

Final author draft only. The copyedited article may differ from this manuscript version. The details of the article are as follows:

Humphries, J. E., & Flowe, H. D. (2015). Receiver operating characteristic analysis of age-related changes in lineup performance. *Journal of Experimental Child Psychology*, 132, 189-204. doi: 10.1016/j.jecp.2014.12.009

Receiver Operating Characteristic Analysis of Age-Related
Changes in Lineup Performance

Humphries, J. E.¹ & Flowe, H. D.²

1. Dr Joyce Humphries
Department of Psychology
Edge Hill University
Ormskirk, UK

2. School of Psychology
University of Leicester
Leicester, UK

Abstract

In the basic face memory literature, support has been found for the *late maturation hypothesis*, which holds that face recognition ability is not fully developed until at least adolescence. Support for the late maturation hypothesis in the criminal lineup identification literature, however, has been equivocal, because of the analytic approach that has been used to examine age-related changes in identification performance. Recently, Receiver Operator Characteristics (ROC) analysis was applied for the first time in the adult eyewitness memory literature to examine whether memory sensitivity differs across different types of lineup tests (Gronlund et al. 2012; Mickes, Flowe, & Wixted, 2012). ROC analysis allows for the separation of memory sensitivity from response bias in the analysis of recognition data. Here, we make the first ROC-based comparison of adults' and children's (5-6 years and 9-10 years) memory performance on lineups by reanalysing data from Humphries, Holliday, & Flowe

(2012). In line with the late maturation hypothesis, memory sensitivity was significantly greater for adults compared to young children. Memory sensitivity for older children was similar to adults. The results indicate that the late maturation hypothesis can be generalized to account for age-related performance differences on an eyewitness memory task. The implications for developmental eyewitness memory research are discussed.

Keywords: eyewitness identification, child eyewitness, late maturation, receiver operating characteristic (ROC), development

Receiver Operating Characteristic Analysis of Age-Related Changes in Lineup Performance

Children are often the sole witness to a crime and their identification evidence can play a pivotal role in criminal investigations (Pike, Bryce, & Kynan, 2002; Pozzulo, 2007). Eyewitness identification evidence is obtained by showing the eyewitness a lineup of faces; a lineup is composed of a suspect and a number of fillers, or persons known by the police to be innocent. If the suspect is the actual perpetrator, the lineup is target present, whereas if the lineup does not contain a guilty suspect, the lineup is target absent. The majority of research indicates that when the perpetrator is present, children as young as age 5 are just as likely as adults to make a correct identification decision (e.g., Goodman & Reed, 1986; Lindsay, Pozzulo, Craig, Lee & Corber, 1997; Marin, Holmes, Guth, & Kovac, 1979; Parker, & Carranza, 1989; Parker & Ryan, 1993; Pozzulo & Balfour, 2006; Pozzulo & Lindsay, 1998). In stark contrast, however, when shown a target absent lineup (i.e. a lineup containing only innocent persons), children are significantly more likely than adults to incorrectly identify someone as the perpetrator (Beal, Schmitt, & Dekle, 1995; Davies, 1996; Dekle, Beal, Elliott, & Huneycutt, 1996; Parker & Carranza, 1989; Parker & Ryan, 1993; Pozzulo & Balfour, 2006; Pozzulo & Dempsey, 2006). By adolescence (10-14 years), some research indicates that identification accuracy in target-absent lineups reaches adult levels (Pozzulo & Lindsay, 1997; Pozzulo & Warren, 2003, Experiment 2; for reviews, see, Havard, 2013; Pozzulo, 2007; Pozzulo & Lindsay, 1998).

Children may perform less accurately than adults on lineups because their memory sensitivity, or their ability to encode and/or detect the perpetrator in the lineup, is poorer. The *late maturation hypothesis* posits that children do not remember faces as accurately as adults

because face processing ability does not fully develop until at least adolescence (see Carey & Diamond, 1977; Mondloch, Le Grand, & Maurer, 2002). Evidence from the basic face processing literature suggests that children's sensitivity may be poorer because they tend to engage in more feature-based rather than configural-based processing (Mondloch, Geldart, Maurer, & Le Grand, 2003; Mondloch, et al., 2002; Schwarzer, 2000). The recognition of faces using configural information may not reach adult levels until adolescence (Bruce et al., 2000; Carey, Diamond, & Woods, 1980; Mondloch, et al., 2003; Mondloch, et al., 2002). Alternatively, age-related differences in memory sensitivity may arise because more basic memory mechanisms have not yet matured (see Crookes & McKone, 2009). In the eyewitness identification literature, some studies have concluded that the ability to accurately remember faces improves with age (Brigham, Van Verst, & Bothwell, 1986; Goodman & Reed, 1986; Karageorge & Zajac, 2011; Leippe, Romanczyk & Manion 1991).

Hypotheses about age-related differences in lineup decision strategies have also been put forward as an explanation for children's poorer identification performance. In particular, children appear more willing than adults to guess under conditions of uncertainty (Ceci & Bruck, 1993; Hughes & Grieve, 1980). In line with this, the positive identification rate for children is larger compared to adults (Lindsay et al., 1997; Parker & Carranza, 1989, Parker & Ryan, 1993; Pozzulo & Lindsay, 1998). Children may be just as able to encode and retrieve the perpetrator from memory as adults, but they may false alarm more often because they set a comparatively low response threshold (Lindsay et al., 1997; Parker & Caranza, 1989). Researchers have proposed that the mere presentation of a lineup may suggest to a child that the perpetrator is present in the lineup, and therefore, that a positive identification has to be made (Ceci, Ross, & Toglia, 1987; Gross & Hayne, 1996). Children may also feel more social pressure to positively identify a face and feel greater reluctance to declare uncertainty compared to adults (Beal, et al., 1995; King & Yuille, 1987; Ricci, Beal, &

Dekle, 1996). Children may also be less aware of the consequences of making a false identification (Brewer, Weber, & Semmler, 2005; Dekle, et al., 1996; Pozzulo & Lindsay 1997; Spring, Saltzstein, & Peach, 2013). Other researchers have proposed that memory sensitivity and response bias differ with age. Namely, children may set a lower response threshold because they have a more difficult time detecting the target from the lures (Leippe, et al., 1991; Pozzulo & Dempsey, 2006; Pozzulo & Lindsay, 1998).

Taken together, there is much evidence that children and adults adopt different strategies during a lineup test. In the next section, we will make the case that in order to adequately test the late maturation account, age-related differences in response thresholds (or response bias) have to be taken into account. In particular, we will use Receiver Operator Characteristics (ROC) analysis to separate response bias from memory sensitivity to test the late maturation hypothesis.

ROC Approach to Examining Age-Related Differences in ID Accuracy

To date, developmental research has compared the performance of children and adults by analysing the hit rate when the target is present in the lineup, and the false alarm rate when the target is absent. This approach, however, is inadequate for assessing the late maturation hypothesis, because it does not separate response bias from memory sensitivity (for reviews, see Gronlund & Neuschatz, 2014; Gronlund, Wixted, & Mickes, 2014; Yonelinas & Parks, 2007). As a hypothetical example, assume that 90% of children in a lineup experiment correctly identified the perpetrator when he was present in the lineup. Does this mean that the children were exceptionally good at remembering the perpetrator? In order to answer this question, children's performance would also have to be evaluated when the target is absent from the lineup. If we found in this experiment that 10% of children identified the innocent suspect in the target absent condition, this would suggest high memory sensitivity, or that children were able to accurately distinguish guilty from innocent suspects. On the other hand,

if the innocent suspect was identified by 30% of the children, it would be less clear as to whether memory sensitivity was high, or whether children were applying a liberal response standard that led them to identify guilty as well as innocent suspects.

Although developmental lineup studies often do include a target absent condition, more often than not, researchers separately analyze target present and target absent lineup conditions. Separate analyses of hit and false alarm rates do not allow for drawing conclusions about how performance differs in when the target is present versus absent. As pointed out by Gronlund and colleagues (2012), performance in target present and target absent lineups must be analysed simultaneously in order to draw inferences about memory processes (also see Gelman & Stern, 2006; Nieuwenhuis, Forstmann, & Wagenmakers, 2011). It is not correct, analytically speaking, to conclude that there are age-related performance differences when the target is absent, and not when the target is present, unless performance in target present and target absent lineups are examined in a factorial analysis.

As a means of disentangling differences in memory performance arising from sensitivity versus response bias, researchers (e.g., Spring et al., 2013) sometimes use SDT estimates of memory sensitivity (d' , A_z , A'), which are estimated from the hit and false alarm rate. However, the accuracy of SDT memory estimates rest on several assumptions that may not hold in the context of lineups (Mickes, Moreland, Clark, & Wixted, 2014). For example, SDT indices assume that the target and lure distributions have equal variances, which in practice is seldom the case (Verde, MacMillan, & Rotello, 2006), especially in studies of recognition memory (Ratcliff, Sheu, & Gronlund, 1992). SDT estimates of memory sensitivity will not be accurate if these assumptions are violated. Finally, the diagnosticity ratio, which is the ratio of guilty to innocent suspect identifications (correct ID rate/false ID rate), has been typically used by researchers to evaluate lineup performance. Using this method, adults' lineup performance has been associated with a higher diagnosticity ratio than

children's (6 – to- 12 years), and therefore linked with better lineup performance (e.g., Fitzgerald, Whiting, Therrien, & Price, 2014). However, the diagnosticity ratio conflates memory sensitivity and response bias (see Mickes, et al., 2014).

Recently, Receiver Operator Characteristic (ROC) analysis was introduced in the adult eyewitness memory literature (Gronlund et al., 2012; Mickes, Flowe, & Wixted, 2012) to address these issues. ROC analysis does not rely on any distribution assumptions, and allows one to examine memory sensitivity across age groups, regardless of whether responding is liberal or conservative. In this paper, we describe the approach, and apply it to test whether the late maturation hypothesis holds in the eyewitness identification context. To illustrate how to conduct an ROC analysis (also see Gronlund, et al., 2014), imagine we have run an eyewitness memory experiment with adults and children. For half of the participants, a previously studied target was present in the lineup and for the other half, he was absent. A lineup identification response (i.e., selecting a particular face, or not selecting anyone) and a confidence rating were obtained from each participant. Confidence estimates the strength of the memory signal on which the witness' identification is based. Participants provide this information using a scale that ranges from 0% to 100% certainty. An eyewitness who makes a positive identification and reports a relatively low confidence level is indicating that their memory signal is relatively weak; consequently, we would know that this eyewitness has made a positive identification using a relatively lenient response criterion.

Using the identification response and confidence data, we can construct an ROC curve for each age group, and compare the curves to assess age-related performance differences. To construct the ROC curve for a given age group, first the hit and false alarm rate is determined at each confidence level. For each confidence level, we calculate the proportion of people who identified the perpetrator in the target present condition, and the proportion of people who identified the innocent suspect in the target absent condition. The

cumulative hit rate and the cumulative false alarm rate are then obtained for every confidence level, starting with the highest confidence level. To display the ROC, the cumulative hit and false alarm rate data are plotted for each confidence level. The resulting ROC is the family of hit and false alarm rates that are obtained across the full range of the confidence scale.

The two hypothetical ROC curves are presented in Figure 1, Panel A, illustrate what the results of an experiment might look like if memory sensitivity was greater for adults compared to children. The hit rate is plotted on the x-axis, and the false alarm rate is plotted on the y-axis. Each data point represents a level of confidence. As we move from left to right on the x-axis, the confidence level decreases (i.e., the response criterion used to make the identification becomes more lenient), and the false alarm rate increases. Note that the false alarm rate on the x-axis is truncated and does not extend to 1.00. This is because in a lineup experiment in which fillers sufficiently resemble the suspect, participants choose the innocent suspect and fillers more often when the response threshold becomes more lenient. Consequently, the false alarm rate for innocent suspects will not reach 1.00. In the hypothetical example shown in Figure 1, Panel A, as we move along the curves from left to right, we can see that the hit rate is always larger for adults compared to children. Put differently, the hit rate for adults is always higher compared to children, holding the confidence level constant. Memory sensitivity is better the further away the ROC bows away from the diagonal line, which represents chance performance. We can statistically compare these differences by evaluating whether the area under the curves is larger for adults compared to children (see Gronlund et al, 2014, for a tutorial).

To further demonstrate ROC analysis and how it can be applied to assess age-related differences in lineup performance, the hypothetical data in Figure 1, Panel B, illustrate what the ROCs might look like if memory sensitivity did not significantly differ by age group. Instead, here the key difference between age groups is that the ROC for children is shifted

rightwards on the x-axis relative to the ROC for adults, reflecting that children adopted a more lenient decision standard relative to adults on average. At every confidence level, the hit rate does not markedly differ by age group; instead, the false alarm rates are larger for children compared to adults. We would expect these results if children and adults differed only in terms of response bias on the lineup test.

Study Aims

To date, ROC analysis has not been used to compare the identification performance of children and adults. Our primary aim was to test using ROC analysis whether the late maturation hypothesis can be generalized to an eyewitness memory context. Toward this end, we reanalyzed the data from Humphries et al. (2012). The researchers found, in keeping with other studies in the literature, that adults performed better than children on target absent lineups. Additionally, adults and children performed similarly in target present lineups, except when the lineup was presented in a sequential format. Hence, their data seem to support a response bias interpretation, rather than the late maturation hypothesis. However, they separately analysed the target present and the target absent conditions. What is more, they did not use ROC analysis, and therefore, response bias and memory sensitivity are conflated. In this study, we will construct the ROC curves using the identification data and confidence ratings that they obtained from each participant; heretofore, the confidence ratings for this dataset have not been reported. Importantly, we believe the analysis will demonstrate the utility of the ROC approach in testing developmental theories about memory in the eyewitness memory domain.

Our second aim was to assess the relationship between confidence and accuracy as a function of age group. This analysis has not been previously reported for this data set. ROC analysis is able to answer whether there are age-related differences in discriminating the target from fillers; it is not, however, necessarily informative with respect to whether

confidence is predictive of accuracy (see Diamond, 1992). If we found, for instance, that memory sensitivity for children does not differ from adults, this would not necessarily mean that children are well-calibrated in their confidence assessments.

Developmental research generally indicates that children starting from around 8 years old are able to effectively monitor the accuracy of recall (Howie & Roebbers, 2007; Roebbers, 2002; Roebbers & Howie, 2003) and recognition (Hiller & Webber, 2013; Roebbers, 2002). Yet, there is also evidence that younger children can also monitor the accuracy of their judgments (Ackerman & Koriat, 2011; Howie & Roebbers, 2007; Roberts & Powell, 2005; von der Linden & Roebbers, 2006). Evidence of children's ability to monitor the accuracy of their memory in an eyewitness identification context is sparse, however. Research using the point-biserial correlation coefficient to examine the relationship between confidence and accuracy has indicated that children (up to 12 years old) have greater difficulty monitoring the correctness of their answers relative to adults (Leippe et al., 1991; Parker & Carranza, 1989; Paker, Haverfield & Baker-Thomas, 1986; Parker & Ryan, 1993) and adolescents (Brewer & Day, 2005). However, this method is known to underestimate the confidence-accuracy relationship (see Juslin, Olsson, & Winman, 1996). Research using an alternative method, the calibration approach, which assesses whether confidence ratings provide a probabilistic estimate of accuracy, children's meta-cognitive judgments have also been found to be less accurate than adults (Keast, Brewer, & Wells, 2007). Yet, for adult witnesses, when confidence is measured immediately following the lineup identification decision, confidence can predict accuracy (see Mickes et al., 2012, and Roediger, Wixted, & DeSoto, 2012, for a review). In view of this literature, we expected to find age-related changes in the relationship between confidence and accuracy, with the association being weaker in children compared to adults.

Method

Participants

A total of 180 adults (18-49 years of age, $M = 20$ years, $SD = 3.41$) were recruited from a psychology undergraduate course and they received course credit for their participation. A total of 360 children were also recruited; young children ($n = 180$, 5- to 6-years old, $M = 5.8$ years, $SD = .3$ years) and older children ($n = 180$, 9- to 10-years old, $M = 9.8$ years, $SD = .3$ years) were recruited from primary schools in the South East of England. Consent from head teachers and legal guardians were obtained.

Design

A 3 (video lineup presentation: simultaneous, sequential, elimination) x 3(Age: 5- to 6-year-olds, 9- to 10-year olds, adults) x 2 (lineup type: target-present, target-absent) between-subjects design was employed. Dependent measures were lineup identification responses and decision accuracy confidence judgments.

Materials and Procedure

Event: A videotaped nonviolent theft was utilised. The film depicted a young male in his early 20s acting suspiciously whilst browsing in a clothing store. Also depicted in the film was a female (22 years old); the young woman's face was never visible, her back was always to the camera. The young man, noticing that the woman's bag was open, reached in and stole the woman's purse and placed it into his own bag before leaving the shop in full view of the camera. The event lasted for 75 s. The culprit remained in view for the entire recording and was filmed from various angles including: close-up, left and right profile, front and back profiles.

Lineup construction and procedures: Lineup foils were selected based on their match to the description of the perpetrator. In target-present lineups, the position of the target was

held constant across participants as evidence indicates that the position of the target may be critical, specifically in sequential lineups (Ebbesen & Flowe, 2002; Clark & Davey, 2005; Carlson, Gronlund, & Clark, 2008; Gronlund, Carlson, Dailey, & Goodsell, 2009). The target always appeared in position four. In target-absent lineups all 6 lineup members were foils, there was no designated innocent suspect. Tredoux's E (1998) analysis was used to assess the fairness of the lineups (please see Humphries et al., 2012 for full details of the lineup construction procedure).

Each lineup consisted of six moving image profiles of the individual lineup members. Image clips were colour, head-and-shoulder shots. The clips depicted a face moving in a 180 degree motion, from a full frontal pose to left profile pose, right profile pose and a final full frontal pose position. The entire film sequence lasted approximately 15 s for each clip. Three lineup procedures were employed, a traditional sequentially administered video lineup, and two novel lineups, a simultaneous video lineup and a simultaneously administered video elimination lineup. Previous research has shown that children, but particularly younger children have difficulty with the sequential lineup procedure (Lindsay et al., 1997). A procedure that has been shown to be effective at improving children and adults identification performance is the elimination lineup procedure (Pozzulo & Balfour, 2006; Pozzulo, Dempsey, & Crescini, 2009; Pozzulo & Lindsay, 1999,). However, research suggests that the elimination lineup must be simultaneously administered for this beneficial effect to be observed (see Beresford & Blades, 2006). To enable the simultaneous administration of video lineups and to ensure that each lineup member's video clip was viewed in their entirety a number of modifications to the traditional simultaneous lineup procedure were introduced.

In the simultaneous video lineup condition, the six image clips were presented in a 3 x 2 array. A static full frontal picture of each lineup member was immediately visible to

participants. A corresponding number (1 - 6) and the word play appeared beneath each clip.

Each participant was provided with the following verbal instructions:

“I am now going to show you some pictures. The man from the film who stole the woman’s purse may be in the pictures that I show you or he may not. I am first going to ask you to look at each picture separately. Each of the pictures is a moving picture. Please look at each picture very carefully as pictures may look very similar to each other.”

After receiving these instructions the experimenter proceeded to play each lineup member’s video clip sequentially in ascending numerical order. When each individual image clip had finished, a full frontal static picture of each lineup member was again visible to and participants were asked to provide their identification decision. The procedures for the video elimination lineups were identical to the simultaneous video lineup procedure apart from the addition of the two-stage identification process. As in the simultaneous condition, each video clip presented in a 3 x 2 array was played in ascending numerical order. When each video clip had finished a static, full frontal image of each lineup member was again visible to participants. After all video clips had been viewed, participants were asked to select (from the visible static images) the person that looked most like the man from the film. Once a lineup member was selected, the remaining five lineup member’s images were removed so that only the static image of the chosen lineup member remained visible. Lineup instructions were based on those used by Pozzulo and Lindsay (1997, 1999); some minor adjustments were made to make them more appropriate for use with video lineups. Instructions provided to participants in the video elimination lineup were as follows:

“I am now going to show you some pictures. The man from the film who stole the woman’s purse may be in the pictures that I show you

or he may not. I am first going to ask you to look at each picture separately. Each of the pictures is a moving picture. Please look at each picture very carefully as pictures may look very similar to each other. When you have seen all of the pictures I want you to tell me which one of the pictures looks most like the man in the film who stole the woman's purse."

Next, with only the participant's chosen lineup members' image visible, participants then received the following verbal instructions and were asked to provide their identification decision:

"I asked you to pick the picture that looked most like the man in the film who stole the woman's purse. Now I am going to let you see the clip again and I would like you to think carefully about what the man in the film looked like. If this is the man from the film who stole the woman's purse then I would like you to tell me that it is him. If it is a picture somebody who just looks like him, I want you to tell me that it is not the man from the film."

In the sequential video lineup individual lineup member clips were presented one-at-a-time and the lineup procedure was terminated following a positive identification, otherwise a "not there" response was recorded. Participants in the sequential video lineup received the following verbal instructions:

"I am now going to show you some pictures. The man from the film who stole the woman's purse may be in the pictures that I show you or he may not. You will be shown one picture at a time. Each of the pictures is a moving picture. Please look at each picture carefully as you will only get to see each picture once and you will not be able to

go back through the pictures. Therefore, it is important that you tell me as soon as you think you see a picture of the man from the film. Once you have chosen a picture, I will stop, and you will not be allowed to see any additional pictures.”

Participants viewed the film separately. Participants were asked to watch the video carefully as they would be asked some questions about the film later. Immediately after viewing the film, participants completed a filler task (pencil-and-paper puzzles) for approximately 10 minutes. Next, participants viewed either a simultaneous, elimination, or sequential video lineup procedure, that was either target-present or target-absent. For all lineup procedures participants were not permitted to, a) to pause a video clip, b) play an individual lineup member’s video clip more than once, or c) to view the video clips out of numerical sequence. Following the lineup identification procedure, witnesses were asked to indicate their confidence in their lineup identification decision. Specifically, children and adults were asked to indicate how sure they were that they had made the correct decision. This was recorded using a 7-point scale (1 = not at all confident, 7 = extremely confident) for the adults and a 4-point scale (1 = very unsure, 4 = very sure) for the children. We used different confidence scales for children and adults so that the scales implemented were age appropriate, particularly for younger children (see Howie & Roebbers, 2007; von der Linden & Roebbers, 2006). We are unaware of any reason why memory sensitivity in children and adults cannot be compared under these circumstances. Research with adult participants using a continuous confidence measure has shown that sensitivity does not vary across reduced compared to full rating scales (see Benjamin, Tullis, & Lee, 2013, Figure 2, 4- versus 8-point rating scale; also see Hadjiiski, Chan, Sahiner, Helvie, & Roubidoux, 2007).

For children, a piece of card displaying 4 distinct clearly labelled “smiley” faces was presented to children (see Figure 2, adapted from Roebbers, 2002). This scale is similar to those implemented in studies where younger and older children have been shown to successfully monitor the accuracy of their reports (e.g., Howie & Roebbers, 2007; Roebbers, 2002; Roebbers & Howie, 2003; von der Linden & Roebbers, 2006). To ensure that children understood the confidence rating scale a number of example questions were included. For example, children were asked the question, “How old are you?”, and were then asked to indicate how sure they were that their answer was correct by pointing to the appropriate confidence rating (cf. Roebbers, 2002). If children did not point to the most appropriate confidence rating, the experimenter showed children the correct confidence rating and provided an explanation for this decision. Older children learnt the appropriate use of the confidence scale relatively quickly, but younger children generally required additional training questions. Only when children had learned the appropriate use of the confidence scale were they asked to rate their post-identification confidence rating.

Results

Age-related Differences in Performance

Table 1 provides the identification responses that were obtained for adults and children in the target present and target absent conditions, collapsed across lineup type. ROC curves were generated using the identification response and confidence data. We calculated the proportion of hits (computed by dividing the number of target identifications by the number of participants in the target present condition) and proportion of false alarms (computed by dividing the number of filler identifications in the target absent condition by the number of participants in the target absent condition) that were made at each confidence level. We then cumulated the proportion of hits and false alarms across confidence levels, starting with the highest level of confidence and ending with lowest level of confidence. The

false alarm rates obtained at each confidence level were divided by the total number of faces in the lineup (i.e., 6) to obtain an estimate of what the false alarm rate would have been if an innocent suspect had been designated, following Mickes et al. (2012). The resulting ROCs for each age group collapsed across lineup condition are shown in Figure 3.

We examined the late maturation hypothesis by comparing the performance of adults to young children, adults to older children, and young children to older children. In making these pairwise performance comparisons, partial area under the curve analysis (pAUC) was undertaken to test whether there were differences in memory sensitivity. The analysis was performed with pROC (Robin et al., 2011), a data analysis pack available for R. pAUC is calculated as the area under the ROC curve between two false alarm rates. pAUC analysis is appropriate because the false alarm rate is constrained to be less than 1 in a fair lineup. In a fair lineup, the odds that the suspect is selected based on chance alone are reduced by the inclusion of fillers. In any given pAUC we undertook, the area of the ROC space was defined by a value less than or equal to the larger of the two false alarm rates that were obtained for the two groups being compared. If the pAUC comparison did not reach statistical significance, thereby indicating that there were no differences across groups in memory sensitivity, we then tested whether there were differences across the two age groups in response bias. This was done by comparing the false alarm rates (i.e., the choice rate in the target absent condition) using z tests. The alpha level was Bonferroni-corrected to .008 to control for Type I errors because there were six comparisons (three ROC comparisons and potentially three false alarm rate comparisons) that were going to be made across the age groups.

Young children versus adults. The pAUC for young children was significantly smaller compared to adults (pAUC = .18 versus .45, respectively, $D = 6.62$, $p < .00001$). Hence, memory sensitivity was poorer for young children compared to adults.

Young children versus older children. The pAUC for young children was significantly smaller compared to older children (pAUC = .18 versus pAUC = .36, respectively, $D = 4.28$, $p < .00001$). Therefore, memory sensitivity for young children was also poorer compared to older children.

Older children versus adults. The pAUC for older children did not differ compared to adults (.19 versus .25, respectively), $D = 1.90$, $p = .03$, indicating similar levels of memory sensitivity across the age groups. The false alarm rate in the target absent condition for older children compared to adults also did not differ (.29 versus .44, respectively, $z = 2.10$, $p = .04$, two-tailed), indicating that response bias was not reliably different across the age groups.

Age-related performance across lineup conditions. One might wonder whether there was any evidence that age-related differences in sensitivity should be qualified by the type of lineup procedure conducted. For descriptive purposes, Figure 4 presents the ROCs for each age group by lineup condition. As can be seen, performance for young children fell on the lowest ROC curve, performance for adults fell on the highest ROC curve, and performance for older children was intermediate, falling between the other two age groups. Thus, the age-related pattern of performance was consistent across all of the lineup procedures. However, inferences should not be drawn from Figure 4 about whether or not age-related performance differences in sensitivity differ depending on lineup type. The sample size within lineup condition was low; hence, inferential statistical comparisons are not presented.

Confidence and Accuracy

The relationship between confidence and accuracy for choosers by age group is shown in Figure 4, which plots average accuracy against participants' confidence ratings. Due to few responses in the first and last confidence categories for adults, data were collapsed across confidence levels 1 and 2, and 6 and 7. Following the approach used by Brewer and Day (2005), we analysed the correlation between confidence and accuracy for

each age group using r_{pb} . These results should be interpreted cautiously, as r_{pb} is known to underestimate the relationship between confidence and accuracy (see Juslin, et al., 1996). As can be seen in Figure 4, confidence and accuracy were significantly associated for adults ($r_{pb} = .24, p < .05$, two-tailed) and older children ($r_{pb} = .37, p < .001$, two-tailed), whereas they were not associated for younger children ($r_{pb} = .08, p > .05$). The strength of the association between confidence and accuracy did not differ for older children compared to adults.

Taken together, the results are consistent with the late maturation hypothesis, with poorer memory sensitivity found for young children compared to older children and adults. Additionally, the analysis of confidence and accuracy for choosers indicated that young children's confidence ratings were not predictive of accuracy.

Discussion

Researchers most often report that while children do not differ from adults on target present lineups, they perform less accurately than adults on target absent lineups (for reviews see, Havard, 2013; Pozzulo, 2007; Pozzulo & Lindsay, 1998). It is not clear, however, whether these findings can be accounted for by the late maturation hypothesis, because previous analyses have not properly taken into account developmental differences in response bias. We report here the first confidence-based ROC analysis of age-related changes in lineup identification performance. ROC analysis separates memory sensitivity from response bias, thereby allowing researchers to compare age-related lineup performance differences.

Our results indicated that the late maturation hypothesis generalises to an eyewitness memory context. We found, using the ROC approach, that younger children's memory sensitivity was indeed poorer compared to older children and adults. Older children, however, did not differ from adults with respect to memory sensitivity and response bias. Therefore, young children seem to have deficiencies in discriminating the target from the fillers on a lineup test compared to adults, while older children do not appear to have such deficiencies.

Recent evidence from the developmental face recognition literature indicates that aspects of face perception are qualitatively adult-like early in childhood, but efficiency and speed of these processes may improve with age (Crookes & McKone, 2009; Jeffery, Rathbone, Read, & Rhodes, 2013). In the present study, older children may have been more able than younger children to effectively use the additional information provided by the moving images to improve their performance to adult-like levels (see Havard, Memon, Clifford, & Gabbert, 2010). While nothing can be done necessarily to help children better encode the perpetrator, further research studies employing ROC analyses could establish whether there are procedural variations in the administration of lineups that significantly enhance younger children's ability to discriminate the target from the fillers (for a review of lineup identification procedures see Pozzulo, 2007; Pozzulo & Lindsay, 1998).

In addition to having poorer memory sensitivity than adults, young children's confidence ratings were not as predictive of their accuracy. These findings are in line with other eyewitness studies, which have largely indicated that children (up to 12 years old) have greater difficulty monitoring the correctness of their answers, relative to adults (Leippe et al., 1991; Parker & Carranza, 1989; Parker, et al., 1986; Parker & Ryan, 1993). In work that has employed a calibration approach, deficits have also been found in in children's (11- year olds) ability to discriminate between correct and incorrect decisions, relative to adults (Keast, et al., 2007). However, modest correlations between confidence and accuracy were observed for 11 year old and adult choosers in the same study. The pattern of results reported here are consistent with the notion that the ability to accurately assess memory strength may not emerge until around 8 – 10 years of age (e.g. Veenman, Van Hout-Wolters, & Afflerbach, 2006).

A reduced confidence scale was used in the present study, wherein the rating scale had four points and omitted the mid-point option (*'I don't know'*). Although the size of the

rating scale (i.e., number of options) and removal of the '*I don't know*' option have not been found to influence older children's observed metacognitive abilities (e.g., Roebbers 2002, Roebbers & Howie, 2003; Roebbers von der Linden, & Howie, 2007), additional research is needed to ascertain whether the number of points on the confidence scale affects the shape of the ROC. Additionally, researchers have questioned the appropriateness of comparing confidence judgments made by children to other age groups, as it is unclear if confidence judgments measure the same construct in children and adults (e.g., Roebbers, Krebs, & Roederer, 2014). The results of this study suggest that ROC analyses can be undertaken with children's confidence ratings to examine age-related changes in memory sensitivity. As can be seen in Figure 3, at every confidence level, the hit rate is larger for adults compared to young children, reflecting that adults have greater memory sensitivity. More work, however, is clearly needed to determine better methods for measuring confidence in children.

Rather than rely on confidence ratings, researchers can instead use instructional bias manipulations (e.g., instructions that affect participants' willingness to guess from the lineup) to experimentally control response bias. By doing so, the study would be attempting to equate response bias levels across age groups so that age-related differences in memory sensitivity can be examined. d' can be calculated from the hit and false alarm rate within each instruction condition for each age group. Mickes and colleagues (2014) provide evidence from adult participants that d' can provide a valid measure of memory sensitivity in the context of lineups, although they point out that an ROC is preferable, because it does not rely on any underlying assumptions about the data being analysed.

Like other developmental research in the eyewitness area, we used a single target (Beresford & Blades, 2006; Lindsay & Wells, 1985; Pozzulo & Balfour, 2006; Pozzulo & Dempsey, 2006; Karageorge & Zajac, 2011), raising the possibility that our findings could be specific to the target used. A single event was also used, and as others have pointed out, the

type of event may influence motivation, attention and ultimately memory strength (Pozzulo, Crescini, & Panton, 2008; Roebbers, Gelhaar, Schneider, 2004; Thierry & Spence, 2004). On the other hand, the pattern of hit and false alarm rates we obtained by age group (see Table 1) is consistent with the overwhelming majority of studies reported in the literature, indicating that our findings in all likelihood are probably not specific to the stimuli that we used.

Additionally, whilst event-related factors that affect memory strength impact accuracy, there is no reason to think that these factors would differentially impact children compared to adults. Finally, the present study employed moving test images, and further ROC studies are needed to assess children's and adult's performance on photographic lineups. Employing moving as opposed to static photographic lineups is the recommended practice for lineups conducted in the UK, in Australia and in some jurisdictions in the US. Research examining the effects of movement on children's (Beresford & Blades, 2006; Havard, et al., 2010) and adult's (Cutler & Fisher, 1990; Darling, Valentine, & Memon, 2008; Valentine, Darling, & Memon, 2007) identification performance has produced mixed results. In the present study, it is possible that viewing moving faces may have provided children and adults with additional information that may have enhanced memory retrieval. Movement was not directly manipulated, however. What is more, we had to collapse across lineup procedures in our statistical analyses due to sample size limitations. Although this approach is not ideal, it can be seen from Figure 4 that the pattern of results was consistent across lineup procedures.

The National Academy of Sciences recently concluded that ROC analysis is a more comprehensive approach to analysing lineup identification data compared to other measures of performance that have been used (National Research Council, 2014). In the main, the approach allows one to examine the contribution of response bias and memory sensitivity to eyewitness identification responses. This is especially critical in developmental research because response bias varies with age. While ROC analysis offers developmental researchers

several advantages over other methods commonly used in the lineup literature (i.e., age comparisons that involve the diagnosticity ratio, or comparisons of the overall hit and false alarm rates), there are caveats to consider. First, one particularly important issue is that confidence should be predictive of accuracy. In the adult eyewitness literature, confidence, if it is taken immediately after the lineup identification decision, is predictive of accuracy (see Mickes et al., 2012, and Roediger, Wixted, & DeSoto, 2012, for a review). The evidence is mixed in children, however, with some studies indicating that confidence can be predictive of accuracy (e.g., Leippe et al., 1991; Parker & Carranza, 1989; Parker & Ryan, 1993), yet others finding that children's confidence is not predictive of accuracy (e.g., Brewer & Day, 2005; Keast et al., 2007). In the present study, memory sensitivity was higher for adults compared to young children at every confidence level, indicating strong support for the late maturation hypothesis. However, we also found that confidence was predictive of accuracy for older children and adults, but not for young children. Further work is needed with young children to examine other methods for measuring response bias besides confidence. Second, there are potentially other approaches besides pAUC analysis that have yet to be tested by eyewitness researchers to evaluate eyewitness performance, and further research is needed to evaluate these alternatives (see Hand, 2009, and National Research Council, 2014, for a fuller discussion of the advantages and limitations of ROC analysis).

In summary, first, our results indicate that the late maturation hypothesis generalizes to an eyewitness memory task. Young children's memory sensitivity in lineups was poorer compared to older children and adults. In contrast, memory sensitivity for older children did not differ from adults. Second, our results indicated that not only was memory sensitivity poorer in young children, but their confidence ratings were not predictive of accuracy. Finally, our results provide a compelling demonstration that confidence-based ROC analysis should be employed to test developmental theories about eyewitness memory.

References

- Ackerman, R., & Koriat, A. (2011). Response latency as a predictor of the accuracy of children's reports. *Journal of Experimental Psychology: Applied*, *17*(4), 406-417. doi:10.1037/a0025129
- Beal, C. R., Schmitt, K. L., & Dekle, D. J. (1995). Eyewitness identification of children: Effects of absolute judgements, nonverbal response options, and event encoding. *Law and Human Behavior*, *19*, 197-216. doi:10.1007/BF01499325
- Benjamin, A. S., Tullis, J. G., & Lee, J. (2013). Criterion noise in ratings-based recognition: Evidence from the effects of response scale length on recognition accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1601-1608. doi:10.1037/a0031849
- Beresford, J., & Blades, M. (2006). Children's identification of faces from lineups: The effects of lineup presentation and instructions on accuracy. *Journal of Applied Psychology*, *91*, 1102-1113. doi:10.1037/0021-9010.91.5.1102
- Brewer, N., & Day, K. (2005). The confidence-accuracy and decision latency-accuracy relationships in children's eyewitness identification. *Psychiatry Psychology and Law*, *12*, 119-128. doi:10.1375/pplt.2005.12.1.119
- Brewer, N., Weber, N., & Semmler, C. (2005). Eyewitness identification. In N. Brewer & K. D. Williams (Eds.), *Psychology and Law: An Empirical Perspective* (pp. 177-221), New York: Guilford.
- Brigham, J. C., Van Verst, M., & Bothwell, R. K. (1986). Accuracy of children's eyewitness identifications in a field setting. *Basic and Applied Social Psychology*, *7*(4), 295-306. doi:10.1207/s15324834basp0704_4

- Bruce, V., Campbell, R. N., Doherty-Sneddon, G., Import, A., Langton, S., McAuley, S., & Wright, R. (2000). Testing face processing skills in children. *British Journal of Developmental Psychology, 18*, 319–333. doi:10.1348/026151000165715
- Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science, 195*(4275), 312–314. doi:10.1126/science.831281
- Carey, S., Diamond, R., & Woods, B. (1980). Development of face recognition: A maturational component? *Developmental Psychology, 16*, 257–269. doi:10.1037/0012-1649.16.4.257
- Carlson, C. A., Gronlund, S. D., & Clark, S. E. (2008). Lineup composition, suspect position, and the sequential lineup advantage. *Journal of Experimental Psychology: Applied, 14*(2), 118–128. doi:10.1037/1076-898X.14.2.118
- Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113*, 403–409. doi:10.1037/0033-2909.113.3.403
- Ceci, S. J., Ross, D. F., & Toglia, M. P. (1987). Suggestibility of children's memory: Psycholegal implications. *Journal of Experimental Psychology: General, 116*, 38–49. doi:10.1037/0096-3445.116.1.38
- Clark, S. E., & Davey, S. L. (2005). The target-to-foils shift in simultaneous and sequential lineups. *Law and Human Behavior, 29*, 151–172. doi:10.1007/s10979-005-2418-7
- Crookes, K., & McKone, E. (2009). Early maturity of face recognition: No childhood development of holistic processing, novel face encoding, or face-space. *Cognition, 111*, 219–247. doi:10.1016/j.cognition.2009.02.004
- Cutler, B. L., & Fisher, R. P. (1990). Live lineups, videotaped lineups and photoarrays. *Forensic Reports, 3*, 439–448.
- Darling, S., Valentine, T., & Memon, A. (2008). Selection of lineup foils in operational contexts. *Applied Cognitive Psychology, 22*, 159–169. doi:10.1002/acp.1366

- Davies, G. M. (1996). Children's identification evidence. In S. L. Sporer, R. Malpass, & G. Kohnken (Eds.), *Psychological Issues in Eyewitness Identification* (pp. 233-258). Mahwah, NJ: Erlbaum.
- Dekle, D. J., Beal, C. R., Elliott, R., & Huneycutt, D. (1996). Children as witnesses: A comparison of lineup versus showup identification methods. *Applied Cognitive Psychology, 10*, 1-12. doi:10.1002/(SICI)1099-0720(199602)10:1<1::AID-ACP354>3.0.CO;2-Y
- Diamond, G. A. (1992). What price perfection? Calibration and discrimination of clinical prediction models. *Journal of Clinical Epidemiology, 45*, 85-89. doi:10.1016/0895-4356(92)90192-P
- Ebbesen, E. B., & Flowe, H. D. (2002). *Simultaneous v. sequential lineups: What do we really know?* Retrieved from <http://www.psy.ucsd.edu/%7eeeb-bes-en/Sim-Seq.htm>
- Fitzgerald, R. J., Whiting, B. F., Therrien, N. M., & Price, H. L. (2014). Lineup Member Similarity Effects on Children's Eyewitness Identification. *Applied Cognitive Psychology, 28*(3), 409-418. doi:10.1002/acp.3012
- Gelman, A. & Stern, H. (2006). The difference between 'significant' and 'not significant' is not itself statistically significant. *American Statistician, 60*, 328-331.
- Goodman, G. S., & Reed, R. S. (1986). Age differences in eyewitness testimonies. *Law and Human Behavior, 10*, 317-332. doi:10.1007/BF01047344
- Gronlund, S. D., Carlson, C. A., Dailey, S. B., & Goodsell, C. A. (2009). Robustness of the sequential lineup advantage. *Journal of Experimental Psychology. Applied, 15*(2), 140-152. doi:10.1037/a0015082
- Gronlund, S. D., Carlson, C. A., Neuschatz, J. S., Goodsell, C. A., Wetmore, S., Wooten, A. & Graham, M. (2012). Showups versus lineups: An evaluation using ROC analysis.

Journal of Applied Research in Memory and Cognition, 1(4), 221-228.

doi:<http://dx.doi.org/10.1016/j.jarmac.2012.09.003>

- Gronlund, S. D. & Neuschatz, J.S. (2014). Eyewitness identification discriminability: ROC analysis versus logistic regression. *Journal of Applied Research in Memory and Cognition*, 3, 54-57. <http://dx.doi.org/10.1016/j.jarmac.2014.04.008>
- Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating Eyewitness Identification Procedures Using Receiver Operating Characteristic Analysis. *Current Directions in Psychological Science*, 23(1), 3-10. doi:10.1177/0963721413498891
- Gross, J., & Hayne, H. (1996). Eyewitness identification by 5- to 6-year-old children. *Law and Human Behavior*, 20, 359-373. doi:10.1007/BF01499028.
- Hadjiiski L., Chan H. P., Sahiner B., Helvie M. A., & Roubidoux, M. A. (2007). Quasi-continuous and discrete confidence rating scales for observer performance studies: effect on ROC analysis. *Academic Radiology*, 14, 38-48.
- Hand, D. J. (2009). Measuring classifier performance: A coherent alternative to area under the ROC curve. *Machine Learning*, 77, 103-123.
- Havard, C. (2013). Are children less reliable at making visual identifications than adults? A Review. *Psychology, Crime & Law*. doi:10.1080/1068316X.2013.793334
- Havard, C., Memon, A., Clifford, B., & Gabbert, F. (2010). A comparison of video and static photo lineups with child and adolescent witnesses. *Applied Cognitive Psychology*, 24, 1209-1221. doi:10.1002/acp.1645
- Hiller, R. M., & Weber, N. (2013). A comparison of adults' and children's metacognition for yes/no recognition decisions. *Journal of Applied Research in Memory and Cognition*, 2(3), 185-191. doi:10.1016/j.jarmac.2013.07.001

- Howie, P., & Roebbers, C. M. (2007). Developmental progression in the confidence-accuracy relationship in event recall: Insights provided by a calibration perspective. *Applied Cognitive Psychology, 21*, 871–893. doi:10.1002/acp.1302
- Huges, M., & Grieve, R. (1980). On asking children bizarre questions. *First Language, 1*, 149-160. doi:10.1177/014272378000100205
- Humphries, J. E., Holliday, R. E., & Flowe, D. H. (2012). Faces in motion: Age-related changes in eyewitness identification performance in simultaneous, sequential, and elimination video lineups. *Applied Cognitive Psychology, 26*, 149-158. doi:10.1002/acp.1808
- Jeffery, L., Rathbone, C., Read, A., & Rhodes, G. (2013). Children's face identity representations are no more view specific than those of adults. *Journal of Experimental Psychology: Human Perception and Performance, 39*(2), 450-463. doi:10.1037/a0029141
- Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence–accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, And Cognition, 22*(5), 1304-1316. doi:10.1037/0278-7393.22.5.1304
- Karageorge, A., & Zajac, R. (2011). Exploring the effects of age and delay on children's person identifications: Verbal descriptions, lineup performance, and the influence of wildcards. *British Journal of Psychology, 102*(2), 161-163. doi:10.1348/000712610X507902
- Keast, A., Brewer, N., & Wells, G. L. (2007). Children's metacognitive judgments in an eyewitness identification task. *Journal of Experimental Child Psychology, 97*, 286–314. doi:10.1016/j.jecp.2007.01.007

- King, M. A., & Yuille, J. C. (1987). Suggestibility and the child witness. In S. J. Ceci, M. P. Toglia & D. F. Ross (Eds), *Children's Eyewitness Memory* (pp. 24-35). New York:
- Leippe, M. R., Romanczyk, A., & Manion, A. P. (1991). Eyewitness memory for a touching experience: Accuracy differences between child and adult witnesses. *Journal of Applied Psychology, 76*, 367–379. doi:10.1037/0021-9010.76.3.367
- Lindsay, R. C. L., Pozzulo, J., Craig, W., Lee, K., & Corber, S. (1997). Simultaneous lineups, sequential lineups, and showups: Eyewitness identification decisions of adults and children. *Law and Human Behavior, 21*(4), 391-404. doi:10.1023/A:1024807202926
- Lindsay, R. C., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70*(3), 556-564. doi:10.1037/0021-9010.70.3.556children. *Law and Human Behavior, 21*, 391–402. 10.1023/A:1024807202926
- Marin, B. V., Holmes, D. L., Guth, M., & Kovac, P. (1979). The potential of children as eyewitnesses. *Law and Human Behavior, 3*, 295–305. doi:10.1007/BF01039808
- Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: comparing diagnostic accuracy of simultaneous vs. sequential lineups. *Journal of Experimental Psychology: Applied, 18*(4):361-76. doi:10.1037/a0030609.
- Mickes, L., Moreland, M. B., Clark, S. E., & Wixted, J. T. (2014). Missing the information needed to perform ROC analysis? Then compute d' , not the diagnosticity ratio. *Journal of Applied Research in Memory & Cognition, 3*(2), 58-62. doi:10.1016/j.jarmac.2014.04.007
- Mondloch, C. J., Geldart, S., Maurer, D., & Le Grand, R. (2003). Developmental changes in face processing skills. *Journal of Experimental Child Psychology, 86*, 67–84. doi:10.1016/S0022-0965(03)00102-4

- Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception, 31*, 553–566.
doi:10.1068/p3339
- National Research Council (2014). *Identifying the Culprit: Assessing Eyewitness Identification*. Washington, DC: The National Academies Press.
- Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience, 14*(9), 1105-1107. doi:10.1038/nn.2886
- Parker, J. F., & Carranza, L. E. (1989). Eyewitness testimony of children in target-present and target-absent lineups. *Law and Human Behavior, 13*, 133-149.
doi:10.1007/BF01055920
- Parker, J. F., Haverfield, E., & Baker-Thomas, S. (1986). Eyewitness testimony of children. *Journal of Applied Social Psychology, 16*(4), 287-302. doi:10.1111/j.1559-1816.1986.tb01141.x
- Parker, J. F., & Ryan, V. (1993). An attempt to reduce guessing behavior in children's and adults' eyewitness identifications. *Law and Human Behavior, 17*, 11-26.
doi:10.1007/BF01044534
- Pike, G., Brace, N., & Kynan, S. (2002). *The visual identification of suspects: Procedures and practice* (Briefing Note 2/02). London: Home Office.
- Pozzulo, J. D. (2007). Lineup identification by children. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology: Memory for people* (Vol. 2, pp. 283–307). Mahwah, NJ: Lawrence Erlbaum and Associates.
- Pozzulo, J. D., & Balfour, J. (2006). Children's and adults' eyewitness identification accuracy when a culprit changes his appearance: Comparing simultaneous and

- elimination line-up procedures. *Legal and Criminological Psychology*, 11, 25-34.
doi:10.1348/135532505X52626
- Pozzulo, J. D., Crescini, C., & Panton, T. (2008). Does methodology matter in eyewitness identification research?: The effect of live versus video exposure on eyewitness identification accuracy. *International Journal of Law and Psychiatry*, 31(5), 430-437.
doi:10.1016/j.ijlp.2008.08.006
- Pozzulo, J. D., & Dempsey, J. (2006). Biased lineup instructions: Examining the effects of pressure on children's and adults' eyewitness identification accuracy. *Journal of Applied Social Psychology*, 36, 1381-1394. doi:10.1111/j.0021-9029.2006.00064.x
- Pozzulo, J. D., & Dempsey, J., & Crescini, C. (2009). Preschoolers' person description and identification accuracy: A comparison of the simultaneous and elimination lineup procedures. *Journal of Applied Developmental Psychology*, 30(6), 667-676.
doi:10.1016/j.appdev.2009.01.004
- Pozzulo, J. D., & Lindsay, R. C. L. (1997). Increasing correct identifications by children. *Expert Evidence*, 5, 126-132. doi:10.1023/A:1008875802767
- Pozzulo, J. D., & Lindsay, R. C. L. (1998). Identification accuracy of Children versus Adults: A meta-analysis. *Law and Human Behavior*, 22, 549-570.
doi:10.1023/A:1025739514042
- Pozzulo, J. D., & Lindsay, R. C. L. (1999). Elimination lineups: an improved identification procedure for child eyewitnesses. *Journal of Applied Psychology*, 84, 167-176.
doi:10.1037/0021-9010.84.2.167
- Pozzulo, J. D., & Warren, L. (2003). Descriptions and identifications of strangers by youth and adult eyewitnesses. *Journal of Applied Psychology*, 88, 315-323.
doi:10.1037/0021-9010.88.2.315

- Ratcliff, R., Sheu, C., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, *99*(3), 518-535. doi:10.1037/0033-295X.99.3.518
- Ricci, C. M., Beal, C. R., & Dekle, D. J. (1996). The effect of parent versus unfamiliar interviewers on children's eyewitness memory and identification accuracy. *Law and Human Behavior*, *20*(5), 483-500. doi:10.1007/BF01499037
- Roebbers, C. M. (2002). Confidence judgments in children's and adult's event recall and suggestibility. *Developmental Psychology*, *38*, 1052-1067. doi:10.1037/0012-1649.38.6.1052
- Roebbers, C. M., Gelhaar, T., & Schneider, W. (2004). 'It's magic!' The effects of presentation modality on children's event memory, suggestibility, and confidence judgments. *Journal of Experimental Child Psychology*, *87*(4), 320-335. doi:10.1016/j.jecp.2004.01.004
- Roebbers, C. M., & Howie, P. (2003). Confidence judgments in event recall: Developmental progression in the impact of question format. *Journal of Experimental Child Psychology*, *85*, 352-371. doi:10.1016/S0022-0965(03)00076-6
- Roebbers, C. M., von der Linden, N., & Howie, P. (2007). Favourable and unfavourable conditions for children's confidence judgments. *British Journal of Developmental Psychology*, *25*(1), 109-134. doi:10.1348/026151006X104392
- Roediger, H. L., Wixted, J. T. & DeSoto, K. A. (2012). The curious complexity between confidence and accuracy in reports from memory. In Nadel, Lynn, and Sinnott-Armstrong, Walter, (Eds.). *Memory and Law* (pp. 84 - 118). Oxford University Press, New York.
- Roberts, K. P., & Powell, M. B. (2005). Evidence of metacognitive awareness in young children who have experienced a repeated event. *Applied Cognitive Psychology*, *19*(8), 1019-1031. doi:10.1002/acp.1145

- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J., & Muller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12 (77), 1-8.
- Roebbers, C. M., Krebs, S. S., & Roderer, T. (2014). Metacognitive monitoring and control in elementary school children: Their interrelations and their role for test performance. *Learning & Individual Differences*, 29, 141-149. doi:10.1016/j.lindif.2012.12.003
- Schwarzer, G. (2000). Development of face processing: the effect of face inversion. *Child Development*, 71, 391-401. doi:10.1111/1467-8624.00152
- Spring, T., Saltzstein, H. D., & Peach, R. (2013). Children's eyewitness identification as implicit moral decision making. *Applied Cognitive Psychology*, 27, 139-175. doi:10.1002/acp.2871
- Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior*, 22, 217-237. doi:10.1023/A:1025746220886
- Thierry, K. L., & Spence, M. J. (2004). A Real-life Event Enhances the Accuracy of Preschoolers' Recall. *Applied Cognitive Psychology*, 18(3), 297-309. doi:10.1002/acp.965
- Valentine, T., Darling, S., & Memon, A. (2007). Do strict rules and moving images increase the reliability of sequential lineup procedures? *Applied Cognitive Psychology*, 21, doi:933-949. doi:10.1002/acp.1306
- Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and Learning: Conceptual and Methodological Considerations. *Metacognition and Learning*, 1, 3-14. doi:10.1007/s11409-006-6893-0
- Verde, M. F., Macmillan, N. A., & Rotello, C. M. (2006). Measures of sensitivity based on a single hit rate and false alarm rate: The accuracy, precision, and robustness of d' ,

A_z, and A'. *Perception & Psychophysics*, 68(4), 643-654.

doi:10.3758/BF03208765

von der Linden, N., & Roebbers, C. M. (2006). Developmental changes in uncertainty monitoring during an event recall task. *Metacognition and Learning*, 1, 213-228.

doi:10.1007/s11409-006-9001-6

Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin*, 133(5), 800-832.

doi:10.1037/0033-2909.133.5.800

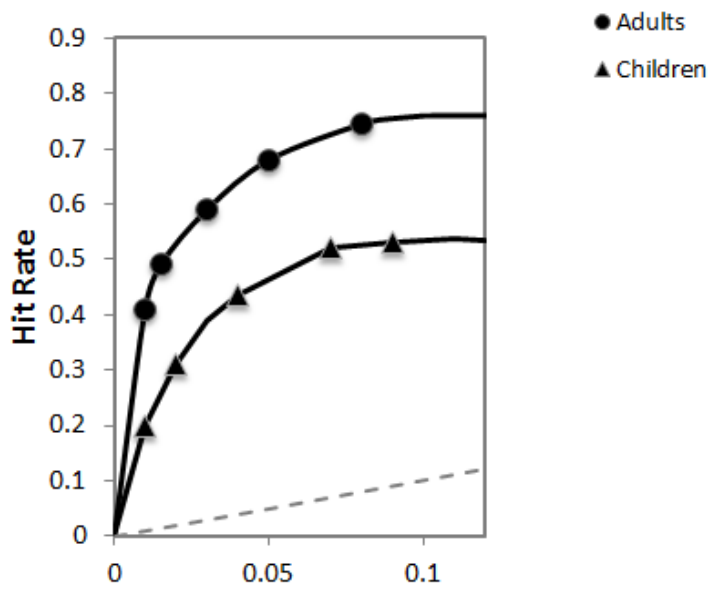
Table 1

Proportion (n) of correct and incorrect identification decisions as a function of lineup composition (TP, TA) and witness age (5- 6 year olds, 9 – 10 year olds, Adults)

	Age Group		
	Adults	9 – 10 years	5 – 6 years
Target-Present	<i>n</i> = 90	<i>n</i> = 90	<i>n</i> = 90
Correct identification	0.72 (65)	0.61 (55)	0.49 (44)
False alarm	0.09 (8)	0.21 (19)	0.41 (37)
Incorrect rejection	0.19 (17)	0.18 (16)	0.10 (9)
Target-Absent	<i>n</i> = 90	<i>n</i> = 90	<i>n</i> = 90
Correct rejection	0.71 (64)	0.56 (50)	0.32 (29)
False alarm ^a	0.29 (26)	0.44 (40)	0.68 (61)

^aNote: False alarm rate is the rate of filler identifications

Panel A:



Panel B:

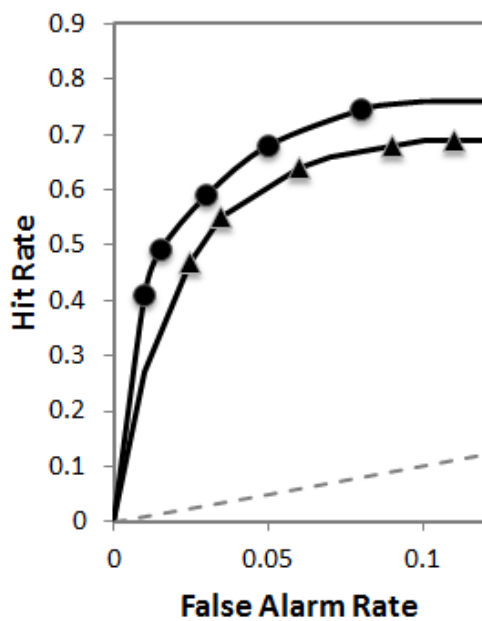


Figure 1. Panel A is a hypothetical illustration of what the results of a lineup experiment might look using ROC analysis if the late maturation hypothesis is correct. Here, adult performance falls on a higher ROC relative to young children, indicating adults have greater memory sensitivity. Panel B illustrates an alternative outcome, whereby the difference between age groups is better characterised as a difference in response bias.

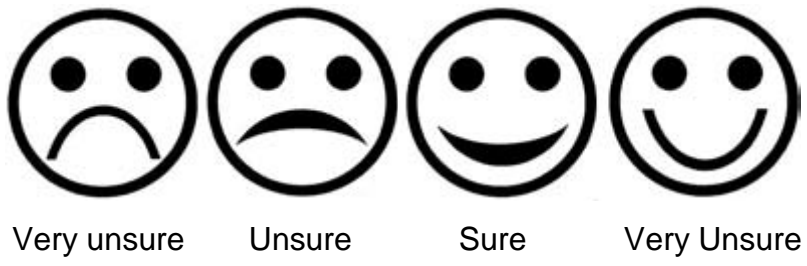


Figure 2. Confidence scale implemented for children in the present study.

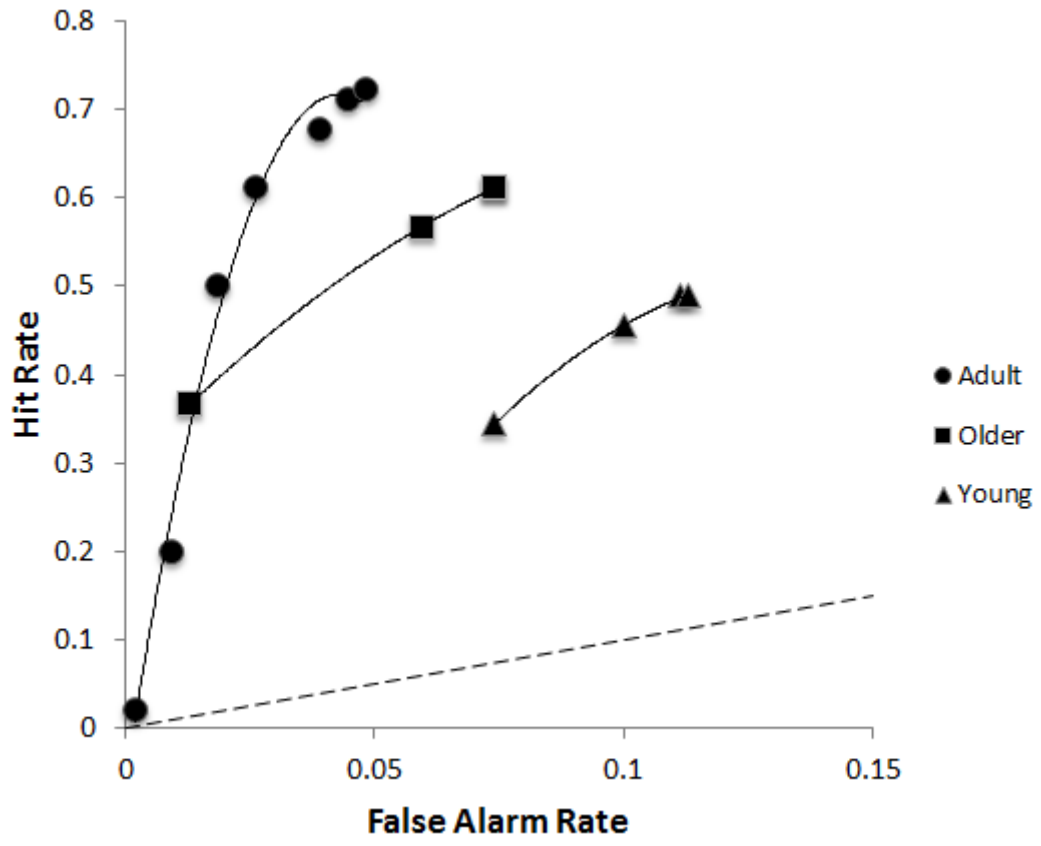


Figure 3. ROC analysis of memory performance for adults, older children and young children, collapsed across lineup condition.

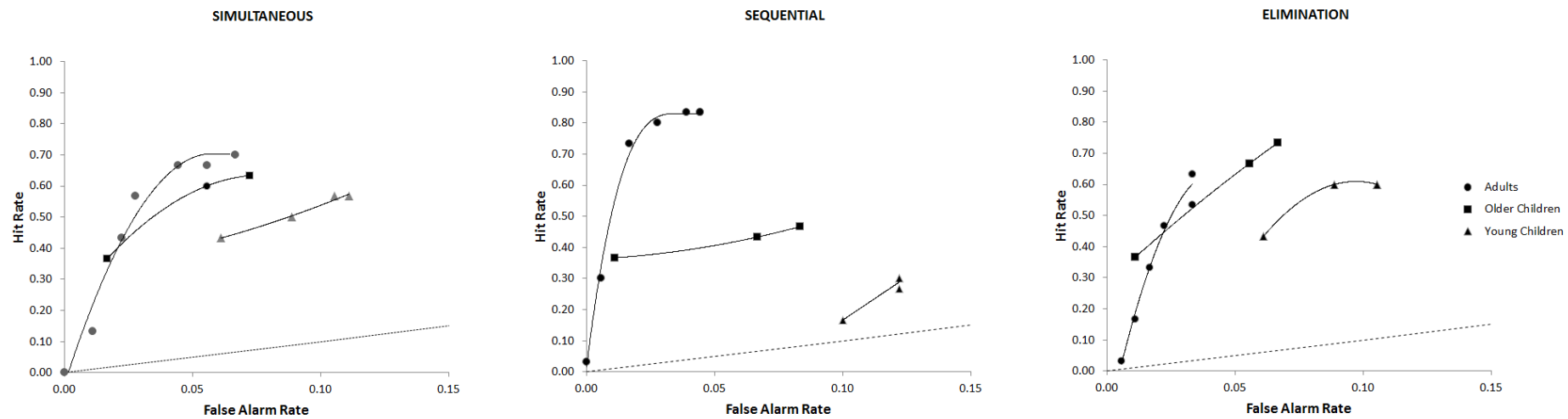


Figure 4. ROC analysis of memory performance for adults, older children and young children by lineup condition.

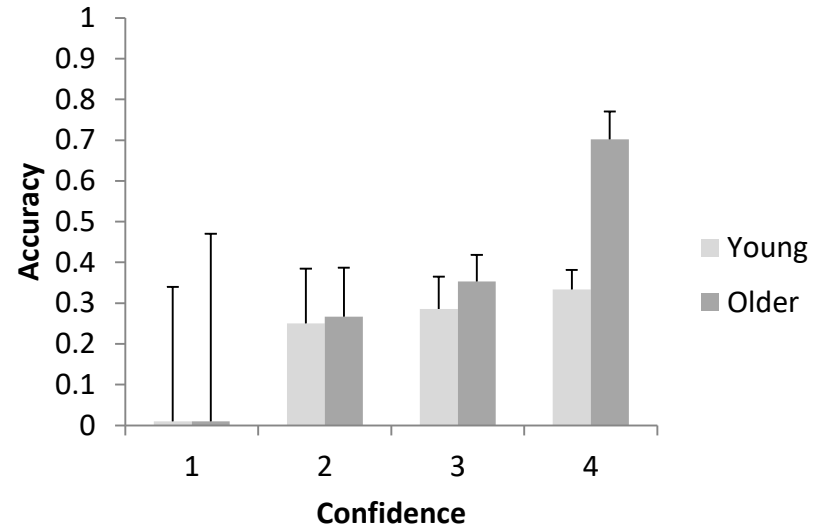
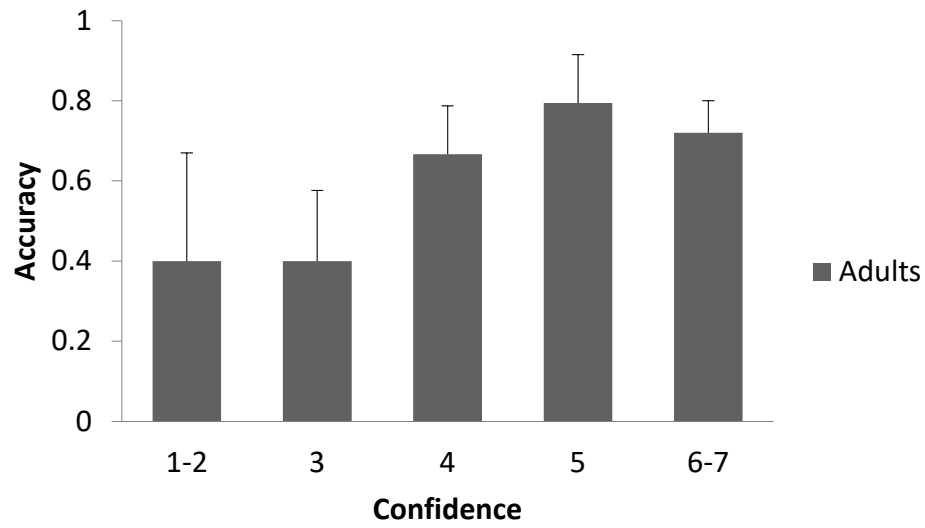


Figure 5. Mean accuracy (+1 SEM) by confidence level and age group for choosers. Note that mean accuracy for both young and older children at confidence level 1 was .00; jiggle was added to the x-axis to make it easier to see the results.