

Supplementary materials

Appendix 1 - The details of the matching procedures

The analyses presented in the current paper are based on the biggest possible subsamples from the group of Polish-English bilinguals and Polish monolinguals (see: Haman et al., 2017). To ensure that the compared groups were internally homogeneous and similar in terms of the controlled variables, we carefully selected participants for the current analysis from the entire tested sample. In the first step, we excluded children who did not match the profile of a typically developing bilingual or monolingual: children who had hearing problems (six bilinguals, 3.5% of the bilingual sample; nine monolinguals, 2.9% of the monolingual sample). Additionally, from the bilingual group we excluded children who were effectively trilingual (15 children, 8.7% of the bilingual sample; see also Mieszkowska et al., 2017); from the monolingual group we excluded those who were found to be bilingual (three children, 1% of the monolingual sample, were living in Poland). We also excluded children at risk of developmental language disorder, as indicated by parental concerns reported in the questionnaires (four bilinguals, 2.3% of the bilingual sample; three monolinguals, 1% of the monolingual sample).

In the next step, we considered only those bilingual children who had undertaken the TRT in Polish (95 bilinguals), i.e., those whose parents indicated Polish as the child's dominant language in the consent form. From this group, we excluded children with missing data in control variables: four who were not tested for non-verbal intelligence; 15 whose parents did not complete a background questionnaire, which is necessary to control for SES and language exposure; and 15 who did not perform word comprehension tests in L1 and/or L2. From the group of Polish monolinguals ($N=268$), we also excluded all children whose parents did not complete the background questionnaire ($n=129$). Since we aimed to examine ToM abilities in the dominant Polish language and compare ToM results for Polish monolinguals and Polish-English bilingual children, we decided to include only those whose Polish language abilities were in the range of typical development. Thus, we excluded another five bilinguals who scored below the fifth percentile in a Polish word comprehension test. This left us with a subsample of 56 Polish-English bilinguals. In the last step, we excluded five bilingual children from mixed families, i.e., where one of the parents was not Polish¹; this

¹ Because our entire sample included only five simultaneous bilinguals (children who were exposed to two languages at home from birth (Polish mother and English father)), we excluded them from further analyses in order to ensure group homogeneity.

left us with a subsample of 51 sequential bilinguals, all typically developing, dominant in Polish, and growing up in families with two Polish parents living in the UK.

We subsequently used an algorithm (a program written in Python) to match each of the 51 bilinguals with a similar Polish monolingual child. The algorithm treated the bilingual group as a reference group and searched through the group of 138 monolinguals to find the best match for a particular bilingual child. Children were matched one-to-one on the following variables: age, gender, SES (years of mother's education), non-verbal intelligence (Raven's score), and L1 proficiency (percentile score). Each of the variables was given an equal importance (weight). First, the variables are standardized within each group, and then for each of the participants in the reference group (here: bilinguals). Then, the algorithm chooses a random participant from the monolingual group and calculates the overall distance (discrepancy) between the two groups. The algorithm performs 10,000 such cycles (random pairings for all participants) and eventually chooses the matching that has the lowest overall distance. The matching procedure served to ensure that any differences between the groups could be attributed to language status (bilingual or monolingual) but not so much to other factors, i.e., age, gender, SES, children's language or cognitive abilities in general (see Kail, 2000), as we know that these factors affect performance in ToM tasks (e.g., Ebert et al., 2017; Farrar & Maag, 2002; Jenkins & Astington, 1996).

Appendix 2 – list of all tasks used in the project

Tasks used in the BI-SLI-PL project, part of COST Action IS1804 (LITMUS battery of tasks). Tasks that were included in the current analyses are **bolded** and marked with an asterisk (*).

Language development	Receptive vocabulary*	EN: BPVS III Dunn, Dunn et al. 2009)* PL: OTSR (Haman & Fronczyk, 2012)*
	Productive vocabulary	EN: EVT-2 (Williams, 2007) PL: Zadanie Nazywania Obrazków (Haman & Smoczyńska, 2010)
	Syntax – comprehension	EN: TROG (Test of Reception of Grammar) (Bishop, 2003) PL: TROG (trans. Smoczyńska, 2005)
	Syntax – production	EN: SRT (Sentence Repetition Task) (Marinis, Chiat, Armon-Lotem, 2010) PL: SRT (Banasik, Haman, Smoczyńska, 2011)
	Phonology	EN: NWR (Non Word Repetition) (Szewczyk, Wodniecka, 2012) PL: NWR (Szewczyk et al., 2012)
	Narratives	EN: MAIN (Multilingual Assessment Instrument for Narratives) (Gagarina et al., 2012) PL: MAIN (Kiebzak-Mandera, et al., 2012)
Cognitive development	Non-verbal IQ*	PL or EN: Raven Scale (Jaworowska & Szustrowa, 2003)*
	Theory of Mind*	PL or EN: TRT (Test of Reflection on Thinking) (Bialecka-Pikul et al., 2018)*
	Executive functions	PL or EN: Airplanes (Senderecka et al., 2012), Simon says (Stromment, 1973), ANT (Attentional Network Test) (Ruedia et al., 2004), Corsi blocks (Kessels et al., 2000), Digits (Wechsler, 1997).

Table S1. A sample story from the Test of Reflection on Thinking (TRT, Bialecka-Pikul et al., 2018), with the accompanying pictures:

The unexpected transfer test (version for boys)



Once Zach and Evan were sitting on the sofa reading a book.



When they got bored, they put the book into a large cupboard and went outside.



Zach came back first; he took the book out of the big cupboard and started reading.



Then he put the book into the little cupboard and went to the kitchen.



Then Evan came back to the room.



1



2



3



4



5

14

Questions asked after the above story is presented:

1. *Where will Evan think the book is?* The experimenter waits ca. 3 seconds for the child to point to a place in the pictures or to give an answer.
2. If the child answers correctly to 1.: *Ok, so where will Evan look for the book?*
If the child answers incorrectly to 1.: *Where will Evan look for the book?*
3. *Why will Evan look there?*

If in the previous questions the child did not say where the book was, ask the control question: *OK, and where is the book now?*

Table S2. *Tasks that make up the Reflection on Thinking Test: a short description of the storyline, the testing questions, and the control questions.*

Name of the task	Description of the storyline	Question relating to accuracy index	Question relating to justification index	Control questions
Understanding of unexpected identity (Appearance-reality task) ^a	Zach shows Evan a sponge which, from a distance, looks like a stone. Zach asks Evan what the object is.	1. What will Evan say it is if he looks at it from a distance?	2. Why will Evan say that?	---
	Zach gives Evan the sponge to hold and asks him again what the object is.	3. What will Evan then say it is?	4. Why will Evan say that? (not evaluated)	---
Understanding of first-order false beliefs (Unexpected change task)	Zach and Evan put a book in place A. When Evan is gone, Zach moves the book to place B. The child is asked about Evan's thoughts/behavior with regard to the book when he is back.	1. Where will Evan think the book is? Where will Evan look for the book?	2. Why will Evan look there?	3. Ok, and where is the book now?
Understanding of first-order false beliefs (Unexpected content task)	Evan puts some candies into a chocolate box. When Zach enters the room, he sees the chocolate box on the table. Evan is asked what Zach will think is in the box.	1. What does Zach think is in the box?	2. Why does Zach think so?	3. What is really in the box? 4. What does Evan think is in the box?
	Then Zach asks Evan to share the chocolates with him.	5. What will Evan answer?	6. Why will Evan say that?	---
Understanding of interpretation	Evan draws a flower with triangular petals. It is partly obscured, therefore the whole picture cannot be seen. Evan asks Zach what the object depicts.	1. What will Zach answer?	2. Why will Zach say that?	---
Understanding of deception	Evan hides some color pencils from Zach. When Zach enters the room and asks about the pencils, Evan decides not to tell him the truth.	1. Where will Evan tell Zach the pencils are?	2. Why will Evan say that?	3. Ok, and where are the pencils?
Understanding of ambiguity	Zach draws a duck. When he shows the picture to Evan, Evan says it is a rabbit.		1. Why did Evan say that?	---
Understanding of emotions resulting from a false belief	Zach sees white flakes falling from the sky, so he gets his sledge and goes outside. Then, he notices Evan shaking a big feather pillow above him.	1. How did Zach feel? Was he sad, surprised, or happy?	2. Why did Zach feel that way?	---
Understanding of the seeing-knowing relationship and the	Evan is drawing with a crayon, then he puts the crayon in a box and leaves the room. Zach takes the crayon and puts it in a basket, but he	1. Where does Zach think Evan will look for the	2. Why does Zach think that?	3. Did Evan see what Zach did

second-order beliefs that result from it	does not know that Evan is watching him. After a while, Evan comes back into the room.	crayon when he comes back to the room?	with the crayon?
Understanding of second-order beliefs	In the park, Evan and Zach meet a man with balloons who says he is going to place A to sell the balloons. Evan goes home to get some money for the balloons; while he is gone, Zach finds out that the man is going to place B instead. Zach also goes home to get some money. On his way to place B, the man with the balloons meets Evan and tells him he is going to sell the balloons at place B.	1. Where does Zach think Evan will look for the man with the balloons?	2. Why does Zach think so? 3. Did Evan meet and talk to the balloon man on his way home?

Note: ^a Training task, not evaluated.

Table S3. Hierarchical linear regression for the monolingual group.

	ToM: TRT overall accuracy index			ToM: TRT overall justification index		
	<i>Adj. R</i> ²	<i>B(SE)</i>	β	<i>Adj. R</i> ²	<i>B(SE)</i>	β
Step 1	.43			.23		
Gender		0.67 (.41)	.17		0.59 (0.77)	.10
Age		0.07 (.03)	.34*		0.04 (0.05)	.13
Non-verbal IQ		0.17 (.05)	.41**		0.26 (0.10)	.41**
SES		0.05 (.07)	.07		0.19 (0.13)	.18
Step 2	.49			.30		
Gender		0.50 (.40)	.13		0.30 (0.74)	.05
Age		0.09 (.03)	.45**		0.08 (0.05)	.25
			*			
Non-verbal IQ		0.11 (.05)	.28*		0.17 (0.10)	.27
SES		0.00 (.07)	.00		0.10 (0.13)	.10
L1 word comprehension index		0.03 (.01)	.29*		0.04 (0.02)	.32*

Table S4. *Logistic regressions for the monolingual group.*

	ToM: TRT first-order false beliefs index			ToM: TRT second-order false belief index		
	<i>AIC</i>	<i>B (SE)</i>	<i>Odds Ratio</i>	<i>AIC</i>	<i>B (SE)</i>	<i>Odds Ratio</i>
Model 1: sociodemographic and cognitive variables	39.62			51.49		
Intercept		-7.51 (4.62)			-9.23 (4.32)	
Gender		0.33 (0.99)	1.40		1.71 (0.84)	5.55*
Age		0.06 (0.06)	1.06		0.09 (0.06)	1.09
Non-verbal IQ		0.27 (0.13)	1.31*		0.26 (0.12)	1.29*
SES		0.00 (0.19)	1.00		-0.14 (0.14)	0.87
Model 2: L1 word comprehension	25.02			53.11		
Intercept		-17.28 (9.07)			-9.84 (4.55)	
Gender		-0.35 (1.53)	0.70		1.65 (0.85)	5.21*
Age		0.30 (0.17)	1.35		0.11 (0.06)	1.11
Non-verbal IQ		0.67 (0.50)	1.95		0.24 (0.12)	1.27*
SES		-1.28 (1.06)	0.28		-0.17 (0.15)	0.85
L1 word comprehension index		0.31 (0.19)	1.36		0.01 (0.02)	1.01

Table S5. Hierarchical linear regression for the bilingual group: sociodemographic variables and receptive vocabulary.

	ToM: TRT overall accuracy index			ToM: TRT overall justification index		
	<i>Adj. R²</i>	<i>B(SE)</i>	<i>β</i>	<i>Adj. R²</i>	<i>B(SE)</i>	<i>B</i>
Step 1	.26			.41		
Gender		0.77 (.52)	.19		1.26 (.74)	.19
Age		0.09 (.03)	.38**		0.13 (.04)	.36**
Non-verbal IQ		0.10 (.05)	.26		0.25 (.07)	.42**
SES		0.03 (.08)	.04		0.16 (.12)	.15
Step 2	.34			.48		
Gender		0.76 (.49)	.18		1.24 (.69)	.19
Age		0.11 (.03)	.47**		0.16 (.04)	.45**
Non-verbal IQ		0.03 (.05)	.09		0.16 (.08)	.26*
SES		0.06 (.08)	.09		0.20 (.11)	.19
L1 word comprehension index		0.03 (.01)	.33*		0.05 (.02)	.32**
Step 3	.36			.54		
Gender		0.77 (.49)	.19		1.28 (.66)	.19
Age		0.12 (.03)	.55** *		0.20 (.04)	.56** *
Non-verbal IQ		0.03 (.05)	.07		0.14 (.07)	.23
SES		0.06 (.08)	.09		0.20 (.11)	.20
L1 word comprehension index		0.03 (.01)	.34*		0.05 (.02)	.33**
L2 word comprehension index		0.02 (.01)	.17		0.04 (.02)	.26*

Table S6. Hierarchical linear regression for the bilingual group: sociodemographic variables and length of English exposure.

	ToM: TRT overall accuracy index			ToM: TRT overall justification index		
	<i>Adj. R²</i>	<i>B(SE)</i>	β	<i>Adj. R²</i>	<i>B(SE)</i>	β
Step 1	.26			.41		
Gender		0.77 (.52)	.19		1.26 (.74)	.19
Age		0.09 (.03)	.38**		0.13 (.04)	.36**
Non-verbal IQ		0.10 (.05)	.26		0.25 (.07)	.42**
SES		0.03 (.08)	.04		0.16 (.12)	.15
Step 2	.27			.41		
Gender		0.84 (.52)	.20		1.34 (.75)	.20
Age		0.07 (.03)	.32*		0.12 (.05)	.32*
Non-verbal IQ		0.10 (.05)	.27		0.26 (.07)	.43**
SES		0.00 (.09)	.00		0.12 (.13)	.12
L2 length of time of exposure		0.02 (.02)	.15		0.02 (.02)	.11
Step 3	.36			.49		
Gender		0.83 (.49)	.20		1.33 (.70)	.20
Age		0.09 (.03)	.41**		0.14 (.05)	.40**
Non-verbal IQ		0.04 (.05)	.10		0.16 (.08)	.27*
SES		0.03 (.08)	.04		0.16 (.12)	.16
L2 length of time of exposure		0.02 (.01)	.17		0.02 (.02)	.13
L1 comprehension		0.03 (.01)	.34*		0.05 (.02)	.32**

Table S7. Hierarchical linear regression for the bilingual group: sociodemographic variables and language cumulative exposure.

	ToM: TRT overall accuracy index			ToM: TRT overall justification index		
	<i>Adj. R²</i>	<i>B(SE)</i>	β	<i>Adj. R²</i>	<i>B(SE)</i>	β
Step 1	.26			.41		
Gender		0.77 (.52)	.19		1.26 (.74)	.19
Age		0.09 (.03)	.38**		0.13 (.04)	.36**
Non-verbal IQ		0.10 (.05)	.26		0.25 (.07)	.42**
SES		0.03 (.08)	.04		0.16 (.12)	.15
Step 2	.27			.41		
Gender		0.78 (.52)	.19		1.28 (.74)	.19
Age		0.11 (.04)	.47**		0.16 (.05)	.44**
Non-verbal IQ		0.11 (.05)	.28*		0.26 (.07)	.44**
SES		0.01 (.09)	.01		0.13 (.12)	.12
Cumulative exposure to L1		-0.01 (.00)	-.18		-.001 (.01)	-.15
Step 3	.26			.41		
Gender		0.80 (.52)	.19		1.32 (.74)	.20
Age		0.10 (.04)	.44*		0.14 (.06)	.38*
Non-verbal IQ		0.11 (.05)	.29*		0.27 (.07)	.45**
SES		0.00 (.09)	.00		0.11 (.12)	.10
Cumulative exposure to L1		-0.01 (.01)	-.16		-0.01 (.01)	-.10
Cumulative exposure to L2		0.00 (.00)	.07		0.01 (.01)	.13
Step 4	.36			.51		
Gender		0.80 (.48)	.19		1.32 (.67)	.20
Age		0.13 (.04)	.56**		0.18 (.05)	.49**
Non-verbal IQ		0.04 (.05)	.11		0.17 (.07)	.28*
SES		0.02 (.08)	.03		0.14 (.11)	.14
Cumulative exposure to L1		-0.01 (.00)	-.21		-0.01 (.01)	-.15
Cumulative exposure to L2		0.00 (.00)	.10		0.01 (.01)	.16

L1 comprehension

0.03 (.01) .37**

0.05 (.02) .35**

Table S8. *Logistic Regressions for the Bilingual Group.*

	ToM: TRT first- order false beliefs index			ToM: TRT second-order false beliefs index		
	<i>AIC</i>	<i>B (SE)</i>	<i>Odds Ratio</i>	<i>AIC</i>	<i>B (SE)</i>	<i>Odds Ratio</i>
Model 1:						
sociodemographic and cognitive variables	39.58			72.8		
Intercept		-1.83 (5.75)			-8.27 (3.75)	
Gender		2.70 (1.26)	14.94*		0.90 (0.66)	2.46
Age		-0.01 (0.06)	0.99		0.06 (0.04)	1.06
Non-verbal IQ		0.18 (0.12)	1.20		0.05 (0.06)	1.05
SES		-0.04 (0.18)	0.96		0.21 (0.11)	1.23
Model 2: L1 word comprehension	41.18			67.89		
Intercept		-2.91 (5.93)			-11.96 (4.57)	
Gender		2.72 (1.28)	15.25*		1.09 (0.73)	2.96
Age		0.01 (0.07)	1.01		0.10 (0.05)	1.10*
Non-verbal IQ		0.15 (0.13)	1.16		-0.04 (0.08)	0.96
SES		-0.03 (0.18)	0.97		0.28 (0.13)	1.32*
L1 word comprehension		0.02 (0.30)	1.02		0.05 (0.02)	1.05*
Model 3: L1 and L2 word comprehension	33.60			69.36		
Intercept		-25.24 (13.81)			-13.15 (4.96)	
Gender		5.87 (2.80)	355.61*		1.08 (0.73)	2.95
Age		0.22 (0.14)	1.25		0.11 (0.05)	1.12*
Non-verbal IQ		0.06 (0.15)	1.06		-0.05 (0.08)	0.95
SES		0.30 (0.26)	1.34		0.28 (0.13)	1.32*
L1 word comprehension		0.03 (0.05)	1.03		0.05 (0.02)	1.05*

L2 word comprehension		0.21 (0.10)	1.23*	0.01 (0.02)	1.01
Model 4: length or time of L2 exposure	41.26			70.39	
Intercept		-1.51 (5.74)	0.22	-8.05 (3.79)	0.00
Gender		2.77 (1.28)	16.01	1.07 (0.69)	2.92
Age		-0.02 (0.07)	0.98	0.03 (0.04)	1.03
Non-verbal IQ		0.19 (0.12)	1.20	0.06 (0.07)	1.06
SES		-0.07 (0.18)	0.94	0.15 (0.12)	1.16
L2 length of time of exposure		0.02 (0.03)	1.02	0.04 (0.02)	1.04*
Model 5: cumulative language exposure to L1	40.92			71.66	
Intercept		-1.35 (5.74)		-8.24 (3.86)	
Gender		2.82 (1.30)	16.81*	1.02 (0.69)	2.78
Age		0.02 (0.08)	1.02	0.11 (0.06)	1.11
Non-verbal IQ		0.21 (0.13)	1.23	0.06 (0.07)	1.06
SES		-0.06 (0.18)	0.94	0.18 (0.12)	1.19
L1 cumulative language exposure		-0.01 (0.01)	0.99	-0.01 (0.01)	0.99
Model 6: cumulative language exposure to L1 and L2	42.78			73.38	
Intercept		-1.10 (5.83)		-8.15 (3.83)	
Gender		2.85 (1.31)	17.27*	1.03 (0.69)	2.80
Age		0.01 (0.09)	1.01	0.10 (0.06)	1.10
Non-verbal IQ		0.21 (0.13)	1.23	0.06 (0.07)	1.06
SES		-0.07 (0.18)	0.93	0.17 (0.12)	1.18
L1 cumulative language exposure		-0.01 (0.01)	0.99	-0.01 (0.01)	0.99
L2 cumulative language exposure		0.00 (0.01)	1.00	0.00 (0.01)	1.00