the british psychological society
promoting excellence in psychology

British Journal of Developmental Psychology (2020), 38, 580–593

www.wileyonlinelibrary.com

Theory of Mind and diverse intelligences in 4-year-olds: Modelling associations of false beliefs with children's numerate-spatial, verbal, and social intelligence

Christopher Osterhaus[1]*, Diane L. Putnick[2], Susanne Kristen-Antonow[1], Daniela Kloo[1], Marc H. Bornstein[2,3] and Beate Sodian[1]

[1]Ludwig-Maximilians-Universität München, Germany
[2]Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, Maryland,
[3]Institute for Fiscal Studies, London, UK

Theory of Mind (ToM) and the structure of intelligence were investigated in 115 4-year-olds. Specifically, we asked whether children's intelligence involves both general and specific aspects and whether standard ToM measures of false belief can serve as indicators of social intelligence. Psychometric intelligence and children's domain-specific understanding of number concepts and of mental states (false belief) were measured in the laboratory; communication and social skills were assessed through mothers' report. A confirmatory factor analysis revealed poor fit for a one-factor model, but good fit for a model with three correlated factors, suggesting that children's intelligence involves both general and specific aspects. Numerate-spatial and verbal intelligence were correlated (.70), and social intelligence correlated to a stronger degree with verbal (.66) than with numerate-spatial intelligence (.37). Laboratory assessment of false belief and mothers' reports about children's social skills loaded on a single factor, pointing to real-world consequences of ToM abilities.

Statement of contribution

What is already known on this subject?

- The structure of intelligence in 4-year-olds comprises domain-general and domain-specific dimensions.
- Some domain-specific dimensions are numerate-spatial, verbal, and social intelligence.

What does this study add?

- Theory of Mind emerges as an aspect of children's social intelligence.
- Social intelligence (including Theory of Mind) is related to children's numerate-spatial abilities.

*Correspondence should be addressed to Christopher Osterhaus, Department of Psychology, Ludwig-Maximilians-Universität München, Leopoldstraße 13, 80802 München, Germany (email: c.osterhaus@psy.lmu.de).

DOI:10.1111/bjdp.12336

Three decades of Theory of Mind (ToM) research have led to rich results and insights into children's developing understanding of their own and others' minds (for a review, see Wellman, 2017). A milestone in children's developing ToM is false belief understanding (Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983). Understanding false belief rests on the insight that mental states (e.g., beliefs) are representations that can differ from the actual state of affairs in the external world. False belief understanding emerges around the age of 4 years (Flavell, 2004; Sodian, 2005; Wellman, 2002).

Because of its foundational importance for human intelligence, false belief understanding has been viewed as a 'cognitive watershed' in child development (see, e.g., Wellman & Liu, 2004). Consistent with this view, marked improvements in other aspects of cognitive functioning, such as language (embedded sentences) and executive functions, are associated with the development of false belief understanding (Carlson & Moses, 2001; Kloo, Kristen-Antonow, & Sodian, 2019; Milligan, Astington, & Dack, 2007). Individual differences in ToM abilities in preschool-age children predict academic and social functioning in school (Lecce, Caputi, & Pagnin, 2014; Lockl, Ebert, & Weinert, 2017). Despite its importance for cognitive development, false belief understanding, to date, has not played a role in psychometric models of intelligence in early childhood, nor do intelligence tests assess it. Children's acquisition of central ToM concepts has been modelled in a scale (see Wellman & Liu, 2004), but the interrelations of ToM and domain-specific as well as domain-general intelligences have not been addressed through psychometric modelling.

An entirely novel approach to modelling relations of domain-general and domain-specific dimensions of intelligence in early childhood was taken by Bornstein and Putnick (2019). Bornstein and Putnick assessed 318 4-year-olds with a newly developed test battery that measures children's psychometric intelligence, their numeracy, verbality, mechanical problem-solving, artistry, motoricity, and sociability (assessed directly) as well as children's adaptive behaviours (assessed in interviews with children's mothers). Using the data obtained with this battery, Bornstein and Putnick tested several theoretical models of the structure of intelligence: a monolithic model, which assumes that the correlations between distinct measures of cognitive development are due to the involvement of one core component, or general intelligence – *g* (e.g., Spearman, 1904, 1927); a multidimensional model, which emphasizes the operation of separate faculties – *F*s and the existence of specific subcomponents of intelligence (e.g., Thurstone, 1938; Thurstone & Thurstone, 1941); and a hierarchical model, which regards cognitive abilities to reflect both general and specific intelligences (e.g., Carroll, 1993). The hierarchical model emerged as the best-fitting model, suggesting that 4-year-olds' intelligence involves both a domain-general component (i.e., shared variances between the tasks) and domain-specific aspects (i.e., specific variance components). The domain-specific dimensions were numerate-spatial, verbal, motor, and interpersonal intelligence.

In the present paper, we follow up on Bornstein and Putnick's (2019) intelligence model by testing a model that includes ToM as a central component of intellectual functioning in 4-year-olds. Bornstein and Putnick's (2019) interpersonal intelligence was composed of children's conceptual understanding of interpersonal relationships, measured by children's adaptive socialization behaviours (assessed in the Vineland Adaptive Behavior Scales; Sparrow, Cicchetti, & Balla, 2005, 2008) and a friendship interview (Furman & Bierman, 1983). The friendship interview taps children's social perspective-taking abilities in a complex, real-world social context. Because of its verbal demands, however, it often yields low-level responses in kindergarten and young elementary school children (Bornstein & Putnick, 2019; Gurucharri & Selman, 1982; Marcone *et al.*, 2015). False belief tasks address the foundational distinction of beliefs and reality which underlies perspective-taking abilities. False belief tasks pose fairly minimal verbal and memory demands and are therefore considered a valid measure of perspective understanding. False belief understanding is considered a foundational aspect of social intelligence and is associated with, and sometimes seen as necessary for, the development of interpersonal relationships (Etel & Slaughter, 2019; Fink, Begeer, Peterson, Slaughter, & de Rosnay, 2015; Imuta, Henry, Slaughter, Selcuk, & Ruffman, 2016; Peterson, Slaughter, Moore, & Wellman, 2016; Peterson, Slaughter, & Paynter, 2007; Slaughter *et al.*, 2015). Therefore, the present study replaces the friendship interview with three measures of false belief understanding and re-examines relations to domain-general and domain-specific aspects of children's intelligence.

Investigating the general structural model of intelligence (Bornstein & Putnick, 2019), the present study comprises a sample of 115 German 4-year-olds. Like Bornstein and Putnick, we assessed children's psychometric intelligence using three verbal (Information, Arithmetic, and Similarities) and two performance subtests (Block design and Picture completion) from the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III German version; Petermann, 2009). In addition, we measured children's ability to use number concepts with a counting game (Bornstein & Putnick, 2019), and we used the Vineland Adaptive Behavior Scales (VABS; Sparrow *et al.*, 2005, 2008) to obtain mothers' assessment of their children's communication and social relation skills. (Our analysis does not include children's artistry and motoricity.)

To investigate the structure of children's intelligence, we conducted several confirmatory factor analyses (CFAs) and fitted the monolithic model as well as several multidimensional models to the data, asking (1) if false belief understanding can be modelled as an aspect of children's interpersonal (social) intelligence (specificity of false belief understanding), (2) if the multidimensional model (numerate-spatial, verbal, and social intelligence) fits the data better than the monolithic, single-factor model (structure of intelligence), and (3) if children's social intelligence (including their false belief understanding) is related to the numerate-spatial and verbal factors (shared variance and relations between domain-specific dimensions).

Associations between false belief and children's verbal abilities are consistently reported in the literature (Milligan *et al.*, 2007), and hence, we expected to find a significant correlation between children's social and verbal intelligences. Although the involvement of a second-order intelligence factor g in both numerate-spatial and social intelligences may not be surprising, the hypothesis that there is a direct link between these domain-specific intelligences is less straightforward. To test whether this link holds and if the correlation contributes to model fit, we fit factor models that express the shared variance as between-factor correlations, rather than as resulting from a second-order g factor (as done by Bornstein and Putnick). This approach allows comparison of the fit of unconstrained multidimensional models to constrained ones (i.e., models that constrain some or all between-factor correlations to be 0), thereby allowing a test of the assumption that all domain-specific dimensions share a substantial amount of variance that contributes to model fit.

Methods

Participants

The participants were N = 115 (52 girls, 63 boys) healthy 4-year-olds (M = 47.7 months, SD = 9 days at the assessment of all variables except for ToM; M = 50.0 months, SD = 23 days at the assessment of ToM). Children were recruited from birth records and came from middle-income backgrounds in Germany. Altogether, 25.4% of the mothers had attended secondary school up to grade 10 (not college-bound degree), 22.8% had attended secondary school up to grade 12 (college-bound degree), 50.9% had a bachelor or a master degree, and 0.9% had a PhD. Participation was voluntary. The university ethics committee approved the study.

Procedure

Assessments were carried out in a university laboratory (in German), and tasks were video-recorded. All children were accompanied by a caregiver. Upon arrival, each child received a warm-up to become familiar with the surroundings. Prior to testing, parents signed informed consent forms.

Child measures

Psychometric intelligence

Children's psychometric intelligence was assessed with a standardized measure of intelligence for children from 3 to 6.5 years, the German version of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III; Petermann, 2009). Following Bornstein and Putnick (2019), we chose three verbal subtests (Information, Arithmetic, and Similarities) and two performance subtests (Block design and Picture completion). Due to the small age range, raw sum scores were used for all subtests.

Counting game

The Counting Game (called The Child's Walk; Bornstein & Putnick, 2019) measures children's ability to use the following number concepts: (1) one-to-one correspondence, (2) stable-order principle, (3) cardinality principle, and (4) the ability to integrate number with another dimension (direction). In part A (training), children count 3 and 5 dots on a die; in part B (Game 1), children move a game piece step-by-step across the squares on a path drawn on a game board with a house (the goal) at one end and trees at the other, matching the number rolled on a die (1, 3, or 5); and in part C (Game 2), children simultaneously integrate a second dimension, direction, which is determined by a second die that shows either a house or trees.

Trained coders watched the video record of the task. In Game 1, the children received 1 point for the correct number of squares that they moved the game piece and an additional point for the correct number of jumps (resulting in max. two points per turn) because sometimes children jumped the correct number of times while moving an incorrect number of squares (e.g., jumped twice on the same square or skipped one). An average score was computed by dividing the sum score across all turns by the number of turns. For Game 2, the child received one point for moving the game piece in the correct direction and one point for moving it the correct number of squares (resulting in max. two points per turn). Analogous to Game 1, an average score across all turns was computed.

Reliability of the German coder with one of the authors of the coding system on U.S. videos (n = 18) was ICC = .98 for Game 1 and .99 for Game 2. Reliability of the entire Counting Game was Cronbach's α = .76, and so scores from Games 1 and 2 were standardized (M = 0, SD = 1) and averaged to yield a single Counting Game Scale Score.

ToM: False belief understanding

Children's false belief understanding was measured using three well-established tasks: two *unexpected contents tasks* and one *unexpected location task*. For all three measures, we followed standard procedures as published in the literature. The first *unexpected contents task* (Hofer & Aschersleben, 2005; Perner, Leekam, & Wimmer, 1987) presented children with a Smarties box. The children are asked what they believe is in the box (typically, children answer saying Smarties or chocolate). The experimenter then shows the children that, actually, a little toy piglet is in the box. After putting the piglet back into the box, the experimenter presents the children with a toy figure (Lucas) and explains that Lucas has never looked in the box. The test question asks the children what Lucas believes is in the box, Smarties or chocolate (correct, 1 point) or a piglet (incorrect, 0 points).

The second *unexpected contents* task was presented in a moral context (Killen, Mulvey, Richardson, Jampol, & Woodward, 2011). The children learn that a classroom helper (the accidental transgressor) accidently throws a paper bag found on a table into the trash. The paper bag contains another child's cupcake; the other child is playing outside. To assess participants' understanding of false belief, we asked children about the belief of the accidental transgressor regarding the contents of the bag (Does he believe that the bag contains trash = 1 point; or a cupcake = 0 points).

The *unexpected location* task is assessed within the same cover story (Killen *et al.*, 2011). After being asked about the agent's belief regarding the content of the bag, the experimenter asks the children where the owner of the cupcake will look for the paper bag when coming back to the classroom (Will he look for it on the table = 1 point; or in the trash bin = 0 points).

The three ToM measures vary in difficulty (i.e., proportion of children who solve the task), with the unexpected contents task in the moral context (M = 0.13, SD = 0.34; 0 = incorrect, 1 = correct) being more difficult than the two other false belief tasks (both M = 0.39, SD = 0.49; see also Killen *et al.*, 2011, for a similar finding). This difference in difficulty is likely to arise from the task content: In the unexpected contents (Smarties) task, the experimenter explicitly states what the agent believes to be in the box, whereas in the unexpected contents task in the moral context, children need to infer the agent's belief from their action and from the context.

The bivariate correlation between the unexpected contents (Smarties) task and the unexpected location task was r = .30, p = .002; the correlation between Smarties and unexpected contents tasks in the moral context was r = .33, p = .001. There was no significant correlation between the two false belief tasks in the moral context, r = .03, p = .767. Of all children, 47 (43%) answered all three tasks incorrectly, 35 (32%) gave a correct answer to a single task, 22 (20%) to two, and 6 (5.5%) to all three tasks. A scale analysis of the data (Osterhaus *et al.*, 2019) showed that all three ToM tasks assess the same underlying construct: In particular, they fit a unidimensional first-order ToM scale, with all infit and outfit mean square statistics indicating a good fit (i.e., they were between .8 and 1.2). The scores for the three ToM measures were, therefore, summed to create a single composite ToM score.

Mothers' report

Adaptive behaviour

The German translation of the Vineland Adaptive Behavior Scales – Interview Edition Survey Forms (VABS; Sparrow *et al.*, 2005, 2008) was used to obtain mothers' estimates of children's (1) receptive, expressive, and written communication skills and (2) their social relation skills. The relevant scales of the VABS consist of 54 items (communication scale) and 59 items (social relations scale).

After a debriefing about the test, parents were made familiar with the question and answer format. Items consisted of short descriptions of specific behaviours (e.g., talks in full sentences), and parents were asked to indicate whether their child displays the behaviour *not at all* (score = 0), *sometimes* (score = 1), or *clearly* (score = 2). Some items could only be scored with 0 or 2. When 7 consecutive questions were rated with 0, the interview for the respective scale ended. Reliabilities: Cronbach's α = .81 (communications scale) and α = .76 (social relations scale). Based on the scores from the single items ranging from 0 to 2, sum scores were computed for the subscales. Raw scores were used in all analyses.

Results

Core performance

Descriptive statistics for all measures are reported in Table 1, and correlations are given in Table 2.

The measurement model: The domain-specific component model

We conducted several CFAs in Mplus 7 (Muthén & Muthén, 2012), using a robust maximum likelihood estimator with Huber/Pseudo ML/sandwich corrections (MLR). The

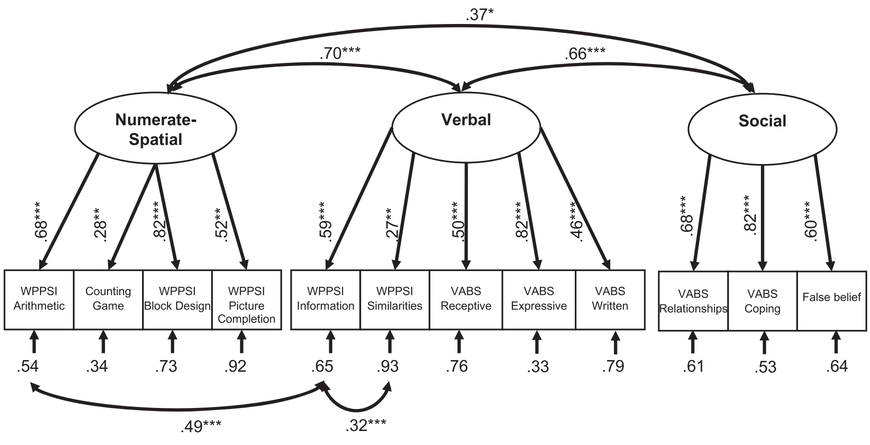
Table 1. Descriptive statistics

	N	*M*	*SD*	Range
WPPSI				
Information	112	16.84	2.92	7–23
Similarities	108	6.70	3.87	0–15
Arithmetic	111	7.96	2.32	1–15
Block design	109	23.71	4.74	12–32
Picture completion	115	14.50	5.26	3–26
VABS (communication)				
Receptive	109	24.78	0.93	22–26
Expressive	109	49.48	4.52	30–60
Written	109	2.25	1.71	0–9
VABS (social relation skills)				
Relationships	108	40.26	3.43	31–50
Coping	108	14.42	4.88	6–29
Counting game (*Z* score)	101	−0.01	0.90	−1.82 to 1.37
Game 1	105	1.39	0.58	0–2
Game 2	101	0.86	0.68	0–2
False belief understanding	126	0.85	0.91	0–3

Table 2. Correlations between all variables in the model

	WPPSI Arithmetic	Counting game	WPPSI Block design	WPPSI Picture completion	WPPSI Information	WPPSI Similarities	VABS Receptive	VABS Expression	VABS Written	VABS Relationships	VABS Coping
Counting game	.560***										
WPPSI Block design	.318**	.424***									
WPPSI Picture completion	.204*	.233*	.230*								
WPPSI Information	.548***	.259*	.159	.165							
WPPSI Similarities	.144	.080	.193*	.126	.400***						
VABS Receptive	.087	.365***	.182	−.027	.164	.139					
VABS Expression	.379***	.399***	.324**	.209*	.456***	.225*	.403***				
VABS Written	.326**	.350***	.324**	.080	.252**	.047	.152	.378***			
VABS Relationships	.014	.195	−.026	−.104	.174	.122	.293**	.335***	.156		
VABS Coping	.134	.264**	−.030	.021	.117	.051	.352***	.326**	.131	.460***	
False belief understanding	.242*	.226*	.090	.049	.317**	.185	.242*	.358***	.108	.281**	.427***

Notes.. ***$p < .001$; **$p < .01$; *$p < .05$.



$\chi^2(49) = 60.53$, $p = .13$, CFI = .96, SRMR = .07, RMSEA = .04 [.00 .07]

Figure 1. Domain-specific component model.

MLR estimator can handle missing data patterns, of which there were 20 in the present study (containing missing values for one or more variables). Missing values were missing at random, as indicated by Little's missing completely at random (MCAR) test: $\chi^2(272) = 253.12, p = .79$.

The primary CFA model that we fitted is the domain-specific component model, favoured based on prior findings (Bornstein & Putnick, 2019): This model includes three separate factors (numerate-spatial, verbal, and social intelligence), and ToM is modelled as an indicator of children's social intelligence. We included correlated error terms for some of the verbal and performance subtests of the WPPSI-III, as was indicated by modification indices (all subsequent CFAs include these correlated errors). No further constraints were imposed, and the factors were allowed to correlate freely. The results revealed a good fit of the model, with $\chi^2(49) = 60.59$, $p = .12$, a comparative fit index [CFI] = .96, Tucker-Lewis Index [TLI] = .94, and root-mean-square error of approximation [RMSEA] = .04 [.00, .07]. All factor loadings (see Figure 1) were significant; the average factor loading was .57. With the exception of three factor loadings (Counting game: .28; WPPSI-III similarities: .27; and VABS communication written: .46), all factor loadings were above .50.

False belief understanding as an indicator of social intelligence

To answer our first question (specificity of false belief understanding as an indicator of social intelligence), we conducted two additional CFAs, in which false belief was modelled as an indicator of verbal intelligence (Comparison Model A) or numerate-spatial intelligence (Comparison Model B). The domain-specific component model fitted the data best, as indicated by chi-square statistics and fit indices (see Table 3). The factor loadings for false belief understanding were .53 ($p < .001$) and .32 ($p = .02$) in Comparison Models A and B, respectively; in the favoured model, the loading of false belief understanding on the factor social intelligence was .60 ($p < .001$), showing that false belief understanding can best be modelled as a specific indicator of children's social intelligence.

Table 3. Model comparisons between models that include false belief understanding (FBU) as an indicator of social, verbal, or numerate-spatial intelligence

Model	χ^2	df	*p*	CFI	TLI	RMSEA	SRMR
Domain-Specific Component Model 3 Factors: FBU as Social Intelligence	60.53	49	.13	.96	.94	.04 [.00 .07]	.07
Comparison Model A 3 Factors: FBU as Verbal Intelligence	62.83	49	.09	.95	.93	.05 [.00 .08]	.07
Comparison Model B 3 Factors: FBU as Numerate-Spatial Intelligence	74.33	49	.01	.91	.87	.07 [.03 .10]	.09

The structure of intelligence: Domain-specific dimensions

To address our second question and to test whether the domain-specific component model fits the data better than a one-factor (monolithic) model, we fitted the one-factor model to our data. As Table 4 indicates, the domain-specific component model fits the data significantly better than the one-factor (monolithic) model (Comparison Model C). In addition, the three-factor model is supported by parallel analysis (Dinno, 2009), which gives a more conservative estimate of the number of factors by correcting for the effects of sampling error on eigenvalues. This analysis revealed three corrected eigenvalues larger than the retention criterion of 0, showing that the domain-specific component model is to be preferred over the monolithic model.

Shared variance and the domain-general dimension

With respect to our third question (relations between dimensions), significant correlations emerged between all factors in the domain-specific component model, suggesting that the three domain-specific components share substantial amounts of variance (see Figure 1). To test the superiority of this model with both domain-general and domain-specific dimensions of intelligence, we conducted several additional CFAs (Comparison Models D to G; see Table 4). In these, some or all factor correlations were constrained to be equal to 0, allowing a test of whether the constraints result in decreased model fit. All comparison models fit significantly worse than the domain-specific component model, suggesting that the domain-specific components share a significant amount of variance that represents the domain-general dimension of children's intelligence.

Discussion

Theory of Mind, and in particular children's understanding of false belief, has been viewed as a 'cognitive watershed' in child development due to their foundational importance for human intelligence. The present study investigated whether false belief understanding can be modelled as a specific indicator of social intelligence in a multidimensional model of intelligence and it re-examines relations to domain-general and domain-specific aspects of children's intelligence.

We found that the three-factor domain-specific component model was clearly superior to a one-factor monolithic model, confirming in an independent sample that by age 4, specific aspects of intelligence can be distinguished (Bornstein & Putnick, 2019). The three false belief measures (assessing conceptual understanding of others' minds) formed together with the level of adaptive functioning (i.e., social skills in the real world)

Table 4. Model comparisons between the domain-specific component model and alternative models of the structure of child intelligence

Model	χ^2	Df	*p*	CFI	TLI	RMSEA	SRMR	$\Delta \chi^{2a}$	*p*
Domain-Specific Component Model 3 Correlated Factors: Numerate-Spatial vs. Verbal vs. Social Intelligence	60.53	49	.13	.96	.94	.04 [.00 .07]	.07		
Comparison Model C (Monolithic Model) 1 Factor: General Intelligence	109.82	52	.00	.79	.73	.09 [.07 .12]	.09	49.96	.00
Comparison Model D 3 Factors, 2 Uncorrelated: Numerate-Spatial and Verbal	99.07	50	.00	.82	.76	.09 [.07 .12]	.14	15.56	.00
Comparison Model E 3 Factors, 2 Uncorrelated: Numerate-Spatial and Social	66.62	50	.06	.94	.92	.05 [.00 .09]	.11	6.09	.01
Comparison Model F 3 Factors, 2 Uncorrelated: Verbal and Social	84.40	50	.00	.87	.83	.08 [.05 .11]	.13	33.30	.00
Comparison Model G 3 Uncorrelated Factors	121.31	52	.00	.74	.67	.11 [.08 .13]	.18	48.96	.00

[a]Satorra–Bentler scaled chi-square difference test.

the domain-specific factor 'social intelligence'. Including false belief as an indicator of children's numerate-spatial or verbal intelligences resulted in reduced model fits. This pattern of results suggests that individual differences in false belief understanding are specific for children's social intelligence. This novel finding supports the idea that is worthwhile to include false belief understanding in models of childhood intelligence.

Children's social intelligence was interrelated with, but separable from, both their verbal and numerate-spatial intelligences. Relations between ToM and verbal abilities have consistently been reported in the literature (Milligan *et al.*, 2007), and therefore, it may not be surprising that the present study revealed a high correlation between children's verbal and social intelligences. It is worth noting, however, that the present study included false belief understanding and mothers' reports about their children's social relationship skills as indicators of social intelligence. The typically high correlation between measures of verbal ability and ToM may, in part, be attributed to the linguistic nature of ToM tasks (such as false belief). This interpretation, however, does not pertain to mothers' reports, thus giving independent evidence that children's verbal and social intelligences are closely linked.

The laboratory-based measures of false belief understanding and mothers' reports about their children's social relationship skills (VABS) loaded on the same factor, social intelligence. This finding accords with results from meta-analyses of the association between ToM and social consequences (Imuta, Henry, Slaughter, Selcuk, & Ruffman, 2016; Slaughter *et al.*, 2015), and work that demonstrates links between ToM and peer preference as well as prosocial tendencies (Eggum *et al.*, 2011; Slaughter *et al.*, 2002). Finding that both false belief understanding and mothers' reports about their children's adaptive social behaviour tap the same underlying construct – despite method variance resulting from the use of report and laboratory-based measures – supports the idea that children's false belief understanding is related to real-world social consequences and behaviours that are noticed and consequently reported by mothers.

The multidimensional model of intelligence that was supported in the present study further corroborates work, indicating that early ToM and social intelligence are related to later school achievement. For example, social–emotional competence is important for early academic progress in kindergarten (Ladd, Birch, & Buhs, 1999), and there is some evidence that preschool ToM abilities predict children's academic achievement (e.g., Lecce, Caputi, & Pagnin, 2014; Lockl, Ebert, & Weinert, 2017).

Our findings with respect to the relations between false belief understanding and numerate-spatial abilities are novel, except for some evidence that suggests that ToM predicts math abilities in elementary school (Kloo, Sodian, Kristen-Antonow, & Osterhaus, 2018; Lockl, Ebert, & Weinert, 2017). Relations between false belief and numerate-spatial skills are theoretically interesting. In particular, understanding representations may be an important developmental precursor of math performance, in accordance with the importance of the use of representations in school mathematics (Vanbinst, Ghesquiere, & De Smedt, 2012).

The present study replicated – in an independent German sample – the general factor structure of domain-specific and domain-general child intelligence proposed by Bornstein and Putnick (2019). The factor structure was very similar to the one obtained in the original study. The sample in the present study is less heterogeneous with respect to mothers' socioeconomic background and level of education than the U.S. sample in Bornstein and Putnick (2019). Considering the restricted variance in key background variables (age and SES), it is worth noting that the correlations that emerged in the present study are very similar to those reported by Bornstein and Putnick for their more diverse

U.S. sample. Our study supports the view that very young children's intelligence comprises both domain-general and domain-specific dimensions and provides initial cross-cultural evidence for the validity of the domain-specific component model. However, more evidence (especially from non-Western samples) is needed to draw conclusions with respect to the universality of the proposed hierarchical model of intelligence.

Acknowledgements

We are grateful to all participants for their friendly collaboration and support of this research. This research was supported by the German Research Council (DFG SO 213/27-1,2); the Intramural Research Program of the NIH/NICHD, USA; and an International Research Fellowship at the Institute for Fiscal Studies (IFS), London, UK, funded by the European Research Council (ERC) under the Horizon 2020 research and innovation programme (grant agreement No 695300-HKADeC-ERC-2015-AdG).

Conflicts of interest

All authors declare no conflict of interest.

Author contribution

Christopher Osterhaus (Conceptualization; Formal analysis; Methodology; Writing – original draft; Writing – review & editing), Diane L. Putnick (Formal analysis; Methodology; Writing – original draft; Writing – review & editing), Susanne Kristen-Antonow (Data curation; Investigation; Project administration; Writing – original draft; Writing – review & editing), Daniela Kloo (Writing – original draft; Writing – review & editing), Marc H. Bornstein (Conceptualization; Writing – original draft; Writing – review & editing), Beate Sodian (Conceptualization; Funding acquisition; Writing – original draft; Writing – review & editing).

Data availability statement

Data are available on request due to privacy/ethical restrictions.

References

Bornstein, M. H., & Putnick, D. L. (2019). *The architecture of the child mind: g, Fs, and the hierarchical model of intelligence*. New York, NY: Routledge. https://doi.org/10.4324/9780429027307

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's Theory of Mind. *Child Development*, *72*, 1032–1053. https://doi.org/10.1111/1467-8624.00333

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, UK: Cambridge University Press.

Dinno, A. (2009). Exploring the sensitivity of Horn's parallel analysis to the distributional form of random data. *Multivariate Behavioral Research*, *44*, 362–388. https://doi.org/10.1080/00273170902938969

Eggum, N. D., Eisenberg, N., Kao, K., Spinrad, T. L., Bolnick, R., Hofer, C., . . . Fabricius, W. V. (2011). Emotion understanding, theory of mind, and prosocial orientation: Relations over time in early childhood. *The Journal of Positive Psychology*, *6*, 4–16. https://doi.org/10.1080/17439760.2010.536776

Etel, E., & Slaughter, V. (2019). Theory of mind and peer cooperation in two play contexts. *Journal of Applied Developmental Psychology*, *60*, 87–95. https://doi.org/10.1016/j.appdev.2018.11.004

Fink, E., Begeer, S., Peterson, C. C., Slaughter, V., & de Rosnay, M. (2015). Friendlessness and theory of mind: A prospective longitudinal study. *British Journal of Developmental Psychology*, *33*, 1–17. https://doi.org/10.1111/bjdp.12060

Flavell, J. H. (2004). Theory-of-mind development: Retrospect and prospect. *Merrill-Palmer Quarterly*, *50*, 274–290. https://doi.org/10.1353/mpq.2004.0018

Furman, W., & Bierman, K. L. (1983). Developmental changes in young children's conceptions of friendship. *Child Development*, *54*, 549–556. https://doi.org/10.2307/1129129

Gurucharri, C., & Selman, R. L. (1982). The development of interpersonal understanding during childhood, preadolescence, and adolescence: A longitudinal follow-up study. *Child Development*, *53*, 924–927. https://doi.org/10.2307/1129129

Hofer, T., & Aschersleben, G. (2005). *„Theory of Mind"-Skala für 3- bis 5-jährige Kinder* (2nd edition) [Theory of Mind Scale for 3- to 5-year-olds, 2nd edition]. Munich, Germany: Max-Planck-Institut für Kognitions- und Neurowissenschaften.

Imuta, K., Henry, J. D., Slaughter, V., Selcuk, B., & Ruffman, T. (2016). Theory of mind and prosocial behavior in childhood: A meta-analytic review. *Developmental Psychology*, *52*, 1192–1205. https://doi.org/10.1037/dev0000140

Killen, M., Mulvey, K. L., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition*, *119*, 197–215. https://doi.org/10.1016/j.cognition.2011.01.006

Kloo, D., Kristen-Antonow, S., & Sodian, B. (2019). Progressing from an implicit to an explicit false belief understanding: A matter of executive control? *International Journal of Behavioral Development*, *44*, 107–115. https://doi.org/10.1177/0165025419850901

Kloo, D., Sodian, B., Kristen-Antonow, S., & Osterhaus, C. (2018). *Theory of Mind, working memory, and math performance: A longitudinal study*. X. Dubrovnik Conference on Cognitive Science, Dubrovnik, Croatia.

Ladd, G., Birch, S., & Buhs, E. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence? *Child Development*, *70*, 1373–1400. https://doi.org/10.1111/1467-8624.00101

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. https://doi.org/10.1080/17405629.2013.821944

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. https://doi.org/10.1016/j.lindif.2016.11.007

Marcone, R., Caputo, A., & della Monica, C. (2015). Friendship competence in kindergarten and primary school children. *European Journal of Developmental Psychology*, *12*, 412–428. https://doi.org/10.1080/17405629.2015.1031215

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. https://doi.org/10.1111/j.1467-8624.2007.01018.x

Muthén, L., & Muthén, B. (2012). *Mplus user's guide. Version 7*. Los Angeles, CA: Muthén & Muthén.

Osterhaus, C., Kloo, D., Kristen-Antonow, S., & Sodian, B. (2019). Scaling of (advanced) ToM and modeling children's ToM competencies: Longitudinal findings in 4- to 8-year-olds. In C. Osterhaus (Chair), *Social cognition beyond preschool: Advances and issues in measuring advanced theory of mind. Symposium to be conducted at the 2019*. Biennial Meeting of the Society for Research in Child Development, Baltimore, MD.

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. https://doi.org/10.1111/j.2044-835X.1987.tb01048.x

Petermann, F. (2009). *Wechsler Preschool and Primary Scale of Intelligence–III (WPPSI-III, German version)*. Frankfurt a. M., Germany: Pearson.

Peterson, C. C., Slaughter, V., Moore, C., & Wellman, H. M. (2016). Peer social skills and theory of mind in children with autism, deafness, or typical development. *Developmental Psychology*, *52*, 46–57. https://doi.org/10.1037/a0039833

Peterson, C. C., Slaughter, V. P., & Paynter, J. (2007). Social maturity and theory of mind in typically developing children and those on the autism spectrum. *Journal of Child Psychology and Psychiatry*, *48*, 1243–1250. https://doi.org/10.1111/j.1469-7610.2007.01810.x

Slaughter, V., Dennis, M. J., & Pritchard, M. (2002). Theory of mind and peer acceptance in preschool children. *British Journal of Developmental Psychology*, *20*, 545–564. https://doi.org/10.1348/026151002760390945

Slaughter, V., Imuta, K., Peterson, C. C., & Henry, J. D. (2015). Meta-analysis of theory of mind and peer popularity in the preschool and early school years. *Child Development*, *86*, 1159–1174. https://doi.org/10.1111/cdev.12372

Sodian, B. (2005). Theory of mind—The case for conceptual development. In W. Schneider, R. Schumann-Hengsteler & B. Sodian (Eds.), *Young children's cognitive development. Interrelationships among working memory, theory of mind, and executive functions* (pp. 95–131). Hillsdale, NJ: Erlbaum.

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland Adaptive Behavior Scales: Second edition (Vineland II), survey interview form/ caregiver rating form*. Livonia, MN: Pearson Assessments.

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2008). *Vineland Adaptive Behavior Scales: Second edition (Vineland II), The expanded interview form*. Livonia, MN: Pearson Assessments.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology*, *15*, 201–292. https://doi.org/10.2307/1412107

Spearman, C. (1927). *The abilities of man*. New York, NY: Macmillan.

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago, IL: University Press.

Thurstone, L. L., & Thurstone, T. G. (1941). Factorial studies of intelligence. *Psychometric Monographs*, *2*, 94.

Vanbinst, K., Ghesquiere, P., & De Smedt, B. (2012). Numerical magnitude representations and individual differences in children's arithmetic strategy use. *Mind, Brain, and Education*, *6*, 129–136. https://doi.org/10.1111/j.1751-228X.2012.01148.x

Wellman, H. M. (2002). Understanding the psychological world: Developing a theory of mind. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 167–187). Oxford, UK: Blackwell.

Wellman, H. M. (2017). The development of theory of mind: Historical reflections. *Child Development Perspectives*, *11*, 207–214. https://doi.org/10.1111/cdep.12236

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*, 655–684. https://doi.org/10.1111/1467-8624.00304

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*, 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Received 18 December 2019; revised version received 16 March 2020