S2: Assessment of Memory and Language: Details of tests and further information.

i) For adults, the Wechsler Memory Scale-III UK (WMS, Wechsler 1997) was administered. 

*Primary Verbal Memory Index* scores were calculated from subtests: i) *Logical Memory* that comprised two stories with 25 ‘idea’ units verbally presented once only for immediate recall, with delayed recall after approximately 30 minutes, and ii) *Verbal Paired Associate Learning* that comprised verbal learning of eight word pairs presented for immediate recall over four consecutive trials, with delayed recall after approximately 30 minutes. *Primary Nonverbal Memory Index* scores were calculated from subtests: i) *Face Recognition* that comprised initial presentation of 24 faces at the rate of one every two seconds, immediately followed by presentation of a second series of 48 faces. For immediate recognition, participants were required to indicate whether they had seen each face before, with delayed recognition of the same series of 48 faces following a filled delay of approximately 30 minutes and ii) *Family Pictures* that comprised presentation of four scenes for ten seconds each: each picture featured specific standard characters engaged in activities. Verbally mediated recall of the character, their spatial location (using a 2x2 grid) and their activity were requested immediately, with filled delay of approximately 30 minutes.

ii) For children, the Children’s Memory Scale (CMS, Cohen, 1997) was administered, with distinct versions according to age-group (CMS ages 5-8 years administered to V1, 1; CMS ages 9-16 administered to V,8, and V,13). *Primary Verbal Memory Index* scores were calculated from raw and age-scaled scores of the following verbal memory subtests: i) *Logical Memory* that comprised two stories containing ‘idea’ or story units (5-8 years, N= 57; 13-16 years, N=88) verbally presented only for immediate recall, followed by delayed recall, without prompts, after a filled delay of approximately 25-30 minutes and ii) *Verbal Paired Associate Learning* that comprised verbal learning of ten word pairs (or fourteen pairs from the 9-16 version) presented for immediate recall over three consecutive trials, and for delayed recall without prompts after a filled delay of approximately 25-30 minutes. *Primary Visual Memory Index* scores were calculated from raw and age-scaled scores of the following nonverbal memory subtests: i) *Face Recognition* where the procedure and scoring procedures were identical to that reported for the WMS (Maximum total N=36 for V1,1 and N=48 for V,8, and V,13) and ii) *Dot Locations* where dots were presented in a fixed pattern in various locations across a grid on each trial. Following removal of the dots, each participant was given plastic chips to indicate their immediate recall of the presented pattern. Three learning trials were administered, and correct positioning of the chips totalled across trials to give an immediate recall raw score (Maximum total N=24 for V1,1 and N=32 for V,8, and V,13). Delayed recall of the pattern was initiated after a filled delay of approximately 30 minutes. In addition, the Family Pictures subtest of the CMS was included, with a similar procedure to the adult version, apart from the use of a family ‘portrait’ of characters to aid recognition of identity, plus a visual scene with no family members present to aid recall of spatial location.

iii) For both adults and children, the (optional) Word List Learning subtest was presented. This was with a multi-item word list (of variable length, according to test version) presented in identical order over four consecutive trials and subsequent delayed recall. Maximum delay total scores varied according to age in the CMS (N=10 for V1,1; N=14 for V,8 and V,13) and comprised N=12 in the adult version.

iv) Establishing a memory-IQ discrepancy on individual cases is a standard neuropsychological method. Three reasons suggest cautious interpretation. 1 - Verbal IQ scores should not be calculated where discrepant performance exists across the component IQ subtests: several affected cases demonstrated a significant WM> VCI discrepancy, 2 - The simple difference method (CMS, notably for Cases VI.1, V.8) is less reliable than the predicted difference method (Bornstein, Chelune et al. 1989; Hawkins and Tulsy, 2001) and 3 - Discrepancy scores have questionable validity when IQ scores are at either tail of the normal distribution (such as Performance IQ in the JR cases), unless the validation sample has been further streamed for ability or socioeconomic variables.
v) The Test for the Reception of Grammar-2 (Bishop, 2003) assessed sentence-picture matching across twenty different grammatical constructions (n=20 blocks, four trials per block, n=80 sentences total).

vi) The Clinical Evaluation of Language Fundamentals-III (CELF\textsuperscript{UK3} Semel, Wiig & Secord, 1995) was administered using six subscales. 

\textit{Organisation and knowledge of word meaning} was assessed by the Word Class subscale where participants pair two related words from an array (maximum score = 35), and the Word Associations subscale where participants generate category exemplars for three target categories (e.g. animals) in one minute. 

\textit{Semantic and syntactically-guided encoding} was assessed through the Semantic Relationships subscale where participants (aged > nine years), select and verify two (visually presented) factual statements, after listening to a set of (orally presented) facts (maximum score = 22) and \textit{Listening to Paragraphs} subscale where participants listen to two short paragraphs (read aloud) and answer literal and inferential questions related to content (maximum score = 10). 

\textit{Sentence-level expression} was tapped by the Sentence Assembly subscale where two sentences are constructed from an unordered array of words (maximum score = 20), the Formulated Sentences subscale where spoken sentences about a picture are constructed from target words (maximum score = 22) and the Sentence Repetition subscale that requires verbatim recall of single sentences (maximum score = 26). Only raw scores were calculated as since age-scaled scores were available only for individuals < 21 years.

vii) The Children’s Test of Nonword Repetition (CNRep; Gathercole & Baddeley, 1996) assessed phonological memory. Participants repeat forty nonsense items (e.g. blonterstaping). There are ten items at each of two, three and five syllable lengths with typical stress for the English language. All items were presented in a fixed random order via digitised voice samples. Responses were recorded and raw scores calculated in line with scoring instructions.

viii) The ‘Camel and Cactus’ Test (CCT; Bozeat, Lambon-Ralph, Patterson, Garrard & Hodges et al. 2000) assessed conceptual and associative knowledge. The Pyramids and Palm Trees (PPT: Howard & Patterson, 1992) was not used, as pilot work with 3 JR cases indicated ceiling performance on both word and picture versions. Two forms of the CCT were administered: i)\textit{Picture version}: The stimuli consisted of 64 test target items that were presented above four pictures. The subject was required to select a single picture that had an associative relationship with the target item. The pictures shown with the target were all taken from the same category of item e.g. camel was presented with cactus (the correct response), tree, sunflower and rose (all types of vegetation). ii) \textit{Word version}: The same 64 target items were presented as words, and all of the four possible responses also as words in the same location as the picture version. Participants selected a single word that had an associative relationship with the target item.

ix) Action & Object naming Stimuli consisted of 520 black and white line drawings of common objects and 275 transitive and intransitive actions used by(Szekely, D’Amico et al. 2005) Americanised object names were replaced with the anglicised version e.g. ‘stroller’ was changed to ‘pushchair’. Picture stimuli were downloaded from \url{http://www.crl.ucsd.edu/~askekely/imnp/actobj.html}. Pictures files were edited to be identical size at presentation. Self-paced presentation was programmed in Java, and designed to run on a Mac or PC. Older subjects needed to take small breaks in the middle of stimuli sets. The stimuli were split into 5 object naming blocks (4 blocks of 100 stimuli and one block of 120 stimuli) and 3 action naming blocks (two blocks of 100 stimuli and one block of 75). Stimuli were pseudo randomised, and each participant received them in the same order. For each trial, the participant saw a fixation point for 500 milliseconds follow by an image of an object or an action that they were required to name as quickly as possible.

x) The Graded Naming test (McKenna and Warrington 1983) consisted of 30 line drawn pictures graded for difficulty of naming. If the participant was not clear of the exact item to be named) the experimenter pointed to the specific item after the more general name had been given. In addition to the original task administration, they were then asked to define the item they were shown to name. No discontinue rules were applied.