Brief Report. Eyewitness identification in child witnesses on the autism spectrum

Rachel Wilcock[a], Laura Crane[b,c], Zoe Hobson[b], Gilly Nash[b], Mimi Kirke-Smith[b] & Lucy A. Henry[b]

[a] Department of Psychology, University of Winchester, Sparkford Road, Winchester, S022 4NR, UK

[b] Division of Language and Communication Science, City, University of London, London, EC1V 0HB

[c] Present address for Laura Crane: Centre for Research in Autism and Education (CRAE), UCL Institute of Education, London, WC1H 0NU, UK

Corresponding author:

Dr Rachel Wilcock

Department of Psychology

University of Winchester

Sparkford Road

Winchester

S022 4NR, UK.

E-mail: Rachel.Wilcock@winchester.ac.uk

Abstract

Background. Although there is increasing interest in the capabilities of children with autism at different stages of the criminal justice process, there is little research into how well this group perform when asked to identify perpetrators from identification lineups. This is despite theoretical and empirical literature suggesting that autistic children experience face recognition memory difficulties.

Method. As part of a broader study into eyewitness memory skills, 50 children with autism and 162 children with typical development (TD) (all with IQs > 69) watched a mock crime event (either live or on a video) involving two male perpetrators. One week later, their eyewitness identification skills were compared, with children asked to identify the perpetrators from two ecologically valid video lineups. The children were also assessed on a standardised face memory task.

Results. When asked to identify perpetrators in the video lineups, in many respects the autistic children performed at an equivalent level to the TD children. This was despite the TD children outperforming the autistic children on the standardized face memory task.

Conclusions. These preliminary findings suggest that group differences between autistic and TD children may not always emerge on an ecologically valid, real world eyewitness identification lineup task, despite autistic children showing poorer performance on a standardized face memory task. However, as identification performance in both groups was low, it remains important for future research to identify how to scaffold eyewitness identification performance in both children with and without an autism diagnosis.

Keywords. autism; eyewitness memory; identification lineup; face memory; child witnesses.

Brief Report. Eyewitness identification in child witnesses on the autism spectrum

Research exploring the eyewitness capabilities of children with autism[1] (see Maras & Bowler, 2014, for a review of research on this topic in adults) has tended to focus on performance at interview, in which children are asked to provide narrative accounts ('free recall') and provide answers to questions ('cued recall') about witnessed events. This work has demonstrated that whilst autistic children often recall less information overall than children with typical development (TD), this information is usually just as accurate (e.g., Almeida, Lamb, & Weisblatt, 2018; Henry, Crane, Nash, Hobson, Kirke-Smith, & Wilcock, 2017a; Henry, Messer, Wilcock, Nash, Kirke-Smith, Hobson, & Crane, 2017b; Mattison, Dando, & Ormerod, 2015; 2018 [true for cued but not free recall]). Yet there has been limited research examining the performance of autistic child witnesses during other stages of the justice process. Eyewitness identification requires witnesses to identify a perpetrator from a lineup (if the perpetrator is present in the lineup), or to reject a lineup if the perpetrator is absent (e.g., if the police have erroneously included an innocent suspect). The current study examined lineup accuracy in autistic and TD children, on perpetrator present and perpetrator absent lineups.

Theoretically, one may expect autistic children to show difficulties in eyewitness identification. Chevallier, Kohls, Troiani, Brodkin and Schultz (2012) proposed that diminished social motivation has downstream effects on various aspects of social cognition (e.g., face recognition memory) in autism. This may be compounded by the complexity of faces, since autistic individuals often experience difficulties processing stimuli as task demands increase (Williams, Minshew, & Goldstein, 2015).

Confirming these theoretical predictions, Weigelt, Koldewyn and Kanwisher (2012) comprehensively reviewed studies on face identity recognition in autistic children and adults. They found no strong evidence for *qualitative* differences in how faces are processed in those with autism (e.g., autistic people show typical face inversion effects). However, they identified *quantitative* differences between those with and without autism regarding how well faces are remembered and recognised. Even with minimal demands on memory (i.e., very short delays between encoding and retrieval), autistic individuals were found to show poorer performance on face memory tasks.

[1] There is debate regarding the way autism is – and should be – described. In this article we use both identity-first (i.e. autistic children) and person-first (i.e. children with autism) language to respect this diversity of views (Kenny et al., 2016).

Nevertheless, it is difficult to extrapolate these findings to an eyewitness context since empirical research on face memory often involves: (a) static photographs (at encoding and retrieval); (b) large numbers of photographs (at encoding and retrieval); and (c) target faces being shown for short periods of time (often seconds) (e.g., Boucher & Lewis, 1992; de Gelder, Vroomen, & van der Heide, 1991; Gepner, de Gelder, & de Schonen, 1996; Hauck, Fein, Maltby, Waterhouse, & Feinstein, 1998; Scherf, Behrmann, Minshew, & Luna, 2008; Serra et al., 2003; Wilson, Brock, & Palermo, 2010; Wolf et al., 2008). In contrast, identification lineups traditionally involve dynamic staged events (usually lasting a few minutes) and include a small number of perpetrators. Then, at retrieval, 'witnesses' need to identify each perpetrator from a limited section of options, shown via dynamic video presentation (e.g., Wilcock, Crane, Hobson, Nash, Kirke-Smith, & Henry, 2018).

The aim of this preliminary study was to examine performance on perpetrator present and absent identification lineups in autistic and TD children (all with IQs > 69) using an ecologically valid methodology. Theoretical and empirical literature would suggest poorer lineup performance in children with autism. However, as lineup tasks have not been administered to autistic children previously, this prediction was tentative.

Method

Design

As part of a larger investigation into the eyewitness skills of children with and without autism (see Henry et al., 2017a; 2017b; Wilcock et al., 2018), 72 children with autism and 202 children with TD (6 to 11 years of age) were recruited. Children viewed a staged event in which two male actors gave a short talk about what school was like in Victorian times. Towards the end of the talk, a minor crime involving theft of keys/phone occurred. Due to practical constraints when collecting data, some children were unable to view the event live (as we had initially intended; for further details see Henry et al., 2017a). As a result, 144 children saw the event live (16 autistic children, 128 TD children) and 68 children saw the event on video (34 autistic children, 34 TD children). Therefore, event medium (i.e., whether the children viewed the event live or on video) was controlled in our analyses, but it was not a focus of our investigation.

Children were questioned about what they saw immediately following the event (Henry et al., 2017b) and again around one week later (Henry et al., 2017a). At the second recall, children also viewed two video lineups and were asked to identify the two actors who gave the talk. In the current paper, we report novel data concerning the lineup identification performance of children with autism, relative to comparison children with TD drawn from the larger study (see Wilcock et al., 2018, for additional data regarding the performance of the TD children).

Ethical approval was obtained from the University at which the research was carried out. Children had informed, written parental consent and gave their own written and verbal assent to participate.

Participants

Fifty children with autism (43 boys, mean age = 9yrs 2m, SD=19m) and 162 children with TD (84 boys, mean age = 8yrs 5m, SD=13m) were included in these analyses[2] (see Table 1 for participant details). For a chi-square test examining the main effect of the critical independent variable (autism/TD) on lineup accuracy (correct/incorrect), a post hoc power analysis was conducted on the whole sample (N=212) using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). The recommended effect sizes were: small *w*=.10, medium *w*=.30, large *w*=.50 (see Cohen, 1988). The alpha level for this analysis was $p < .05$. Post hoc analyses showed that the statistical power for this study was .31 for detecting a small effect, .99 for detecting a medium effect and .99 for detecting a large effect (i.e., sufficient power for detecting medium and large effect sizes). Given the unequal sample sizes between the autistic (N = 50) and TD (N =162) children, post hoc analyses were conducted on a sample of 100 (assuming 50 autistic and 50 TD children). This showed that the statistical power was .17 for detecting a small effect, .85 for detecting a medium effect and .99 for detecting a large effect (i.e., still sufficient power for detecting medium and large effect sizes).

[insert Table 1 about here]

All participants had an IQ of 70 or above, to ensure they had satisfactory cognitive abilities to complete the tasks, and all autistic children had a formal autism diagnosis from a clinical professional (obtained independently of the research study and confirmed by parents and/or schools).

Control and individual differences measures

---

[2] In this paper we report data for all eligible children drawn from the larger investigation who experienced identical PACE Code D compliant procedures regarding the administration of the lineup. Data were excluded for those children who received an adapted lineup procedure as part of the Registered Intermediary condition (see Henry et al. 2017a, for details of the full data set).

Intelligence: Two subtests ('Vocabulary' and 'Matrix Reasoning') of the Wechsler Abbreviated Scale of Intelligence (WASI-II; Wechsler & Zhou, 2011) were used to confirm eligibility into the study.

Social Communication Questionnaire (SCQ, Rutter et al., 2003): To provide further independent information about the diagnostic status of participants, parents were asked to complete the SCQ. Completed questionnaires were received for 36 of the autistic children (72% of the autism sample) and 128 of the TD children (79% of the TD sample). The mean SCQ scores were 19.42 (SD = 6.64) for the autism group and 4.90 (SD = 4.27) for the TD group.

Memory: The Facial Memory subtest from the Test of Memory and Learning 2 (TOMAL-2; Reynolds & Voress, 2007) provided an indication of face memory performance on a standardised task[3].

Lineup Procedure

Children viewed two lineups – one for each actor (henceforth, 'Perpetrators 1 and 2') in the staged event. Produced by the UK's Metropolitan Police Service (as per PACE Code D, 2011), each lineup contained nine video images: colour head and shoulders frontal perspectives of each man, before they moved from the left to the right profile, and then back to a frontal perspective. The 'foils' within each lineup (videos of men who were not in our staged event) were chosen by experienced police employees using their national database, PROMAT™. A measure of lineup bias confirmed the lineups were not biased toward the perpetrators or innocent replacements[4] (Malpass & Lindsay, 1999).

In actual criminal investigations, the suspect may be guilty or innocent. To simulate this in empirical studies, participants can be presented with 'perpetrator present' (PP) lineups (the 'suspect' featured in the lineup is guilty), or 'perpetrator absent' (PA) lineups (the 'suspect' featured in the lineup is an innocent suspect). Here, each child viewed one PP lineup and one PA lineup, and approximately half saw Perpetrator 1 and half saw Perpetrator 2. In accordance with PACE Code D, all children were informed prior to the lineups that the perpetrator may or may not be present in the

[3] Whilst other measures of language, memory and attention were administered, they are not presented here as they are not directly relevant to the current investigation (see Henry et al., 2017b, for further details).

[4] For Perpetrator 1, four of 30 mock witnesses (.13) randomly chose the perpetrator, and for Perpetrator 2, five of 30 mock witnesses randomly chose the perpetrator (.16); both of which are only slightly higher than what would be expected by chance (.11). Similar results were also found when randomly selecting an 'innocent suspect' to replace Perpetrator 1 (.10) and an 'innocent suspect' to replace Perpetrator 2 (.16) in the 'perpetrator absent' lineups.

lineup. PP and PA lineups were randomised: some children viewed the PP lineup for Perpetrator 1 and the PA lineup for Perpetrator 2, and vice versa.

For PP lineups, possible lineup responses were: correct hits (the child correctly identified the perpetrator); foil identification (the child incorrectly identified a foil, i.e., not the perpetrator); or incorrect rejection (the perpetrator was in the lineup, but the child responded that he was not). For PA lineups, possible responses were: correct rejections (the child correctly responded that the perpetrator was not present); or foil identification (the child incorrectly identified a foil as the perpetrator).

Results

First, potential group differences in cognitive variables (age, IQ, and Facial Memory) that might impact on witness performance were assessed. Table 1 includes mean ages for participants in each condition, and standardised/scaled scores (and SDs) for IQ and Facial Memory. Age differed between the two groups, as did full scale IQ. Consistent with previous literature, the autism group performed more poorly than the TD group on the face memory task. We checked for correlations between lineup accuracy for Perpetrator 1 and 2, and age, IQ and Facial Memory (for all participants and for the autism and TD groups separately) and no significant correlations were found; see Table 2.

In line with recommendations on group matching in autism research (e.g., Burack, Iarocci, Flanagan, & Bowler, 2004; Jarrold & Brock, 2004), we controlled for variables thought to be important for identification performance that differed between the groups – age and IQ (see Table 1), but did not control for abilities that are associated with an autism diagnosis (Facial Memory). In addition, because event medium (live or video) is likely to affect lineup performance, this was included in the model. Two hierarchical logistic regressions were conducted (one for each perpetrator). For both perpetrators, logistic regression analyses were performed with lineup accuracy (correct or incorrect) as the dependent variable. Variables entered into the model were age (in months), full scale IQ, perpetrator presence/absence (PP/PA), event medium (live/video), and group (autism/TD). We also included two interaction terms to determine whether or not certain variables had different effects on each group (perpetrator presence/absence x group; event medium x group). At step 1, age and IQ were entered; at step 2, perpetrator presence/absence and event medium were entered; at step 3, group was entered; and at step 4, perpetrator presence/absence x group and event medium x group interactions were entered.

For Perpetrator 1, at step 1, the overall model did not significantly predict lineup accuracy (omnibus χ^2 = .65, DF 2, *p* = .72) and age and IQ were not significant predictors (age *p* = .66; IQ = *p* = .48). The model accounted for between 0.3 and 0.4% of the variance, with 21.8% of incorrect performance and 81.1% of correct performance successfully predicted. At step 2, the overall model significantly predicted lineup accuracy (omnibus χ^2 = 40.20, DF 2, *p* <.001), and there were significant effects of event medium (*p* <.001) and perpetrator presence/absence (*p* <.001). The model accounted for between 17.5% and 23.4% of the variance, with 72.3% of incorrect performance and 66.7% of correct performance successfully predicted. At step 3, the overall model did not significantly predict lineup accuracy (omnibus χ^2 = 3.13, DF 1, *p* = .08) and group was not significant (*p* = .08). The model accounted for between 18.7% and 25% of the variance, with 76.2% of incorrect and 64.9% of correct performance successfully predicted. At step 4, the overall model did not significantly predict lineup accuracy (omnibus χ^2 = .20, DF 2, *p* = .90) and neither of the interactions were significant (event medium x group *p* = .68; perpetrator presence x group *p* = .83). The model accounted for between 18.8% and 25.1% of the variance with 75.2% of incorrect and 65.8% of correct performance successfully predicted. Table 3 provides the log likelihood (R^2), coefficients, the Wald statistic, associated degrees of freedom, and probability values for each of the predictor variables, at each step, for Perpetrator 1.

The significant factors identified in the logistic regression were followed up using odds ratios and chi-squared analysis. Odds ratio values of less than 1 are indicative of worsening odds of the event given the tested condition, whilst values over 1 are indicative of increasing odds of the event given the tested condition. The odds ratio of an accurate lineup response occurring on the PP lineup was .53, whereas it was 2.4 on the PA lineup; therefore, participants were more likely to be correct on the PA lineup. The odds ratio of an accurate lineup response occurring after viewing the event on video was .62, whereas it was 1.44 when viewing the event live; thus, participants viewing the event live were more likely to be correct. Follow up chi-squared analysis confirmed a significant association between perpetrator presence and accuracy χ^2(1, N=212) = 27.53, *p* <.001, ϕc = .36. This remained the case for the autism group separately χ^2(1, n= 50) = 5.27, *p* = .02, ϕc = .33, where the odds ratio of an accurate lineup response occurring was .23 on the PP lineup and 1.0 on the PA lineup. The same association existed for the TD children separately χ^2(1, n=162) = 23.32, *p* <.001, ϕc = .38, where the odds ratio of an accurate lineup response occurring was .66 on the PP lineup, whereas it was 3.39 on the PA lineup. Therefore, both groups were more likely to be correct on the PA lineup. Follow up chi-squared analysis confirmed a significant association between event medium and accuracy χ^2(1, N=212) = 8.01, *p* = .005, ϕc = .19. However, event medium was not significant when examining the data separately for the autism and TD groups: χ^2(1, n=50) = 1.00, *p* = .32, ϕc = .14, and χ^2(1, n=162) = 2.13, *p* = .15, ϕc = .12, respectively. Event medium remained significant for participants viewing the PP lineup χ^2(1, n=109) = 4.43, *p* = .03, ϕc = .20, where the odds ratio of an accurate lineup response occurring was .25 following the video event, but .70 following the live event. For the PA lineup, χ^2(1, n=103) = 5.53, *p* =.02, ϕc = .23, the odds ratio of an accurate lineup response occurring was 1.2 following the video event, whereas it was 3.6 following the live event. Thus, performance on the PA lineup indicated accuracy was likely to increase following the live event; for the PP lineup, accuracy was likely to decrease following the video event.

For Perpetrator 2, at step 1, the overall model did not significantly predict lineup accuracy (omnibus χ^2 = .58 DF 2, *p* = .75) and age and IQ were not significant (age *p* = .51; IQ = *p* = .66). The model accounted for between 0.3 and 0.4% of the variance, with 2.2% of incorrect performance and 98.3% of correct performance successfully predicted. At step 2, the overall model significantly predicted lineup accuracy (omnibus χ^2 = 13.59, DF 2, *p* = .001) and there was a significant effect of event medium (*p* < .001) but no significant effect of perpetrator presence/absence (*p* = .76). The model accounted for between 6.5% and 8.7% of the variance, with 45.2% of incorrect performance and 79% of correct performance successfully predicted. At step 3, the overall model did not significantly predict lineup accuracy (omnibus χ^2 = .02, DF 1, *p* = .88) and group was not a significant predictor (*p* = .88). The model accounted for between 6.5% and 8.7% of the variance, with 45.2% of incorrect and 79% of correct performance successfully predicted. At step 4, the overall model significantly predicted lineup accuracy (omnibus χ^2 = 8.54, DF 2, *p* = .01). The interaction between group and perpetrator presence was significant (*p* = .006) but there was no significant interaction between group and event medium (*p* = .46). The model accounted for between 10.2% and 13.6% of the variance, with 39.8% of incorrect and 84.9% of correct performance successfully predicted. Table

4 gives the log likelihood (R^2), coefficients, the Wald statistic, associated degrees of freedom, and probability values for each of the predictor variables[5].

The odds ratio of an accurate lineup response occurring after viewing the event on video was .62, whereas it was 1.8 when viewing the event live; thus, participants viewing the event live were more likely to be correct. Follow up chi-squared analysis confirmed a significant association between event medium and accuracy, $\chi^2(1, N=212) = 13.02$, $p < .001$, $\phi c = .25$. This remained the case for the autism group separately $\chi^2(1, n=50) = 4.90$, $p = .03$, $\phi c = .31$: the odds ratio of an accurate lineup response occurring after viewing the event live was .77, whereas it was .42 having viewed the event on video. Thus, autistic children were less likely to be accurate following the video event. The same association existed for the TD group, $\chi^2(1, n=162) = 5.83$, $p = .02$, $\phi c = .19$, where the odds ratio of an accurate lineup response occurring after seeing the event live was 1.56, whereas it was .89 after viewing the event on video. The TD group were more likely to be accurate on the lineup after viewing the event live. Event medium remained significant for participants viewing the PP lineup $\chi^2(1, n=103) = 29.61$, $p < .001$, $\phi c = .54$ where the odds ratio of an accurate lineup response occurring following the video event was .20, whereas it was 2.77 following the live event (suggesting greater lineup accuracy following the live event). Event medium was not significant for those viewing the PA lineup $\chi^2(1, n=109) = .15$, $p = .70$, $\phi c = .04$. Thus, the advantage of seeing the event live only existed for participants viewing the PP lineup.

Chi-square analysis focusing on the significant interaction between group and perpetrator presence confirmed a significant effect of group on the PP lineup $\chi^2(1, n=103) = 9.25$, $p = .002$, $\phi c = .30$. The odds ratio of TD children being correct on the PP lineup was 1.69, whereas the odds ratio for autistic children was .39. Thus, TD children were more likely to be accurate on the PP lineup than the autistic children. There was no significant effect of group on the PA lineup, $\chi^2(1, n=109) = .51$, $p = .47$, $\phi c = .07$.

[insert Tables 3, 4 and 5 about here]

In view of missing SCQ data for some children, and given that other children (eight children in the autism group and four children in the TD group) did not reach the SCQ cut-off for the group they

[5] Given that there was a significant difference between groups on Facial Memory, we ran both logistic regressions again while controlling for Facial Memory. The results were the same as those reported above for Perpetrator 1 and 2.

were assigned to, we ran the logistic regressions for Perpetrators 1 and 2 excluding these participants. This resulted in a total sample of 28 autistic children (mean SCQ score = 22.07, SD = 4.55) and 124 TD children (mean SCQ score = 4.45, SD = 3.43). The analyses with this reduced sample showed an identical pattern of results to those outlined for the full sample (please see Supplementary Materials for full details).

Discussion

This preliminary study was the first to look at identification lineup performance in autistic child witnesses relative to TD comparison witnesses (all with IQs > 69) on perpetrator present and perpetrator absent lineups. Despite the autistic children's poorer performance on a standardized face memory task, results demonstrated that the lineup performance of autistic children was mostly equivalent to that of their TD peers: For Perpetrator 1, there were no significant group differences in performance on either perpetrator present or absent lineups; for Perpetrator 2, there was no significant group difference on the perpetrator absent lineup, although children with autism were less likely to make a correct identification on the perpetrator present lineup.

How can we account for these findings? The Perpetrator 1 lineup may have been more difficult than the Perpetrator 2 lineup, perhaps due to the high level of similarity between Perpetrator 1 and the lineup foils (see also Wilcock et al., 2018). Because of this, all children (across the autism and TD groups) struggled to correctly identify Perpetrator 1 (and there were no group differences observed). In contrast, the Perpetrator 2 lineup may have been easier (with a lower level of similarity between Perpetrator 2 and the lineup foils). As such, the TD children performed better than the autistic children when the perpetrator was present, although not when he was absent. It is often found that vulnerable groups show varied performance on perpetrator present lineups and this may, in part, be dependent on lineup difficulty (Wilcock, Bull, & Milne, 2008). Therefore, although children with autism may be as accurate as children with TD on some lineups, task difficulty may be an important qualifying factor. Further, although not a focus for the present investigation, we also acknowledge that event medium had differing and somewhat complex effects on performance. This could be a focus for future research.

The relatively good lineup performance in our autism group contrasts with the robust group differences reported in performance on standardized face memory tasks (e.g., Wiegelt et al., 2012). Indeed, our autistic children performed more poorly than their TD counterparts on a standardised

Facial Memory task in the current study. There are a number of possible reasons for the relatively better performance seen on lineups. First, the lineup task is an ecologically valid test of face recognition. A key strength of this research was that, for the first time, a mock witness paradigm was used: perpetrators were viewed for an appropriate length of time; lineups were produced by a major UK police service; and there were realistic delays between encoding and retrieval. This is very different from typical laboratory face recognition tasks. For example, in the Facial Memory task used in the current study, the images comprised up to 12 black and white static photographs that were presented for a very short exposure period before being immediately selected from an array of up to 30 images. This constitutes a highly demanding task that may pose particular difficulty for children with autism (see Williams et al., 2015). The disparity between the demands of the lineup identification and standardised face memory tasks may also account for the low correlation between face memory performance and lineup accuracy (for both perpetrators) in our sample.

Whilst these findings suggest autistic children fare relatively well on some lineup identification tasks compared to TD children, they may still need support in such contexts for several reasons. First, autistic children had difficulty with one of the two perpetrator present lineups, potentially due to higher levels of similarity between the perpetrator and lineup foils (which could be challenging for children with autism). Second, the children in the current study all had IQs within the average or above average range and these results may not extend to children with autism who also have intellectual disabilities. Third, the results presented here were based on an empirical study involving a mild ‘mock’ theft, and may not translate to real-life cases with associated anxiety and distress for child witnesses, particularly those with autism. Finally, children in general are more likely to make errors on lineup identification tasks (relative to young adults, see Fitzgerald & Price, 2015). Consistent with this finding, lineup performance in both the TD and autism groups was, whilst equivalent, fairly low (cf. Fitzgerald & Price, 2015).

A vital next step will be to identify how to scaffold eyewitness identification performance in autistic and TD children, particularly when task demands are high (e.g., with multiple perpetrators or with there being a substantial delay between viewing the crime and see the lineups). In England and Wales (and internationally), the Registered Intermediary (RI) role has already been developed to support communication in vulnerable witnesses (including children, as well as those with autism). RIs are trained professionals who facilitate communication between vulnerable witnesses and members of the Justice System (see Cooper & Mattison, 2017; Plotnikoff & Woolfson, 2015). RIs have been shown to bolster eyewitness identification performance in TD children by including adaptations that improve accuracy (Wilcock et al., 2018), and RIs already support children with and without autism in justice system contexts. Therefore, evaluating the empirical evidence for the utility of RIs facilitating lineup performance with autistic children is a crucial next step.

Conflict of interest: None

Acknowledgements

We would like to express our thanks to Superintendent John Armstrong for developing the identification lineups, as well as the schools, teachers, parents, and children who kindly assisted with the research. This work was supported by the Economic and Social Research Council [Grant: ES/J020893/2].

References

Almeida, T.S., Lamb, M.E. & Weisblatt, E.J. (2018). Effects of delay, question type, and socioemotional support on episodic memory retrieval by children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-018-3815-3

Boucher, J., & Lewis, V. (1992). Unfamiliar Face Recognition in Relatively Able Autistic Children. *Journal of Child Psychology and Psychiatry, 33*(5), 843-859. doi: 10.1111/j.1469-7610.1992.tb01960.x

Burack, J.A., Iarocci, G., Flanagan, T.D., & Bowler, D.M. (2004). On mosaics and melting pots: conceptual considerations of comparison and matching strategies. *Journal of Autism and Developmental Disorders, 34*(1), 65-73. doi: 10.1023/B:JADD.0000018076.90715.00

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E.S., & Schultz, R.T. (2012). The Social Motivation Theory of Autism. *Trends in Cognitive Sciences, 16*(4), 231-239. doi: 10.1016/j.tics.2012.02.007

Cooper, P. & Mattison (2017). Intermediaries, vulnerable people and the quality of evidence: An international comparison of three versions of the English intermediary model. *The International Journal of Evidence & Proof, 21*(4) 351–370. doi: 10.1177/1365712717725534

de Gelder, B., Vroomen, J., & van der Heide, L. (1991). Face recognition and lip-reading in autism. *European Journal of Cognitive Psychology, 3*, 69-86. doi: 10.1080/09541449108406220

Fitzgerald, R.J. & Price, H.L. (2015). Eyewitness identification across the lifespan: A meta-analysis of age differences. *Psychological Bulletin, 141,* 1228-1265. doi: 10.1037/bul0000013

Gepner, B., de Gelder, B., de Schonen, S. (1996). Face processing in autistics: evidence for a generalised deficit? *Child Neuropsychology, 2*, 123-139. doi: 10.1080/09297049608401357

Hauck, M., Fein, D., Maltby, N., Waterhouse, L., & Feinstein, C. (1998). Memory for Faces in Children with Autism. *Child Neuropsychology, 4*(3), 187-198. doi: 10.1076/chin.4.3.187.3174

Henry, L.A., Crane, L., Nash, G., Hobson, Z., Kirke-Smith, M., & Wilcock, R. (2017a). Verbal, visual, and intermediary support for child witnesses with autism during investigative interviews. *Journal of Autism and Developmental Disorders, 47*, 2348-2362. doi: 10.1007/s10803-017-3142-0

Henry, L.A., Messer, D., Wilcock, R., Nash, G., Hobson, Z., Kirke-Smith, M., & Crane, L. (2017b). Do measures of memory, language, and attention predict eyewitness memory in children with and without autism spectrum disorder? *Autism and Developmental Language Impairments, 2*. doi: 10.1177/2396941517722139

Jarrold, C. & Brock, J. (2004). To match or not to match? Methodological issues in autism related research. *Journal of Autism and Developmental Disorders, 34,* 81-86. doi.org/10.1023/B:JADD.0000018078.82542.ab

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism, 20*(4), 442–462. doi.org/10.1177/1362361315588200.

Malpass, R.S., & Lindsay, R.C.L. (1999). Measuring Line-up Fairness. *Applied Cognitive Psychology, 13*, S1-S7. doi: 10.1002/(SICI)1099-0720(199911)13:1+<S1::AID-ACP678>3.0.CO;2-9

Maras, K.L., & Bowler, D.M. (2014). Eyewitness testimony in autism spectrum disorder: a review. *Journal of Autism and Developmental Disorders, 44*(11), 2682-2697. doi: 10.1007/s10803-012-1502-3

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

Mattison, M., Dando, C.J., & Ormerod, T. (2018). Drawing the answers: Sketching to support free and probed recall by child witnesses and victims with autism spectrum disorder. *Autism, 22*(2), 181-194. doi: 10.1177/1362361316669088

Police and Criminal Evidence Act (1984). Code D code of practice for the identification of persons by police officers (2011 edition). Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/253831/pace-code-d-2011.pdf

Plotnikoff, J. & Woolfson, R. (2015). *Intermediaries in the criminal justice system: Improving communication for vulnerable witnesses and defendants.* Bristol: Policy Press.

Reynolds, C. & Voress, J.K. (2007). *Test of Memory and Learning: Second Edition.* Austin, TX: Pro-Ed.

Scherf, K.S., Behrmann, M., Minshew, N.J., & Luna, B. (2008). Atypical development of face and greeble recognition in autism. *Journal of Child Psychology and Psychiatry, 49*(8), 838-847. doi: 10.1111/j.1469-7610-2008.01903.x

Serra, M., Althaus, M., de Sonneville, L.M.J., Stant, A.D., Jackson, A.E., & Minderaa, R.B. (2003). Face Recognition in Children with a Pervasive Developmental Disorder Not Otherwise Specified. *Journal of Autism and Developmental Disorders, 33*(3), 303-317. doi: 10.1023/A:1024458618172

Wechsler, D. & Zhou, X. (2011). *Wechsler Abbreviated Scale of Intelligence – Second Edition*. Bloomington, MN: Pearson.

Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face Recognition Deficits in Autism Spectrum Disorders Are Both Domain Specific and Process Specific. *PLOS One, 8*(9), e74541. doi: 10.1371/journal.pone.0074541

Wilcock. R., Bull, R., & Milne, R. (2008). *Witness identification in criminal cases: Psychology and practice.* Oxford: Oxford University Press.

Wilcock, R., Crane, L., Hobson, Z., Nash, G., Kirke-Smith, M., & Henry, L.A. (2018). Supporting child witnesses during identification lineups: exploring the effectiveness of Registered Intermediaries. *Applied Cognitive Psychology, 32*(3), 367-375. doi: 10.1002/acp.3412

Williams, D.L., Minshew, N.J., Goldstein, G. (2015). Further understanding of complex information processing in verbal adolescents and adults with autism spectrum disorders. *Autism, 19*(7), 859-867. doi: 10.1177/1362361315586

Wilson, C.E., Brock, J., Palermo, R. (2010). Attention to social stimuli and facial identity recognition skills in autism spectrum disorder. *Journal of Intellectual Disability Research, 54*, 1104-1115. doi: 10.1111/j.1365-2788.2010.01340.x

Wolf, J.M., Tanaka, J.W., Klaiman, C., Cockburn, J., Herlihy, L., Brown, C., South, M., McPartland, J., Kaiser, M.D., Phillips, R., & Schultz, R.T. (2008). Specific impairment of face-processing abilities in children with autism spectrum disorder using the Let's Face It! Skills battery. *Autism Research, 1*, 329-340. doi: 10.1002/aur.56

SUPPLEMENTARY MATERIALS: Analyses excluding participants without SCQ data and participants not reaching the appropriate SCQ cut off.

The same logistic regression models were run excluding the participants without SCQ data and participants who failed to reach the appropriate SCQ cut off for their group; one for each perpetrator (n = 152). Predictor variables were age (in months), full scale IQ, lineup perpetrator presence/absence (PP/PA), event medium (live/video), and group (autism/TD). We also included two interaction terms to determine whether or not certain variables had different effects on each group (perpetrator presence/absence x group; event medium x group). At step 1, age and IQ were entered; at step 2, perpetrator presence and event medium were entered; at step 3, group was entered; at step 4, event medium x group and perpetrator presence x group interactions were entered.

For Perpetrator 1, at step 1, the overall model did not significantly predict lineup accuracy (omnibus χ^2 = .99, DF 2, p = .61), and age and IQ were not significant (age p = .44; IQ = p = .46). The model accounted for between 0.6 and 0.9% of the variance with 44.7% of incorrect performance and correct performance successfully predicted. At step 2, the overall model significantly predicted lineup accuracy (omnibus χ^2 = 29.72, DF 2, p <.001) and there was a significant effect of event medium (p =.001) and perpetrator presence/absence (p <.001). The model accounted for between 18.3% and 24.4% of the variance, with 73.7% of incorrect performance and 60.5% of correct performance successfully predicted. At step 3, the overall model did not significantly predict lineup accuracy (omnibus χ^2 = 2.20, DF 1, p = .14) and group was not significant (p = .15). The model accounted for between 19.5% and 25.9% of the variance, with 77.6% of incorrect and 57.9% of correct performance successfully predicted. At step 4, the overall model did not significantly predict lineup accuracy (omnibus χ^2 = 2.73, DF 2, p = .26) and neither of the interactions were significant (event medium x group p = .27; perpetrator presence x group p = .38). The model accounted for between 20.9% and 27.9% of the variance, with 71.1% of incorrect and 67.1% of correct performance successfully predicted. For Perpetrator 1, the odds ratio of an accurate lineup response occurring on the PA lineup was 2.14, whereas on the PP lineup it was .54; therefore, participants were more likely to be correct on the PA lineup. The odds ratio of an accurate lineup response occurring after viewing the event live was 1.40 whereas viewing it on video was .48; thus, participants viewing the event live were more likely to be correct.

For Perpetrator 2, at step 1, the overall model did not significantly predict lineup accuracy (omnibus χ^2 = .57 DF 2, *p* = .75), and age and IQ were not significant (age *p* = .58; IQ *p* = .56). The model accounted for between 0.4 and 0.5% of the variance, with 7.5% of incorrect performance and 95.3% of correct performance successfully predicted. At step 2, the overall model significantly predicted lineup accuracy (omnibus χ^2 = 7.86, DF 2, *p* = .02) and there was a significant effect of event medium (*p* = .01) but no significant effect of perpetrator presence/absence (p = .41). The model accounted for between 5.4% and 7.2% of the variance with 43.3% of incorrect performance and 77.6% of correct performance successfully predicted. At step 3, the overall model did not significantly predict lineup accuracy (omnibus χ^2 = .00, DF 1, *p* = .96) and group was not significant (*p* = .97). The model accounted for between 5.4% and 7.2% of the variance, with 43.3% of incorrect and 77.6% of correct performance successfully predicted. At step 4, the overall model significantly predicted lineup accuracy (omnibus χ^2 = 7.38, DF 2, *p* = .02). The interaction between group and perpetrator presence was significant (*p* = .015) but there was no significant interaction between group and event medium (*p* = .41). The model accounted for between 9.9% and 13.2% of the variance, with 40.3% of incorrect and 84.7% of correct performance successfully predicted. The odds ratio of an accurate lineup response occurring after viewing the event live was 1.71 whereas on video it was .69; thus, participants viewing the event live were more likely to be correct. Chi-square analysis focusing on the significant interaction between group and perpetrator presence revealed that group was a significant predictor on the PP lineup χ^2(1, n=68) = 8.38, *p* = .004, ϕc = .35. The odds ratio of a TD child being correct on the PP lineup was 1.5 whereas the odds ratio value for the autism group was .18. TD children were more likely to be accurate on the PP lineup than the autistic children. There was no significant effect of group on the PA lineup, χ^2(1, n=84) = .39, *p* = .53, ϕc = .07.

In summary, the results were the same whether children with or without SCQ data were included.

Table 1
Mean scores (standard deviations) on background variables for children in each interview condition.

Variables:	Autism (n=50, 7 girls)	TD (n=162, 78 girls)	Group differences
Age	9yrs 2m (19m)	8yrs 5m (13m)	*t*(210) = 3.79, *p* =.001, *g* = .61
WASI-II Full Scale[1]	99.66 (13.98)	108.02 (13.25)	*t*(210) = -3.85, *p* =.001, *g* = .62
TOMAL2 Facial Memory[2]	8.10 (2.87)	10.64 (2.97)	*t*(209) = -5.28, *p* =.001, *g* = .86

[1]Standardised scores (mean = 100, SD = 15); [2]Scaled scores (mean = 10, SD = 3)

Table 2

Correlations between perpetrator 1 and 2 lineup accuracy and age, IQ and facial memory.

		Age	WASI-II Full Scale IQ	TOMAL2 Facial Memory
All participants (N =212)	Perpetrator 1 Accuracy	*r* = .027, *p* = .69	*r* = -.05, *p* = .50	*r* = -.11, *p* = .13
	Perpetrator 2 Accuracy	*r* = .04, *p* = .53	*r* = -.03, *p* = .70	*r* = .02, *p* = .77
Autism (n = 50)	Perpetrator 1 Accuracy	*r* = -.05, *p* = .71	*r* = -.18, *p* = .21	*r* = -.08, *p* = .59
	Perpetrator 2 Accuracy	*r* = .04, *p* = .76	*r* = .00, *p* = 1.0	*r* = .01, *p* = .97
TD (n = 162)	Perpetrator 1 Accuracy	*r* = -.12, *p* = .85	*r* = .06, *p* = .42	*r* = -.03, *p* = .70
	Perpetrator 2 Accuracy	*r* = .00, *p* = .97	*r* = .000, *p* = .97	*r* = .08, *p* = .33

Table 3: Logistic regression predictors for Perpetrator 1 accuracy (N = 212).

Step	R^2	*B*	Wald	Degrees of Freedom	Significance	Exp *B*	95% confidence interval	
							Lower	Upper
Constant	293.42	.09	.47	1	.492	1.10		
Step 1								
Constant	292.77	.12	.69	1	.406	1.12		
Age		-.00	.19	1	.660	1.00	.98	1.01
WASI Full Scale		.01	.49	1	.484	1.01	.99	1.03
Step 2								
Constant	252.56	-1.47	17.88	1	.001	.23		
Age		.01	.32	1	.573	1.01	.97	1.03
WASI-II Full Scale IQ		.02	2.23	1	.136	1.02	1.00	1.04
Event Medium		1.16	10.74	1	**.001**	3.19	1.60	6.39
Perpetrator Presence		1.67	28.24	1	**.001**	5.30	2.87	9.80
Step 3								
Constant	249.44	-1.14	8.41	1	.004	.32		
Age		.01	.81	1	.369	1.01	.99	1.03

WASI-II Full Scale IQ		.01	.57	1	.449	1.01	.99	1.03
Event Medium		.89	5.34	1	**.021**	2.44	1.15	5.22
Perpetrator Presence		1.68	28.01	1	**.001**	5.36	2.88	9.97
Group		-.77	3.10	1	.078	.47	.20	1.09
Step 4								
Constant	249.24	-1.24	7.35	1	.007	.29		
Age		.01	.87	1	.352	1.01	.99	1.03
WASI-II Full Scale IQ		.01	.62	1	.432	1.01	.99	1.03
Event Medium		1.0	4.53	1	**.033**	2.72	1.08	6.83
Perpetrator Presence		1.72	22.78	1	**.001**	5.60	2.76	11.35
Group		-.51	.47	1	.492	.60	.14	2.57
Event Medium x Group		-.34	.17	1	.681	.72	.14	3.54
Perpetrator Presence x Group		-.17	.05	1	.826	.85	.19	3.70

Table 4: Logistic regression predictors for Perpetrator 2 accuracy (N = 212).

Step	R^2	*B*	Wald	Degrees of Freedom	Significance	Exp *B*	95% confidence interval	
							Lower	Upper
Constant	290.70	.25	3.17	1	.075	1.28		
Step 1								
Constant	290.12	.27	3.58	1	.058	1.31		
Age		-.01	.43	1	.514	.99	.98	1.01
WASI-II Full Scale IQ		.00	.19	1	.664	1.00	.99	1.02
Step 2								
Constant	276.53	-.57	3.60	1	.058	.57		
Age		.00	.17	1	.685	1.00	.99	1.02
WASI-II Full Scale IQ		.01	.81	1	.368	1.01	.99	1.03
Event Medium		1.16	12.75	1	**.001**	3.19	1.69	6.02
Perpetrator Presence		.89	.09	1	.760	1.09	.62	1.92
Step 3								
Constant	276.51	-.60	2.84	1	.092	.55		

Age		.00	.14	1	.707	1.00	.99	1.02
WASI-II Full Scale IQ		.01	.80	1	.372	1.01	.99	1.03
Event Medium		1.18	10.68	1	**.001**	3.27	1.61	6.64
Perpetrator Presence		.09	.09	1	.763	1.09	.62	1.92
Group		.06	.02	1	.880	1.06	.48	2.35
Step 4								
Constant	267.96	-.28	.47	1	.492	.76		
Age		.00	.00	1	.984	1.00	.98	1.02
WASI-II Full Scale IQ		.01	.73	1	.393	1.01	.99	1.03
Event Medium		1.07	6.04	1	**.014**	2.92	1.24	6.88
Perpetrator Presence		-.36	1.16	1	.281	.70	.37	1.34
Group		-1.24	2.96	1	.085	.29	.07	1.19
Event Medium x Group		.63	.56	1	.455	1.89	.36	9.98

Perpetrator Presence x Group	2.08	7.55	1	**.006**	8.02	1.82	35.39

Table 5: Identification performance for Perpetrator 1 and Perpetrator 2 by condition, event medium and perpetrator presence/absence. Frequencies shown in parentheses.

			Perpetrator 1				
			Perpetrator Present			Perpetrator Absent	
		n	Hit (%)	Foil ID (%)	Incorrect Rejection (%)	Correct rejection (%)	Foil ID (%)
Autism	Live	16	12.5 (1)	37.5 (3)	50 (4)	75 (6)	25 (2)
Autism	Video	34	22.2 (4)	16.7 (3)	61.1 (11)	37.5 (6)	62.5 (10)
TD	Live	128	44.8 (30)	29.9 (20)	25.4 (17)	78.7 (48)	21.3 (13)
TD	Video	34	18.8 (3)	25 (4)	56.3 (9)	72.2 (13)	27.8 (5)
			Perpetrator 2				
			Perpetrator Present			Perpetrator Absent	
		n	Hit	Foil ID	Incorrect Rejection	Correct rejection	Foil ID
Autism	Live	16	62.5 (5)	0 (0)	37.5 (3)	75 (6)	25 (2)
Autism	Video	34	11.8 (2)	41.2 (7)	47.1 (8)	58.8 (10)	41.2 (7)
TD	Live	128	75 (45)	18.3 (11)	6.7 (4)	54.4 (37)	45.6 (31)
TD	Video	34	22.2 (4)	33.3 (6)	44.4 (8)	62.5 (10)	37.5 (6)