

## **Theory of mind (ToM)**

### **False belief (FB)**

This was a summary score of total performance across a combined 1<sup>st</sup> and 2<sup>nd</sup> order FB task and a more complex 2<sup>nd</sup> order FB task. The combined task was based on the chocolate story described by Sullivan et al., (1994) and also used by Hughes et al. (2000). Participants were read a story about two characters, Mary and John, accompanied by cartoon illustrations. They were told that Mary and John hide some chocolate in the kitchen fridge but whilst Mary was outside, John removes the chocolate and places it in his bag. The 1<sup>st</sup> order FB question asks: “Where does Mary think the chocolate is?”. The participant is also asked a justification question (“Why does Mary think the chocolate is in the \_\_\_\_\_?”) and a control question (“Where has John put the chocolate really?”). For the second order part of the story, the participant is told that when John was hiding the chocolate Mary saw him from the kitchen window, but that John did not see Mary looking. The 2<sup>nd</sup> order FB question asks: “Where does John think Mary will look for the chocolate?”. The participant is also asked a justification question (“Why does John think Mary will look for the chocolate in the \_\_\_\_\_?”) and three control questions (“Did John see Mary watching him through the window?”; “Where is the chocolate really?”; “Where was the chocolate first of all?”). For the 1<sup>st</sup> order section, the participant scored one point for passing the FB question correctly and an additional point for passing the justification question correctly (score range 0-2). For the 2<sup>nd</sup> order section, the participant scored one point for passing the FB question correctly and a score of one or two for passing the justification question correctly (1 = a correct justification but no use of mental state language e.g. ‘John didn’t see Mary looking’; and 2 = a correct justification with use of mental state language e.g. ‘He thinks she didn’t see him’) (score range 0-3). If the participant failed the FB question then the overall score was automatically set to 0.

The 2<sup>nd</sup> order FB task (see Bowler, 1992 for a full transcript) was verbally more demanding. Participants who failed the two false belief questions in the combined story automatically received a score of 0 for the 2<sup>nd</sup> order FB task and it was not administered, unless they passed an additional 1<sup>st</sup> order story (not reported here). For the 2<sup>nd</sup> order FB task, participants listened to the experimenter reading a story about two characters, Peter and Jane, who visit two shops at lunchtime to find Peter a new coat; they identify a preferred coat in one of the shops and plan to purchase it that evening. Later that day, both characters independently

discover that the coat Peter had wanted to purchase has sold out. The participants are tested on their understanding of Jane's 2<sup>nd</sup> order FB that Peter will have gone to the shop they had both planned to go to, and not to the other shop ("Where does Jane think Peter has gone to buy his coat?"). This is followed by a justification question ("Why?"), reality question ("Where has Peter really gone to buy his coat?") and memory question ("In which shop did Peter see the coat he liked the best?"). Scoring was identical to the 2<sup>nd</sup> order FB questions above (score range 0-3). A combined summary score ranging from 0-8 was calculated based on performance across both tasks.

### **Strange stories (Happé, 1994)**

This was included as more general measure of mental state understanding and required understanding of the effect of and motivation for double bluffs, misunderstanding, lies and persuasion. The participants were read a series of six stories, which were simultaneously provided in written form and accompanied by an appropriate illustration (for further details of our protocol see Ricketts et al., 2012). At the end of each story, which included at least one character, they were asked a question about the text. The four stories with a ToM component, where the participant had to demonstrate understanding of the character's thoughts, feelings and intentions, were included in the current analysis. A 0-1-2 scoring system was used, with 0 representing an incorrect or "don't know" response and 2 representing a full and explicitly correct answer. Performance across the four stories was averaged.

### **Frith- Happé animations (Abell et al., 2000; Castelli et al., 2000)**

This was included as more general measure of mental state understanding and required recognition and understanding of the intent of one character to manipulate the mental state of another character. Participants viewed a series of six silent 2D animations (average length 41 seconds, SD 4.1 s) featuring two interacting 'characters', a big red triangle and a small blue triangle, moving within a white framed space (for further details of our protocol see Jones et al., 2011; Ricketts et al., 2012). The participant had to watch the animation and describe what they thought the two triangles were doing, with their verbal response recorded for later transcription and scoring. Analysis focussed on the four animations that included an interaction with a mentalising (ToM) component, with participants scored for their understanding of intentionality. This was scored from 0-5, with 0=no mental state language i.e. describing a non-deliberate action, and 5=sophisticated use of mental state language i.e.

describing deliberate action with goal of affecting mental state). Performance across the four animations was averaged.

### **Reading the mind in the eyes task – children’s version (RMET; Baron-Cohen et al., 2001)**

This task was included for its ability to capture the understanding of another person’s mental state based on their eye expression. Participants were shown black and white photographs of the eye region of the face of 28 people. They were asked to pick which of four words best described what the person in the photo was thinking or feeling. The children’s version of the task was selected because the language level of the words was more accessible. The outcome variable was the total number of items correct. Score range: 0-28 (high score = better performance).

### **Executive functions (EF)**

#### **Opposite Worlds (Manly et al., 2001)**

This test, from the Test of Everyday Attention for Children (TEA-Ch; Manly et al., 2001), was included as a measure of inhibition of a verbal response. The participant was told there was a ‘Same World’, where everything was the same as this world, and an ‘Opposite World’, where you had to say the opposite of what you would here. For the Same World they had to say the number (1 or 2) contained within each square on an A4 sheet of paper. For the Opposite World they had to say the opposite number to the one printed in each square. The time taken to complete each sequence of numbers was recorded, with two trials of each World presented. The outcome variable was the subtraction of the mean Same Worlds score from the mean Opposite Worlds score. Score range: time difference in milliseconds (high score = worse performance)

#### **Card Sort (Tregay et al., 2009)**

This was included as a measure of cognitive set-shifting. Three independent trials were presented in which participants had to sort a pack of cards into a character’s (represented by a photo) favourite cards. The cards depicted single objects that varied on three dimensions: colour, shape and size. Participants were instructed: “I will show you the cards one at a time and I will ask you if you think it is one of [character]’s favourite cards. There is a rule about

which cards [character] likes best. I would like you to try to work out the rule and remember it.” The participant was told that it was fine to guess to begin with and that they would be told if they were correct or not. A counterbalanced system was used to pre-determine the sorting dimension, with the participant’s decision on the first sort taken as the correct answer (i.e. the character’s favourite). The sorting continued until the participant was correct on six consecutive sorts or after 20 sorts had elapsed. The outcome variable was the number of errors on each trial, summed to create a total error score.

### **Category fluency (Benton et al., 1994)**

This task was included as a verbal measure of generativity. Participants were given one minute to think of as many words that belonged to the category of animals. This was followed by the category of food. Before the task started they were given a short practice with the category of items of clothing, which was terminated after a few correct responses had been given. For the outcome variable, the total number of valid responses across the three trials was calculated. Non-valid items included items not belonging to the category, whole word repetition and non-words.

### **Design fluency (Jones-Gotman & Milner, 1977)**

This task was included as a non-verbal measure of generativity. Participants were asked to think of as many made up drawings as they could in two minutes. They were given two rules: i) that the drawings must be completely made up so could not be recognisable objects, ii) the drawings could not just be scribbles. The experimenter then created some made-up drawings, as well as ‘banned’ drawings of a flower and a scribble. They were presented with sheets of A4 paper with a 5x3 grid of squares and told to put a new drawing in each square. The outcome variable was the total number of valid responses, which excluded nameable items, scribbles and perseverations of earlier designs.

### **Digit span (Cohen, 1997)**

This task was an assessment of digit span taken from the Children’s Memory Scale (Cohen, 1997) and was included as a measure of verbal working memory. Participants listened to the experimenter say a list of one-digit numbers at a rate of 1 per second and then had to recall

them in reverse formation as accurately as possible. The number of trials completed correctly was used as a measure of working memory (i.e. the raw score).

### **Planning drawing (Booth et al., 2003)**

This task was a measure of both central coherence and planning; only planning was examined here. Participants were asked to copy a simple line drawing of a house, snowman and ship, presented in a pseudo-random order. After they had completed their drawing they were asked to re-draw it but with an additional feature. For the house two more windows needed to be added, for the snowman teeth needed to be added, and for the ship people had to be added to the portholes. Successful inclusion of the new feature required the participant to modify their original drawing (e.g. enlarging the drawing, altering other features). The degree of modification was coded on a 0-2 scale, with 2 indicating a good allowance and 0 indicating very little or no allowance. Thus, a high score indicated better planning. The outcome variable was the summed score across the three trials, range 0-6.

- Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, 15, 1-16.
- Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage*, 12, 314-325.
- Benton, A. L., Hamsher, K. deS, & Sivan, A. B. (1994). *Multilingual aphasia examination*. Iowa City, IA: AJA Associates.
- 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(2), 241-251.
- Booth, R., Charlton, R., Hughes, C., & Happé, F. (2003). Disentangling weak coherence and executive dysfunction: planning drawing in autism and attention-deficit/hyperactivity disorder. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1430), 387-392. doi: 10.1098/rstb.2002.1204
- Bowler, D.M. (1992). "Theory of mind" in Asperger's syndrome. *Journal of Child Psychology and Psychiatry*, 33, 877-893
- Cohen, M. (1997). *Children's Memory Scale (CMS)*. San Antonio, Texas: The Psychological Corporation.
- Happé, F.G. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, 24, 129-54.
- Hughes, C., A. Adlam, F. Happe, J. Jackson, A. Taylor, & A. Caspi (2000). Good test--retest reliability for standard and advanced false-belief tasks across a wide range of abilities. *Journal of Child Psychology and Psychiatry*, 41, 483-90

- Jones, C.R.G., Swettenham, J., Charman, T., Marsden, A.J.S., Tregay, J., Baird, G., Simonoff, E., & Happé. (2011). No evidence for a fundamental visual motion processing deficit in adolescents with autism spectrum disorders. *Autism Research*, 4, 347-357.
- Jones-Gotman, M., & Milner, B. (1977). Design fluency: the invention of nonsense drawings after focal cortical lesions. *Neuropsychologia*, 15(4-5), 653-674.
- Manly, T., Anderson, V., Nimmo-Smith, I., Turner, A., Watson, P., & Robertson, I. H. (2001). The differential assessment of children's attention: the Test of Everyday Attention for Children (TEA-Ch), normative sample and ADHD performance. *Journal of Child Psychology and Psychiatry*, 42(8), 1065-1081.
- Ricketts, J., Jones, C.R.G., Happé, F., & Charman, T. (2013). Reading comprehension in autism spectrum disorders: The role of oral language and social functioning. *Journal of Autism and Developmental Disorders*, 43, 807-816.
- Sullivan, K., D. Zaitchik, H. & Tager Flusberg (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology*, 30, 395-402
- Tregay, J., Gilmour, J., & Charman, T. (2009). Childhood rituals and executive functions. *British Journal of Developmental Psychology*, 27(Pt 2), 283-296.

Supplementary Table 1: Correlations between cognitive tasks.

	False belief	Strange stories	F-H animations	RMET	Opposite worlds	Card sort	Category fluency	Design fluency	Digit span	Planning drawing
Strange stories	.43*** (.06)									
Animations	.33** (.04)	.20 (-.03)								
RMET	.45*** (-.08)	.29** (-.01)	.45*** (.31**)							
Opposite worlds	.25* (-.08)	.23* (.06)	.13 (.03)	.21* (.03)						
Card sort	.63*** (.29**)	.29** (-.07)	.34** (.07)	.22* (-.19)	.31** (.11)					
Category fluency	.09 (-.32**)	.13 (-.08)	.45*** (.35**)	.33** (.15)	.29** (.20)	.34** (.21)				
Design fluency	.29** (-.05)	.24* (.02)	.27* (.10)	.39*** (.21)	.30** (.16)	.27** (.02)	.32** (.28*)			
Digit span	.44*** (.01)	.37*** (.10)	.30* (.08)	.38*** (.06)	.36*** (.23*)	.51*** (.20)	.28** (.14)	.27* (.02)		
Planning drawing	.29** (.06)	.26* (.15)	.01 (-.14)	.05 (-.11)	.20* (-.00)	.38*** (.24*)	.13 (-.02)	.07 (-.03)	.18 (-.02)	
FSIQ	-.73***	-.58***	-.47***	-.55***	-.38***	-.65***	-.34**	-.40**	-.58***	-.35***

Animations, Frith-Happé animations; FSIQ, full-scale IQ; RMET, Reading the mind in the eyes task. \*\*\* p < .001 \*\* p < .01, \* p < .05

Correlations in brackets are partialled for FSIQ. Dashed squares indicate correlations within cognitive category (theory of mind; executive function).