and Individual Differences, 53*, 93–102.
Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development, 74*, 1130–1144.
Low, J., & Simpson, S. (2012). Effects of labeling on preschoolers' explicit false belief performance: Outcomes of cognitive flexibility or inhibitory control? *Child Development, 83*, 1072–1084.
Marx, A., & Stanat, P. (2009). Entwicklung eines Hörverstehenstest für Jugendliche [Development of an oral text comprehension test for adolescence]. Paper presented at the 72. Tagung der Arbeitsgruppe für empirische pädagogische Forschung (AEPF), Landau, Germany.
Melchers, P., & Preuß, U. (2003). *Kaufman Assessment Battery for Children (K-ABC)* (German ed., 6th ed.). Frankfurt, Germany: Swets & Zeitlinger.
Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin, 135*, 749–773.

Miller, S. A. (2016). *Parenting and theory of mind*. New York: Oxford University Press.
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*, 622–646.
Muthén, L. K., & Muthén, B. (2012). *Mplus users guide* (7th ed.). Los Angeles: Muthén & Muthén.
Nelson, K. (2005). Language pathways into the community of minds. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 26–49). New York: Oxford University Press.
Pelletier, J., & Astington, J. W. (2004). Action, consciousness and theory of mind: Children's ability to coordinate story characters' actions and thoughts. *Early Education and Development, 15*, 5–22.
Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.
Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief—The case for a conceptual deficit. *British Journal of Developmental Psychology, 5*, 125–137.
Peterson, C., & Wellman, H. M. (2019). Longitudinal theory of mind (ToM) development from preschool to adolescence with and without ToM delay. *Child Development, 90*, 1917–1934.
Rakoczy, H., Harder-Kasten, A., & Sturm, L. (2012). The decline of theory of mind in old age is (partly) mediated by developmental changes in domain-general abilities. *British Journal of Psychology, 103*, 58–72.
Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory-of-mind understanding. *Child Development, 73*, 734–751.
Slade, L., & Ruffman, T. (2005). How language does (and does not) relate to theory of mind: A longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology, 23*, 117–141.
Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*, 934–947.
Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology, 30*, 395–402.
Tellegen, P. J., Winkel, M., Wijnberg-Williams, B. J., & Laros, J. A. (2005). *Snijders-Oomen non-verbaler Intellingenztest für Kinder von 2½–7 Jahre (SON-R 2½-7) [Snijders–Oomen Nonverbal Intelligence Test for 2½- to 7-year-old children]*. Lisse, The Netherlands: Swets & Zeitlinger.
Warnell, K. R., & Redcay, E. (2019). Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition, 191* 103997.
Weimer, A. A., Dowds, S. J. P., Fabricius, W. V., Schwanenflugel, P. J., & Suh, G. W. (2017). Development of constructivist theory of mind from middle childhood to early adulthood and its relation to social cognition and behavior. *Journal of Experimental Child Psychology, 154*, 28–45.
Wellman, H. M. (2014). *Making minds: How theory of mind develops*. New York: Oxford University Press.
Wellman, H. M., & Peterson, C. (2013). Theory of mind, development, and deafness. In S. Baron-Cohen, H. Tager-Flusberg, & M. V. Lombardo (Eds.), *Understanding other minds: Perspectives from developmental social neuroscience* (third ed., pp. 51–71). New York: Oxford University Press.
White, S., Hill, E., Happe, F., & Frith, U. (2009). Revisiting the Strange Stories: Revealing mentalizing impairments in autism. *Child Development, 80*, 1097–1117.