

Supplementary material

Appendix 1 Pilot study with NT adults

Prior to the assessment of neurotypical (NT) and autistic children, a pilot study with 40 NT adults was conducted to verify that NT adults performed at ceiling and that test items worked as expected.

Participants

Forty neurotypical adult speakers of English ($N = 10$), French ($N = 10$), Italian ($N = 10$), and German ($N = 10$) ranging in age from 20 to 35 years, participated in the pilot study. Twenty-six identified as female, fourteen as male. All adult participants reported no history of language or cognitive delays or impairments and no history of a diagnosis of ASD or Asperger (three participants had to be excluded beforehand because of a history of language or cognitive delays). The language of testing was the language in which the participant was most fluent in. Adult participants were recruited through flyers, advertisement emails, and Facebook at the University of Fribourg, Switzerland, through flyers at the University of Geneva, Switzerland, and through the participant recruitment platform “SONA” at the University of Konstanz, Germany. All participants gave written informed consent and received monetary compensation for their participation (50CHF/50€ for the first testing point in person, 10CHF/10€ for the second testing point online).

Table 1. Neurotypical adults

	Neurotypical adults ($N = 40$)
Age	
Mean in months (SD)	300 (51.6)
Mean in years; months (range)	25;0 (20;4 – 35;9)
Gender	
Diverse (N, in %)	0 (0.0%)
Female (N, in %)	26 (65.0%)
Male (N, in %)	14 (35.0%)
Own/parental educational level	
Mean (SD; range)	4.25 (1.06; 1 – 5)
Non-verbal IQ (RPM, z-score)	
Mean (SD; range)	107 (10.4; 85 – 134)
Receptive vocabulary (PPVT, %-score (adults), z-score (children))	
Mean (SD; range)	0.92(0.06; 0.9 – 0.99)

Testing procedure

The adult participants were invited between January and February 2023 via email to first complete two online questionnaires: one background questionnaire on their personal history and the self-report form of the “Quantifying Bilingual Experience” (Q-BEx) questionnaire (De Cat et al., 2022) on their language experiences. After completion, the participants were invited once (for 160 minutes) or twice (for 80 minutes each time) to the ABCCD lab at the University of Fribourg, Switzerland, or the Psycholinguistics lab at the University of Konstanz, Germany.

The ABCCD ToM measure was administered through an application on iPads (version: 10.2" 9th 256Go). All participants completed Block 1 (Diverse Desires), Block 2 (First-order False Beliefs), and Block 3 (Second-order False Beliefs) with short breaks in between the blocks. The measure lasted about 30 minutes, including breaks. It was followed by other tasks,

including a short form of the Raven’s Progressive Matrices (RPM; Raven et al., 2018) to control for nonverbal cognitive abilities, as well as the Peabody Picture Vocabulary Test (PPVT) in the respective languages (French: Dunn et al., 1993, Italian: 2016; English: Dunn and Dunn, 2007; German: Lenhard et al., 2015) to control for receptive vocabulary. The PPVT was conducted in the language in which the ToM measure was administered. The adult participants were invited again 14 months later, in March 2024, via email to participate in a retest of the ABCCD ToM measure, using the online platform “Gorilla Experiment Builder” (Anwyl-Irvine et al., 2020) at home on their computer or laptop at a convenient time. This retest session, which 14 adults completed, lasted approximately 30 minutes.

Analysis

In the NT adult group, performance was close to ceiling, and, as shown in Table 2, with most test items receiving correct responses in above 95% of the cases. This was expected since the task was meant to be easy for NT adults.

Table 2. Percentages of correct responses to individual items in Blocks 1, 2, and 3 in NT adults

Block 1 Diverse Desires N = 40		Block 2 First-order False Beliefs N = 40		Block 3 Second-order False Beliefs N = 40	
Test item 1	100.0%	Test item 1	97.5%	Test item 1	97.5%
Test item 2	100.0%	Test item 2	90.0%	Test item 2	95.0%
Test item 3	95.0%	Test item 3	97.5%	Test item 3	97.5%
Test item 4	100.0%	Test item 4	95.0%	Test item 4	92.5%

Test-retest analyses revealed that all participants ($N = 14$) in a retest of Blocks 2 and 3 one year later scored correctly on all items. Because data points do not present variation, test-retest correlations of individual items could not be calculated.

In the adult participant group, results confirmed our hypothesis, with performance reaching ceiling levels at both test and retest phases, suggesting that the measure is easily understood and completed by NT adults.

Appendix 2 Explanation on ToM toolkit by Marinis and colleagues (2023)

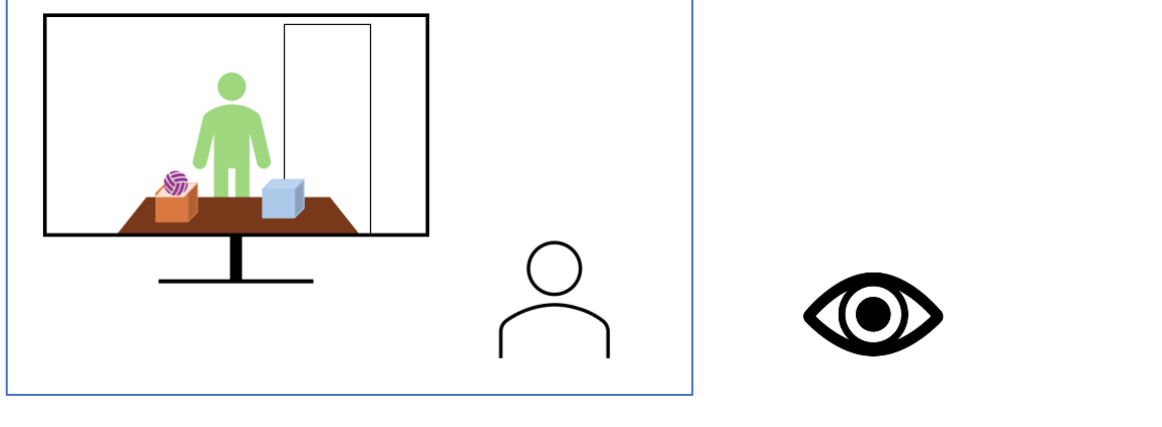
Description:

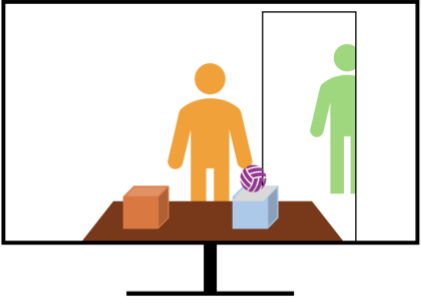

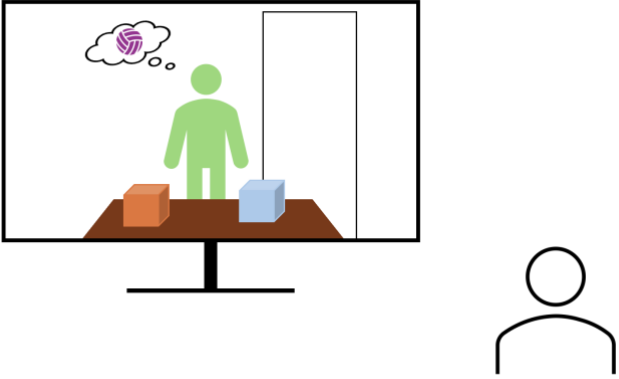

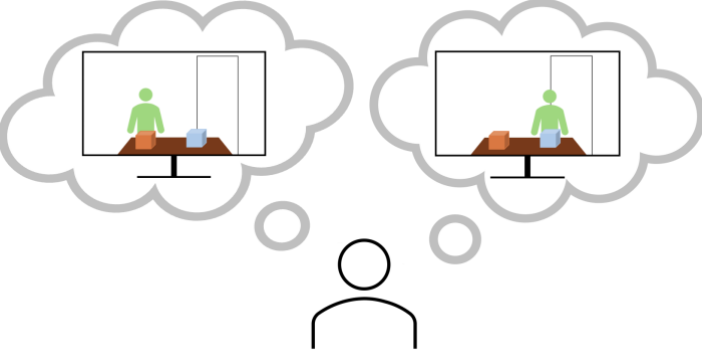
The toolkit includes second-order False Belief tasks, exemplified by the following scenario: An observer (O) watches a television showing a man (M) standing behind a table with two boxes. The man (M) sees that there is a ball in the orange box. Meanwhile, the observer (O) steps away and does not see the subsequent events: a woman (W) arrives, and changes the location of the ball from the orange to the blue box, while being observed by the returning man (M). However, the woman (W) does not know that the man has observed the change. Both the man (M) and the woman (W) leave before the observer (O) resumes watching. Then the man (M) comes back and searches for the ball.

The test question is: What does the observer think: Where will the man (M) search for the teddy?

Thus, this second-order false belief task does assess whether the participants are able to take the observer's perspective about the man's perspective (**BoB_M**).

Table 3

Time	Illustration	Explanation of perspectives
T ₀		<p>Reality/participant's/observer's/man's perspective/observer's perspective about the man's perspective: The ball is in the orange box. → B_P = B_O = B_M = BoB_M</p>

T ₁			<p>Reality/participant's/man's/woman's perspective: The ball is in the blue box. → $B_P = B_W = B_M$</p> <p>Observer's perspective/observer's perspective about the man's perspective: The ball is in the orange box. → $B_o = B_o(B_M)$</p>
T ₂			
T ₃			<p>The expected second-order reasoning answer (the observer's perspective about the man's perspective) here is: "orange"</p> <p>However, this same answer can also be drawn based on first-order reasoning, namely by taking the observer's perspective which would also lead to the answer "orange"</p>

Appendix 3 Mean performances of NT and autistic children in individual items

NT children

Table 4. Percentages of correct responses to individual items in Blocks 1, 2, and 3 in NT children between 4;0-4;11 years

Block 1 Diverse Desires <i>N = 22</i>		Block 2 First-order False Beliefs <i>N = 20</i>		Block 3 Second-order False Beliefs <i>N = 3</i>	
Test item 1	95.5%	Test item 1	15.0%	Test item 1	0%
Test item 2	77.3%	Test item 2	20.0%	Test item 2	66.7%
Test item 3	90.9%	Test item 3	25.0%	Test item 3	33.3%
Test item 4	86.4%	Test item 4	35.0%	Test item 4	33.3%

Table 5. Percentages of correct responses to individual items in Blocks 1, 2, and 3 in NT children between 5;0-5;11 years

Block 1 Diverse Desires <i>N = 32</i>		Block 2 First-order False Beliefs <i>N = 28</i>		Block 3 Second-order False Beliefs <i>N = 15</i>	
Test item 1	90.6%	Test item 1	25.0%	Test item 1	33.3%
Test item 2	87.5%	Test item 2	28.6%	Test item 2	13.3%
Test item 3	87.5%	Test item 3	42.9%	Test item 3	20.0%
Test item 4	84.4%	Test item 4	53.6%	Test item 4	13.3%

Table 6. Percentages of correct responses to individual items in Blocks 1, 2, and 3 in NT children between 6;0-6;11 years

Block 1 Diverse Desires <i>N = 22</i>		Block 2 First-order False Beliefs <i>N = 22</i>		Block 3 Second-order False Beliefs <i>N = 12</i>	
Test item 1	95.5%	Test item 1	40.9%	Test item 1	58.3%
Test item 2	95.5%	Test item 2	40.9%	Test item 2	41.7%
Test item 3	95.5%	Test item 3	36.4%	Test item 3	33.3%
Test item 4	95.5%	Test item 4	54.5%	Test item 4	25.0%

Table 7. Percentages of correct responses to individual items in Blocks 1, 2, and 3 in NT children between 7;0-7;11 years

Block 1 Diverse Desires <i>N = 21</i>		Block 2 First-order False Beliefs <i>N = 20</i>		Block 3 Second-order False Beliefs <i>N = 15</i>	
Test item 1	95.2%	Test item 1	50.0%	Test item 1	66.7%
Test item 2	95.2%	Test item 2	65.0%	Test item 2	73.3%
Test item 3	95.2%	Test item 3	60.0%	Test item 3	60.0%
Test item 4	95.2%	Test item 4	70.0%	Test item 4	40.0%

Table 8. Percentages of correct responses to individual items in Blocks 1, 2, and 3 in NT children between 8;0-8;11 years

Block 1 Diverse Desires <i>N = 22</i>		Block 2 First-order False Beliefs <i>N = 22</i>		Block 3 Second-order False Beliefs <i>N = 21</i>	
Test item 1	100.0%	Test item 1	81.8%	Test item 1	80.9%
Test item 2	95.5%	Test item 2	81.8%	Test item 2	80.9%
Test item 3	100.0%	Test item 3	81.8%	Test item 3	76.2%
Test item 4	100.0%	Test item 4	90.9%	Test item 4	71.4%

Table 9. Percentages of correct responses to individual items in Blocks 1, 2, and 3 in NT children between 9;0-9;11 years

Block 1 Diverse Desires <i>N</i> = 13		Block 2 First-order False Beliefs <i>N</i> = 13		Block 3 Second-order False Beliefs <i>N</i> = 12	
Test item 1	100.0%	Test item 1	84.6%	Test item 1	83.3%
Test item 2	100.0%	Test item 2	76.9%	Test item 2	91.7%
Test item 3	100.0%	Test item 3	92.3%	Test item 3	75.0%
Test item 4	100.0%	Test item 4	92.3%	Test item 4	75.0%

Table 10. Percentages of correct responses to individual items in Blocks 1, 2, and 3 in NT children between 10;0-10;11 years

Block 1 Diverse Desires <i>N</i> = 20		Block 2 First-order False Beliefs <i>N</i> = 20		Block 3 Second-order False Beliefs <i>N</i> = 18	
Test item 1	100.0%	Test item 1	85.0%	Test item 1	94.4%
Test item 2	100.0%	Test item 2	90.0%	Test item 2	88.9%
Test item 3	100.0%	Test item 3	80.0%	Test item 3	72.2%
Test item 4	100.0%	Test item 4	80.0%	Test item 4	72.2%

Autistic children

Table 11. Percentages of correct responses to individual items in Blocks 1, 2, and 3 in autistic children between 4;0-4;11 years

Block 1 Diverse Desires <i>N</i> = 7		Block 2 First-order False Beliefs <i>N</i> = 3		Block 3 Second-order False Beliefs <i>N</i> = 1	
Test item 1	42.9%	Test item 1	33.0%	Test item 1	0%
Test item 2	57.1%	Test item 2	0%	Test item 2	0%
Test item 3	57.1%	Test item 3	33.0%	Test item 3	0%
Test item 4	57.1%	Test item 4	33.0%	Test item 4	100%

Table 12. Percentages of correct responses to individual items in Blocks 1, 2, and 3 in autistic children between 5;0-5;11 years

Block 1 Diverse Desires <i>N</i> = 10		Block 2 First-order False Beliefs <i>N</i> = 5		Block 3 Second-order False Beliefs <i>N</i> = 2	
Test item 1	40.0%	Test item 1	60.0%	Test item 1	50.0%
Test item 2	60.0%	Test item 2	40.0%	Test item 2	100%
Test item 3	60.0%	Test item 3	20.0%	Test item 3	50.0%
Test item 4	40.0%	Test item 4	40.0%	Test item 4	50.0%

Table 14. Percentages of correct responses to individual items in Blocks 1, 2, and 3 in autistic children between 6;0-6;11 years

Block 1 Diverse Desires <i>N</i> = 11		Block 2 First-order False Beliefs <i>N</i> = 7		Block 3 Second-order False Beliefs <i>N</i> = 5	
Test item 1	53.8%	Test item 1	0.0%	Test item 1	40.0%
Test item 2	61.5%	Test item 2	16.7%	Test item 2	40.0%
Test item 3	61.5%	Test item 3	16.7%	Test item 3	40.0%
Test item 4	61.5%	Test item 4	33.3%	Test item 4	60.0%

Table 15. Percentages of correct responses to individual items in Blocks 1, 2, and 3 in autistic children between 7;0-7;11 years

Block 1 Diverse Desires <i>N = 11</i>		Block 2 First-order False Beliefs <i>N = 7</i>		Block 3 Second-order False Beliefs <i>N = 0</i>	
Test item 1	63.6%	Test item 1	28.6%	Test item 1	40.0%
Test item 2	72.7%	Test item 2	57.1%	Test item 2	40.0%
Test item 3	81.8%	Test item 3	42.9%	Test item 3	40.0%
Test item 4	72.7%	Test item 4	57.1%	Test item 4	60.0%

Table 16. Percentages of correct responses to individual items in Blocks 1, 2, and 3 in autistic children between 8;0-8;11 years

Block 1 Diverse Desires <i>N = 10</i>		Block 2 First-order False Beliefs <i>N = 7</i>		Block 3 Second-order False Beliefs <i>N = 3</i>	
Test item 1	70.0%	Test item 1	42.9%	Test item 1	0.0%
Test item 2	70.0%	Test item 2	28.6%	Test item 2	0.0%
Test item 3	70.0%	Test item 3	57.1%	Test item 3	33.3%
Test item 4	70.0%	Test item 4	42.9%	Test item 4	0.0%

Table 17. Percentages of correct responses to individual items in Blocks 1, 2, and 3 in autistic children between 9;0-9;11 years

Block 1 Diverse Desires <i>N = 11</i>		Block 2 First-order False Beliefs <i>N = 9</i>		Block 3 Second-order False Beliefs <i>N = 6</i>	
Test item 1	72.7%	Test item 1	55.6%	Test item 1	100%
Test item 2	81.8%	Test item 2	55.6%	Test item 2	100%
Test item 3	81.8%	Test item 3	66.7%	Test item 3	100%
Test item 4	77.8%	Test item 4	66.7%	Test item 4	66.7%

Table 18. Percentages of correct responses to individual items in Blocks 1, 2, and 3 in autistic children between 10;0-10;11 years

Block 1 Diverse Desires <i>N = 20</i>		Block 2 First-order False Beliefs <i>N = 17</i>		Block 3 Second-order False Beliefs <i>N = 9</i>	
Test item 1	90.0%	Test item 1	52.9%	Test item 1	100%
Test item 2	95.0%	Test item 2	52.9%	Test item 2	88.9%
Test item 3	90.0%	Test item 3	76.5%	Test item 3	55.6%
Test item 4	85.0%	Test item 4	52.9%	Test item 4	66.7%