

Age Differences in Lineup Identification Accuracy: People Are Better With Their Own Age¹

Daniel B. Wright^{2,4} and Joanne N. Stroud³

Previous research has reported that young adults are better at eyewitness face recognition than are older adults. However, these studies have used young adults as culprits and fillers. We explore how the relative ages of the witness and the culprit influence eyewitness accuracy in 2 experiments. In the first experiment, young (18–25 years old) and older (35–55 years old) adults each saw 4 crime videos. In 2 the culprit was a young adult and in 2 the culprit was an older adult. Participants were more accurate at identifying the culprit when viewing culprit present lineups comprising people of their own age: an “own age bias” analogous to the own race bias. In the 2nd experiment, using a similar procedure, young (18–33 years old) and older (40–55 years old) adults viewed both culprit present and culprit absent lineups. The results of the first experiment were replicated for the culprit present lineups. However, no own age bias was found for the culprit absent lineups. Implications for police procedures dealing with cross-generation identifications are discussed.

KEY WORDS: eyewitness testimony; memory; own age bias; lineups; age differences.

In criminal cases, eyewitness testimony can be the most important evidence and in some countries the testimony of a single eyewitness can be enough to convict a person. From cases where convicted people have been exonerated by DNA evidence (Connors, Lundregan, Miller, & McEwan, 1996), there is a growing belief that the majority of false convictions are due to errant eyewitness testimony. Eyewitness evidence is often derived from police lineups where the witness views several members of the public (called *fillers*, *foils*, or *distracters*) and the suspect. The witness is told that the culprit may or may not be in the lineup. The witness is then asked to choose the culprit if she or he feels the culprit is present. The problem is that witnesses often choose someone other than the culprit. In about 20% of lineups reported in

¹These studies were previously reported at the Society for Applied Research in Memory and Cognition (SARMAC) conferences in Toronto (1997) and Boulder, CO (1999).

²Psychology Group, School of Cognitive and Computing Sciences, University of Sussex, Brighton, United Kingdom.

³Department of Experimental Psychology, University of Bristol, United Kingdom.

⁴To whom correspondence should be addressed at Psychology Group, School of Cognitive and Computing Sciences, University of Sussex, Brighton BN1 9QH, United Kingdom; e-mail: danw@cogs.susx.ac.uk.

Wright and McDaid's survey of lineups in London, witnesses chose a filler (Wright & McDaid, 1996). Although the filler is not charged with any crime, innocent suspects, who arguably are in a similar situation to fillers, could be convicted on this evidence.

Eyewitness research can be divided into two categories: system research and estimator research (Wells, 1978; see Wright & Davies, 1999, for a review). System research investigates the ways in which the judicial system can be changed to improve the reliability of eyewitness reports. This includes how to interview eyewitnesses and how to conduct lineups. Based in part on psychology research, the Technical Working Group for Eyewitness Evidence (1999) produced guidelines for interviewing witnesses and conducting lineups. This has been a great advance not just for eyewitness researchers, but for psychology as a whole (Wells et al., 2000). Estimator research is involved with estimating when eyewitness testimony is likely to be reliable and when it is not. Examples include that reliability is poor with short exposure times, when the lighting is poor, when the eyewitness has heard misleading information, and when the culprit and the eyewitness are of different races. The last of these, called the *own race bias*, has been the focus of much research (see Meissner & Brigham, 2001, for a meta-analysis of past research).

The own race bias refers to the finding that people tend to be better at identifying faces of people of their own race than people of other races. For example, Wright, Boyd, and Tredoux (2001) had a male confederate—someone working for researchers but pretending to be a normal member of the public—approach either a Black or a White person in shopping centers in either South Africa or England. Half of the confederates were Black and half were White. The confederate asked the person a couple of questions, for example, where the cinema was located. A few minutes later the experimenter approached and asked the person to try to identify the confederate with whom she or he had just spoken from a 10-person culprit present lineup. Black participants were more accurate recognizing the Black confederate than the White confederate, while the opposite was true for the White participants. Evidence for the own race bias is now so well established that approximately 90% of experts agree that it is reliable enough for expert scientific testimony (Kassin, Tubb, Hosch, & Memon, 2001). In this research we examine whether another own group memory bias occurs: an own age bias.

AN OWN AGE BIAS?

Is eyewitness accuracy related to the relative ages of the witness and the culprit? For very young and very old people, the age of the witness is clearly related to memory. Memory for faces develops through childhood and possibly into adolescence (George, Hole, & Scaife, 2000). Similarly, face memory (and memory more generally) declines as people grow older (Grady & Craik, 2000). Here we avoid groups where there may be substantial differences in cognitive ability. We explore lineup accuracy for two groups of people, young and middle-aged adults, when viewing culprits and fillers from these people's age groups.

Some research has explored accuracy in lineup identifications for witnesses of approximately these ages. Adams-Price (1992) compared the memories of

20–35-year-olds with 40–55-year-olds for two simulated thefts. She found that the 40–55-year-olds did worse. However, her fillers and culprits were of college age. Similarly, O'Rourke, Penrod, Cutler, and Stuve (1989) had participants from 18 to 74 years of age view a robbery of a store. They found identification accuracy declined with age, but did not state the age of the robber. Yarmey (1993) also compared memory abilities for people of different ages. He had one of two young (21 and 24 years old) women walk up to members of the public and ask them for help to find some lost jewelry. Two minutes later the other woman asked them a series of questions about the stranger with whom they had just spoken. Overall, the young (18–29 years old) people had better memories. As with Adams-Price's study, however, Yarmey could not claim that the young people were better in general, only that they were better with the young culprits: "since the two culprits were young adults, it is possible that a same age effect occurred" (Yarmey 1993, p. 1929).

List (1986) has looked at this possibility. She had 10-year-olds, college age students, and older adults (65–70-year-olds) view several shoplifting videos. In these, the culprits were either college age or middle age females (no specific ages were given). She found that the older adults had poorer recognition memory for the college age culprits than did the other two age groups, but were better for the middle age culprits (List, 1986). List's study looked at a number of different issues, in particular, what people thought normal shoplifting incidents were like (List, 1986). She argued that the older participants may have felt that the middle age culprits were "more relevant to themselves because self-reference enabled more elaborate or multiple encodings of this information for older adults" (p. 56). To examine this, List should have used culprits of the witnesses' own generation. There were higher order interactions in List's study, between other variables and the age of witness by age of culprit interaction, which makes the interpretation of this interaction less clear (List, 1986). Some research outside of eyewitness testimony has found age of participant by age of face interactions (Bartlett & Leslie, 1986; Fulton & Bartlett, 1991; Mason, 1986). These studies have used more standard recognition paradigms, but suggest that an own age bias may occur.

The studies that we report are similar in design to some of the own race bias studies. Although there is no agreed-upon explanation for the own race bias, one explanation is that people become experts identifying groups within whom they have much contact. If people become experts at recognizing faces of people in their age group, we would expect an own age bias. There is some evidence that this is true for remembering faces of infants (Chance, Goldstein, & Andersen, 1986). An own age bias would be shown by an interaction between the age of the witness and the age of the culprit. This would mean young people were relatively more accurate with young faces than with older faces, compared with the older sample.

PILOT STUDIES

Prior to conducting our main studies, two pilot studies were conducted. The first made sure that culprits' and the fillers' ages were perceived to be in the correct age category. The second was to make sure that there were not large biases present in our lineups.

Table 1. The Descriptions of the Culprits

	Young (car)	Older (car)	Young (TV)	Older (TV)
Height	6'1"	5'11"	5'10"	5'9"
Weight (lb)	154	169	140	182
Hair	Brown/auburn	Brown but greying	Brown	Dark
Build	Slim	Medium	Slim	Large
Age	23	51	21	48

Four videos were created. Two showed a car being stolen and two showed a television being stolen. The car videos lasted approximately 2 min. During this time, a woman drove into a car park and parked at the end of a row of cars. She left the window on the driver's side open and walked away. The culprit, who was a 23-year-old man in one video and a 51-year-old man in the other, walked past six cars, checking to see whether the doors were unlocked on two cars, and then spotted the open window. He unlocked the door, hot-wired the car, and drove off. He was in view for 1 min and 10 s, with his face in view for 6 s. The television videos showed either a 21-year-old or a 48-year-old male walk out of the side gate of a house carrying a television. He walked down the driveway checking to see if anyone was looking, put the television into the back of a car, and drove away. He is shown throughout the video (total time 1 min and 5 s) with a frontal view of his face for 8 s. Descriptions of all the culprits are given in Table 1. Males were used as the culprits in all videos because males commit most crimes. Fillers were volunteers approached on the street with the requirement that they fit the description of the culprit given in Table 1 and were in the appropriate age group (18–25 or 35–55 years old). The photos were taken with people standing in front of a light plain background. They were of the person's face taken from directly in front of them. The top of their shirts were visible. The fillers were different for each lineup.

The first pilot study was to verify that one of the dependent variables, how old the culprit and the fillers appear, was properly manipulated. To test this, five participants of different ages were asked to estimate the ages of the fillers and the culprits from the photographs that were used in the main studies. The means for the people in the young lineups ranged from 19.0 to 24.2 years old and the means for the people in the older lineups were between 44.4 and 53.4 years old. In total, there were 70 ratings for the young lineups and 70 ratings for the older lineups. There was no overlap in these ratings. In fact, none of the ratings were in the 30–39 years old range. Therefore, the culprits' and fillers' apparent ages were appropriate.

The second pilot study investigated whether the culprit stood out in the lineup. There are numerous cases in legal history where a lineup is clearly biased and there continue to be lineups that are sufficiently biased to be of much concern (see Lindsay, 1994, for discussion). As we are comparing two groups of people for each lineup, it is not vital to have an unbiased lineup to examine the effects of interest, but we felt it was worth making sure that the target did not stand out. There are several ways of measuring bias in lineups (Tredoux, 1998, 1999). Ten participants were shown the four descriptions in Table 1 and then asked to choose the culprit from the seven person culprit present lineups used for each of the four videos. They were asked to

try to guess who the target was. If the lineups were unbiased, then we would expect about one seventh of the choices to be of the culprit. No culprit was chosen more than twice, which is not statistically different from this. While this does not mean that the lineups are unbiased (with a total of 40 trials overall, the width of the confidence interval is about 25%), it does show that participants in the main studies should not be choosing the target because he stands out. Further, no foils were chosen at a level significantly above chance.

EXPERIMENT 1

Method

Participants

This study had two between subjects variables. Potential participants were asked to volunteer if they fit into one of the age groups. They were approached at their workplace, at leisure areas, or around the university. Participants (total $n = 113$) were either between 18 and 25 or between 35 and 55 years old, and were randomly assigned either to a 1-day condition (young, $n = 30$; older, $n = 29$) or to a 1-week condition (young, $n = 26$; older, $n = 28$). The imbalance in the number of participants for the four conditions was due to seven people not being able to make the second session. The participants were all Caucasian males. They were unpaid volunteers and did not know any of the culprits or fillers. The younger participants were students at City University in London, England. The older participants were recruited from various locations in Essex, England. None of the participants had any experience with lineups.

Procedure

Participants were individually shown the four videos described earlier. The order of these videos was counterbalanced to prevent an order effect confounding those effects of interest. After viewing all the videos, participants filled out a short crime questionnaire about the seriousness of each of the crimes they had witnessed and some of their views toward sentencing and the prison population. This was done to detract momentarily from the main purpose of the study. Participants were told that they would be contacted again in either 1 day or 1 week and asked to try to identify the culprits they had just seen from lineups.

The lineups were simultaneous and consisted of the seven photographs (six fillers, one culprit). The position of the fillers and the culprit in the lineup was random for each participant. Participants were read instructions that asked them to “identify the person you saw in each video,” and told to take as much time as they wanted. They were told “that the person you saw may or may not be here” and instructed “if you cannot make a positive identification you should say so.” The instructions were based on those used in the lineup suites in London. Participants who made an identification were asked to rate their confidence on a 1–7 scale from *I’m highly unsure whether it’s him* to *I’m positively sure it’s him*. Participants who did not make an identification did not rate their confidence because we felt that doing this might confuse them, creating problems for their other confidence ratings. The lineups were presented in the same order as the videos for each participant.

Table 2. Frequencies and Percentages of Choosing the Culprit, Choosing a Filler, and Making No Identifications for Experiment 1

	Identification			Odds ratio
	Positive ID	Filler ID	No ID	
Young culprits				
1 day young Ss	28 (47%)	24 (40%)	8 (13%)	2.75
1 day older Ss	14 (24%)	26 (45%)	18 (31%)	
1 week young Ss	15 (29%)	31 (60%)	6 (12%)	1.66
1 week older Ss	11 (20%)	30 (54%)	15 (27%)	
Total	68 (30%)	111 (49%)	47 (21%)	
Older culprits				
1 day young Ss	22 (37%)	29 (48%)	9 (15%)	1.50
1 day older Ss	27 (47%)	26 (45%)	5 (9%)	
1 week young Ss	11 (21%)	32 (62%)	9 (17%)	1.13
1 week older Ss	13 (23%)	32 (57%)	11 (20%)	
Total	73 (32%)	119 (53%)	34 (15%)	
Grand total	141 (31%)	230 (51%)	81 (18%)	

Note. The odds ratios are found by dividing the odds of correct own age identification by the odds of a correct identification of someone from the other age group.

Results

The frequencies and percentages for choosing the culprit, for choosing a filler, and for making no identifications are shown in Table 2. There were no significant effects for whether it was a television being stolen or an automobile being stolen and so the data in this table are collapsed across this variable. There were more fillers chosen and fewer nonidentifications than found in a survey of real lineups (Wright & McDaid, 1996), where approximately 20% of the time a filler was chosen and approximately 40% of the time no identification was made. However, the situation is very different. Because the culprit is present in all the lineups, positive identifications are the correct responses. After a 1-day delay there is a crossover interaction, with both groups performing better with their own age group. The older age group performs particularly poorly with the young culprits. After 1 week, performance is much worse overall. If people were randomly guessing, we would expect the culprit to be chosen approximately 1/7th of the time, or 14%.

A useful statistic in these situations is the odds ratio (Wright, 2002; see Swets, 1986, for alternatives). It is the ratio of the odds for a correct own age identification divided by the odds for a correct identification of someone from the other age group. For the young participants attempting to identify the young culprit after 1 day, the odds of making a correct identification is the number of correct identifications (28) divided by the number of incorrect identifications (32), which equals 0.88. An odds of 1 means that a correct identification and an incorrect identification are equally likely. Because this is less than 1 it means more people were incorrect than were correct. The odds for older participants correctly identifying a young culprit is 14/44 or 0.32. As it is smaller than the odds for the young participants, it means that the probability of a correct identification was lower than for the older participants in this condition. The odds ratio of these is the odds for the young participants divided by the odds for the older participants, which is 2.75. An odds ratio of 1 means both sets

of participants are equally likely to identify the culprit correctly. Here, the odds for the younger participants identifying a young culprit are 2.75 times more than for the older participants identifying a young culprit. The other odds ratios are reported in Table 2. Values above 1 correspond to more accurate own age identifications than identifications of the other age group. The higher the odds ratio, the stronger the own age effect. The own age bias was present in all conditions. The effect was largest after just 1 day and for the young culprits.

As the data are binary and include two within-subject factors, the most appropriate statistical technique to analyze these data would be a repeated measures form of logistic regression. However, because this is a relatively uncommon technique, and that an ANOVA gives a good approximation in these circumstances (with approximately equal cell sizes and percentages above 20%; Lunney, 1970), we report the results of an ANOVA for ease of communication (the logistic regressions led to the same conclusions for all the analyses).

An ANOVA shows that these data yield a significant interaction between age of participants and age of culprits, $F(1, 109) = 6.03, p = .02, \eta^2 = .05$. This shows that a statistically significant own age effect was observed. There is also the predicted main effect by whether there was a 1-day or a 1-week delay, $F(1, 109) = 10.53, p = .002, \eta^2 = .09$; people were more accurate after a day than they were after a week. The next largest effect, for people being slightly more accurate for the TV stealing videos, was nonsignificant, $F(1, 109) = 2.52, p = .12, \eta^2 = .02$. The results were analyzed separately for the young and the older targets. For the younger targets, the interaction was statistically significant, $F(1, 109) = 6.70, p = .01, \eta^2 = .06$. For the older targets, the interaction was in the predicted direction, but was nonsignificant, $F(1, 109) = 0.77, p = .38, \eta^2 = .01$. Thus, the effect is statistically significant for younger adults performing better than older adults with young targets, but we cannot reject the hypothesis that participants perform equally well with the older targets.

Confidence ratings were made each time a participant made an identification, whether correct or incorrect. In order to estimate the relationship between confidence and accuracy, a regression was run using the binary variable of whether the participant correctly chose the culprit to predict confidence. Because of the repeated nature of the design and that no confidence ratings were asked for after nonidentifications, an alternative to the standard statistical procedures was necessary. The data were reorganized as a multilevel data set, with individual responses nested with the individual participants. The package MLwiN (Rasbash, Healy, Browne, & Cameron, 1998) was used to calculate the differences between correct and incorrect identification (see Wright, Boyd, & Tredoux, in press, for methods using multilevel modelling for assessing the confidence/accuracy relationship). Figure 1 shows these differences in confidence ratings for accurate and inaccurate identifications with their 95% confidence intervals. All the mean differences are positive, indicating that accurate decisions tend to be made with more confidence, as has been found in previous research (Sporer, Penrod, Read, & Cutler, 1995). The 95% confidence intervals show that each is reliably above zero. Reading from left to right, the χ^2 and p values for the difference between accurate and inaccurate responses are 3.72 ($p = .05$), 16.34 ($p < .001$), 13.65 ($p < .001$), 12.64 ($p < .001$), 5.32 ($p = .02$), 3.74 ($p = .05$), 5.96 ($p = .01$), and 7.26 ($p = .01$). Thus, the confidence/accuracy relationship was

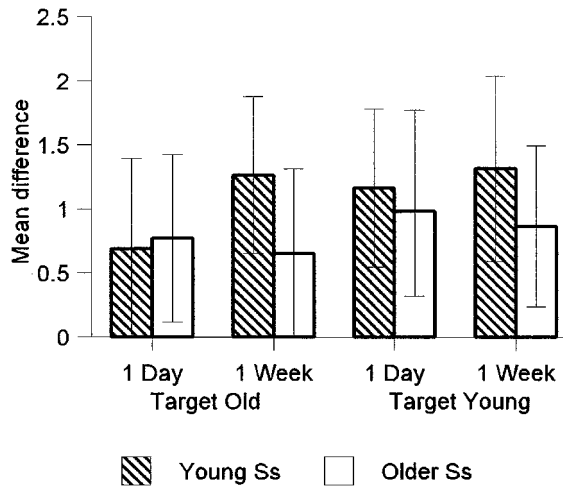


Fig. 1. The mean difference in confidence, with the 95% confidence intervals, between accurate and inaccurate responses. All means are above zero, showing that accurate identifications tend to be more confident than inaccurate identifications.

significant and positive for each condition. More important than the significance of any single test is the consistency across tests. One difference among the comparisons is that the younger participants had, overall, stronger confidence/accuracy associations. No other patterns emerged.

Discussion

Much research (e.g., Adams-Price, 1992; List, 1986; O'Rourke et al., 1989; Yarmey, 1993) has examined eyewitness accuracy across adults of different ages and the findings have been that older adults perform worse. However, these studies tend to use young adults as the culprits and fillers. Our results confirm this. When young culprits and fillers were used, the young adults outperformed the older adults. However, when the culprits and fillers were older adults, the younger participants did not do better than the older adults. The older adults, in fact, performed slightly better than the younger adults.

The observed interaction for the accuracy of face recognition in lineups is important for theories of face recognition and has implications for police lineups. With respect to theories of face recognition, this shows that adults of different age groups are better able to recognize people of their own age groups. Our pilot work showed that differences between the two age groups are easily detected. However, the differences may not be as large as between adults and children, between young and elderly adults, between males and females, or even between races. The size of the own age appears smaller than the own race bias. The odds ratios for own age bias (see Table 2) ranged from approximately change (1.13) to a fairly strong effect (2.75). Using a similar design, Wright et al. (2001) observed from own race odds ratios from 2.02 to 5.54.

While the experiment demonstrates an important effect, it leaves an important question unanswered. All of the lineups included the culprit. What happens with culprit absent lineups? This is the main focus of our second experiment. Searcy, Bartlett, and Memon (1999, 2000) have found that older participants make more false identifications than younger participants in memory recognition studies. Therefore, investigating the own age bias in situations where the only possible identifications are misidentification is of great importance. Further, as Experiment 1 was the first time the own age interaction has been observed in an eyewitness context, it is best to be cautious interpreting this effect without replication.

EXPERIMENT 2

Our second study was designed to replicate the own age bias in culprit present lineups and to test whether it occurs also in culprit absent lineups. If a manipulation increases correct identifications in culprit present lineups, it is possible that it will increase incorrect identifications in culprit absent lineups. This would occur if the manipulation lowered the necessary threshold for saying that the culprit was present. If this were the case then the manipulation may increase the chance of an innocent person being identified as the culprit. Therefore in the present study a culprit absent condition is included.

Method

Participants

The study had one between-subjects variable with two levels. Participants (total $n = 180$) were either between 18 and 33 or between 40 and 55 years old. Unlike the first study, we made both these ranges 16 years. Also, because the effect size was largest for the 1-day delay in our first study, to increase power, all participants here were tested with a 1-day delay. There were two within-subject variables. One was the type of crime witnessed and the other the age of culprit and fillers. The participants were all unpaid volunteers. None knew any of the culprits or fillers, had any experience with lineups, or took part in our earlier studies.

Materials

The four crime videos were the same as those used in the first experiment. The order of presentation was random. Four seven-person simultaneous photograph lineups were the same as those in the first experiment. The position of the culprit within the lineup was random.

Procedure

The experiment was run in two sessions. Participants were shown the stimuli and tested individually. During the first session participants were shown the four crime videos. The order of the videos was random. After seeing all the videos, participants were asked to fill out the same distracter questionnaire that was used in Experiment 1.

The next day participants were asked to give descriptions of each of the four culprits they had previously seen in the videos. They were told to give five points for each culprit. This was done to simulate that real eyewitness situations, where eyewitness are asked to describe the culprit. We recognize this could produce some verbal overshadowing effects (Schooler & Engstler-Schooler, 1990), but as all participants took part in this task it should not confound our results. Participants were then shown the four lineups in the same order as they saw the videos and instructed to identify the person they saw in each video. They were told that the culprit may or may not be in the lineup and if they could not make a positive identification then they should say so. They were allowed to take as much time as they wanted. Culprit absent lineups were made by removing the culprit from half of the lineups and replacing the culprit with another filler. The extra filler matched the appropriate description from Table 1. This was done randomly and therefore participants did not necessarily have two culprit absent and two culprit present lineups. We recognize that using two culprit absent and two culprit present lineups would have simplified the analysis.

Results

Table 3 shows the frequencies and percentages for the identifications from the different lineups, broken down by the age of the participants and the age of the culprits. Looking first at the culprit present lineups, about 40% of the time, the culprit was correctly identified, but also about 40% of the time a filler was chosen. The percentages for choosing the culprit and making a false identification varied across the different conditions. Overall, people were accurate about 10% more often when the culprit and fillers were of their own age. The odds of making a correct choice are 1.68 times higher for own age identifications than other age identifications. People were about 10% less likely to choose incorrectly a filler of their own age than a filler of the other age. The odds of an incorrect identification are 1.50 times higher if the culprit is of a different age than the participant. The percentages for no identification are approximately the same across the conditions. This replicates our findings of an own age bias for culprit present lineups.

Because we randomly assigned each trial as either a culprit present lineup or a culprit absent lineup the design is more complex than Experiment 1. Conventional statistical procedures cannot be easily used for inferential statistics here. We were

Table 3. The Frequencies and Percentages of Identification for Culprit Present and Culprit Absent Lineups, Broken Down by the Age of the Participant and the Age of the Culprit (and Fillers) for Experiment 2

	Lineup type				
	Culprit present			Culprit absent	
	Culprit	Filler	No ID	Filler	No ID
Young participant/young culprit	44 (49%)	27 (30%)	18 (20%)	58 (64%)	33 (36%)
Young participant/older culprit	33 (36%)	37 (41%)	21 (23%)	60 (67%)	29 (33%)
Older participant/young culprit	30 (32%)	42 (45%)	22 (23%)	60 (67%)	30 (33%)
Older participant/older culprit	40 (43%)	33 (36%)	19 (21%)	59 (64%)	33 (36%)
Total	40%	38%	22%	65%	35%

required to conduct a multilevel logistic regression (see Wright, 1998a, for an introduction) and used the package MLwiN for the analysis.⁵ Several models were investigated to predict accuracy. We found no effect for whether the scene was of a television or an automobile being stolen, $\chi^2(1) = 0.68$, $p = 0.41$. The main effects were not significant for either the age of the participants, $\chi^2(1) = 1.26$, $p = 0.26$, or for the age of the culprit, $\chi^2(1) = 0.02$, $p = 0.90$. The interaction term, which is the important term for this research because it measures the own age bias, was statistically significant, $\chi^2(1) = 7.12$, $p = 0.008$, and in the predicted direction, thus supporting the own age bias.

Now we turn to the culprit absent lineups. The first thing to note about the culprit absent conditions is that 65% of the participants incorrectly identified a filler. This is much higher than in the culprit present conditions, thus replicating Wells' finding that removing the culprit from a lineup often increases the number of false identifications in simultaneous lineups (Wells, 1993). False identifications occurred in about 3% more trials when the culprit and the participants were of different ages. This is a small difference and represents an odds ratio of only 1.15. It follows that correct responses, making no identification, were only 3% more common for own age trials, with an odds ratio of 1.15. While in the correct direction of an own age bias, this effect is much smaller than for the culprit present conditions, and as shown below is nonsignificant.

Multilevel logistic regressions were conducted also for the culprit absent lineups. There were no main effects for the age of the participant, $\chi^2(1) = 0.001$, $p = .97$, the age of the culprit, $\chi^2(1) = 0.012$, $p = .91$, or for what was being stolen, $\chi^2(1) = 0.599$, $p = .44$. The important effect was whether adding the interaction between the age of the participant and of the culprit improved the fit of the model. This effect was small (approx. 3%, odds ratio = 1.15) and nonsignificant, $\chi^2(1) = 0.382$, $p = .54$. The effect is too small for us to claim that an own age bias occurs with culprit absent lineups.

Discussion

When young culprits and fillers are used in culprit present lineups, younger participants are more accurate at lineup identifications than older participants. This replicates past research. However, this does not mean that the younger participants are better in general. When culprits and fillers are older, younger participants were not more accurate than older participants and again, as with Experiment 1, older participants were marginally more accurate with culprits of approximately their own age. This replication of the results from the first experiment supplies strong evidence for the existence of an own age bias.

⁵Some technical details are necessary to mention. Since there are few trials per participant the data set is what statisticians call *sparse*. Some estimation techniques have problems estimating such models when the response variable is binary (Rodríguez & Goldman, 1995). However, Goldstein and Rasbash (1996) and Rodríguez and Goldman (2001) have shown that more recent techniques lessen these problems. They show that second-order Taylor expansion with penalized quasi-likelihood estimation greatly improves the estimates (see Goldstein, 1995, for mathematical details). These were used for our analyses. We also allowed for extra-binomial variation, which tends to be underestimated when the data set is sparse (Wright, 1997).

The most important result of this experiment relates to the culprit absent condition. We observed the own age bias only with culprit present lineups, not with culprit absent lineups. This has important implications for mistaken identifications. It appears that older witnesses will not be more likely than younger participants to identify an innocent young suspect, but they will be more likely to fail to identify a guilty young culprit.

GENERAL DISCUSSION

These two experiments have demonstrated the existence of an own age bias in culprit present lineups. Older participants are better at identifying older culprits than younger culprits. Younger participants are better at identifying younger culprits than older culprits. The culprit absent condition yielded no significant effects. This suggests that false identifications do not increase in line with the own age bias.

The implications of the observed interaction for eyewitness testimony parallel those of cross-racial research. Cross race research has been important for eyewitness testimony in two ways. First, the results have been used to warn jurors that people have poorer memory for people of other races. The present research say that this advice can also be given for inter-generational identification accuracy. The testimony of adult witnesses of the same age as the suspect may be given more credence, than of a witness of a different age. Similarly, all other things being equal, a young witness would be more likely to identify a culprit of the person's own age than of another age.

Second, cross-racial research has been helpful for police officers recruiting fillers of people from other races. In England this was one of the main reasons for the introduction of lineup suites (Wright & McDaid, 1996). It has helped to make officers aware that differences, which they may not perceive as large, may be sizeable for people from other racial groups. These difficulties exist in much of the United States and other countries. Wells and Olson (2001) have argued how the reliability of cross-race identifications could be improved by, for example, using more fillers in cross-race lineups. Similar arguments might be made for cross-generational identifications. For example, an elderly person could help to construct a lineup if the suspect and fillers are elderly, and alternatively a young person could help with composing a lineup of younger people. In deciding how to assist the police, it is important to take into account that own age bias only appears in culprit present lineups.

This research is the first to show an own age bias. As such, further replications are necessary to determine the robustness and size of the effect. This is necessary before allowing expert scientific testimony on the topic and before advising police to alter policy. It is also premature to speculate on why the effect occurs. Future research should include different methods of presenting the stimuli. Here, videotaped incidents were shown. This has the advantage of having control over stimuli, but is not as realistic as having live incidents, a method used in some field studies. Further, only four culprits were used in these studies. It is important to also use methods, like those common in much face recognition research, where many target faces are shown. This makes it easier to generalize the results (Wells & Windschitl, 1999; Wright, 1998b). Finally, archival studies of real cases are necessary to show that the effect translates

into cases and this is critically important for influencing the public policy (Wells et al., 2000).

In summary, there are large and justified concerns about the accuracy of lineup identifications. Eyewitness errors account for a large number of innocent people being imprisoned. It is also likely that a large number of guilty people are being set free. We know about several characteristics of the lineup situation that influence accuracy. Here we explored the relative ages of the witnesses and the culprits. Several studies have previously examined the age of witnesses, and these studies, for the most part, had found that the older adults were worse than younger adults. We believe this finding only holds for identifications where the culprit is also young. Within the age groups we examined, we found that the accurate identifications were the highest when the witnesses were viewing culprits and fillers of the same age.

ACKNOWLEDGMENTS

This research was funded by the United Kingdom's Economic and Social Research Council (ESRC).

REFERENCES

- Adams-Price, C. (1992). Eyewitness memory and aging: Predictors of accuracy in recall and person recognition. *Psychology and Aging, 7*, 602–608.
- Bartlett, J. C., & Leslie, J. E. (1986). Aging and memory for pictures of faces. *Memory and Cognition, 14*, 371–381.
- Chance, J. E., Goldstein, A. G., & Andersen, B. (1986). Recognition memory for infant faces: An analog of the other-race effect. *Bulletin of the Psychonomic Society, 24*, 257–260.
- Connors, E., Lundregan, T., Miller, N., & McEwan, T. (1996). *Convicted by juries, exonerated by science: Case studies in the use of DNA Evidence to establish innocence after trial*. NIJ Research Report. U.S. Department of Justice.
- Fulton, A., & Bartlett, J. C. (1991). Young and old faces in young and old heads: The factor of age in face recognition. *Psychology and Aging, 6*, 623–630.
- George, P. A., Hole, G. J., & Scaife, M. (2000). Factors influencing young children's ability to discriminate unfamiliar faces by age. *International Journal of Behavioral Development, 24*, 480–491.
- Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.
- Goldstein, H., & Rasbash, J. (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society: A, 159*, 505–514.
- Grady, C. L., & Craik, F. I. (2000). Changes in memory processing with age. *Current Opinion in Neurobiology, 10*, 224–231.
- 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. *American Psychologist, 56*, 405–416.
- Lunney, G. H. (1970). Using analysis of variance with a dichotomous dependent variable: An empirical study. *Journal of Educational Measurement, 7*, 263–269.
- Lindsay, R. C. L. (1994). Biased lineups: Where do they come from? In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 182–200). Cambridge, UK: Cambridge University Press.
- List, J. (1986). Age and schematic differences in the reliability of eyewitness testimony. *Developmental Psychology, 22*, 50–57.
- Mason, S. E. (1986). Age and gender as factors in facial recognition and identification. *Experimental Aging Research, 12*, 151–154.
- Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*, 3–35.
- O'Rourke, T. E., Penrod, S. D., Cutler, B. L., & Stuve, T. E. (1989). The external validity of eyewitness identification research: Generalizing across subject populations. *Law and Human Behavior, 13*, 385–395.

- Rasbash, J., Healy, M., Browne, W., & Cameron, B. (1998). *MLwiN* (version 1.00). Multilevel Models Project, Institute of Education, London.
- Rodríguez, G., & Goldman, N. (1995). An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society: A*, *158*, 73–89.
- Rodríguez, G., and Goldman, N. (2001). Improved estimation procedures for multilevel models with binary response: A case study. *Journal of the Royal Statistical Society: A*, *164*, 339–356.
- Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, *22*, 36–71.
- Searcy, J. H., Bartlett, J. C., & Memon, A. (1999). Age differences in accuracy and choosing rates on face recognition and eyewitness identification tasks. *Memory and Cognition*, *27*, 538–552.
- Searcy, J. H., Bartlett, J. C., & Memon, A. (2000). Relationship of availability, lineup conditions and individual differences to false identification by young and older eyewitnesses. *Legal and Criminological Psychology*, *5*, 219–236.
- Sporer, S. L., Penrod, S., Read, J. D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, *118*, 315–327.
- Swets, J. A. (1986). Indexes of discrimination or diagnostic-accuracy: Their ROCs and implied models. *Psychological Bulletin*, *99*, 100–117.
- Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement*. Washington, DC: United States Department of Justice, Office of Justice Programs.
- Tredoux, C. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior*, *22*, 217–237.
- Tredoux, C. (1999). Statistical considerations when determining measures of lineup size and lineup bias. *Applied Cognitive Psychology*, *13*, S9–S26.
- Wells, G. L. (1978). Applied eyewitness testimony research: System variables versus estimator variables. *Journal of Personality and Social Psychology*, *36*, 1546–1557.
- Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, *48*, 553–571.
- Wells, G. L., Malpass, R. S., Lindsay, R. C. L., Fisher, R. P., Turtle, J. W., & Fulero, S. M. (2000). From the lab to the police station: A successful application of eyewitness research. *American Psychologist*, *55*, 581–598.
- Wells, G. L., & Olson, E. A. (2001). The other-race effect in eyewitness identification. What do we do about it? *Psychology, Public Policy, and Law*, *7*, 230–246.
- Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling in social psychological experimentation. *Personality and Social Psychology Bulletin*, *25*, 1115–1125.
- Wright, D. B. (1997). Extra-binomial variation in multilevel logistic models with sparse structures. *British Journal of Mathematical and Statistical Psychology*, *50*, 21–29.
- Wright, D. B. (1998a). Modelling clustered data in autobiographical memory research: The multilevel approach. *Applied Cognitive Psychology*, *12*, 339–357.
- Wright, D. B. (1998b). People, materials and situations. In J. A. Nunn (Ed.), *Laboratory psychology* (pp. 97–116). Hove, England: Erlbaum.
- Wright, D. B. (2002). *First steps in statistics*. London: Sage.
- Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2001). A field study of own-race bias in South Africa and England. *Psychology, Public Policy, and Law*, *7*, 119–133.
- Wright, D. B., Boyd, C. E., & Tredoux, C. G. (in press). Inter-racial contact and the own race bias for face recognition in South Africa and England. *Applied Cognitive Psychology*.
- Wright, D. B., & Davies, G. M. (1999). Eyewitness testimony. In F. T. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay, & M. T. H. Chi, (Eds.), *Handbook of applied cognition* (pp. 789–818). West Sussex, UK: Wiley.
- Wright, D. B., & McDaid, A. T. (1996). Comparing system and estimator variables using data from real line-ups. *Applied Cognitive Psychology*, *10*, 75–84.
- Yarmey, A. D. (1993). Adult age and gender differences in eyewitness recall in field settings. *Journal of Applied Social Psychology*, *23*, 1921–1932.