Applied Cognitive Psychology, *Appl. Cognit. Psychol.* (2011)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/acp.1804

Delay and Age Effects on Identification Accuracy and Confidence: An Investigation Using a Video Identification Parade

BRIAN R. CLIFFORD[1,2]*, CATRIONA HAVARD[2], AMINA MEMON[3] and FIONA GABBERT[4]

[1]*School of Psychology, University of East London, London, UK*
[2]*University of Aberdeen, Aberdeen, UK*
[3]*Royal Holloway, University of London, Egham, Surrey, UK*
[4]*University of Abertay Dundee, Dundee City, UK*

Summary: A total of 236 children (7–8 -year -olds and 13–14-year -olds) attempted identification from either target-present (TP) or target-absent (TA) video parades, after either two days or two weeks, following exposure to a live target. It was found that with TP line-ups, the most frequent response was a correct identification, but in the TA condition, the most frequent response was a false identification. Under both TP and TA conditions, the 3-way interaction of delay, age and response was significant. Further analysis revealed that delay was the major contributor of variance, causing an increase in errors for both age groups in TP delayed line-ups but only for the younger age group in TA delayed line-ups. Confidence varied as a function of type of identification response made, with the highest confidence being given to a correct identification in a TP line-up. However, response confidence entered into a number of interactions involving both delay and age. Copyright © 2011 John Wiley & Sons, Ltd.

This paper addresses three issues that are problematic for both law and psychology within the area of identification: Delay effects, confidence and children. Although Wells, Memon and Penrod (2006) assert that common sense tells us that memory declines over time, Dysart and Lindsay (2007) argue that it is risky to assume that we currently know what impact delay, actually, has on the accuracy of eyewitness identification. This same uncertainty exists within the legal system. Although both the courts in the USA (e.g. Neil v Biggers, 1972) and the UK (R v Turnbull, 1977) stress delay as a factor for consideration when juries evaluate identification evidence, such considerations are frequently discounted by other judicial factors. In addition, mock jurors do not seem to weigh adequately delay effects on memory (e.g. Bradfield & Wells, 2000), because, as Cutler, Penrod and Dexter (1990) have shown, mock jurors are just as willing to convict on the bases of identification made 14 days after a crime as compared with 2 days after the crime.

Although, as Dysart and Lindsay (2007) point out, delay as an issue has not been investigated systematically within the eyewitness literature, there are three approaches that appear relevant: archival studies, experimental studies and meta-analytical surveys of the research to date. Because our focus is on identification as a result of delay in children and adolescents, whereas the majority of studies have looked at adults, these adult results will be presented in tabular form, simply to demonstrate the uncertainty that resides in this area. (Table 1)

All three approaches have their attendant problems, nevertheless, a tentative conclusion concerning delay from archival studies could be that while, we cannot say that correct (or suspect) identification rates decline with time (because there is no ground-truth), what we can say is that identification errors—incorrect identifications or misses—appear to increase with delay. Conversely, the experimental studies seem to suggest delay has little effect on identification, irrespective of the length of delay employed, or method of testing, although the manipulation of the line-ups, with either the target present (TP) or with the target absent (TA), does seem to be important. Lastly, the meta-analyses, while mixed, do seem to suggest that there is a statistically reliable association between longer retention intervals and degraded identification accuracy.

Do these patterns and conclusions found mainly, if not solely, with adults also hold true for children? This is an important question because of the legal belief that young children's memories are, particularly, sensitive to the passage of time (see Flin, Boon, Knox, & Bull, 1992). In terms of age-related differences in identification accuracy, following a meta-analysis comparing the identification abilities of four age groups of children with adults, Pozzulo and Lindsay (1998) concluded that the identification literature suggests that from age five onwards, children's correct identification rates are equivalent to those of adults. Indeed, their identification can often be better than adults (see Parker & Carranza, 1989). Within the category of children itself, younger children can sometimes produce better performance than older children. For example, in Memon, Havard, Clifford, Gabbert and Watts' (in press) survey of 1718 identifications from video parades in the UK, 474 witnesses fell within the 5–11 and 12–15 years age group. It was found that 71% of the younger age group made a suspect identification and 14% a foil identification, whereas the older age group produced rates of 57% and 34%, respectively. Although impressive, numerically, these differences were not statistically reliable. By way of contrast, Pike, Brace and Kynan (2002), again based on a survey of real police identifications, found that under 10-year-olds performed more poorly than 11–15-year-olds (42% compared with 45%, respectively), but not reliably so.

The child–adult equality (Pozzulo, 2007), however, is somewhat attenuated when we consider the method of testing identification accuracy. Children and young adolescents are

*Correspondence to: Brian R. Clifford, School of Psychology, University of Aberdeen, Kings College, Old Aberdeen, AB24 2UB, UK.
E-mail: b.r.clifford@uel.ac.uk

Table 1. The effect of delay on identification accuracy in three types of study

Type of study	Reference	Delay period	Delay outcome
Archival studies			
	Tollestrup, Turtle, and Yuille (1994)	1 day–120 days	Delay effect
	Behrman & Davey (2001)	<7 days–>7 days	Delay effect
	Memon et al. (in press)	1 week–1 month	No delay effect
		<1 week–>1 month	Delay effect
	Valentine, Pickering, and Darling (2003)	1 week–6 months	No delay effect
	Wright & McDaid (1996)	1.49–1.9 week (log transformed)	Delay effect for foil choices
Experimental studies			
	Shepherd (1983)	1 week–1 month–3 months	No delay effect
		11 months	Delay effect
	Cutler, Penrod, O'Rourke, and Martens (1986)	7–28 days	Delay effect
	Yarmey, Yarmey, and Yarmey (1996)	0 minutes–24 hours	Delay effect
	Egan, Pittner, and Goldstein (1977)	2–56 days	No delay effect for correct identification Delay effect for false identification
	Laughery, Fessler, Lenorovitz, and Yoblick (1974)	4 minutes–1 week	No delay effect
	Mauldin & Laughery (1981)	30 minutes–2 days	No delay effect
	Cutler, Penrod, and Martens (1987)	2 days–2 week	No delay effect
	Read, Hammersley, Cross-Calvert, and McFadzen (1989)	20 minutes–1 week	No delay effect
	Yarmey (2004)	0 minutes–4 hours	No delay effect
	Dysart (1999)	0 minutes–3 weeks	No delay effect
	Krafka & Penrod (1985)	2 hours–24 hours	No delay effect
Meta-analysis			
	Steblay et al. (2001)		No delay effect
	Steblay et al. (2003)		No delay effect
	Shapiro & Penrod (1986)	Average delay of 4.5 days	Delay effect
	Deffenbacher (1986)		Delay effect
	Deffenbacher et al. (2008)		Delay effect

more likely than adults to make a positive identification (i.e. to choose) from a line-up, particularly when presented with a culprit—or target—absent line-up. It would seem that all children have difficulty with TA line-up procedures (e.g. Brewer, Keast, & Sauer, 2009; Havard, Memon, Clifford, & Gabbert, 2010; Keast, Brewer, & Wells, 2007; Parker & Carranza, 1989; Parker & Ryan, 1993; Pozzulo, 2007; Pozzulo & Lindsay, 1998). Given positive identification of an innocent suspect is known to be a major factor in wrongful convictions (Innocence Project, 2010); incorporation of TA line-ups is an important consideration in any investigation involving children's identification accuracy.

Unfortunately, as far as children and delay is concerned, there is very little published work. An interesting early study that looked at children and delay on identification was conducted by Ellis and Flin (1990) using a face recognition paradigm. They found in their first study with children of 7 and 10 years of age, with face exposure durations of 1 or 3 seconds, and testing recognition accuracy either immediately or 1 week later, that recognition accuracy was less for the younger children at immediate testing, but did not differ after a week's delay, due to the decrease in accuracy with delay for the older age group. In the second experiment, with faces exposed for 2 or 6 seconds and delays of immediate, 1 day or 1 week, 7-year-old children's recognition accuracy was unaffected by delay interval, but the 10-year-old's recognition was better at immediate testing than at either 1 day or 1 week testing. From this study, it would seem that older children are more susceptible to delay effects than are younger children.

Just as courts have recognised the possible importance of delay in recognition and identification, so they have argued that the confidence espoused by a witness is an important factor that jurors should consider. However, the evidence underlying this assumption is mixed. Intuitively, it would seem that memories about which one feels confident should be more accurate than less certain memories, (e.g. Leippe & Eisenstadt, 2007). However, with adults, depending upon the methods used, the confidence/accuracy relationship is either low or non-existent (correlational studies) or positive but not massive (calibration studies: e.g. Brewer, Keast, & Rishworth, 2002; Brewer & Wells, 2006; Juslin, Olsson, & Winman, 1996; Keast et al., 2007; Olsson, 2000; Weber & Brewer, 2003).

This uncertainty concerning the confidence/accuracy relationship is magnified in children because, as Shaw, McClure and Dykstra (2007) point out, research indicates that children are especially vulnerable to factors that can affect eyewitness accuracy and confidence. Several eyewitness identification studies have reported data on the confidence/accuracy relationship for children, with modest point-biserial correlations being generally reported (e.g. Leippe, Romanczyk, & Manion, 1991; Parker & Carranza, 1989; Parker, Haverfield, & Baker-Thomas, 1986; Parker & Myers, 2001; Parker & Ryan, 1993). More recently, Brewer and Day (2005) compared children's (10-year-olds) and adolescents' (16-year-olds) identification accuracy and certainty. Although the confidence/accuracy correlation for both groups were similarly low (.26 and .33, respectively); the accuracy/certainty relation differed markedly in other ways.

At each level of certainty (measured on a five-point scale), children's identification was much lower than that of adolescents. Thus, young children were much more overconfident than were adolescents, so much so that when children were 'really sure' about their accuracy, the proportion correct was only .49 for children compared with .93 for adolescents. One drawback of this study is that a TA condition was not included. This omission was corrected in a later study by Keast et al. (2007). These researchers compared 11–12-year-olds' and adult's identification accuracy and confidence in TP and TA line-ups. They found that for choosers, but not for non-choosers, meaningful confidence/accuracy relations were observed for adults but not for children. On the basis of two experiments, these researchers concluded that children's confidence provided no useful guide for investigators about the likely guilt or innocence of an identified suspect.

The present study employed the Video Identification Parade Electronic Recording (VIPER) video parade technology currently employed by UK police forces (see Memon et al., in press) to investigate with specific reference to children and adolescents these two issues—delay and confidence—that are problematic for the criminal justice system. In the light of the literature reviewed above, we employed delay periods of 2 days and 2 weeks, as the UK national VIPER bureau indicated to us that they can create line-ups within a 24–48-hour period, and 2 weeks are more of an average length of delay between witnessing an event and being presented with a video line-up in terms of real-life criminal investigations. Both TP and TA VIPER video line-ups were employed as children's performance has been shown to differ markedly as a function of line-up type. The specific age groups chosen were designed to compare pre-adolescent and adolescent children, as the research on their respective performance on TA and TP line-ups is conflicting. As an example, Memon et al. (in press) found that 5–11-year-olds performed better than 12–15-year-old on a TP line-up, but Havard et al. (2010) found that a 7–9-year-old age group, while performing better, but not reliably so, on a TP video line-up, performed more poorly than a 13–15-year-old age group on a TA line-up.

We hypothesised that delay effects would occur for both age groups of children; that older children would perform better than younger children, on TA line-ups; and that confidence would be poorly correlated with accuracy in the more difficult situation of TA line-ups, especially for our younger children.

METHOD

Participants

A total of 236 children were recruited from state run primary and secondary schools in Aberdeen, Scotland, with the consent of the legal guardian of each child. There were 115 participants aged between 7 and 8 years ($M = 7.8$ years, $SD = 0.7$ years; 58 females and 57 males) and 121 participants aged between 13 and 14 years ($M = 13.2$ years, $SD = 0.4$ years; 62 females and 59 males).

Materials

The target (actor) was a young male Caucasian, aged 27 years. The target was filmed at a VIPER suite at a local police station in order that the line-up met the standard specific VIPER content. Once the film had been made, it was sent to the VIPER headquarters for quality control purposes before being approved. VIPER line-ups are compiled in a standardised format comprising 15 second clips of each person shown in a sequence one after another. Each clip starts with a head and shoulders shot of the person looking directly at the camera, who then slowly turns their head to present a full right profile followed by a full left profile. Finally, the person returns to looking directly into the camera in a full face pose. Each face is accompanied by a number for identification purposes.

To control for factors that might affect identification accuracy, the video images used in the line-ups, including the targets, were rated by 31 individuals who did not take part in the experiment proper. There were two groups of raters: 12 were aged 6 to 9 years ($M = 8.25$ years) and 19 were aged 21 to 55 years ($M = 32.1$ years). Each face was rated on a 1–7 scale for distinctiveness, that is, 'if you had to pick this person out of a crowd at a railway station, how easy would it be?' The ratings found no significant differences in the target and foils (F (99, 261) = 11.57, $p = 0.12$).

Four 9-person line-ups were created, half TP and half TA according to the VIPER specifications. The positions of the substitute target foil (TA line-up) and target (TP line-up) were manipulated so that for half of the relevant line-ups (TP or TA), they appeared at position 4, hereafter referred to as Line-up A, and for the other half at position 6, hereafter referred to as Line-up B. This manipulation was designed to detect any bias for choosing early or late in the sequence of faces.

Design

The study employed a 2 (witness age: 7–8 years vs. 13–14 years) × 2 (delay: 2 days vs. 2 weeks) × 2 (line-up type: TP vs. TA) between-subjects design. There were 29 participants in each of the eight cells created by this design, except for the 2 days × 7/8 years × TP cell that contained 28 participants, the 2 days × 13/14 years × TA cell that contained 30 participants and the 2 weeks × 13/14 years × TP and TA cells that both contained 31 participants. The major dependent variable was the line-up identification decisions. For the TP line-ups, there were three possible responses: a correct identification (hit), a foil identification (false positive) or an incorrect rejection (miss). For the TA line-ups, responses were either a correct rejection or a false identification. Data from the TP and TA line-ups were analysed separately, unless stated otherwise. The second dependent variable was confidence in the response-choice made, ranging from 1 'very unsure', through 2 'unsure', 3 'middle', 4 'sure', to 5 'very sure'.

The study took place in two stages. Phase 1 involved exposure to a live interaction with a single target. Phase 2 took place either 2 days or 2 weeks later and involved identification from either a TP or TA line-up and an immediate rating of confidence in the choice made.

Procedure

In the first phase, groups of children (ranging in size from 10 to 15) witnessed a live event where an actor claimed to be a researcher from the university and described and illustrated what types of research were carried out in Psychology. This was a very general presentation where the actor described how Psychologists study things like memory and how the brain works. The presentation lasted precisely 2 minutes and he did not interact with the children in any other way.

In the second phase, 2 days or 14 days later, children were tested individually and carried out the identification task. The participants were told that they would be shown a series of faces on a computer screen one at a time and that the man who came to talk to them previously 'may or may not be present'. They were shown the line-up twice and were told that they could pause the line-up at any time and could go back and see any face again before making a decision. In Scotland, the Lord Advocate's guidelines on Visual Identification Parades (2007) *state* and in England and Wales recent revisions to Pace [the Police and Criminal Evidence Act (1984) (2008)] *require* that a video parade should be viewed twice before the witness makes a decision.

After the second viewing, the participants were asked if they wanted to view all or any part of the line-up again. They were, then, asked if the person they had seen and who had talked to them about psychology was in the line-up. If they identified the person (by indicating the number), they were shown the relevant video clip of the designated line-up member (not a photograph as is sometimes done by police using VIPER) and asked, 'is this the person you saw?' If, having viewed the numbered face, the participant said 'no' then they viewed the video again, and the procedure was repeated. If, after viewing the video as prescribed by VIPER operators, they did not identify any line-up member, they were asked if any of the line-up members looked like the man who had previously given the talk, and if so, which person was it and in what way or ways did they look like him. This part of the procedure again follows the Lord Advocates' guidelines. All the responses were recorded. However, in order to conform to normal identification experimentation, these latter responses will not be commented on further.

After the participants had made a decision, they were then asked how sure they were of their answer and shown a card which had 'very sure' 'sure' 'in the middle' 'unsure' and 'very unsure' and were asked to point to one of the responses. All the responses were recorded and converted into a 5-point rating scale (from 1 = very unsure to 5 = very sure). Finally, the children were asked if they remembered the experimenter saying 'the person may or may not be present'. Once they had responded, they were thanked and debriefed.

RESULTS

As preliminary analyses revealed that the specific placement of the target or replacement foil (Line-ups A and B) had no effect on either identification choices (all χ^2 tests <1) or on confidence in response choice (all *t-test*s <1), all subsequent analyses were conducted collapsed over this factor.

In terms of overall responses, it can be seen from Figure 1 that in TP line-ups, the most frequent response was a correct identification (53%), followed by foil identification (30%) and then incorrect rejection (17%), whereas in TA line-ups the more frequent response was a false identification (60%) followed by correct rejection (40%). Separate analysis on the two types of line-up revealed that these decision differences were significantly different: TP χ^2 (2) = 22.847, $p < .001$; TA χ^2 (1) = 5.342, $p < .021$

Employing hierarchical log linear analysis (HILOG), conducted separately on the two types of line-up (TP and TA), where the variables were delay, age and response type, it was found that for both analyses, the three-way interaction of age, delay and response was the best predictor of the cell frequencies. For the TP analysis, χ^2 (2) = 9.153, $p = .010$, and for the TA analysis χ^2 (1) = 5.294, $p = .021$.

The three-way interaction in the TP line-up is best explained by the fact that while delay had an effect on both age groups by increasing the number of foil identifications and incorrect rejections and decreasing the number of correct identifications, the effect was much more marked in the younger age group. Specifically, the percentage of correct identifications dropped much more precipitously for the 7/8-year-olds than for the 13/14-year-olds, whereas the percentage of incorrect rejections rose much more steeply for the 7/8-year-olds than for the 13/14-year-olds, over the period of delay.

In the TA condition, the three-way interaction indicates that while delay had little effect on the 13/14-year-old age group, it clearly influenced responding in the younger age group. Thus, while delay only marginally influenced correct rejection and false identification rates with the 13/14-year-olds, by increasing accuracy and decreasing inaccuracy, delay had a much more dramatic effect on the 7/8-year-olds. Here, delay massively decreased correct rejections and increased foil identifications.

If the higher order interaction in HILOG is significant, then one should not really be interested in the lower order effects because they are confounded with the higher order effects. That is, with categorical data, all lower-order effects are consumed within higher-order effects. However, partial association analyses indicated that the delay × response interaction was significant within both the TA analysis (χ^2 (1) = 4.503, $p = .034$) and the TP analysis (χ^2 (2) = 9.532, $p = .009$).

Following these various HILOG and partial association analyses, effect sizes were investigated by means of odds ratio analyses. To compute these, the possible responses in the TP line-ups, of which there were three, were collapsed into a dichotomous correct and incorrect response, by collapsing foil identifications and incorrect rejection. Odds ratios were calculated for TP and TA line-ups separately. Analysing by age group separately, it was found that for delay under TA conditions, the odds ratio for the younger children being correct at a 2-day delay compared with a 2-week delay was 5.96, but for the older children it was only 0.97. Thus, the younger children were some six times more likely to be correct at the 2-day delay compared with the 2-week delay, whereas the older children were marginally less likely to be correct at the shorter delay than the longer delay.

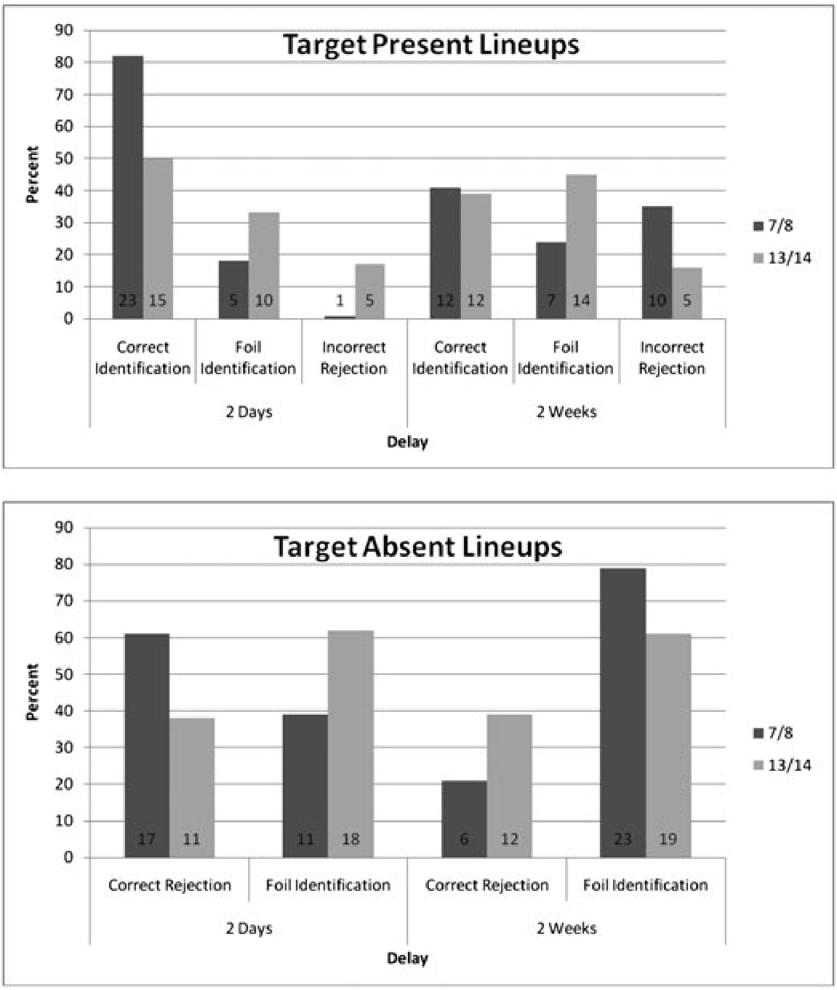

Figure 1. Percentage response choice (frequencies within bars) in the three-way interaction of age by delay and by response, for target present line-ups (top panel) and target absent line-ups (bottom panel)

This same pattern emerged for the TP condition. Here, younger children were again some five times (5.39) more likely to be correct at the shorter delay than at the longer, whereas the older children were only 1.59 times as likely to be more correct at the shorter than the longer retention interval.

When we analysed by delay separately, looking at age and correctness of identification choice, it was found that for the TA condition the younger children were some 2.54 times more likely to be correct compared with the older children at 2-days delay, but less likely to be correct (0.41) than the older children at the 2-week delay period. For the TP condition, again the same pattern emerged: the younger children were some four times (3.83) more likely to be correct compared with the older children at 2 days delay, but only some 1.13 times more likely to be correct at the 2-week delay period.

Confidence ratings

In terms of the confidence in response choices, shown in Table 2, a $2 \times 2 \times 5$ analysis of variance was conducted, with delay (2 days, 2 weeks), age (7/8, 13/14) and response type [correct identification, foil identification, incorrect rejection (TP), correct rejection, false identification (TA)] as between subject factors. The only main effect to attain significance was response (F (4, 215) = 10.836, p = .001, partial η^2 = .168). Post hoc analyses revealed that the confidence rating for correct identification was significantly higher than for incorrect rejection (p = .001), foil identification (p = .006), correct rejection (p = .001) and false identification (p = .001). Confidence ratings for the latter three response categories did not differ significantly. Incorrect rejection choices had the lowest confidence rating of all choices and differed significantly from all other responses except correct rejections (see Table 1). Three first order interactions were significant. The delay × response interaction was significant (F (4, 215) = 3.551, p = .008, partial η^2 = .062), due to the fact that while confidence ratings for both correct identification and false identifications decreased with delay, incorrect rejections, correct rejections and foil identifications all increased with delay. The age × response interaction was also significant (F (4, 215) = 2.770, p = .028, partial η^2 = .049) accounted for by the correct identification, foil identification and the false identification confidence ratings being similar for both age groups, but the incorrect rejection confidence rating being higher in the older age group than the younger age group, but the correct rejection confidence being higher in the younger age group than the older group. Lastly, the delay × age interaction was significant (F (1, 215) = 5.672, p = 018, partial η^2 = .026), best explained by the fact that while the younger age group increased their

Table 2. Mean confidence ratings as a function of delay, age, line-up type and response choice (standard deviations in parenthesis)

Delay	Age group	Line up type				
		Target absent		Target present		
		Correct rejection	False ID	Correct ID	Foil ID	Incorrect rejection
2 days	7/8	2.94	3.55	4.22	3.25	1.00
		(0.97)	(0.93)	(1.09)	(0.50)	(0.00)
	13/14	2.73	3.58	4.20	3.55	3.20
		(1.00)	(0.60)	(0.56)	(0.69)	(0.84)
2 weeks	7/8	3.83	3.22	3.79	3.71	2.80
		(0.98)	(0.99)	(0.82)	(0.85)	(0.79)
	13/14	2.88	3.16	3.71	3.29	3.20
		(0.80)	(0.77)	(0.69)	(0.47)	(0.45)
Grand mean		3.09	3.38	3.98	3.45	2.55
		(0.94)	(0.82)	(0.79)	(0.63)	(0.54)

ID, Identification.

confidence ratings at the longer delay, the older age group decreased their confidence with time. The higher-order interaction of delay × age × response was non-significant ($F < 1$).

We next explored the confidence/accuracy relationship in our children. When we correlated accuracy and confidence in the TP situation, it was found that there was a significant and positive correlation [r ($N = 118$) = .452, $p = .001$]. This indicated that our children were most confident when making a correct identification. When we analysed the confidence/accuracy relationship in the TA line-up situation, we obtained a significant but negative correlation [r ($N = 117$) = −.192, $p = .0$ 48]. In this case, our data suggest that the participants were more confident when making a false identification rather than a correct rejection. Age and confidence were not correlated in either TP or TA line-ups; however, delay and confidence were correlated negatively in TP line-ups (r ($N = 118$) = −.251, $p = .005$), indicating that as retention interval increased so confidence levels decreased. No such relationship was found for TA line-ups.

After the children had made their line-up choices and indicated their confidence in the choices made, we were keen to find out if they remembered our instruction that the man who talked to them about psychology ‘may or may not’ be present in the line-up they were about to see. In the event, the participants preponderantly remembered these unbiased instructions, and there was no difference in this remembering as a function of either line-up type (Yes, TP: 83%; Yes, TA: 86%); age (Yes, 7/8: 80%; Yes, 13/14: 85%) or delay (Yes, 2-days: 77%; Yes, 2-weeks: 88%). Clearly then, our children had attended to and retained the unbiased instruction concerning the possible presence or absence of the target in the line-ups they were to be exposed to.

DISCUSSION

This study has revealed delay effects in children both in terms of identification responses and the confidence held in these choices. However, these main effects were frequently qualified by the specific age range of the children making these choices and confidence responses, and the context under which choices and ratings were made, i.e. whether identifications were being made in a TP or a TA line-up.

Delay had a much more marked effect on our younger children, causing a substantial drop in both correct identification (TP line-ups) and correct rejections (TA line-ups). Delay had little or no effect on our 13/14-year-old’s correct rejection rates under TA line-ups, and only a small effect upon correct identification in TP line-ups. In terms of identification errors, again delay caused our 7/8-year-old participants to increase false identifications in TA line-ups and both foil identifications and incorrect rejections in TP line-ups. Once again, the effect of delay was less marked in our 13/14 age group, with delay having no effect on the rate of false identification in TA line-ups and no effect on incorrect rejections in TP line-ups. Only in the proportions of foil identifications under TP line-ups did delay effects reveal themselves in this age group.

This variability of delay effects in children in the present study mirrors the presence of null or negative studies, in the various meta-analyses conducted mainly on adult data that have investigated retention interval effects (Deffenbacher, 1986; Deffenbacher, Bornstein, McGorty, & Penrod, 2008; Shapiro & Penrod, 1986; Steblay, Dysart, Fulero, & Lindsay, 2001, 2003). Why should delay effects appear so mercurial?

One possibility is that it is important just which retention intervals are manipulated. The purpose of Deffenbacher et al.’s (2008) meta-analysis was to confirm or deny the reality of an Ebbinghausian-shaped (1913) forgetting curve for the human face. Despite some 93% of experts in the field agreeing that the rate of memory loss for an event was greatest right after an event and then levels off over time to some asymptote (Kassin, Tubb, Hosch, & Memon, 2001), and thus averring that an Ebbinghaus curve exists, Deffenbacher et al. (2008) pointed out that little direct evidence was available to justify such a belief.

Deffenbacher et al. then went on to provide such a justification, based on Wickelgren’s (1974) single-trace fragility theory of recognition, which posits an interference-free, time-decay process that produces rapid forgetting in the first seconds and minutes of the retention interval because initial trace fragility is very high. As consolidation sets in so trace fragility decreases such that the rate of forgetting slows in a negatively accelerating fashion, and therefore less is forgotten per unit time. In a similar but essentially simpler version of Wickelgren's (1974) theory, Wixted and Carpenter (2007) concur with the idea of rapid initial forgetting and then a slower rate of forgetting thereafter. Both Deffenbacher et al. (2008) and Wixted and Carpenter (2007) demonstrated the curve-fitting potential of Wickelgren's theory to several extant experimental findings (see also Wixted & Ebbesen, 1991).

For our purposes, our choice of 2-day and 2-week retention intervals, although motivated by ecological validity issues and current police records, clearly places those intervals well beyond the initial rapid forgetting phase of the forgetting curve, thus any marked effects of delay will be greatly attenuated, because, our two intervals are located on the negatively accelerating 'tail' of the forgetting curve. This will hold true whether we adopt the three parameter solution preferred by Deffenbacher et al. (2008)—initial memory strength, time-decay and interference, or the two parameter solution of Wixted and Carpenter (2007)—initial memory strength and trace decay. Given this restricted difference range between our two retention intervals, because of their positioning on the 'curve', it is not surprising that delay was found to interact with other variables manipulated in this study.

If we look at age as a moderator of delay effects, then we found our younger children were more affected than our older children. This interaction could be explained by our older children having better or more efficient encoding strategies (initial memory strength), better storage dynamics (memory decay rates) or retrieval resources. If this is the case, then delay should have less effect on them compared with the younger children.

This lesser susceptibility to delay effects in our older participants could explain a somewhat surprising finding. Although our 7/8-year-olds produced better correct identification in TP line-ups than our 13/14-year-olds (82% vs. 50%), and correct rejections in TA line-ups (61% vs. 38%) at our 2-day delay period, this initial superiority reverse dramatically at the later delay interval of 2 weeks. Why?

Let us begin with the initial superiority of the younger age group. The first point to make is that better performance by a young participant relative to an older participant is not new. As an example, in their meta-analysis Pozzulo and Lindsay (1998) found 5–6-year-olds had higher identification rates than adults (.71 vs. .54). In an experimental study, Parker and Carranza (1989) observed children 8–10 years-of-age produced more correct identifications than adults. When we look at pre-adolescents versus adolescents, the focus of the present study, in their survey of 445 real-life child witnesses, Memon et al. (in press) found younger children 5–11 exhibited increased suspect identification (74%) compared with 12–15-year-olds (54%). Conversely, Pike et al. (2002) found under 10-year-olds performed more poorly on suspect identification (42%) than 11–15-year-olds (45%), again in a survey of real police identifications. In both, these studies, the differences were statistically unreliable.

When we turn to experimental studies with this age group, only one study is conceptually similar to the present study. Havard et al. (2010) were concerned to compare pre-adolescent and adolescent identification under TP and TA video and static photo line-up conditions, following exposure to a live target, after a 2/3-day delay. They were not concerned with manipulating delay as a factor. Focussing only on the video identification part of their study, Havard et al. found that 7–9-year-olds out-performed 13–15-year-olds (74% vs. 64%, respectively), as in the present study. Although this was not a statistically reliable difference in the Havard et al. study, it is in the present study. Thus, the two studies are in agreement. However, unlike the present study, Havard et al. found that younger children performed more poorly on TA video line-ups (27.6% vs. 76%), which was statistically reliable. The discrepancy between the present study and Havard et al. in terms of TA line-ups, which is contended to be a difficult condition, may be a function of (i) different exposure durations to the target (2 minutes in the current study, 3 minutes in Havard et al.) in interaction with (ii) the mode of stimulus presentation (static, lecture-type presentation in the current study, where all participants could see the actor all the time versus dynamic, movement round the room, hence out of eye-line of all participants for some of the time, in the Havard et al. study). The content of the actor's presentation in the two studies also differed, and thus may have been of differential interest (psychology as an interesting topic of study in the current experiment, exploration of the type of shoe worn by children in the Havard et al. study). Because Havard et al. were not concerned to explore delay effects, it is unclear whether their study would support or contradict our delay findings, and the interaction of delay, age and type of line-up effects.

If one assumes for a moment, the featural versus holistic processing distinction and its correlation with age—as children develop, they move from a feature-based to a holistic-based perceptual strategy (e.g. Carey, Diamond, & Woods, 1980; Maurer, Le Grand, & Mondloch, 2002; Tanaka & Farah, 1993; Yovel & Duchaine, 2006), then it could be that our younger participants rely on one or at most a few features which, if present in a later identification task, eventuates in a high correct identification rate. However, should such a feature be forgotten, (over a delay period) then identification at delay will be poor. This account explains both the TP correct identification and the TA correct rejection profile of our 7/8-year-olds at 2 days. In the first case, identification follows from the remembered feature being present in the line-up. In the latter case, the line-up is rejected because the remembered feature is not present. If our 13/14-year-olds are employing either multiple featural encoding, or more likely, holistic/configural processing, then because of first-order and second-order relation encoding, the faces of line-up members will be more homogeneous, thus more prone to incorrect identification than an identification based solely on the presence or absence of a single feature. This apparent ability to draw upon several dimensions of facial features protects against the 'ravages of time' because several potential sources of comparison can be drawn upon over time. This should result in stability of performance over time, although the bases of performance may involve different sources of facial encodings being utilised to make line-up choices.

An alternative to explanations based upon the featural—holistic dichotomy, is one based on face-space theory

(Humphreys & Johnson, 2007; Nishimura, Maurer, Jeffery, Pellicano, & Rhodes, 2008; Valentine, 1991). This posits that as children grow older their face-space becomes richer and more features of any one face can be processed, stored and retrieved. This theory also accounts for our data: our 7/8-year-olds only access or store one or at most a few features; our 13/14-year-olds access or store many features. The focus on one feature eventuates in a 'go—no go' situation depending upon the presence or absence of that feature in the line-up personnel. The focus on multiple features lends itself to interference effects, because any one line-up participant may have some quantum of these features, but over time (with delay) a proportion of features will be retained and thus serve to either identify the culprit or falsely identify a foil in a TP line-up or fail to correctly reject in a TA line-up.

Of late, however, a spate of papers has appeared which calls into question the whole explanation of improved face recognition with the transition from featural to holistic visual processing (Crookes & McKone, 2009; Ge et al., 2008; Nishimura, Maurer, & Gao, 2009; Picozzi, Cassia, Turati, & Vescovo, 2009; Rakover, 2002). These authors converge on the conclusion that by the age of 5 years of age, and perhaps even younger, full quantitative and qualitative maturity of face-specific perceptual development is present, and that the improvements seen in face memory with age is due entirely to the development of general cognitive factors, such as memory ability per se, ability to use deliberate task strategies, employment of meta-memory judgements, ability to concentrate, ability to avoid distraction, or the ability to narrow the focus of visual attention. All these factors are known to improve substantially across childhood and most improve further into adolescence (Betts, McKay, Maruff, & Anderson, 2006; Bjorklund & Douglas, 1997; Flavell, 1985; Kail, 1991; Skoczenski & Norcia, 2002).This development of general cognitive ability with age would serve to explain the stability of identification choice in our 13/14-year-olds over delay and the more labile performance of our younger 7/8-year-old participants over the same delay period.

Situational and contextual factors clearly play a part in whether initial identification is better by younger or older witnesses. Debriefing questioning indicated that our younger children were more interested in, and motivated by, the target's visual and verbal presentation than were our older children. This motivational factor could account for the initial better performance on both TP and TA line-ups by our younger children. Thereafter, the older children's better memory, rehearsal strategies and possibly better tuned meta-cognitive beliefs about witnessing situations and /or how various factors (such as delay) are likely to affect memory (Brewer, Weber, & Semmler, 2007) results in their equal or better performance in terms of correct identification choices under TP and TA line-up situations, respectively. At this point, it is important to point out that it is unlikely that the actor engaged different groups differentially, or that he became more proficient at engagement over time with different groups or with different delay groups. The actor used a scripted talk that was rehearsed prior to going into schools. In addition, schools were visited on a random basis, and for each school, there was a 2-day and a 2-week delay, randomly chosen. Given these controls, it is unlikely that either different groups were differentially engaged, or that delay was confounded with actor proficiency.

Another issue concerning the actor is whether we should have used more than one actor. Following Wells and Windschitl (1999), however, because the actor was not used to represent a manipulated variable (e.g. age, race or gender) but to present a common stimulus across all conditions of manipulated interest, we feel justified in the use of only one actor. We do, however, accept that such a usage reduces the generalizability of our findings.

The enhanced cognitive resources occasioned by age would also explain the better calibration of confidence and accuracy in our older children compared with our younger children. In terms of confidence, overall, our children had some degree of 'calibration' between choice and certainty, at least in TP line-ups. No such 'calibration' was found, however, in TA line-ups. This suggests that TA formats are problematic for the participants. This is further demonstrated by the fact that while the confidence/accuracy correlation was significant and positive in TP line-up conditions, it was significant but negative in TA line-ups.

If confidence is an index of memorial certainty, then these interaction effects indicate that all children appreciate that memory decreases with delay, but that some identification situations are more difficult than others (TA vs. TP). In these difficult situations, the calibration of confidence and accuracy is most fragile with younger children, relative to older children. The current finding of a delay × age interaction mirrors the finding of Sykes, Gruneberg, and Turner's (1988) that younger children's confidence increased over delay, and Erskine, Markham and Howie's (2001) finding of no association of confidence and accuracy in a difficult recognition condition, which in the current study would be represented by our TA condition. The delay × response interaction mirrors the effect of delay found by Sauer, Brewer, Zweck and Weber (2010) on overconfidence and diagnosticity, using the calibration method of confidence/accuracy computation. In the present study, as delay interval increased so confidence levels in TP line-ups increased for incorrect choices (incorrect rejections and foil identifications) but decreased for correct choices (correct identifications). In TA line-ups, confidence decreased with time for incorrect choices (false identification) but increased for correct choices (correct rejections).

In agreement with several researchers (e.g. Sauer et al., 2010; Sporer, Penrod, Read, & Cutler, 1995), it was found that more reliance can be placed on the confidence/accuracy relationship of choosers than of non-choosers. However, the correlations were modest, and thus it should not be concluded that a consideration of the confidence espoused by 7/8 and 13/14-year-old children is informative for the diagnosis of identification accuracy.

Overall, then, this study has addressed three problematic areas of identification evidence as presented in court—the effect of delay, the reliability of children and adolescent witnesses, and the role of confidence in judging accuracy—and their interaction. It has demonstrated that delay has a marked effect upon the identification accuracy of young children 7–8 years of age but a less pronounced effect upon older children 13–14 years of age. This delay effect can serve to reverse initial better performance by the younger children on both TP and TA line-ups. Delay also serves to decrease the confidence with which correct identification choices are held in TP line-ups, and increases the malleability of the younger age group's confidence estimates concerning their correctness of decision. The confidence-accuracy relationship is particularly unreliable under conditions of increased uncertainty, such as TA line-ups.

The argument is that TA line-ups produce difficulty for all children because they are more likely to make a choice and therefore a false identification, and a suggestion that sequential line-ups, as used here in the VIPER procedure, are especially difficult for children (e.g. Pozzulo, 2007; Pozzulo & Lindsay, 1998). However, the VIPER directives (as used here) do not conform to the strict procedures advocated by Lindsay, Lea and Fulford (1991) and Lindsay and Wells (1985) for sequential line-ups: (i) the witness being unaware of the number of images to be presented, and (ii) the termination of the line-up procedure as soon as an identification of the target (perpetrator) is made. Future research should look at the performance profiles of differently aged children, who presumably have different levels of meta-cognitive monitoring and control, when the VIPER procedure and the strict sequential line-up procedure are compared directly. To our knowledge, there are only three published studies that have used VIPER, and only one with children, and none investigated delay. Given the ambiguity that pertains in legal and psychological circles concerning children, delay and preferred method of testing identification, and their interaction, this is an area of research in urgent need of attention.

ACKNOWLEDGEMENT

This project was supported by funds from the Scottish Institute of Policing Research.

REFERENCES

Behrman, B. W., & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior*, *25*, 475–491.

Betts, J., McKay, J., Maruff, P., & Anderson, V. (2006). The development of sustained attention in children: The effect of age and task load. *Child Neuropsychology*, *12*, 205–221.

Bjorklund, D. F., & Douglas, R. N. (1997). The development of memory strategies. In N. Cowen (Ed.), *The development of memory in childhood* (pp. 201–246). Hove: Psychology Press.

Bradfield, A. L., & Wells, G. L. (2000). The perceived validity of eye witness identification testimony: A test of the five Biggers criteria. *Law and Human Behavior*, *24*, 581–594.

Brewer, N., & Day, K. (2005). The confidence-accuracy and decision latency – accuracy relationship in children's eyewitness identification. *Psychiatry, Psychology and Law*, *12*, 119–128.

Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. *Journal of Experimental Psychology. Applied*, *8*, 44–56.

Brewer, N., Keast, A., & Sauer, J. D. (2009). Children's eyewitness identification performance: Effects of a not sure response option and accuracy motivation. *Legal and Criminological Psychology*, *15*, 261–277. Advanced online publication. DOI: 10.1348/135532509X474822.

Brewer, N., Weber, N., & Semmler, C. (2007). A role for theory in eyewitness research. In R. C. L. Lindsey, D. E. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology: Vol. 2. Memory for people* (pp. 201–218). New Jersey: Laurence Erlbaum Associates.

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology. Applied*, *12*, 11–30.

Carey, S., Diamond, R., & Woods, B. (1980). Development of face recognition: A maturational component? *Developmental Psychology*, *16*(4), 257–269.

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.

Cutler, B. L., Penrod, S., & Dexter, H. R. (1990). Juror sensitivity to eyewitness identification evidence. *Law and Human Behavior*, *14*, 185–191.

Cutler, B. L., Penrod, S., & Martens, T. K. (1987). Improving the reliability of eyewitness identification: Putting context into context. *The Journal of Applied Psychology*, *72*, 629–637.

Cutler, B. L., Penrod, S., O'Rourke, T. E., & Martens, T. K. (1986). Unconfounding the effects of contextual cues on eyewitness identification accuracy. *Social Behavior*, *1*, 113–134.

Deffenbacher, K. A. (1986). On the memorability of the human face. In H. D. Ellis, M. A. Jeaves, F. Newcombe, & A. Young (Eds.), *Aspects of face processing* (pp. 61–70). Dordrecht, the Netherlands: Martinus Nijhoff.

Deffenbacher, K. A., Bornstein, B. H., McGorty, E. K., & Penrod, S. D. (2008). Forgetting the once-seen face: Estimating the strength of an eyewitness's memory representation. *Journal of Experimental Psychology. Applied*, *14*, 2, 139–150.

Dysart, J. E. (1999). The effects of delay on eyewitness identification accuracy. Unpublished Master's thesis, Queens University, Kingston, Ontario, Canada.

Dysart, J. E., & Lindsay, R. C. L. (2007). The effects of delay on eyewitness identification accuracy: Should we be concerned? In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology: Vol. 2. Memory for people* (pp. 361–376). New Jersey: Laurence Erlbaum Associates.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology.* (H. A. Ruger & C. E. Bussenius. Trans) New York: Teachers College, Columbia University.

Egan, D., Pittner, M., & Goldstein, A. G. (1977). Eyewitness identification: Photographs vs. Live models. *Journal of Law and Human Behavior*, *1*, 199–206.

Ellis, H. D., & Flin, R. H. (1990). Encoding and storage effects in 7-year-olds' and 10-year-olds' memory for faces. *British Journal of Developmental Psychology*, *8*, 77–92.

Erskine, A., Markham, R., & Howie, P. (2001). Children's script-based inferences: Implications for eyewitness testimony. *Cognitive Development*, *16*, 871–887.

Flavell, J. H. (1985). *Cognitive development*. Englewood Cliffs, NJ: Prentice Hall.

Flin, R., Boon, J., Knox, A., & Bull, R. (1992). The effect of a five-month delay on children's and adults' eyewitness memory. *British Journal of Psychology*, *83*, 323–336.

Ge, L., Anzures, G., Wang, Z., Kelly, D., Pascals, O., Quinn, P. C., Slater, A. M., Yang, Z., & Lee, K. (2008). An inner face advantage in children's recognition of familiar peers. *Journal of Experimental Child Psychology*, *101*, 124–136.

Havard, K., Memon, A., Clifford, B. R., & Gabbert, F. (2010). A comparison of video and static photo lineups with child and adolescent witnesses. *Applied Cognitive Psychology*, *24*, 1209–1221.

Humphreys, K., & Johnson, M. H. (2007). The development of "face-space" in infancy. *Visual cognition*, *15*, 5, 578–598.

Innocence Project. (2010). 250 exonerated, too many wrongfully convicted: An innocence project report on the first 250 DNA exonerations in the USA. Benjamin N. Cardozo School of Law, Yeshiva University, New York, USA.

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*, 1304–1316.

Kail, R. (1991). Development change in speed of processing during childhood and adolescence. *Psychological Bulletin*, *109*, 3, 490–501.

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research: A new survey of the experts. *The American Psychologist*, *56*, 405–416.

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.

Krafka, C., & Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality and Social Psychology*, *49*, 58–69.

Laughery, K. R., Fessler, P. K., Lenorovitz, D. R., & Yoblick, D. A. (1974). Time delay and similarity effects in facial recognition. *The Journal of Applied Psychology*, *59*, 490–496.

Leippe, M., Romanczyk, A., & Manion, A. P. (1991). Eyewitness memory for a touching experience: Accuracy differences between child and adult witnesses. *The Journal of Applied Psychology*, *76*, 367–379.

Leippe, M. R., & Eisenstadt, D. (2007). Eyewitness confidence and the confidence accuracy relationship in memory for people. In R. C. L. Lindsay, D. F. Ross, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology: Vol. 2. Memory for people* (pp. 377–425). New Jersey: Laurence Erlbaum Associates.

Lindsay, R. C. L., Lea, J. A., & Fulford, J. A. (1991). Sequential lineup presentation: Technique matters. *The Journal of Applied Psychology*, *76*, 741–745.

Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identification from lineups: Simultaneous versus sequential lineup presentation. *The Journal of Applied Psychology*, *70*, 556–564.

Lord Advocates guidelines – the conduct of visual identification procedures (2007). Retrieved from the web on 07. 13. 09. http://www.copfs.gov.uk/resourses/Doc/13547/0000269.pdf

Mauldin, M. A., & Laughery, K. R. (1981). Composite production effects on subsequent facial recognition. *The Journal of Applied Psychology*, *66*, 351–359.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255–260.

Memon, A., Havard, C., Clifford, B. R., Gabbert, F., & Watt, M. (in press). A field evaluation of the VIPER system: A new technique for eliciting eyewitness identification evidence. *Psychology, Crime and Law*.

Neil v. Biggers, 409 U.S. 188 (1972).

Nishimura, M., Maurer, D., & Gao, X. (2009). Exploring children's face-space: A multidimensional scaling analysis of the mental representation of facial identity. *Journal of Experimental Child Psychology*, *103*, 355–375.

Nishimura, M., Maurer, D., Jeffery, L., Pellicano, E., & Rhodes, G. (2008). Fitting the child's mind to the world: Adaptive norm-based coding of facial identity in 8-year-olds. *Developmental Science*, *11*(4), 620–627.

Olsson, N. (2000). A comparison of correlation, calibration and diagnosticity as measures of the confidence-accuracy relationship in witness identification. *The Journal of Applied Psychology*, *85*, 504–511.

Parker, J. F., & Carranza, L. E. (1989). Eyewitness testimony of children in target-present and target-absent line-ups. *Law and Human Behavior*, *13*, 133–149.

Parker, J. F., Haverfield, E., & Baker-Thomas, S. (1986). Eyewitness testimony of children. *The Journal of Applied Psychology*, *16*, 287–302.

Parker, J. F., & Myers, A. (2001). Attempts to improve children's identification from sequential-presentation lineups. *Journal of Applied Social Psychology*, *31*, 796–815.

Parker, J. F., & Ryan, V. (1993). An attempt to reduce guessing behaviour in children's and adult's eyewitness identification. *Law and Human Behavior*, *17*, 11–26.

Picozzi, M., Cassia, V. M., Turati, C., & Vescovo, E. (2009). The effect of inversion on 3- to 5-year-old's recognition of face and non-face visual objects. *Journal of Experimental Child Psychology*, *102*, 487–502.

Pike, G., Brace, N., & Kynan, S. (2002). The visual identification of suspects: Procedures and practice. Home Office Briefing Note. 2/02. Home Office Research, Development and Statistics Directorate.

Police and Criminal Evidence Act (1984). (2008). Codes of practice. Retrieved 07. 13. 09 from: http://police.homeoffice.gov.uk/operational-policing/powers-pace-codes/pace-code-intro/

Pozzulo, J. D. (2007). Person description and identification by child witnesses. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds), *Handbook of eyewitness Psychology: Vol. 2. Memory for people* (pp. 283–307). New Jersey: Laurence Erlbaum Associates.

Pozzulo, J. D., & Lindsay, R. C. L. (1998). Identification accuracy of children versus adults: A meta-analysis. *Law and Human Behavior*, *22*, 549–570.

R. v. Turnbull, QB 224 (1977).

Rakover, S. S. (2002). Featural vs. configural information in faces: A conceptual and empirical analysis. *British Journal of Psychology*, *93*, 1–30.

Read, J. D., Hammersley, R., Cross-Calvert, S., & McFadzen, E. (1989). Rehearsal of faces and details in action events. *Applied Cognitive Psychology*, *3*, 295–311.

Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence-accuracy relationship for witness identification. *Law and Human Behavior*, *34*, 337–347.

Shapiro, P. N., & Penrod, S. D. (1986). A meta-analysis of facial identification studies. *Psychological Bulletin*, *100*, 139–156.

Shaw, J. S., McClure, K. A., & Dykstra, J. A. (2007). Eyewitness confidence from the witnessed event through trial. In M. P. Toglia, J. D. Read, D. F. Ross, & R. C. L. Lindsay (Eds.), *Handbook of eyewitness psychology: Vol 1. Memory for events* (pp. 371–397). New Jersey: Laurence Erlbaum Associates.

Shepherd, J. (1983). Identification after long delays. In S. Lloyd-Bostock, & B. R. Clifford (Eds.), *Evaluating witness evidence: Recent psychological research and new perspectives* (pp. 173–188). Chichester, New York: John Wiley & Sons.

Skoczenski, A. M., & Norcia, A. M. (2002). Late maturation of visual hyperacuity. *Psychological Science*, *13*, 6, 537–541.

Sporer, S. L., Penrod, S. D., Read, D., & Cutler, B. L. (1995). Choosing, confidence and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, *118*, 315–327.

Steblay, N., Dysart, J. E., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior*, *25*, 459–473.

Steblay, N., Dysart, J. E., Fulero, S., & Lindsay, R. C. L. (2003). Eyewitness accuracy rates in police showup and lineup presentations: A meta-analytic comparison. *Law and Human Behavior*, *27*, 523–540.

Sykes, R. N., Gruneberg, M. M., & Turner, T. (1988). Feelings of certainty in child witnesses. In M. M. Grunberg, & P. E. Morris (Eds.), *Practical aspects of memory: Current research and issues*, vol. *1* (pp. 128–134) Oxford: John Wiley & Sons.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*, *46*(2), 225–245.

Tollestrup, P. A., Turtle, J. W., & Yuille, J. C. (1994). Actual victim and eyewitnesses to robbery and fraud: An archival analysis. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: current trends and developments* (pp. 144–160). Cambridge: Cambridge University Press.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. *The Quarterly Journal of Experimental Psychology*, *43*(2), 161–204.

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real line-ups. *Applied Cognitive Psychology*, *17*, 969–993.

Weber, N., & Brewer, N. (2003). The effect of judgement type and confidence scale on confidence-accuracy calibration in face recognition. *The Journal of Applied Psychology*, *88*, 490–499.

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest* *7*(2), 45–75.

Wells, G. L., & Windschitl, P.D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, *25*, 1115–1125.

Wickelgren, W. A. (1974). Single-trace fragility of memory dynamics. *Memory & Cognition*, *2*, 775–780.

Wixted, J. T., & Carpenter, S. K. (2007). The Wicklegren power law and the Ebbinghaus savings function. *Psychological Science*, *18*, 133–134.

Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science*, *2*, 409–415.

Wright, D. B., & McDaid, A. T. (1996). Comparing systems and estimator variables using data from real lineups. *Applied Cognitive Psychology*, *10*, 75–84.

Yarmey, A. D. (2004). Eyewitness recall and photo identification: A field experiment. *Psychology, Crime and Law*, *10*, 53–68.

Yarmey, A. D., Yarmey, M. J., & Yarmey, A. L. (1996). Accuracy of eyewitness identification in showups and line-ups. *Law and Human Behavior*, *20*, 459–477.

Yovel, G., & Duchaine, B. C. (2006). Specialized face perception mechanisms extract both part and spacing information: Evidence from developmental prosopagnosia. *Journal of Cognitive Neuroscience*, *18*(4), 580–593.

