

Do measures of memory, language, and attention predict eyewitness memory in children with and without autism?

Lucy A Henry

City, University of London, London, UK

David J Messer

The Open University, Milton Keynes, UK

Rachel Wilcock

University of Winchester, Winchester, UK

Gilly Nash, Mimi Kirke-Smith and Zoe Hobson

City, University of London, London, UK

Laura Crane

City, University of London, London, UK; University of Winchester, Winchester, UK; University College London, London, UK

Autism & Developmental Language Impairments

Volume 2: 1–17

© The Author(s) 2017

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/2396941517722139

journals.sagepub.com/home/dli



Abstract

Background and aims: There are few investigations of the relationship between cognitive abilities (memory, language, and attention) and children's eyewitness performance in typically developing children, and even fewer in children on the autism spectrum. Such investigations are important to identify key cognitive processes underlying eyewitness recall, and assess how predictive such measures are compared to intelligence, diagnostic group status (autism or typically developing) and age.

Methods: A total of 272 children (162 boys, 110 girls) of age 76 months to 142 months ($M = 105$ months) took part in this investigation: 71 children with autism and 201 children with typical development. The children saw a staged event involving a minor mock crime and were asked about what they had witnessed in an immediate Brief Interview. This focused on free recall, included a small number of open-ended questions, and was designed to resemble an initial evidence gathering statement taken by police officers arriving at a crime scene. Children were also given standardised tests of intelligence, memory, language, and attention.

Results & conclusions: Despite the autism group recalling significantly fewer items of correct information than the typically developing group at Brief Interview, both groups were equally accurate in their recall: 89% of details recalled by the typically developing group and 87% of the details recalled by the autism group were correct. To explore the relationship between Brief Interview performance and the cognitive variables, alongside age, diagnostic group status and non-verbal intelligence quotient, multiple hierarchical regression analyses were conducted, with Brief Interview performance as the dependant variable. Age and diagnostic group status were significant predictors of correct recall, whereas non-verbal intelligence was less important. After age, non-verbal intelligence, and diagnostic group status had been accounted for, the only cognitive variables that were significant predictors of Brief Interview performance were measures of memory (specifically, memory for faces and memory for stories). There was little evidence of there being differences between the autism and typically developing groups in the way the cognitive variables predicted the Brief Interview.

Corresponding author:

Lucy A Henry, Division of Language and Communication Science, City, University of London, 10 Northampton Square, London EC1V 0HB, UK.
Email: Lucy.Henry.1@city.ac.uk



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Implications: The findings provide reassurance that age – the most straightforward information to which all relevant criminal justice professionals have access – provides a helpful indication of eyewitness performance. The accuracy of prediction can be improved by knowing the child's diagnostic status (i.e. whether the child is on the autism spectrum), and further still by using more specific assessments (namely memory for faces and memory for stories), possibly via the input of a trained professional. Importantly, the findings also confirm that whilst children with autism may recall less information than typically developing children, the information they do recall is just as accurate.

Keywords

Eyewitness, interviews, autism, memory, language, attention

Introduction

Evidence from eyewitnesses, and its reliability, is often a key element in judicial processes (Kebbell & Milne, 1998; Wells, Memon, & Penrod, 2006). Historically, child witnesses were thought to be inherently unreliable (Odegard & Toglia, 2013), but the consensus now is that even developmentally young children provide at least some accurate information if interviewed appropriately (Bull, 2010; Lamb, Malloy, & La Rooy, 2011). As children develop, the amount and accuracy of their recall increases (Brown & Lamb, 2015; La Rooy, Malloy, & Lamb, 2011; Odegard & Toglia, 2013), and their suggestibility declines (London, Henry, Conratt, & Corser, 2013). The most reliable evidence from child witnesses is obtained using free recall and open questions (Brown & Lamb, 2015; Bull, 2010; La Rooy et al., 2011), which appears to maximise their recall without compromising accuracy.

In the case of children with autism spectrum disorder (henceforth, autism), a small but growing literature indicates that they remember less about witnessed events relative to typically developing (TD) children of comparable age and intelligence (IQ) (Bruck, London, Landa, & Goodman, 2007; McCrory, Henry, & Happé, 2007), and also when matched for verbal and non-verbal abilities but differing in age (Mattison, Dando, & Ormerod, 2016, 2015). Nevertheless, the information that they do provide is often just as accurate (Bruck et al., 2007; Mattison et al., 2015, 2016 [true for probed but not free recall]; McCrory et al., 2007). Further, children with autism are no more suggestible than their TD peers, and are not more likely to confabulate items of information (Bruck et al., 2007; Mattison et al., 2016, 2015; McCrory et al., 2007). Thus, existing research evidence suggests that children with autism *can* be reliable eyewitnesses, but may provide less information than their TD peers. We sought to add to this evidence base in the current study.

There is, however, considerable variability in the performance of different children – with and without autism – when asked to recall a witnessed event. Whilst some children produce full and accurate descriptions of

events, others provide very sparse accounts, with these variations occurring even among children of similar developmental levels (Chae & Ceci, 2005). The challenge is to identify which variables may explain these differences.

The current investigation explored individual differences factors that could predict performance on a Brief Interview about a witnessed event in 6- to 11-year-old children with and without autism. The use of an immediate Brief Interview was designed to simulate a situation where a police officer arrives at a crime scene to take an initial statement. The focus was on the prediction of Brief Interview performance from easy to obtain variables such as age and diagnostic group status (autism or TD), as well as general ability (as indicated by specific measures of intelligence). We further examined whether individualised assessments of key cognitive abilities, namely standardised measures of memory, language, and attention, could add predictive power to the other variables (age, diagnostic group status, IQ). These variables are theoretically and practically relevant to eyewitness recall in children with or without autism, as outlined next.

Age and intelligence (IQ)

One of the most reliable findings in the literature is that, with increasing age, TD children's volume and accuracy of recall improves (e.g. Brown & Lamb, 2015; Burgwyn-Bailes, Baker-Ward, Gordon, & Ornstein, 2001; Chae, Kulkofsky, Debaran, Wang, & Hart, 2016; La Rooy et al., 2011; Odegard & Toglia, 2013). This is likely because age is related to many cognitive abilities relevant to witness recall. Research on age-related improvements in recall for autistic children is limited, but there is some suggestion that age may be less strongly related to witness performance than in TD children (Bruck et al., 2007). IQ has modest and variable relationships with eyewitness recall in TD children that change with age (e.g. Elischberger & Roebbers, 2001; Geddie, Fradin, & Beer, 2000; Henry & Gudjonsson, 2007; Roebbers & Schneider, 2001). Although it is unclear whether similar relationships

emerge in children with autism, the limited available research suggests they may not (Bruck et al., 2007). We, therefore, investigated the role of IQ as a possible predictor of eyewitness performance. Because verbal IQ and full-scale IQ also assess language ability, and language ability was a further predictor in our study (see below), non-verbal IQ was chosen to be the relevant predictor variable, to minimise shared variance between IQ and language in the predictive analyses (also, see Dawson, Soulières, Gernsbacher, & Motttron, 2007 for a discussion of the issues involved in measuring intelligence in individuals on the autism spectrum).

Memory

Although general memory ability (comprising verbal and visual memory) seems relevant to eyewitness recall, standardised measures of memory have rarely been included in investigations with TD children. When memory has been considered, results have been inconclusive. Baker-Ward, Gordon, Ornstein, Larus, and Clubb (1993), for example, found no consistent relationships between verbal memory and witness recall in 3- to 7-year-old children, whereas Henry and Gudjonsson (2003) reported verbal, but not non-verbal memory, to predict free recall and performance on open-ended questions in 5- to 12-year-old children. These latter results emerged for a repeated interview two weeks after a witnessed event, but were not apparent in an immediate interview. Somewhat stronger relationships between verbal memory and eyewitness performance were reported by Henry and Gudjonsson (2003) in children with intellectual disabilities (11–12 years) for both immediate and delayed recall.

Potential differences between verbal and visual memory and their relationships with witness recall may be particularly relevant to the recall of children with autism. This is because visual, but not verbal, memory difficulties have been reported in both children and adults on the autism spectrum (e.g. Goddard, Howlin, Dritschel, & Patel, 2007; Goddard, Dritschel, Robinson, & Howlin, 2014), although, importantly, Goddard et al. (2014) failed to find relationships between verbal or visual memory and autobiographical memory performance in children with autism. Accordingly, several measures of verbal and non-verbal memory (taken from a standardised test battery) were included in the current study, as these relationships have not been examined in children with autism using a witness recall paradigm. The memory tasks used in the current study ranged from more abstract assessments of memory (for word pairs and pictorial sequences) to processes more closely associated with eyewitness recall (memory for faces and stories).

Language

Language is integral to the development of a child's ability to organise, elaborate on, and recall personally experienced events (Fivush & Nelson, 2004). Relationships between eyewitness memory and language could reflect the ability to encode information in a verbal format, rehearse past experiences effectively, comprehend the interview questions and context, and/or respond to and structure a verbal narrative account. Research suggests that language ability is related to the amount and completeness of information recalled by TD children, although the details of the findings vary. For example, Chae and Ceci (2005) found that verbal intelligence related to open-ended recall of a witnessed event in 5- to 8-year-old children, but this relationship was largely driven by the older children (7–8 years) and was not present for measures of cued recall. Further, Burgwyn-Bailes et al. (2001) reported that receptive vocabulary significantly predicted delayed (but not immediate) memory of an emergency medical procedure; a relationship that was stronger for younger than older children (age range 3–7 years). In a further study, receptive vocabulary was related to performance on general open-ended questions (and errors in both free and general recall), but not to free recall in children between the ages of 8 and 12 years (Henry & Gudjonsson, 2007).

Recent work has focussed on younger TD children using more extensive assessments of language, reporting clearer and more consistent relationships between language and eyewitness recall. Chae, Kulkofsky, Debaran, Wang, and Hart (2014) found that 3- to 5-year-old children with higher expressive and receptive vocabulary skills produced more information about a witnessed event. Similarly, Chae et al. (2016) reported that several measures of language (adaptive language use, receptive and expressive vocabulary, narrative skill) were related to measures of event memory in 3- to 5-year-old children.

Language may be an even stronger predictor of witness recall in children with autism, given the extensive range of structural language difficulties characteristic of this group (e.g. Boucher, 2012). There is little direct evidence about the relationship between witness recall and language for children with autism, but McCrory et al. (2007) reported a correlation (controlling for IQ) between total amount recalled and letter fluency. Goddard et al. (2014) also reported category fluency to be a significant predictor of autobiographical memory in children with autism. Although measures of verbal fluency are often considered to reflect executive functioning (Pennington & Ozonoff, 1996; Smith-Spark, Henry, Messer, & Ziecik, 2017), there is evidence that they may be more strongly related to language ability (e.g. Henry, Messer, & Nash, 2015). Therefore, there is

reason to suppose that language skill might be related to witness recall in children with and without autism.

In the current study, several standardised language measures (from a range of assessment batteries) were included. This was important given: (1) evidence that receptive and/or expressive language skills are related to eyewitness memory; (2) the need to explore these relationships more thoroughly in children (6–11 years); and (3) the fact that language difficulties in children on the autism spectrum can be complex and variable (Boucher, 2012; Taylor, Maybery, & Whitehouse, 2014; Williams, Botting, & Boucher, 2008). Measures were used to assess: receptive vocabulary (to provide a general assessment of semantic knowledge related to objects and events); sentence recall (to provide an assessment of grammar); sequencing ability (to provide an assessment of the ability to generate coherent narratives); and grammatical abilities (including the ability to generate sentences from a list of words, which has similarities with generating sentences about remembered events). All of these measures were relevant in terms of providing a coherent narrative about a to-be-remembered event.

Attention

Attentional processes have rarely been investigated in relation to eyewitness testimony, despite their potential importance to the initial encoding of information. Chae et al. (2016) recently reported that questionnaire measures of ‘attentional focusing’ (i.e. questions about the child’s concentration) and inhibitory control (i.e. whether the child can wait before starting a new activity, if asked to) were positively related to measures of witness recall in 3- to 5-year-old TD children. For children with autism, McCrory et al. (2007) found that response suppression (i.e. inhibition) was correlated with witness recall. These results suggest that measures of attention might be related to eyewitness performance. Further, the documented difficulties with attention for many children with autism (e.g. van der Meer et al., 2012) make this an important area to investigate. However, available evidence is limited and no previous studies have utilised standardised behavioural measures of attention. The current study included measures of sustained, focused, and sustained-divided attention taken from a widely used and reliable standardised test battery.

In summary, the current investigation explored both easy to obtain (age, diagnostic group status) and more detailed cognitive (non-verbal IQ, memory, language, attention) predictors of eyewitness performance in children with and without autism, in relation to an immediate Brief Interview about a witnessed event. Administering a battery of cognitive tasks enabled us

to identify predictors of children’s event memory that could be helpful for professionals in the justice system, thereby highlighting the types of cognitive characteristics that contribute to informative and reliable eyewitness testimony. We first examined whether there were autism/TD group differences in the volume and accuracy of recall in the Brief Interview. Next, we determined the predictive power of the cognitive assessments, alongside age and diagnostic group status, also assessing whether predictive relationships were similar or different across the two groups. The current study represents the first thorough investigation of these issues in children with and without autism, building on and extending previous findings as follows: (1) a large sample of 201 TD children was included to ensure predictive relationships between cognitive variables and eyewitness memory were robust and reliable; (2) 71 children with autism (of the same age and IQ range as the TD children) were assessed to obtain novel data on predictors of eyewitness memory as a function of diagnostic group; and (3) a wider range of predictors was included (age, diagnostic group status, non-verbal IQ, memory, language, and attention) compared to previous research.

Based on previous findings, it was predicted that age would be related to immediate memory for a witnessed event in an open-ended Brief Interview: the relationship is well established in TD groups, but limited previous research made this prediction tentative for the autism sample. It was also expected that group (autism or TD) might be a significant predictor of performance, as individuals on the autism spectrum have been reported as producing fewer correct responses in relation to these types of tasks. Finally, as the existing findings in relation to IQ are variable, it was predicted that relationships with witness recall *may* emerge in one or both samples. For the three cognitive domains, it was expected that at least some memory subtests would be related to recall, particularly those with more relevance for eyewitness skills (e.g. memory for stories and faces). We also predicted that measures of receptive and expressive language would be related to Brief Interview performance. Finally, given the lack of previous evidence on this topic, we made a tentative prediction that attention variables may be related to witness recall.

Method

Participants

A total of 274 children (6–11 years old) were recruited for this study, but two participants were excluded (one from the autism group and one from the TD group) because they had full-scale IQs in the intellectual

Table 1. Mean (SD) scores for age and all cognitive variables for autism and TD groups, together with group differences.

Variables	Autism group (n = 60 to 71)	TD group (n = 201)	Group differences
Age	9yrs 4 m (18 m)	8yrs 7 m (15 m)	$t(103.8) = 3.66^d$, $p < .001$
WASI-II full-scale IQ ^a	97.8 (15.7)	106.8 (13.8)	$t(270) = 4.54$, $p < .001$
WASI-II non-verbal IQ ^b	48.0 (9.3)	50.8 (8.6)	$t(270) = 2.31$, $p = .02$
Memory for Stories ^c	10.2 (3.6)	12.3 (3.0)	$t(103.5) = 4.29^d$, $p < .001$
Facial Memory ^c	8.0 (3.0)	10.9 (3.0)	$t(269) = 6.94$, $p < .001$
Paired Recall ^c	10.0 (3.6)	11.1 (2.8)	$t(97.5) = 2.14^d$, $p = .03$
Visual Sequential Memory ^c	9.7 (3.8)	12.2 (3.1)	$t(99) = 4.97^d$, $p < .001$
BPVS-3 ^a	85.0 (16.4)	92.9 (13.9)	$t(108.3) = 3.66^d$, $p < .001$
CELF-4 Recalling Sentences ^c	7.1 (4.0)	10.9 (3.0)	$t(94.9) = 7.06^d$, $p < .001$
CELF-4 Formulated Sentences ^c	5.3 (3.3)	9.7 (3.1)	$t(265) = 9.72$, $p < .001$
ELT-2 Sequencing ^a	100.7 (14.4)	109.4 (8.6)	$t(85.6) = 4.74^d$, $p < .001$
ELT-2 Grammar and Syntax ^a	96.4 (14.3)	106.5 (10.3)	$t(85.4) = 5.21^d$, $p < .001$
TEA-Ch Sky Search ^a	8.0 (3.8)	9.2 (2.8)	$t(95.7) = 2.35^d$, $p = .02$
TEA-Ch Score! ^c	7.5 (4.1)	9.1 (3.5)	$t(269) = 3.10$, $p < .005$
TEA-Ch Dual Task decrement ^c	4.3 (3.7)	6.3 (3.7)	$t(268) = 3.85$, $p < .001$

BPVS: British Picture Vocabulary Scale; CELF: Clinical Evaluation of Language Fundamentals; ELT: Expressive Language Test; TEA-Ch: Test of Everyday Attention for Children; WASI-II: Wechsler Abbreviated Scale of Intelligence, second edition.

^aStandardised scores (mean 100, SD 15).

^bT-scores (mean 50, SD 10).

^cScaled scores (mean 10, SD 3).

^dEqual variances not assumed.

disability range (i.e. less than 70). The final sample consisted of 272 children (162 boys, 110 girls) between the ages of 76 months and 142 months ($M = 105$ months, standard deviation (SD) = 16 months). Of the 272 children, 201 were in the TD group, whereas 71 children had (prior to taking part in the research) received a formal autism diagnosis from an appropriately qualified clinical professional. This diagnosis was obtained independently of the research study and this information was provided to us by the parents and/or the school. To further confirm the diagnostic status of the participants, the Social Communication Questionnaire (Rutter, Bailey, & Lord, 2003) was sent to all participating parents. These were completed for 203 children (48 from the autism group, 155 from the TD group), and an independent samples t -test revealed higher levels of autism traits on this measure for the autism ($M = 19.81$, $SD = 6.64$) relative to the TD ($M = 5.17$, $SD = 4.31$) groups, $t(59.75) = 14.37$, $p < .001$ (equal variances not assumed).

Note that, in a few instances, it was not possible to carry out all the assessments for the children in the autism group. In this sample, assessments were obtained from 60 to 71 children and, in regression analyses where an assessment was missing, the mean for the group was substituted. Participants attended mainstream primary schools or special schools in Greater London and South East England. Although most were seen in schools, some participated either at the

University at which the research was carried out, or in their homes. Full details of the two groups are given in Table 1.

Materials and procedure

This study was part of a larger investigation of eyewitness performance across several stages (evidence gathering statements, investigative interviews, identification line-ups, cross-examinations), but only the first phase (evidence gathering statements; referred to as 'Brief Interviews') is relevant to the current paper. (Note that the Brief Interviews in the current phase of the research were not directly comparable to the later investigative interviews, because, following the Brief Interviews, children were allocated to one of four different types of interview conditions – see Henry et al., 2017.)

Staged event. Children watched either a live event during school assembly or a high-quality video of the event, which involved two actors giving a talk about what school was like a long time ago in Victorian times.¹ This talk was short (around 3.5 minutes) and contained educational content: several key facts about Victorian schools were given in each talk, with 'props' used to demonstrate key information such as a writing slate or an abacus. The event also included a minor crime, involving the 'theft' of either a phone or a set of keys.

Towards the end of the talk, the 'theft' was explained as a misunderstanding, to avoid exposing the children to high levels of stress or anxiety. Children were randomly assigned to one of two parallel talks that were identical in structure and length, except that each involved slightly different materials and different names for the key actors (Versions A and B) to provide some measure of the generalisability of our findings.²

Evidence gathering statements – 'Brief Interviews'. In empirical research, staged events are usually followed, somewhat later, by a full evidential investigative interview. However, in real-life, police officers typically question and collect initial 'statements' from witnesses immediately after the event. This initial questioning (referred to here as 'Brief Interviews') is critical because performance at this point may determine whether they will proceed to a full investigative interview.

Here, participants witnessed the event and, on the same day (as soon as possible after the event, which was usually seen in the morning), one of a pool of seven interviewers (pre- or post-doctoral research assistants) questioned each child individually. One child with autism failed to complete the Brief Interview and the mean for this group was substituted for their score for the predictive analyses. There was no effect of interviewer on Brief Interview total correct performance for children with autism, $F(5, 65) = 1.20$, $p = .32$, or for the TD children, $F(3, 197) = 2.07$, $p = .11$.

Interviewers followed a standard protocol that began with them asking the child: 'Tell me what you remember about what you just saw' (free recall). A series of follow-up prompts (all open-ended questions: who was there? what did they do? what did they look like? when did it happen? where did it happen?) could be used depending on what was said in response to the initial question (the total number of prompts given was totalled for each child). At the end of the interview, the children were asked if they remembered anything else; a prompt that could have been asked multiple times depending on whether the child recalled additional items of information in response to the 'anything else?' prompt (i.e. the prompt was repeatedly asked until the child could not offer further information). Overall, children with autism ($M = 12.11$, $SD = 5.97$, range 3–31) were given more prompts than TD children ($M = 9.37$, $SD = 3.85$, range 1–22). Hierarchical regression analysis controlling for age and full-scale IQ at Step 1, and including diagnostic group status as a dummy variable at Step 2, was carried out using total number of prompts during the Brief Interview as the dependent variable. The overall model was significant, $F(3, 267) = 21.88$, $p < .001$, accounting for 18.8% (adjusted) of the variance. Group was significant when entered at Step 2, F Change (1, 267) = 19.41,

$p < .001$, and accounted for 5.8% of the variance. This indicated that children in the autism group were given significantly more prompts (Beta group .26, $p < .001$). Age and IQ also had significant Beta values at Step 2 (age .27; IQ .25; $ps < .001$).

Each interview was audio-taped, transcribed, and coded for the total number of correct details recalled: e.g. 'The man (1) with the blonde hair (1), Alex (1), stole (1) the man (1) with the brown hair's (1) keys (1)' = 7 units of correct information. Incorrect items of information (details that were present but wrongly described) and confabulations (details that were not present) were scored using the same principles. Only unique utterances were coded (repeated information was ignored). Further coding was carried out to classify correct details by type (adapted from Memon, Milne, Holley, Bull, & Köhnken, 1997) relating to six key areas: people (descriptions of the men giving the talk, e.g. their names, clothing, appearance); setting (descriptions of the environment in which the event took place, or the time it happened); actions (information about what the men did, e.g. holding X, moving Y); conversations (verbatim accounts of what the men said to the children, e.g. 'Alex said "where's my phone?"'); objects (i.e. names or descriptions of the items the men had); and other information about the event that we classified as 'general' information (e.g. facts about Victorian times that the children were told during the talk, which were not recalled as verbatim conversation items, e.g. 'girls did needlework'). Ten percent of Brief Interview transcripts were double-coded, and Pearson product-moment correlation coefficients were calculated between the two raters for the total numbers of correct, incorrect, and confabulated items of information ($rs = .98$, $.88$, and $.88$, respectively, indicating high agreement).

Cognitive measures. An extensive range of cognitive measures (memory, language, attention, and intelligence) was administered to assess whether these variables related to Brief Interview performance (see Table 1).

Memory. Four of the eight core subtests from the Test of Memory and Learning 2 (TOMAL-2; Reynolds & Voress, 2007) were used to assess verbal and non-verbal memory. Verbal memory tasks included 'Memory for Stories', which assessed the child's ability to recall a series of short passages, and 'Paired Recall', which required the child to learn pairs of words, some already related (e.g. cold-hot) and others unrelated (e.g. girl-flag), over several trials (test-retest reliabilities $.79$ and $.78$, respectively). Non-verbal memory was assessed using 'Facial Memory', which required the recognition of series of previously viewed black and white pictures

of faces, and 'Visual Sequential Memory', which required the child to remember the order of a series of abstract visually presented figures (test-retest reliabilities .72 and .71, respectively). These subtests were chosen because they included both general memory skills and those that were relevant to witness skills. Suitable from five years of age, the subtests took around 25 minutes to administer.

Language. The British Picture Vocabulary Scale, third edition (BPVS-3; Dunn, Dunn, & Styles, 2009) is a well-established test of receptive (hearing) vocabulary for use with children aged 3–16 years (administration time 10–15 minutes). On this task, the experimenter names a word and the child selects (from one of four options) a picture that best represents the word. Two subtests of the Clinical Evaluation of Language Fundamentals, fourth edition (CELF-4 UK; Semel, Wiig, & Secord, 2006) were included: 'Recalling Sentences' assesses the ability to recall a sentence correctly and reflects grammatical understanding (test-retest reliability .90) and 'Formulated Sentences' assesses the child's ability to formulate complete, grammatically correct and meaningful sentences (of increasing length and complexity) about a picture, using specified words (test-retest reliability .86). The CELF-4 UK is reliable and widely used in speech and language therapy settings. Indeed, Recalling Sentences is a potential marker for language impairment (e.g. Conti-Ramsden, Botting, & Faragher, 2001). This test is suitable for use from the age of five and the total testing time (for both subtests) was around 15–20 minutes. Finally, two subtests of the Expressive Language Test 2 (ELT-2, Bowers, Huisinigh, LoGiudice, & Orman, 2010) were used: Sequencing (a test of narrative ability, test-retest reliability .79) and Grammar and Syntax (a test of grammatical morphology, test-retest reliability .83). The ELT-2 provided an indication of the child's ability to use expressive language to produce narratives (potentially relevant for eyewitness recall, which requires providing narratives in response to open-ended questions). It is suitable for children between the ages of 5 and 11 years and the two subtests took approximately 15 minutes to administer in total.

Attention. The Test of Everyday Attention for Children (TEA-Ch; Manly, Robertson, Anderson, & Nimmo-Smith, 1999) was used to assess a range of attention skills. Selective/focused attention was assessed with 'Sky Search', requiring the timed identification of target spaceships whilst controlling for motor speed (test-retest reliability .75). Sustained attention was assessed using 'Score!', which required children to listen for 'scoring' sounds as if they were keeping score on a computer game (percentage test-retest

agreement 76%). Sustained-divided attention was assessed with 'Sky Search Dual Task', a combination of the previous two tests designed to assess dual task decrements (test-retest reliability .81). These tasks took around 15 minutes to administer and are suitable for children of 6–16 years of age.

Intelligence. The second edition of the Wechsler Abbreviated Scale of Intelligence (WASI-II; Wechsler & Zhou, 2011) was used as a well-validated and reliable measure of intellectual ability. Full-scale IQ was estimated based on one subtest from the Verbal Comprehension Index ('Vocabulary') and one subtest from the Perceptual Reasoning Index ('Matrix Reasoning'). Suitable for use from six years of age, the two chosen WASI-II subtests have high split-half (.91 and .87, respectively) and test-retest reliability (.92 and .81, respectively), and (together) can be administered in approximately 15 minutes. As well as using the non-verbal IQ score in the predictive multiple regression analyses, the full-scale IQ score was used to establish suitability for the study and to control for overall intellectual ability when examining TD/autism group differences in Brief Interview performance.

A note on predictor variables. In the regression analyses (used to predict performance at Brief Interview), standardised scores from the above assessments were used. Standardised scores are often used when important decisions are made about children's abilities. They also provide an indication of children's abilities compared to other children of the same age. Age equivalent scores were not used in the present analyses, as these resulted in many children falling in the same age band and were also difficult to calculate accurately for children of low ability. Likewise, raw scores were not used because these are difficult to interpret (they are linked with age, rather than ability) and can have variable scaling across the different measures. It should, however, be recognised that children of different ages can have the same standardised score, but differ in both their competence and raw scores. This potentially reduces the power of these scores as predictors (as discussed later).

General procedure

The study was given full ethical approval at the University at which it was carried out; further, all children had informed parental consent, and gave their own written and oral assent to participate. Data from both samples were collected between April 2013 and January 2016. Children viewed the to-be-remembered event at school (or, occasionally, at home or at the University) and Brief Interviews were administered on

the same day. Cognitive testing took place by a team of post-doctoral researchers and was split over several sessions to fit in with school timetables/family needs, and to ensure the children remained engaged with the tasks.

Results

Sample characteristics and data screening

Table 1 includes mean scores (SDs), and any group differences, for children with and without autism on age and the cognitive variables. As might be expected, the autism group had lower scores than the TD group on all cognitive variables, and the variance in their scores was often larger.

Checks were carried out on normality, and log transformations of variables were used when these improved the distribution of scores. For all regressions reported, key statistical checks (Durbin–Watson, tolerance/VIF statistics, Cook's/Mahalanobis distances) were carried out to ascertain that no individual cases had undue influence on the regressions (Field, 2013); regressions involving interaction terms were excluded from these checks (Allison, 2012). Table 2 provides details of the correlations between the cognitive variables and age in the autism and TD groups.

Were there group differences in Brief Interview performance?

Mean scores on the Brief Interview are presented in Table 3. The total number of correct responses in the Brief Interview was higher in TD children ($M = 34.66$, $SD = 15.12$) than in children with autism ($M = 24.60$, $SD = 14.89$). To explore this group difference, whilst controlling for differences between the autism and TD groups in age and IQ, hierarchical multiple regression was used. In these analyses, we controlled for full-scale IQ to best reflect *overall* cognitive abilities, including both verbal and non-verbal skills. At Step 1, age and IQ were entered, followed at Step 2 by a dummy-coded group diagnostic status variable. The overall model was significant, $F(3, 268) = 37.55$, $p < .001$, accounting for 28.8% (adjusted) of the variance. Importantly, Step 2 was significant, indicating that the groups differed in the number of correct details recalled after controlling for age and IQ, $F \text{ Change } (1, 268) = 29.00$, $p < .001$ (7.6% of the variance accounted for at Step 2). All three predictors had significant Beta values at Step 2 (age .37; IQ .29; Group .30, all $ps < .001$). Despite the autism group recalling fewer items of information than the TD group, both groups were accurate in their recall: 88.9% of details recalled by the TD group ($SD = 8.4$) and 86.8% of the details recalled by the autism group ($SD = 11.6$) were correct (note that one child in the

autism group and one in the TD group scored zero; therefore, no proportion correct score could be calculated). A similar hierarchical regression (using arcsine transformed proportion scores, as recommended by Howell, 2013) indicated that although the full model was significant, $F(3, 266) = 4.69$, $p = .003$, accounting for 4.0% of the variance, there were no significant group differences in proportion correct (no significant R^2 change at Step 2: $F \text{ Change } (1, 266) = 2.25$, $p = .14$). Only age had a significant Beta value at Step 2 (.21, $p = .001$). There were few errors (i.e. incorrect or confabulated items of information) in either the TD or autism groups, although there was high variance in these data. After log transformations to improve error distributions, hierarchical regression revealed a significant overall model for incorrect items, $F(3, 268) = 4.18$, $p = .006$, as well as a significant group difference at Step 2 (the autism group made fewer errors than the TD group, $F \text{ Change } (1, 268) = 7.78$, $p = .006$, explaining 2.8% of the variance). Only group had a significant Beta value (.18, $p = .006$). The regression model was not significant for confabulations, $F(3, 268) = 0.43$, $p = .73$, and nor was there a significant change at Step 2, indicating no group difference in performance.

Breaking down the correct responses into six types of detail (people, setting, actions, conversations, objects, general – see lower portion of Table 3) and using the same hierarchical regressions (on log transformed data to improve data distributions) revealed significant overall models, and importantly, group differences at Step 2 for five types of detail: people, $F \text{ Change } (1, 268) = 12.53$, $p < .001$; setting, $F \text{ Change } (1, 268) = 18.51$, $p < .001$; actions, $F \text{ Change } (1, 268) = 8.33$, $p = .004$; objects, $F \text{ Change } (1, 268) = 15.41$, $p < .001$; and general, $F \text{ Change } (1, 268) = 48.13$, $p < .001$. All five full models were significant (total variance accounted for was between 10.6% and 34.7% – adjusted) and, in each case, children with autism recalled fewer details than TD children (see Table 1). Age was a significant predictor at Step 2 in all five models (Betas .22 – .37, all $ps < .001$); and IQ was a significant predictor in four models (Betas .16 – .34, all $ps < .01$, n.s. for setting details). Reporting of conversation details did not differ by group, $F \text{ Change } (1, 268) = .14$, $p = .71$, and the overall model was not significant, $F(3, 268) = 2.14$, $p = .10$.

We checked that the same results would be found for individually matched samples of children with and without autism. It was possible to match 54 children with autism (49 boys, 5 girls) closely to 54 TD children (34 boys, 20 girls) on age (\pm four months: autism $M = 106.0$, $SD = 16.0$; TD $M = 106.1$, $SD = 15.8$; $t(106) = .02$, $p = .98$) and full-scale IQ (\pm six points: autism $M = 101.1$, $SD = 15.6$; TD $M = 101.2$, $SD = 14.7$; $t(106) = .03$, $p = .98$). As before, the total

Table 2. Bivariate correlations between variables in the TD and autism groups.

	Non-verbal IQ	Age	TOMAL: Memory for Stories	TOMAL: Facial Memory	TOMAL: Visual Sequential Memory	TOMAL: Paired Recall	BPVS-3: Receptive vocabulary	CEL-F-4: Recalling Sentences	CEL-F-4: Formulated Sentences	ELT-2: Sequencing	ELT-2: Grammar and Syntax	TEA-Ch: Sky Search	TEA-Ch: Dual Task decrement
Non-verbal IQ	1	-.16	.27*	.34**	.40**	.50**	.40**	.40**	.40**	.42**	.47**	.29*	.32**
Age	.03	1	.16	-.03	.13	.07	-.05	.16	.22	-.01	.06	.17	.06
TOMAL: Memory for Stories	.20**	.10	1	.32**	.34**	.57**	.53**	.52**	.63**	.50**	.49**	.17	.32**
TOMAL: Facial Memory	.17*	.11	.08	1	.35**	.35**	.24*	.25*	.38**	.40**	.31*	.33**	.23
TOMAL: Visual Sequential Memory	.25**	-.03	.10	.29**	1	.26*	.28*	.33**	.27*	.28*	.26*	.18	.21
TOMAL: Paired Recall	.16*	-.17*	.31**	.17*	.27**	1	.53**	.57**	.55**	.54**	.53**	.28*	.42**
BPVS-3: Receptive Vocabulary	.35**	-.07	.54**	.10	.20**	.32**	1	.58**	.52**	.55**	.51**	.03	.41**
CEL-F-4: Recalling Sentences	.29**	-.006	.36**	.05	.21**	.28**	.56**	1	.77**	.57**	.62**	.17	.42**
CEL-F-4: Formulated Sentences	.31**	-.23**	.30**	.04	.14*	.31**	.42**	.58**	1	.60**	.67**	.22	.42**
ELT-2: Sequencing	.14*	.02	.22**	.14*	.12	.13	.30**	.32**	.29**	1	.62**	.16	.51**
ELT-2: Grammar and Syntax	.33**	-.20**	.35**	.03	.122	.33**	.48**	.49**	.46**	.40**	1	.18	.42**
TEA-Ch: Sky Search	.26**	.003	.03	.18*	.14	.03	.14	.09	.15*	.18*	.21**	.16	.50**
TEA-Ch: Sky Dual Task decrement	.19**	-.003	.06	-.03	.06	.10	.14*	.11	.08	.08	.13	.09	.10
TEA-Ch: Sky Dual Task decrement	.24**	-.01	.19**	.16*	.13	.12	.25**	.24**	.23**	.25**	.33**	.29**	.1

Note: The correlations for the autism group are in the top right segment of the table (in bold), and those for the TD group in the bottom left segment.

BPVS: British Picture Vocabulary Scale; CELF: Clinical Evaluation of Language Fundamentals; ELT: Expressive Language Test; TEA-Ch: Test of Everyday Attention for Children; TOMAL: Test of Memory and Learning.

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

Table 3. Mean (SD) scores on the Brief Interview for autism and TD groups.

Variables	Autism group (n = 70)	TD group (n = 201)
Brief Interview total correct	24.6 (14.9)	34.7 (15.1)
Brief Interview total incorrect	1.9 (1.9)	2.7 (2.0)
Brief Interview total confabulations	1.6 (2.4)	1.6 (2.8)
Brief Interview percentage accurate	86.8 (11.6)	88.9 (8.4)
Correct details – people	8.8 (5.5)	11.7 (6.1)
Correct details – setting	1.9 (1.3)	2.5 (1.2)
Correct details – actions	4.1 (3.7)	5.5 (4.4)
Correct details – conversations	1.3 (1.7)	1.3 (2.1)
Correct details – objects	3.0 (2.5)	4.4 (2.9)
Correct details – general	5.5 (5.9)	9.1 (5.2)

number of correct responses in the Brief Interview was significantly higher in the TD group ($M = 33.04$, $SD = 14.95$) compared to the autism group ($M = 23.55$, $SD = 14.49$), $t(106) = 3.35$, $p = .001$. The groups did not differ on the total number of incorrect items (autism $M = 2.0$, $SD = 1.9$; TD $M = 2.5$, $SD = 2.0$; $t(106) = 1.54$, $p = .13$), although there had been a small group difference on this measure in our original analysis (which had greater power). As before, there were no group differences for total confabulations (log transformed) (autism $M = 1.9$, $SD = 2.5$; TD $M = 1.7$, $SD = 3.0$; $t(106) = 1.06$, $p = .29$), or proportion correct (arcsine transformed) (autism $M = 84.8\%$, $SD = 11.9$; TD $M = 88.6\%$, $SD = 9.3$; $t(105) = 1.81$, $p = .07$). The results (and means) for types of details were also highly similar, the only difference being that the effect of group was no longer significant for action details. These findings provide reassurance that virtually the same results are found regardless of whether the full sample is analysed using regression (i.e. reflecting all children taking part in the study which should enhance transparency and provide greater power to detect effects), or smaller age and IQ matched subgroups are compared.

Group differences and choice of dependent variable

The preceding analyses indicated that, in the Brief Interview, there were group differences in the number of correct responses and errors, but no differences in the proportion of correct responses. From a practical viewpoint, the proportion of correct responses when recalling an event is useful when trying to predict the overall accuracy of what witnesses report. However, the groups were not significantly different on this measure and further inspection of these data revealed that nearly

one-third of children in both groups had more than 94% correct responses (30% of the TD group; 32% of the autism group) and approximately 10% of the children in each group were completely accurate (9% of the TD group; 13% of the autism group). Furthermore, around 10% of the children made no errors. Thus, most of these children's reports were accurate, with minimal errors. Consequently, we decided to use the number of correct responses (which was a more sensitive measure) as the dependant variable in the regression analyses as this would provide information about the ability of children to provide accurate eyewitness reports.

Predicting Brief Interview performance from age, non-verbal IQ, group diagnostic status, and the cognitive variables

To assess relationships between age, diagnostic group status, non-verbal IQ, the three cognitive domains (memory, language, and attention), and Brief Interview performance, data were analysed using hierarchical multiple regressions with separate regressions for memory, language, and attention. Based on the statistical checks, two cases were excluded from the analyses on language and attention (one from each group). The variables were entered in the order that reflects the ease of obtaining the information: Step 1 was age, Step 2 was diagnostic group status (TD/autism entered as a dummy variable), and Step 3 was non-verbal IQ. Non-verbal IQ, rather than full-scale IQ, was used in these analyses because non-verbal IQ was less likely to share variance with the five different measures of language that were being used as predictors. At Step 4, the assessment variables relevant to each cognitive domain were entered separately into each of three regressions (e.g. the four memory variables). At Step 5, to investigate whether there were group differences involving each variable entered at Step 4, dummy variables for each interaction term were entered (e.g. group \times each memory measure such as Memory for Stories); standardised coefficients are not reported for Step 5 as these can be misleading (Preacher, 2003).

The first three steps were common to all three regressions and the entry of each variable produced a significant R^2 change (age, $R^2 = .09$, F Change = 26.748; $p < .001$; group, $R^2 = .14$, F Change = 48.36, $p < .001$; non-verbal IQ, $R^2 = .03$, F Change = 9.59, $p = .002$). Age and group diagnostic status had similar standardised Beta coefficients, with non-verbal IQ being less important (standardised Beta coefficients at Step 3, age .40, $p < .001$; group .36, $p < .001$; non-verbal IQ .17, $p = .002$). For the three regressions concerning each domain, the standardised coefficients indicated that age and group remained significant predictors at

Steps 4 and 5 (see below). However, non-verbal IQ was only a significant predictor at Steps 4 and 5 for the attention domain (see below).

In relation to the domain of memory, at Step 4, the four TOMAL-2 assessments were entered (Facial Memory, Memory for Stories, Paired Recall and Visual Sequential Memory) and there was a significant R^2 change, $R^2 = .09$, F Change = 8.63; $p < .001$. The standardised Beta coefficients identified Facial Memory and Memory for Stories as significant predictors (the standardised Beta coefficients for all variables were: age .36, $p < .001$; group .22, $p < .001$; non-verbal IQ .07; Facial Memory .24, $p < .001$; Memory for Stories .16, $p = .007$; Paired Recall .06, and Visual Sequential Memory .00; only p values $< .10$ reported). At Step 5, the interaction terms produced no further significant R^2 change, Facial Memory and Memory for Stories remained significant predictors, and none of the interaction terms made a significant contribution to the regression equation.

For language (BPVS scores for the TD group were log transformed to improve data distribution), the entry of these variables (BPVS-3 Receptive Vocabulary, CELF-4 Recalling Sentences, CELF-4 Formulated Sentences, ELT-2 Sequencing, and ELT-2 Grammar and Syntax) at Step 4 resulted in a significant R^2 change, $R^2 = .04$, F Change = 2.922; $p = .01$. However, none of the standardised Beta coefficients of the variables entered at Step 4 was significant (the standardised Beta coefficients were: age .41, $p < .001$; group .26, $p < .001$; non-verbal IQ .08; Receptive Vocabulary .12, $p = .08$; Recalling Sentences .02; Formulated Sentences .14; $p = .09$; Sequencing .06; and Grammar and Syntax $-.07$; only p values $< .10$ reported). The entry of the interaction terms at Step 5 did not result in a significant R^2 change, suggesting no overall significant interactions between group and the assessment variables. However, two of the interaction terms showed significant effects, these were group \times Recalling Sentences ($p = .04$) and group \times BPVS-3 Receptive Vocabulary ($p = .02$).

In the case of attention (selective/focussed attention, TEA-Ch Sky Search; sustained attention, TEA-Ch Score!; and sustained-divided attention, TEA-Ch Dual Task decrement), the R^2 change values at Steps 4 and 5 were non-significant. At Step 4, none of the coefficients for the three attention variables were significant (the standardised Beta coefficients for all variables were: age .40, $p < .001$; group .36, $p < .001$; non-verbal IQ .17, $p = .00$; TEA-Ch Sky Search $-.06$; TEA-Ch Score! $-.03$, and TEA-Ch Dual Task decrement .07; only p values $< .10$ reported). At Step 5, the overall R^2 change was non-significant; however, one interaction of group \times Score! was significant ($p = .03$).³

It is noted that although results from standardised tests are often used when interpreting an individual's

abilities relative to others, raw scores provide a better indication of developmental level. To examine whether the effects of age were still present when raw scores rather than standardised scores were entered into the regressions, a further set of analyses were conducted. These analyses produced very similar findings to those using standardised scores, except that, in general, age was a less important predictor and group was a more important predictor. This suggested that the inclusion of raw scores reduced the predictive power of age (see Appendix 1 for details).

Discussion

We evaluated group differences in eyewitness memory in 6–11-year-old children with and without autism using a brief (immediate) interview about a witnessed event. Following this, the prediction of recall from age, diagnostic group status, and non-verbal IQ was considered; we further assessed whether three important areas of cognition (memory, language, attention) added to the accuracy of the predictions.

In line with previous research, children with autism recalled fewer items of correct information about the witnessed event than their TD peers (with age and full-scale IQ controlled) (Bruck et al., 2007; Mattison et al., 2016, 2015; McCrory et al., 2007). Breaking correct recall down into the types of details remembered (people, setting, actions, objects, general, and conversation) indicated that group differences were present for all types of information except conversation (in fact, few relevant details were recalled by either group about conversations). These results were almost identical when sub-samples of children with and without autism (54 in each group) closely matched for age and IQ were compared, adding confidence to the findings.

Children with autism also needed a higher number of prompts during their recall, in line with existing research (e.g. Goddard et al., 2014), which suggests that their narratives may have been even more impoverished had the additional prompts not been provided. In terms of errors, the number of incorrect items was lower in the children with autism in the full sample (although this difference did not reach significance for the matched sample), and there were no group differences in numbers of confabulations. Importantly, both groups recalled a high proportion of information accurately (close to 90%), and did not differ significantly on this measure. These findings accord well with previous reports of group differences regarding the amount of information recalled by children with and without autism, alongside high absolute levels of performance (e.g. Bruck et al., 2007 – 84% accuracy; McCrory et al.,

2007 – over 90% accuracy). These results should reassure criminal justice professionals that children on the autism spectrum (who do not have intellectual disabilities) are able to provide eyewitness evidence that is as accurate as that of TD peers in response to interviews that emphasise free recall narratives and open questions (e.g. Bull, 2010). However, the findings also emphasise that children on the autism spectrum may need more support (e.g. more open-ended prompts or possibly more comprehensive investigative interviews, see Henry et al., 2017) to provide their best evidence.

Regression analyses were conducted to identify important predictors of Brief Interview performance. The first three steps involved the separate entry of age, diagnostic group status, and non-verbal IQ. As expected, all three variables were significant predictors of Brief Interview performance and, together, accounted for about one-quarter of the variance in these scores. Inspection of the standardised Beta coefficients at Step 3 indicated that age and group were strong predictors of performance, with non-verbal IQ being less important. Age remained an important predictor at Steps 4 and 5 in each of the regressions (i.e. once assessments from each cognitive domain had been entered in Step 4, along with the relevant interaction terms in Step 5). These results are consistent with previous research showing strong age effects for witness ability in TD children (Brown & Lamb, 2015; Burgwyn-Bailes et al., 2001; Chae et al., 2016; La Rooy et al., 2011; Odegard & Toglia, 2013). They also contribute novel evidence that age is a strong predictor of recall for children with autism.

Diagnostic group status also remained a significant predictor in Steps 4 and 5 of each regression, reflecting the fact that the autism group had almost one-third fewer correct responses than the TD group. Importantly, this does not reflect poorer accuracy of the information recalled, but criminal justice professionals should be aware that children with autism may recall *fewer* items of information than their peers. Although non-verbal IQ was a significant independent predictor at Step 3, it was less important than age and diagnostic group status (as assessed by the standardised Beta coefficients). Furthermore, at Steps 4 and 5, non-verbal IQ was only a significant predictor for the analysis involving attention, and not for those involving memory or language, confirming that this variable was not an important predictor. This is consistent with previous research showing few strong or reliable relationships between IQ and eyewitness memory in TD children with IQs in the average range (Elischberger & Roebbers, 2001; Geddie et al., 2000; Henry & Gudjonsson, 2007; Roebbers & Schneider, 2001), and additionally confirms that this finding is true for autistic children.

In relation to the three cognitive domains, memory was the most important assessment in improving the explanatory power of the regression models. Both Facial Memory and Memory for Stories were significant independent predictors as shown by their standardised Beta coefficients. These were the only two memory variables to remain significant predictors at Step 5 when the interaction variables were entered, and the absence of significant coefficients for the interaction terms in the analysis of memory suggests that these predictive relationships did not vary by group.

The finding that Facial Memory predicts witness recall is novel because previous studies with TD children have used exclusively verbal memory measures (Baker-Ward et al., 1993), or included abstract measures of non-verbal memory (Henry & Gudjonsson, 2003). In children with autism, no relationships between overall measures of verbal or visual memory and autobiographical memory were found (Goddard et al., 2014). However, the reasons for the relationship between Facial Memory and witness recall are not obvious. There was one element of the Brief Interview that involved describing the ‘actors’. Yet there were many other salient details about the event that had to be recalled, and most children said little about the faces of the actors beyond their hairstyle and hair colour. It is possible that Facial Memory assesses memory abilities related to social stimuli in general and, as with most eyewitness testimony, the to-be-recalled event was situated in a social context. This might be especially important for children with autism, who are known to have difficulties processing social and facial information (Williams, Goldstein, & Minshew, 2005). Another possibility is that memory for faces involves encoding and remembering reasonably detailed information about social stimuli, and that this capacity is useful for being able to encode and recall witnessed events. Further research is needed to investigate this intriguing finding, especially as Facial Memory was a better predictor of interview performance than other variables such as Memory for Stories (which might have been expected to be a stronger predictor).

Nevertheless, the fact that Memory for Stories was also a significant predictor of Brief Interview at Steps 4 and 5 supports previous research on TD children showing relationships (albeit sometimes inconsistent at different age levels) between witness recall and verbal memory (Baker-Ward et al., 1993; Henry & Gudjonsson, 2003). Memory for Stories has important similarities to the experience of recalling a witnessed event and remains a useful assessment tool. Although Goddard et al. (2014) found no relationships between verbal (or visual) memory and autobiographical memory in children with autism, they did not separate out more ‘ecologically relevant’ and more ‘abstract’

memory measures (e.g. they combined memory for stories with paired recall – and combined memory for faces with recalling spatial locations). It is possible that this accounted for the lack of significant relationships.

In the case of the language variables, when these were introduced at Step 4, there was a significant increase in the R^2 value. Contrary to predictions, given the role of language in the development of memory for personally experienced events (Fivush & Nelson, 2004), none of the language variables were significant independent predictors (as indicated by the standardised Beta coefficients). There was some weak evidence suggesting language interacted with diagnostic group status. At Step 5, with the entry of the interaction terms, there were two variables that showed a significant interaction: BPVS-3 Receptive Vocabulary and CELF-4 Recalling Sentences. Additional, exploratory regression analyses on a restricted set of variables indicated that Receptive Vocabulary was more strongly related to Brief Interview performance in the autism group than the TD group, perhaps reflecting the importance of word knowledge when reporting information in children on the autism spectrum. In contrast, Recalling Sentences, which assesses the child's ability to integrate information from verbal short-term memory with semantic and syntactic long-term memory knowledge, was (marginally) more strongly related to Brief Interview performance in the TD than the autism group. This perhaps indicates the importance of recall processes for the TD group. Previous investigations have revealed relationships between language ability measures and witness recall in TD children (Burgwyn-Bailes et al., 2001; Chae & Ceci, 2005; Chae et al., 2014, 2016; Henry & Gudjonsson, 2007), and have hinted at similar relationships in children with autism (albeit using verbal fluency measures: Goddard et al., 2014; McCrory et al., 2007). Nevertheless, our findings can only tentatively suggest that different language processes may be important to autism and TD groups.

Although Chae et al. (2016) found links between witness recall and a questionnaire measure of attention in pre-schoolers, the current behavioural assessments of attention entered at Step 4 were not significant predictors of Brief Interview performance. It may be that children had sufficient attentional resources to process information about the event, i.e. even children with low attention scores were not appreciably disadvantaged. At Step 5, there was some evidence of group differences for TEA-Ch Score!, with the relationship between this variable and Brief Interview performance being marginally stronger in the autism, relative to the TD, group. However, this exploratory finding requires further research.

Across the three regressions, there were few significant interaction terms at Step 5, and in none of the

regressions was the overall change in variance significantly increased by adding the interaction terms. This suggests that the cognitive mechanisms underlying eyewitness recall are likely to be similar across the TD and autism groups (although further research using large samples closely matched for age and IQ is needed to confirm these findings). Furthermore, the analyses failed to identify specific cognitive predictors of eyewitness memory that were superior to age. This suggests that those in the criminal justice system can rely on age as a general indicator of performance in children with and without autism, provided they have adequate levels of intellectual functioning. This has the benefit of being straightforward information to which all relevant criminal justice professionals have access. We have also demonstrated that knowing a child's diagnostic status (i.e. whether they have autism) and obtaining further information that may require expert assessment (namely, memory for faces and stories which could be assessed by a Registered Intermediary, a communication specialist that assists vulnerable witnesses to give best evidence: Cooper & Wurtzel, 2013; Plotnikoff & Woolfson, 2015) can significantly improve the accuracy of prediction. Although non-verbal IQ was a significant predictor of Brief Interview recall in the absence of the memory measures, it became non-significant once they were entered, suggesting that direct measures of relevant memory skills are more useful predictors. Finally, it was expected that assessments of cognitive abilities that could underlie eyewitness recall, such as memory, would be better predictors than age; yet our findings did not support this expectation. This suggests that age, because it is related to the general abilities of the children, is a better predictor than most standardised scores of memory, language, and attention, which give important information about the relative ability of a child compared to their age group, but do not give an indication of the absolute ability of the child.

It is important to note that the constraints involved in conducting a staged event for an experimental study meant that the children were questioned about a non-traumatic (and very mild) crime event, so generalising these findings to real events must be done with caution. Further, our interview was not a full evidential investigative interview, but rather a brief evidence gathering statement (administered on the same day as the witnessed event). This is akin to an interview used by police officers to help determine whether the witness can provide enough evidence to warrant a full evidential interview. It is also important to acknowledge that although the design of the Brief Interview was carefully constructed to replicate experiences related to eyewitness testimony in children, because of ethical and practical considerations, it was not possible to fully replicate the actual experiences of children involved in these

events. Nevertheless, the current study has confirmed previously established group differences in recall between children with and without autism. In addition, it represents the first attempt to examine cognitive predictors of witness performance in samples of children with and without autism, finding many similarities across autism and TD groups. Further research could explore other potential domains to identify variables which could increase the prediction of eyewitness performance, such as anxiety or suggestibility (e.g. Bettenay, Ridley, Henry, & Crane, 2015). (Note that measures of suggestibility were not relevant for the current study as we did not include leading or misleading questions, nor did we use a misinformation paradigm.)

Summary

The findings indicated that children with autism recalled fewer correct details about a witnessed event than TD children, although the accuracy of the information they provided was just as high. In terms of predictive relationships between witness recall and the variables assessed here (age, diagnostic group status, non-verbal IQ, memory, language, and attention), there were clear commonalities across the groups. Age was the most important predictor of Brief Interview performance in child witnesses, both with and without autism. In many ways, this should provide reassurance that this simple metric, readily available to criminal justice professionals, is a useful predictor of eyewitness recall alongside diagnostic group status. Facial Memory and Memory for Stories were also important predictors, emphasising the value of standardised measures of witness-related skills; albeit measures that criminal justice professionals may not be able to access without specialist assistance.

Acknowledgements

The authors would like to express their thanks to those who helped with testing (Richard Batty, Debbie Collins and Genevieve Waterhouse), assisted with data coding (Frances Beddow and Hannah Webb), offered specialist police advice (DC Mark Crane), and helped to script the event (Brian Protheroe). The authors also express their heartfelt thanks to the schools, teachers, parents and children who kindly assisted with the research.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this

article: This work was supported by the Economic and Social Research Council (grant number: ES/J020893/2). Data for this project can be accessed from the UK Data Service ReShare record 852471.

Notes

1. Around two-thirds of the children watched the event live (autism = 22, TD = 167) and the remainder watched it on video (autism = 49, TD = 34). Hierarchical regression analyses controlling for age and full-scale IQ (which were not matched across live/video) at Step 1, and including live versus video presentation as a dummy variable at Step 2, were carried out (for the autism and TD groups separately) using Brief Interview total correct details as the dependent variable. There were no significant differences between presentation modes for either group on Brief Interview total correct details. For the autism group, the overall model was significant, $F(3, 67) = 12.82$, $p < .001$, accounting for 33.6% (adjusted) of the variance. Importantly, Step 2 was not significant, indicating that the groups did not differ in number of correct details recalled depending on video versus live performance. Age and IQ had significant Beta values at Step 2 (age .33; IQ .51; $ps < .001$) but the Beta value for live versus video was not significant (.09, $p = .34$). For the TD group, the overall model was significant, $F(3, 197) = 18.03$, $p < .001$, accounting for 20.3% (adjusted) of the variance. Importantly, Step 2 was not significant, indicating that the groups did not differ in number of correct details recalled depending on video versus live performance. Age and IQ had significant Beta values at Step 2 (age .40; IQ .20; $ps < .001$) but the Beta value for live versus video was not significant ($-.002$, $p = .97$). Data from the two presentation modes were, therefore, combined.
2. A total of 138 children watched Version A (autism = 36, TD = 102) and 134 watched Version B (autism = 35, TD = 99). Hierarchical regression analyses controlling for age and full-scale IQ (which were not matched across Versions A and B) at Step 1, and including Version as a dummy variable at Step 2, were carried out (for autism and TD groups separately) using Brief Interview total correct details as the dependent variable. There were no significant differences between presentation versions for either group on Brief Interview total correct details. For the autism group, the overall model was significant, $F(3, 67) = 13.43$, $p < .001$, accounting for 34.8% (adjusted) of the variance. Importantly, Step 2 was not significant, indicating that the groups did not differ in number of correct details recalled depending on performance version. Age and IQ had significant Beta values at Step 2 (age .38; IQ .49; $ps < .001$) but the Beta value for performance version was not significant (.14, $p = .15$). For the TD group, the overall model was significant, $F(3, 197) = 18.45$, $p < .001$, accounting for 20.7% (adjusted) of the variance. Importantly, Step 2 was not significant, indicating that the groups did not differ in the number of correct details recalled depending on performance version. Age and IQ had significant Beta values at Step 2 (age .40; IQ .20; $ps < .001$) but the Beta value for performance version was

not significant (.06, $p = .32$). Data from the two performance versions were, therefore, combined.

3. To determine the nature of the significant effects with the interaction variables, the analyses were run separately for each group with the entry restricted to those variables that were significant at Step 4 (to reduce the number of variables entered in the analysis). For the autism group, BPVS-3 Receptive Vocabulary was a significant predictor (standardised Beta coefficient .44, $p = .001$), but it was a non-significant predictor for the TD group. For the autism group, Recalling Sentences was a non-significant predictor, but in the TD group, it was a marginally significant predictor (standardised Beta coefficient .14, $p = .07$). In the case of Score!, it was a marginally significant predictor for the autism group (standardised Beta coefficient .20, $p = .09$), and a non-significant predictor in the TD group.

References

- Allison, P. J. (2012). *When can you safely ignore multicollinearity?* Retrieved from <https://statisticalhorizons.com/multicollinearity>
- Baker-Ward, L., Gordon, B. N., Ornstein, P. A., Larus, D. M., & Clubb, P. A. (1993). Young children's long-term retention of a paediatric examination. *Child Development*, 64, 1519–1533. doi:10.2307/1131550
- Bettenay, C., Ridley, A., Henry, L. A., & Crane, L. (2015). Changed responses under cross-examination: The role of anxiety and individual differences in child witnesses. *Applied Cognitive Psychology*, 29, 485–449. doi:10.1002/acp.3125
- Boucher, J. (2012). Research review: Structural language in autistic spectrum disorder – Characteristics and cause. *Journal of Child Psychology and Psychiatry*, 53(3), 219–233. doi:10.1111/j.1469-7610.2011.x
- Bowers, L., Huisinigh, R., LoGiudice, C., & Orman, J. (2010). *The expressive language test 2*. East Moline, IL: LinguiSystems.
- Brown, D. A., & Lamb, M. E. (2015). Can children be useful witnesses? It depends how they are questioned. *Child Development Perspectives*, 9, 250–255. doi:10.1111/cdep.12142
- Bruck, M., London, K., Landa, R., & Goodman, J. (2007). Autobiographical memory and suggestibility in children with autism spectrum disorder. *Development and Psychopathology*, 19(1), 73–95. doi:10.1017/S0954579407070058
- Bull, R. (2010). The investigative interviewing of children and other vulnerable witnesses: Psychological research and working/professional practice. *Legal and Criminological Psychology*, 15, 5–23. doi:10.1348/014466509X440160
- Burgwyn-Bailes, E., Baker-Ward, L., Gordon, B. N., & Ornstein, P. A. (2001). Children's memory for emergency medical treatment after one year: The impact of individual difference variables on recall and suggestibility. *Applied Cognitive Psychology*, 15, S25–S48. doi:10.1002/acp.833
- Chae, Y., & Ceci, S. J. (2005). Individual differences in children's recall and suggestibility: The effect of intelligence, temperament, and self-perceptions. *Applied Cognitive Psychology*, 19, 383–407. doi:10.1002/acp.1094
- Chae, Y., Kulkofsky, S., Debaran, F., Wang, Q., & Hart, S. L. (2014). Low-SES children's eyewitness memory: The effects of verbal labels and vocabulary skills. *Behavioral Sciences and the Law*, 32, 732–745. doi:10.1002/bsl.2145
- Chae, Y., Kulkofsky, S., Debaran, F., Wang, Q., & Hart, S. L. (2016). Low-SES preschool children's eyewitness memory: The role of narrative skill. *Behavioral Sciences and the Law*, 34, 55–73. doi:10.1002/bsl.2242
- Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment. *Journal of Child Psychology and Psychiatry*, 42(6), 741–748. doi:10.1111/1469-7610.00770
- Cooper, P., & Wurtzel, D. (2013). Better the second time around? Department of Justice Registered Intermediaries Schemes and lessons from England and Wales. *NILQ*, 65(1), 39–61.
- Dawson, M., Soulières, M., Gernsbacher, M. A., & Mottron, L. (2007). The level and nature of autistic intelligence. *Psychological Science*, 18(8), 657–662. doi:10.1111/j.1467-9280.2007.01954.x
- Dunn, L. M., Dunn, D. M., & Styles, B. (2009). *British picture vocabulary scale* (3rd ed.). London, UK: GL Assessment.
- Elischberger, H. B., & Roebbers, C. M. (2001). Improving young children's free narratives about and observed event: The effects of nonspecific verbal prompts. *International Journal of Behavioral Development*, 25, 160–166. doi:10.1080/01650250042000203
- Field, A. P. (2013). *Discovering statistics using IBM SPSS Statistics: And sex and drugs and rock 'n' roll* (4th ed.). London: Sage publications.
- Fivush, R., & Nelson, K. (2004). Culture and language in the emergence of autobiographical memory. *Psychological Science*, 15(9), 573–577. doi:10.1111/j.0956-7976.2004.00722.x
- Geddie, L., Fradin, S., & Beer, J. (2000). Child characteristics which impact accuracy of recall and suggestibility in preschoolers: Is age the best predictor? *Child Abuse and Neglect*, 24, 223–235. doi:10.1016/S0145-2134(99)00133-7
- Goddard, L., Dritschel, B., Robinson, S., & Howlin, P. (2014). Development of autobiographical memory in children with an autism spectrum disorder: Deficits, gains and predictors of performance. *Development and Psychopathology*, 26(1), 215–228. doi:10.1017/S0954579413000904
- Goddard, L., Howlin, P., Dritschel, B., & Patel, T. (2007). Autobiographical memory and social problem-solving in young adults with ASD. *Journal of Autism and Developmental Disorders*, 37, 291–300.
- Henry, L. A., Crane, L., Nash, G., Hobson, Z., Kirke-Smith, M., & Wilcock, R. (2017). Verbal, visual, and intermediary support for child witnesses with autism during investigative interviews. *Journal of Autism and Developmental Disorders*, 47(8), 2348–2362. doi:10.1007/s10803-017-3142-0
- Henry, L. A., & Gudjonsson, G. H. (2003). Eyewitness memory, suggestibility and repeated recall sessions in children with mild and moderate intellectual disabilities. *Law and Human Behavior*, 27, 481–505. doi:10.1023/A:1025434022699
- Henry, L. A., & Gudjonsson, G. H. (2007). Individual and developmental differences in eyewitness memory and

- suggestibility in children with intellectual disabilities. *Applied Cognitive Psychology*, 21, 361–381. doi:10.1002/acp.1280
- Henry, L. A., Messer, D. J., & Nash, G. (2015). Verbal fluency and executive functioning in children with language difficulties. *Learning and Instruction*, 39, 137–147. doi:10.1016/j.learninstruc.2015.06.001
- Howell, D. C. (2013). *Statistical methods for psychology* (8th ed.). Wadsworth: Belmont, CA.
- Kebbell, M., & Milne, R. (1998). Police officers' perceptions of eyewitness factors in forensic investigations. *Journal of Social Psychology*, 138, 323–330. doi:10.1080/00224549809600384
- Lamb, M. E., Malloy, L. C., & La Rooy, D. J. (2011). Setting realistic expectations: Developmental characteristics, capacities, and limitations. In M. E. Lamb, D. J. La Rooy, L. C. Malloy & C. Katz (Eds.), *Children's testimony: A handbook of psychological research and forensic practice* (pp. 15–48). West Sussex, UK: Wiley-Blackwell.
- La Rooy, D. J., Malloy, L. C., & Lamb, M. E. (2011). The development of memory in childhood. In M. E. Lamb, D. J. La Rooy, L. C. Malloy & C. Katz (Eds.), *Children's testimony: A handbook of psychological research and forensic practice* (pp. 49–68). West Sussex, UK: Wiley-Blackwell.
- London, K., Henry, L. A., Conradt, T., & Corser, R. (2013). Suggestibility and individual differences in typically developing and intellectually disabled children. In A. M. Ridley, F. Gabbert & D. J. La Rooy (Eds.), *Suggestibility in legal contexts: Psychological research and forensic implications* (pp. 129–148). Oxford, UK: John Wiley & Sons, Ltd.
- Manly, T., Robertson, I. H., Anderson, V., & Nimmo-Smith, I. (1999). *The test of everyday attention for children*. London, UK: Pearson.
- Mattison, M., Dando, C. J., & Ormerod, T. (2016). Drawing the answers: Sketching to support free and probed recall by child witnesses and victims with autism spectrum disorder. *Autism*. Epub ahead of print 14 November 2016. doi:10.1177/1362361316669088
- Mattison, M. L. A., Dando, C. J., & Ormerod, T. (2015). Sketching to remember: Episodic free recall task support for child witnesses and victims with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 45(6), 1751–1765. doi: 10.1007/s10803-014-2335-z
- McCrary, E., Henry, L. A., & Happé, F. (2007). Eye-witness memory and suggestibility in children with Asperger syndrome. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 48(5), 482–489. doi:10.1111/j.1469-7610.2006.01715.x
- Memon, A., Milne, R., Holley, A., Bull, R., & Köhnken, G. (1997). Eyewitness performance in cognitive and structured interviews. *Memory*, 5, 639–656. doi:10.1080/741941481
- Odegard, T., & Togli, M. P. (2013). Children as eyewitnesses: Historical background, and factors affecting children's eyewitness testimony. In R. E. Holliday & T. A. Marche (Eds.), *Child forensic psychology*. London, UK: Palgrave MacMillan.
- Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry*, 37(1), 51–87. doi:10.1111/j.1469-7610.1996.tb01380.x
- Plotnikoff, J., & Woolfson, R. (2015). *Intermediaries in the criminal justice system: Improving communication for vulnerable witnesses and defendants*. Bristol, UK: Policy Press.
- Preacher, K. J. (2003). *A primer on interaction effects in multiple linear regression*. Retrieved from <http://quantpsy.org/interact/interactions.htm>
- Reynolds, C., & Voress, J. K. (2007). *Test of memory and learning* (2nd ed.). Austin, TX: Pro-Ed.
- Roebers, C. M., & Schneider, W. (2001). Individual differences in children's eyewitness recall: The influence of intelligence and shyness. *Applied Developmental Science*, 5, 9–20. doi:10.1207/S1532480XADS0501_2
- Rutter, M., Bailey, A., & Lord, C. (2003). *Social Communication Questionnaire (SCQ)*. Los Angeles: Western Psychological Services.
- Semel, E., Wiig, E. H., & Secord, W. A. (2006). *Clinical evaluation of language fundamentals* (4th UK Ed.). London, UK: Harcourt Assessment.
- Smith-Spark, J. H., Henry, L. A., Messer, D. J., & Ziecik, A. P. (2017). Verbal and non-verbal fluency in adults with developmental dyslexia: Phonological processing and executive control problems? *Dyslexia*. doi:10.1002/dys.1558. [Epub ahead of print]
- Taylor, L. J., Maybery, M. T., & Whitehouse, A. J. O. (2014). Do autism spectrum disorders and specific language impairment have a shared aetiology? In J. Arcuili & J. Brock (Eds.), *Communication in autism: Trends in language acquisition research* (pp. 75–102). Amsterdam: John Benjamins Publishing Company.
- van der Meer, J. M. J., Oerlemans, A. M., van Steijn, D. J., Lappenschaar, M. G. A., de Sonnevill, L. M. J., Buitelaar, J. K., & Rommelse, N. N. J. (2012). Are autism spectrum disorder and attention-deficit/hyperactivity disorder different manifestations of one overarching disorder? Cognitive and symptom evidence from a clinical and population-based sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(11), 1160–1172. doi:10.1016/j.jaac.2012.08.024
- Wechsler, D., & Zhou, X. (2011). *Wechsler abbreviated scale of intelligence* (2nd ed.). Bloomington, MN: Pearson.
- Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, 7(2), 45–75. doi:10.1111/j.1529-1006.2006.0027.x
- Williams, D., Botting, N., & Boucher, J. (2008). Language in autism and specific language impairment: Where are the links? *Psychological Bulletin*, 134(6), 944–963. doi:10.1037/a0013743
- Williams, D. L., Goldstein, G., & Minshew, N. J. (2005). Impaired memory for faces and social scenes in autism: Clinical implications of memory dysfunction. *Archives of Clinical Neuropsychology*, 20, 1–15. doi:10.1016/j.acn.2002.08.001.

Appendix I

Additional analyses examined whether the effects of age were still present when raw scores rather than standardised scores were entered into the regressions. At Step 3, the standardised Beta for age was .30, $p < .001$; for group was .36, $p < .001$; and for non-verbal IQ raw score was .19, $p = .002$. At Step 4, some of the variables from the three cognitive domains were more important predictors than in the analyses of standardised scores, and as at Step 3, group was a more important predictor than age. Most notably at Step 4, Facial Memory had the highest standardised Beta coefficient (Facial Memory = .26, $p < .001$; group = .26, $p < .001$; Memory for Stories = .16, $p = .010$; age = .12, $p = .059$; one TD case was excluded on the basis of Mahalanobis distance). For language, at Step 4, the standardised Beta coefficient for Formulated Sentences was significant (Beta = .22, $p = .030$) as were group and age (group = .24, $p < .001$; age = .16, $p = .028$; one TD case was excluded because of Mahalanobis distance).

Analyses were not carried out on the attention variables as a large number of cases had a high Mahalanobis distance.

The regressions using the raw scores, in general, provided a very similar picture to that provided by the standardised scores in terms of the significant predictors of Brief Interview performance. However, the findings involving raw scores indicated that group was slightly better than age as a predictor. This suggests that raw scores of non-verbal IQ and the three cognitive domains shared variance with age, and their entry reduced the predictive power of age. In particular, at Step 4, Facial Memory had a slightly higher Beta coefficient than age, indicating its importance as a predictor variable. This is likely to be because raw scores provide a measure which corresponds to the ability of a child, whereas standardised scores provide a measure of the ability relative to a child's age group. Thus, in some cases, knowing a child's abilities, and in particular their memory abilities, could prove useful.