Psychology and Aging
2017, Vol. 32, No. 3, 243–258

0882-7974/17/$12.00 http://dx.doi.org/10.1037/pag0000168

A Signal-Detection Analysis of Eyewitness Identification Across the Adult Lifespan

Melissa F. Colloff and Kimberley A. Wade
University of Warwick

John T. Wixted
University of California, San Diego

Elizabeth A. Maylor
University of Warwick

Middle-aged and older adults are frequently victims and witnesses of crime, but knowledge of how identification performance changes over the adult life span is sparse. The authors asked young (18–30 years), middle-aged (31–59 years), and older (60–95 years) adults (N = 2,670) to watch a video of a mock crime and to attempt to identify the culprit from a fair lineup (in which all of the lineup members matched the appearance of the suspect) or an unfair lineup (in which the suspect stood out). They also asked subjects to provide confidence ratings for their identification decisions. To examine identification performance, the authors used a standard response-type analysis, receiver operating characteristic analysis, and signal-detection process modeling. The results revealed that, in fair lineups, aging was associated with a genuine decline in recognition ability—discriminability—and not an increased willingness to choose. Perhaps most strikingly, middle-aged and older adults were generally effective at regulating their confidence judgments to reflect the likely accuracy of their suspect identification decisions. Model-fitting confirmed that the older adults spread their decision criteria such that identifications made with high confidence were likely to be highly accurate, despite the substantial decline in discriminability with age. In unfair lineups, ability to discriminate between innocent and guilty suspects was poor in all age groups. The research enhances theoretical understanding of the ways in which identification behavior changes with age, and has important practical implications for how legal decision-makers should interpret identifications made by middle-aged and older eyewitnesses.

Keywords: eyewitness identification, aging, discriminability, response bias, confidence and accuracy

Supplemental materials: http://dx.doi.org/10.1037/pag0000168.supp

Imagine that you are a police officer investigating a crime. You have only one witness, a 69-year-old, whose ability to recognize the culprit is critical for your case. How might your witness's ability to make an accurate identification from a lineup be different to that of a young or middle-aged adult? Now imagine that you are a judge deliberating the verdict. Can you trust the identification made by this older witness to the same extent that you might trust an identification made by a younger witness? In nearly every country, the proportion of people aged 60 and over is growing faster than any other age group (World Health Organization, 2015), and middle-aged and older adults are frequently witnesses or victims of crime (e.g., Acierno et al., 2010; Willoughby, 2015). Yet, knowledge of how eyewitness identification performance changes with age is limited (Fitzgerald & Price, 2015). In this study, we aimed to learn more about eyewitness identification behavior in middle-aged and older adults by examining their ability to identify culprits and gauge the accuracy of their identification decisions.

A lineup usually contains one police suspect who is either guilty (i.e., the real culprit) or innocent, and a number of other lineup members, called foils, who are known to be innocent. Many eyewitness identification studies have shown that older adults make more mistakes in lineup tasks than do young adults. Older adults are more likely than young adults, for instance, to make an incorrect identification when the real culprit is not in the lineup (see Bartlett & Memon, 2007, for a review). Early studies also found that older adults are more likely to select a person from a lineup than are their young counterparts (see Sporer & Martschuk, 2014, for a review). As a result, many researchers have, explicitly or implicitly, suggested that the age-related decline in identification accuracy occurs because older adults are too willing to make an identification decision (e.g., Sporer & Martschuk, 2014; Wilcock, Bull, & Vrij, 2005). However, attempts to reduce older

Melissa F. Colloff and Kimberley A. Wade, Department of Psychology, University of Warwick; John T. Wixted, Department of Psychology, University of California, San Diego; Elizabeth A. Maylor, Department of Psychology, University of Warwick.

Sections of these data were presented by Melissa F. Colloff at the Society for Applied Research in Memory and Cognition (June 2015), Victoria, BC, Canada.

We thank Sophie Nightingale, Harriet Rowthorn, and James Bartlett for insightful comments, and members of the University of the Third Age for participating in our research.

Correspondence concerning this article should be addressed to Kimberley A. Wade, Department of Psychology, University of Warwick, Coventry, United Kingdom, CV4 7AL. E-mail: k.a.wade@warwick.ac.uk

adults' false identification rates—by reducing proclivity to choose—have not been effective in eradicating the age-related deficit in performance (Memon & Gabbert, 2003; Rose, Bull, & Vrij, 2005; Wilcock et al., 2005). It seems that an increased willingness to choose with age is not the whole story.

Indeed, there are good reasons to expect that aging is associated with a genuine decline in recognition accuracy—also known as *discriminability*—and not just an increased willingness to choose. Healthy aging is associated with a number of changes in memory function, but one prominent theory suggests that people become increasingly reliant on familiarity with age and this tendency promotes memory errors (Healy, Light, & Chung, 2005; Searcy, Bartlett, & Memon, 1999). According to dual-process accounts of memory, recognition is based on two processes: recollection and familiarity (see Mandler, 1980; Yonelinas, 2002, for reviews). Recollection involves retrieving specific contextual information about the original stimulus, such as source, time, place, thoughts, and feelings, whereas familiarity is a sense that the stimulus has previously been encountered without retrieving any contextual details. Evidence from several different paradigms including old/new word recognition studies (Dywan & Jacoby, 1990; Jacoby, 1999; Jennings & Jacoby, 1997), face recognition studies (Bartlett & Fulton, 1991; Bartlett, Strater, & Fulton, 1991; Edmonds, Glisky, Bartlett, & Rapcsak, 2012), and lineup tasks (Searcy et al., 1999; Searcy, Bartlett, & Memon, 2000; Searcy, Bartlett, Memon, & Swanson, 2001), suggest that older adults have deficits in recollecting diagnostic source specific information and, as a result, are more reliant on less diagnostic familiarity processes than are their younger counterparts.

What does this mean for older adults' ability to discriminate between who is innocent and who is guilty in a lineup? Faces in a lineup are highly homogenous (Diamond & Carey, 1986), so even faces that have never been seen before could evoke a feeling of familiarity (Bartlett, Hurry, & Thorley, 1984; Young, Hay, McWeeny, Flude, & Ellis, 1985). Because older adults are poorer at recollecting diagnostic details associated with a previously seen face, they may rely on familiarity to a greater extent than young adults, thereby making it harder for them to tell if a person in the lineup is innocent or guilty.

Indeed, face recognition studies show that discriminability declines with age (e.g., Fulton & Bartlett, 1991; Lamont, Stewart-Williams, & Podd, 2005). Three meta-analyses of lineup research have shown that, compared to young adults, older adults make more false identifications when the culprit is not in the lineup, but also fewer correct identifications when the culprit is in the lineup (Bartlett, 2014; Fitzgerald & Price, 2015; Sporer & Martschuk, 2014). Only three studies, however, have directly measured young and older adults' ability to discriminate between innocent and guilty suspects as well as their willingness to identify the suspect. One study calculated overall choosing rate and signal-detection estimates of discrimination (d') and response bias (c) for 21 published lineup studies. The authors concluded that although older adults do choose from lineups at a higher rate than young adults, it was an impaired ability to discriminate between innocent and guilty suspects that hindered older adults' performance (Fitzgerald & Price, 2015; see also Wylie, Bergt, Haby, Brank, & Bornstein, 2015). By contrast, Key et al. (2015) measured people's ability to discriminate between innocent and guilty suspects in fair lineups (where the foils matched the appearance of the suspect) and unfair lineups (where the suspect stood out because the foils did not match the appearance of the suspect) using receiver operating characteristic (ROC) analysis. Surprisingly, Key et al. found no difference between their young and older samples on either lineup type.

If people's ability to discriminate between innocent and guilty suspects declines with age, should the Criminal Justice System disregard identifications made by older, or even middle-aged, adults? Somewhat surprisingly, merely knowing that older adults have lower discriminability does not provide us with the information needed to answer that question. To answer that question, we need to consider whether older adults can assess the likely accuracy of their memories and assign appropriate confidence judgments (Mickes, 2015). That is, do older adults express high confidence in their decision when their answer is correct, and lower confidence when their answer is incorrect, and do they do so to the same degree as younger people? If they do, then a high-confidence identification from an older adult would be as trustworthy as a high-confidence identification from a younger adult even though older adults exhibit reduced discriminability.

Gauging the Accuracy of Identifications

Eyewitness research on confidence judgments in older adults is mixed. Some lineup studies have found that accuracy and confidence are better correlated in young people than in older people (Adams-Price, 1992; Memon, Hope, Bartlett, & Bull, 2002; Wylie et al., 2015), and a recent review concluded that confidence should not be used as a proxy for accuracy in older adults (Erickson, Lampinen, & Moore, 2016). Also, older adults often make high-confidence errors (Dodson, Bawa, & Krueger, 2007; Dodson, Bawa, & Slotnick, 2007; Dodson & Krueger, 2006), and older adults who rate their memory self-efficacy as higher are more likely to make false identifications (Searcy et al., 2000, 2001). These studies may indicate that older adults tend to be overconfident in the validity of weaker memory signals—they may fail to adjust their confidence judgments appropriately to reflect their lower likelihood of accuracy.

However, it may be premature to conclude that older adults are unable to assign appropriate confidence judgments. Many of the lineup studies (e.g., Adams-Price, 1992; Memon et al., 2002; Wylie et al., 2015) have calculated the correlation coefficient, yet a low correlation coefficient does not necessarily indicate a poor relationship between confidence and accuracy (Juslin, Olsson, & Winman, 1996). Correlation coefficients reflect the relationship between categorical confidence judgments (0, 10, 20, etc.) and binary accuracy (correct or incorrect). When displayed in a graph, confidence is plotted on the x-axis and accuracy (correct or incorrect) on the y-axis, and each point represents the confidence and accuracy of one person. Correlation coefficients fit a straight line through these data, and the distribution of confidence judgments heavily influences the line. Subjects' confidence judgments in empirical studies usually fall within a restricted range (i.e., the distribution of confidence judgments is unimodal) and this serves to underestimate the relationship between confidence and accuracy (Juslin et al., 1996; Lindsay, Read, & Sharma, 1998). Furthermore, because accuracy is plotted as a binary outcome for each person, correlation coefficients do not provide information about the likely accuracy of an identification made with a particular level of confidence (Brewer & Wells, 2006; Juslin et al., 1996). A more suitable statistical technique for testing whether people can assess the likely accuracy of their memories is to plot their *average* accuracy at different levels of confidence—that is, plot confidence-accuracy curves. Only this technique reveals the likely accuracy of an identification made with a particular level of confidence. It also remains unaffected by the distribution of confidence judgments because average accuracy (i.e., probability of a correct identification decision) at a particular level of confidence is the same, regardless of the number of identifications made at that level of confidence (Brewer, Keast, & Rishworth, 2002; Brewer & Wells, 2006; Juslin et el., 1996; Mickes, 2015).

To our knowledge, Key et al. (2015) is the only study to have plotted confidence-accuracy curves for young and older adults in an eyewitness identification paradigm. When older adults made suspect identifications with the highest level of confidence, they were as likely to be correct as were young adults. This finding should be interpreted with caution, though, because the young and older groups were also equivalent in discriminability. Therefore, the study by Key et al. does not tell us whether older adults can assess the accuracy of their memories to the same extent as young adults when their memory ability is worse. Nevertheless, many other eyewitness studies have found that older adults tend to assign lower confidence ratings to their identification decisions on average than young adults, which may suggest that older adults are aware that they are less accurate (Goodsell, Neuschatz, & Gronlund, 2009; Memon et al., 2002; Neuschatz et al., 2005; Searcy et al., 2001; Wylie et al., 2015; but see Havard & Memon, 2009; Searcy et al., 1999). If middle-aged and older adults are able to gauge the likely accuracy of their memories, then they should be as accurate as young adults at each level of confidence, despite any decline in memory ability that occurs with age.

The Current Study

We aimed to answer two questions: (a) Is the age-related decline in accurate identification decisions due to an increased willingness to make an identification, a decline in discriminability, or both? (b) Are middle-aged and older adults able to gauge the likely accuracy of their identification decisions to the same extent as young adults? To answer these questions, we extended Colloff et al. (2016), which examined fair versus unfair lineup techniques for suspects with distinctive features (e.g., scars, piercings; see also Zarkadi, Wade, & Stewart, 2009). In Colloff et al.'s study, subjects watched one of four videos (carjacking, graffiti, mugging, theft) in which different distinctive culprits committed a nonviolent mock crime. After a short delay, subjects were presented with a lineup constructed using one of four techniques and made an identification decision (see Figure 1). Lineups either contained the culprit (target-present) or did not (target-absent). Subjects performed similarly on the three fair lineup techniques, and all three fair lineup techniques enhanced subjects' ability to discriminate between innocent and guilty suspects in comparison with unfair lineups in which the suspect was the only person in the lineup with the distinctive feature. In the current study, we combined a subset of Colloff et al.'s data with newly collected data and we planned to collapse the data over the three fair lineup techniques. Data collection for both studies occurred within a 9-month period.

Method

Subjects

Older adults. We collected data from 1,285 subjects aged over 60 by contacting University of the Third Age groups from around the United Kingdom. Subjects were not paid but were offered the chance to learn about the research. We excluded subjects who failed to report their age ($n = 4$), experienced technical difficulties ($n = 38$), stated they had seen the video before ($n = 9$), or incorrectly answered an attention check question ($n = 10$). Colloff et al. (2016) included 8,925 subjects aged between 16 and 91. Of these, 346 subjects were aged over 60. We added these to our cleaned older sample ($n = 1{,}224$) to make a total of 1,570 older adults.

Young and middle-aged adults. We randomly sampled 1,570 people aged 18–30 and 1,570 people aged 31–59 from Colloff et al.'s (2016) dataset. To the extent possible, we matched the young and middle-aged samples with our older sample on gender and ethnicity in each condition.

Final sample. Although we initially planned to analyze the data from all four videos, ultimately, we only analyzed the data

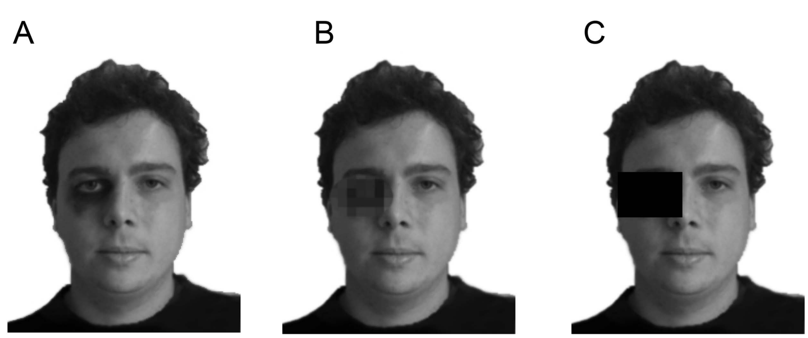

Figure 1. Examples of a suspect whose distinctive feature is visible (A), or has been concealed with pixelation (B) or a block (C). In fair *replication* lineups, all of the lineup members had the same distinctive feature as the suspect. In fair *pixelation* lineups, all of the lineup members had the same area concealed by pixelation. In fair *block* lineups, all of the lineup members had the same area concealed with a black block. In unfair *do-nothing* lineups, the suspect's feature was visible and the other lineup members had no distinctive features.

from the graffiti and mugging videos because identification performance for the other two videos was very low for young subjects, and at floor for older subjects (see online supplemental materials). Limiting the analysis to the graffiti and mugging videos resulted in a final sample size of 890 older adults (163 from Colloff et al., 2016, and 727 new recruits; see online supplement for background performance measures), and 890 middle-aged adults and 890 young adults from Colloff et al. Table 1 shows a demographic breakdown of the final sample. There were between 89 and 117 subjects in each of the eight cells of the design. The Department of Psychology Research Ethics Committee, University of Warwick, approved this research.

Materials

Each video was 30 s in length. In the graffiti scenario, a White male culprit in his early 20s with severe bruising around his right eye approaches a wall while shaking a can of spray paint. He checks for witnesses, and then paints "UNI SUCKS" on the wall. In the mugging scenario, a White male culprit in his early 20s with a facial tattoo on his right cheek approaches and instructs another White male in his late 20s to give him his phone. When the victim refuses, the culprit pushes the victim, snatches the phone and runs away.

Target-present lineups contained the culprit (i.e., the guilty suspect) and five foils, and target-absent lineups contained six foils. Each culprit (one from each video) had a pool of 40 foil faces matching their description, and lineups were randomly generated from these pools (see Colloff et al., 2016, for details about lineup materials and checks). Randomly generated lineups ensure that the findings are not limited to the idiosyncrasies of a small number of guilty-innocent suspect pairs and do not require fairness and bias measures which are known to be unstable (Mansour, Beaudry, Kalmet, Bertrand, & Lindsay, 2017). For replication lineups, the culprit's distinctive feature (i.e., black-eye or tattoo) had been digitally added to each foil in the pool. For pixelation lineups, the culprit's feature had been concealed by pixelating the area, and the same area had been pixelated on each foil in the pool. For block lineups, the culprit's feature had been concealed by overlaying the area with a solid black rectangle, and the same area had been covered with the same shape on each foil in the pool. In target-present do-nothing (unfair) lineups, the culprit's feature was visible and the foils had no distinctive features. In target-absent do-nothing lineups, one foil had the culprit's distinctive feature (i.e., the innocent suspect), while the remaining five foils had no distinctive features (see Figure 1).

Table 1
Demographic Information for the Young, Middle-Aged, and Older Groups

Demographic	Young	Middle-aged	Older
Gender			
Male	311	292	307
Female	579	598	583
Age (years)			
M	22.48	42.49	68.82
SD	3.70	8.27	6.41
Range	18–30	31–59	60–95
Race or ethnicity			
White/European	856	861	853
Latin/Hispanic	1	1	0
Black/African/Caribbean	9	9	8
South Asian	5	8	5
East Asian	0	1	0
Middle Eastern	1	0	1
Mixed	5	5	6
Other	3	3	3
Prefer not to say	10	2	14

Procedure

The eyewitness memory procedure was identical to that of Colloff et al. (2016). Subjects completed the study online and were told that the study was about "personality and perception." Subjects first watched the mock crime video and were instructed to pay close attention because they would be asked questions about it later. Subjects were asked if they had encountered any technical problems while watching the video. Next, the 8-min filler task began: Subjects were given three questionnaires and an anagram puzzle. They were then asked to indicate their confidence that they would be able to recognize the culprit from the video.[1] Following this, subjects were asked to attempt to identify the culprit, and they were warned that the culprit "may or may not be present." The lineup images were presented simultaneously in an array of two rows of three photos. The position of the lineup members in the array was randomly determined for each subject. Subjects had to click on the face who they believed was the culprit, or click on a button labeled "not present" if they believed the culprit was not in the lineup. Subjects rated their confidence in their decision, using a 100-point Likert-type scale ranging from 1 (*completely uncertain*) to 100 (*completely certain*), and then answered an attention check question ("What happened in the video that you watched?") and some demographic questions.

Results

We examined subjects' identification responses, conducted ROC analysis and fit a signal-detection process model of identification performance (Wixted & Mickes, 2014). We also plotted confidence-accuracy curves. Preliminary analyses confirmed that subjects performed similarly on the three fair lineups (see online supplement). Therefore, within each age group, we collapsed the data over the replication, pixelation and block lineups.

Identification Responses

We calculated the proportion of suspect identifications, foil identifications and lineup rejections (i.e., "not present" responses) in the fair and unfair lineups. Figure 2 shows the identification responses made by the young, middle-aged, and older adults in (a) target-present and (b) target-absent lineups, as a function of lineup type. The number of innocent suspect identifications in the unfair target-absent lineups was the number of times the lineup member with the distinctive feature was identified. We estimated the num-

[1] We do not discuss subjects' ratings on this scale because it is outside the scope of the current paper.

ber of innocent suspect identifications in the fair target-absent lineups by dividing the number of false identifications in target-absent lineups by the number of lineup members (i.e., six; Brewer & Wells, 2006; Mickes, 2015). We estimated the number of foil identifications by dividing the total number of false identifications by six (the number of lineup members), and then multiplying by five (the number of lineup members that were not the innocent suspect). In a completely fair target-absent lineup, this estimation technique returns the same mean estimate of the number of innocent suspect identifications as predesignating a single individual to be the innocent suspect (Wixted & Wells, 2016).

We used hierarchical loglinear analysis to examine the identification responses. In loglinear analysis, a two-way interaction indicates that there is a relationship between two of the variables and is conceptually similar to a main effect in a linear model (e.g., ANOVA). Statistical significance is assessed using standardized residuals (z-scores), which are the difference between the frequencies observed and the frequencies that would be expected if there were no relationship between the variables (Field, Miles, & Field, 2012).

Target-present lineups. Figure 2A shows that there was a decline in the number of accurate responses with age. A 3 (age:

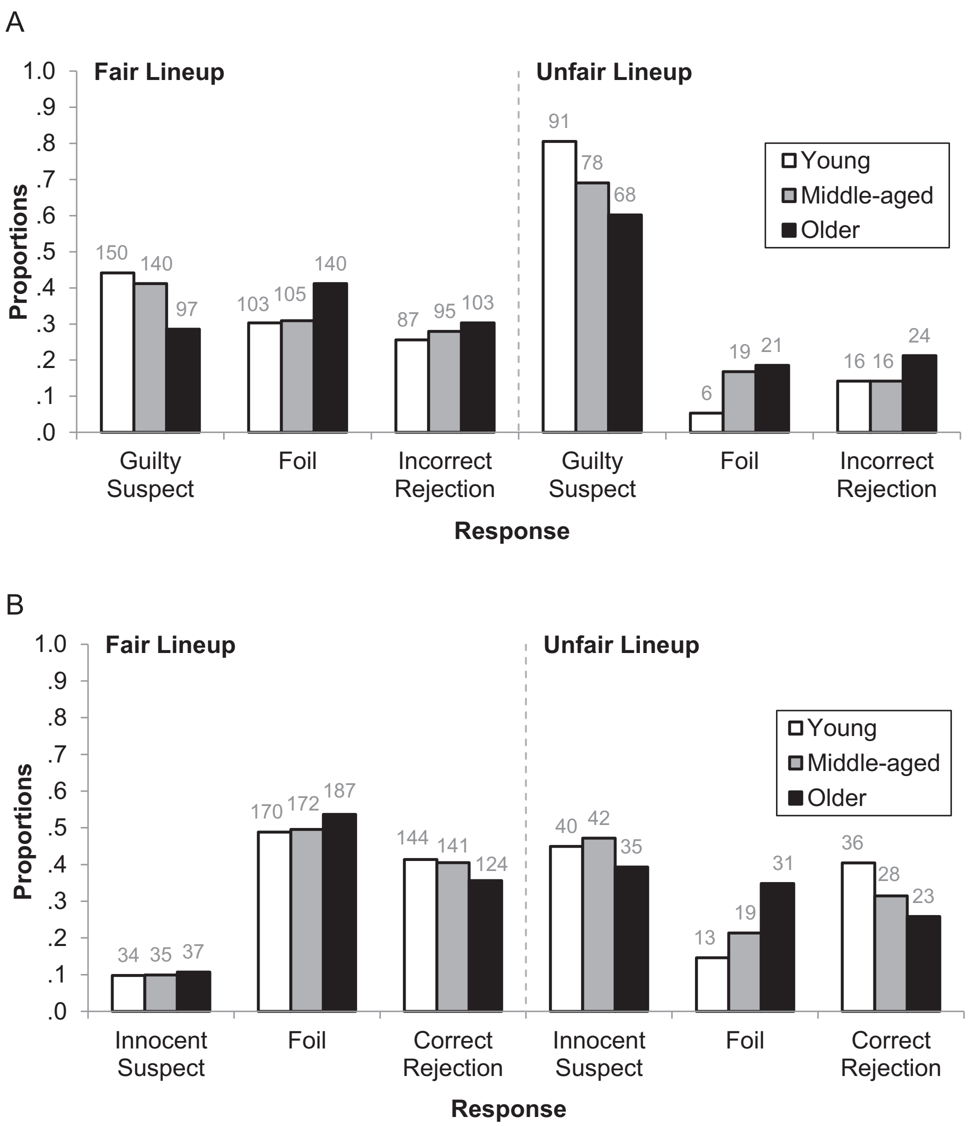

Figure 2. Identification responses made by the young, middle-aged, and older adults in (A) target-present and (B) target-absent lineups, as a function of lineup type. In fair target-absent lineups, the number of innocent suspect identifications was estimated by dividing the total number of false identifications by 6, and the number of foil identifications was estimated by dividing the total number of false identifications by 6 and then multiplying this by 5. In unfair target-absent lineups, the number of innocent suspect identifications was the number of times the person with the distinctive feature was identified and the number of foil identifications was the number of times a foil without the distinctive feature was identified. Data labels are absolute frequencies.

young, middle-aged, older) × 2 (lineup type: fair, unfair) × 3 (identification response: guilty suspect, foil, incorrect rejection) hierarchical loglinear analysis revealed a significant two-way interaction, indicating that age influenced identification responses, $\chi^2(4, N = 1{,}359) = 30.82$, $p < .001$ (likelihood ratio: $\chi^2(8) = 37.10$, $p < .001$). All three age groups made a similar number of lineup rejections, but the number of guilty suspect identifications decreased and the number of foil identifications increased with age. Older adults made fewer guilty suspect identifications ($z = -2.98$, $p < .01$) but more foil identifications ($z = 2.59$, $p < .01$) than expected, and young adults made more guilty suspect identifications ($z = 2.29$, $p < .05$) and fewer foil identifications ($z = -1.95$, $p > .05$) than expected. Three 2 (age) × 2 (identification response: guilty suspect, foil) two-way chi-square analyses indicated that when subjects made a selection from the lineup, older adults were 1.71 times more likely to identify a foil than middle-aged adults, $\chi^2(1, N = 668) = 11.76$, $p < .001$, odds ratio (*OR*) = 1.71, 95% confidence interval (CI) [1.24, 2.37], and 2.15 times more likely to identify a foil than young adults, $\chi^2(1, N = 676) = 23.42$, $p < .001$, *OR* = 2.15, 95% CI [1.56, 2.99]. But middle-aged adults were not significantly more likely to identify a foil than young adults, $\chi^2(1, N = 692) = 2.03$, $p = .15$, OR = 1.26, 95% CI [0.91, 1.75]. In short, older subjects made more incorrect identifications and fewer correct identifications in target-present lineups than middle-aged and young subjects.

The loglinear analysis also revealed a significant two-way interaction indicating that lineup technique influenced identification responses, $\chi^2(4, N = 1{,}356) = 112.05$, $p < .001$ (likelihood ratio: $\chi^2(6) = 118.33$, $p < .001$). Fair lineups led to fewer guilty suspect identifications ($z = -3.76$, $p < .001$), but more foil identifications ($z = 3.04$, $p < .01$) and more rejections ($z = 1.82$, $p > .05$) than expected. Conversely, unfair lineups led to more guilty suspect identifications ($z = 6.52$, $p < .001$), but fewer foil identifications ($z = -5.27$, $p < .001$) and fewer rejections ($z = -3.15$, $p < .01$) than expected. Specifically, subjects were 3.80 times more likely to make a correct identification in the unfair lineups compared to the fair lineups, $\chi^2(1, N = 1{,}356) = 104.72$, $p < .001$, *OR* = 3.80, 95% CI [2.90, 5.00]. This suggests that when the guilty suspect stood out, subjects in all age groups identified the guilty suspect instead of identifying another foil or rejecting the lineup.

Target-absent lineups. Figure 2B shows that there was a decline in the number of accurate (reject) responses with age. We conducted a 3 (age: young, middle-aged, older) × 2 (lineup type: fair, unfair) × 3 (identification response: innocent suspect, foil, correct rejection) hierarchical loglinear analysis. The two-way interaction between age and identification response did not reach statistical significance, $\chi^2(4, N = 1{,}311) = 7.36$, $p = .11$ (likelihood ratio: $\chi^2(8) = 14.31$, $p = .07$), but the numerical trends indicated that older adults tended to make fewer rejections but more foil identifications than expected, and young adults tended to make more rejections and fewer foil identifications than expected.

The loglinear analysis revealed a significant two-way interaction indicating that lineup type influenced identification responses, $\chi^2(4, N = 1{,}311) = 155.01$, $p < .001$ (likelihood ratio: $\chi^2(6) = 162.37$, $p < .001$). Although fair and unfair lineups led to a similar number of lineup rejections (fair: $z = 0.71$, $p > .05$; unfair: $z = -1.40$, $p > .05$), fair lineups led to fewer innocent suspect identifications ($z = -5.38$, $p < .001$), but more foil identifications ($z = 2.65$, $p < .01$) than expected. Conversely, unfair lineups led to more innocent suspect identifications ($z = 10.63$, $p < .001$), but fewer foil identifications ($z = -5.25$, $p < .001$) than expected. Specifically, when subjects made an identification, they were 13.96 times more likely to identify the innocent suspect in the unfair lineups compared to the fair lineups, $\chi^2(1, N = 875) = 259.45$, $p < .001$, *OR* = 13.96, 95% CI [9.69, 20.34]. This suggests that when the innocent suspect stood out, subjects in all age groups shifted their identifications from the other lineup members onto the innocent suspect.

In sum, our results indicate that the number of erroneous identifications increased with age. Unfair lineups also led to more correct identifications in target-present lineups but more incorrect identifications of innocent suspects in target-absent lineups, in all age groups.

ROC Analysis

Next, we conducted ROC analysis to investigate whether the patterns of identification responses were due to changes in subjects' ability to discriminate between guilty and innocent suspects, or subjects' willingness to identify the suspect. Traditional methods of assessing identification accuracy (e.g., by calculating proportions of identifications) confound discriminability and response bias (e.g., Mickes, Flowe, & Wixted, 2012; Wixted & Mickes, 2012). ROC analysis, by contrast, is a theory-free technique that (when applied to studying eyewitness performance) plots correct and incorrect identification rates over decreasing levels of confidence (Gronlund, Wixted, & Mickes, 2014). Confidence is used as a proxy for willingness to choose, so ROC analysis allows us to measure ability to discriminate, independently from willingness to choose (National Research Council, 2014). For fair lineups, ROC analysis can be thought of as measuring either the ability to discriminate guilty suspects from innocent suspects, *or* the ability to discriminate guilty suspects from foils. This is because when the lineup is fair, innocent suspect identifications and foil identifications are one and the same. However, for unfair lineups, it is necessary to define which discriminability one wishes to measure. One can measure the ability to discriminate (a) guilty suspects from innocent suspects, (b) guilty suspects from foils, or (c) innocent suspects from foils. The most important question for applied purposes concerns the ability to discriminate between guilty and innocent suspects, because suspect identifications can result in criminal proceedings, whereas foil identifications do not. Thus, we constructed an ROC that measures ability to discriminate between guilty and innocent suspects, but all three discriminability measures are estimated when we fit a theoretical model to our data (see online supplement).

To construct our ROC curves, we collapsed our data to an 11-point confidence scale by rounding subjects' confidence ratings to the nearest 10, so that each ROC curve would have 11 operating points of decreasing confidence (100, 90, 80, etc.). We calculated the correct identification rate (hit rate; HR) and the false identification rate (false alarm rate; FAR) for each level of decreasing confidence, such that the first HR/FAR pair (plotted on the lower left of each curve) was calculated using subjects who had made an identification with a confidence of 100, the second HR/FAR pair included subjects who had made an identification with a confidence of 100 or 90, and so forth. HR was the number of guilty suspect identifications ÷ number of target-present lineups. FAR was the number of innocent suspect identifications ÷ number of target-absent lineups. Again, the number of innocent suspect identifications in unfair lineups was the number of times the lineup member with the distinctive feature was identified. We estimated the number of innocent suspect identifications in fair target-absent lineups by dividing the number of false identifications by the number of lineup members, that is, six. Figure 3 shows the ROC curves for the fair and unfair lineups in the young, middle-aged and older subjects.

To statistically compare the ROC curves, we used the statistical package *pROC* to calculate the partial area under the curve (*p*AUC) and *D,* a measure of effect size: $D = (\text{AUC1} - \text{AUC2})/s$, where *s* is the standard error of the difference between the two AUCs and is estimated using bootstrapping (Robin et al., 2011).[2]

Fair lineups. Taken together, the results for the fair lineups in Figure 3 suggest that ability to discriminate between innocent and guilty suspects declined with age. The *p*AUC for the older adults (*p*AUC = 0.016, 95% CI [0.011, 0.021]) was not significantly smaller than the *p*AUC for the middle-aged adults (*p*AUC = 0.024, 95% CI [0.017, 0.031], $D = 1.65$, $p = .10$), but it was significantly smaller than the *p*AUC for the young adults (*p*AUC = 0.028, 95% CI [0.022, 0.036], $D = 2.68$, $p = .007$). The *p*AUC for the middle-aged adults was smaller than the *p*AUC for the young adults, but not significantly so ($D = 0.92$ $p = .36$).

Unfair lineups. Considering the unfair lineups in Figure 3, however, the ROC curves for each age group are largely overlapping and close to the dashed chance line. Thus, ability to discriminate between innocent and guilty suspects in unfair lineups was similar, and poor, in all age groups. The *p*AUC for the older adults (*p*AUC = 0.008, 95% CI [0.003, 0.016]) was similar to the *p*AUC for both the middle-aged (*p*AUC = 0.010, 95% CI [0.005, 0.021], $D = 0.37$, $p = .71$) and the young adults (*p*AUC = 0.008, 95% CI [0.005, 0.018], $D = 0.05$, $p = .96$). The *p*AUC for the middle-aged adults was also similar to the *p*AUC for the young adults ($D = 0.37$, $p = .71$). All three age groups were less able to distinguish between innocent and guilty suspects in the unfair lineups than in the fair lineups. The *p*AUC for the unfair lineups was significantly smaller than the *p*AUC for the fair lineups in the young ($D = 3.94$, $p < .001$), middle-aged ($D = 2.53$, $p = .01$), and older adults ($D = 1.96$, $p = .05$). Finally, Figure 3 shows that the ROC curves for the unfair lineups shifted to the right of the ROC curves for the fair lineups, reflecting an increase in both correct and false identifications. In line with the identification response analysis, this indicates that subjects of all ages were more willing to identify the suspect when he was the only person in the lineup with the distinctive feature.

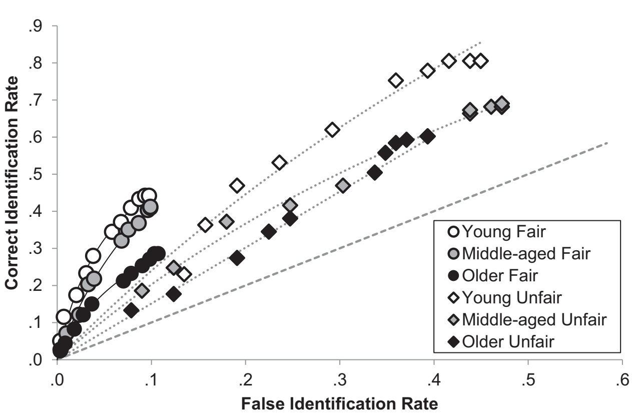

Figure 3. Receiver operating characteristic curves for the fair and unfair lineups for the young, middle-aged, and older adults. The dashed line represents chance-level performance.

In sum, the ROC results indicate that ability to discriminate between guilty and innocent suspects declined with age in fair lineups, but all age groups were poor at sorting guilty and innocent suspects into their appropriate categories in unfair lineups. All subjects were more willing to identify the suspect in unfair lineups compared to fair lineups.

Modeling

To further test these conclusions, we fit a signal-detection process model to our data (Wixted & Mickes, 2014). The results of our model fitting aligned with the results of our ROC analyses (see online supplement),[3] indicating that the findings of our atheoretical *p*AUC analysis map onto measures of underlying memory discriminability (cf. Lampinen, 2016). Here we limit our discussion to our findings when we fit the model to the fair lineups because this also furthers our theoretical understanding of how identification behavior changes with age. The model accounts for all identification decisions: suspect identifications, foil identifications and lineup rejections in both target-present and target-absent lineups. Therefore, the model fitting helps us to understand witnesses' decision-making processes and illustrates how willingness to make identi-

[2] Calculating *p*AUC is appropriate because, for a lineup task, the maximum HR and FAR are both likely to be less than 1 even when responding is infinitely liberal (in which case every witness would make an identification). Unless memory is perfect, some witnesses will fail to recognize the perpetrator from target-present lineups. If they make an identification anyway, this will sometimes land on a foil. Hence, unless every witness recognizes the perpetrator (i.e., unless memory is perfect), the maximum HR will be less than 1. The constraint on the maximum FAR is even more severe. The FAR was calculated using false identifications of innocent suspects (it did not include incorrect identifications of foils). If every witness made a guess in a fair target-absent lineup, the maximum FAR of innocent suspects is 1/n, where *n* is lineup size. Thus, in our case, the maximum FAR was 1/6 = .017. With a maximum HR of less than 1 and a maximum FAR of much less than 1, *p*AUC—rather than the full AUC—must be used as the dependent measure. When calculating *p*AUC, one must define the specificity (1 – FAR), which is the range of the curve that one wishes to measure. We used the FAR range covered by the least extensive curve to set the specificity (1 – .098) to .902 (Colloff et al., 2016, discuss the benefits of this method). Using a FAR range from 0 to .098 means that a *p*AUC of .005 represents chance discrimination (i.e., .098 * .098 * .50 = .005) and a *p*AUC of .098 represents maximum discrimination (i.e., 1 * .098 = .098).

[3] We examined whether the fair lineups enhanced discriminability more than the unfair lineups in each age group. We also examined whether all three age groups had similar discriminability on the unfair lineups (see supplemental materials). Though the trends were always the same in the modeling and ROC analysis, there was one occasion when the statistical significance (*p* value) was different across the two types of analyses. In older adults, the difference in discriminability between the fair and unfair lineups only approached significance in the modeling ($p = .08$), but was marginally significant in the ROC analysis ($p = .05$). Nevertheless, it is important to note that, regardless of which type of analysis we use, our conclusion remains the same: unfair lineups yield poor discriminability in subjects of all ages.

fications (i.e., placement of the decision criterion) changes with differences in discriminability (Palmer & Brewer, 2012).

The model assumes that when a witness views the faces in a lineup, each face has some memory strength value (i.e., degree of familiarity). Guilty suspects, innocent suspects and foils each have memory strength values with Gaussian distributions and means of μ_{guilty}, $\mu_{innocent}$, and μ_{foil}, respectively. In a fair lineup, the innocent suspect is not more similar to the guilty suspect than the other foils, so $\mu_{innocent} = \mu_{foil}$. Therefore, the model for a fair lineup consists of two distributions: one for guilty suspects (μ_{guilty}), and one for innocent suspects and foils ($\mu_{innocent}$). μ_{guilty} lies higher on the decision axis than $\mu_{innocent}$ because, on average, guilty suspects are associated with a greater memory strength (i.e., feel more familiar) than innocent suspects and foils who have not been seen before. The distance between the μ_{guilty} and $\mu_{innocent}$ distributions (d') measures subjects' underlying ability to discriminate between who is guilty and innocent. Smaller values of d' indicate a greater overlap of the μ_{guilty} and $\mu_{innocent}$ distributions and reflect poorer discriminability (see Figure 4).

The model also assumes that there is a set of response criteria that reflect different levels of confidence. To limit the number of parameters, we collapsed our data from the 11-point confidence scale used in the ROC analysis (0, 10, etc.), down to a 5-point scale: 0–20 (c_1), 30–40 (c_2), 50–60 (c_3), 70–80 (c_4), and 90–100 (c_5). These confidence intervals ensured: (a) a similar number of identification decisions at each confidence level in each condition, and (b) consistency throughout our analyses because we also used these intervals to construct confidence-accuracy plots. The model assumes that the lineup is rejected if no face is familiar enough to exceed the lowest decision criterion (c_1). Conversely, an identification is made when the familiarity of one or more faces exceeds c_1, and the face with the highest familiarity value is identified. The confidence in the identification is determined by the highest criterion that is exceeded.

If the increase in erroneous identifications with age is due to impairment in underlying theoretical discriminability, then there

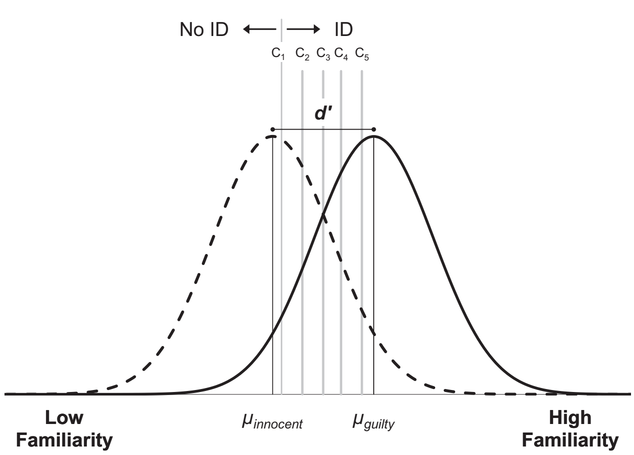

Figure 4. Signal-detection model for a fair lineup. The dashed distribution represents identifications of innocent suspects and foils, and the solid distribution represents identifications of guilty suspects. Adapted from "A signal-detection-based diagnostic-feature-detection model of eyewitness identification," by J. T. Wixted and L. Mickes (2014). Copyright, 2014 by the American Psychological Association.

should be a greater overlap of the guilty and innocent distributions (i.e., d' should decline) with age. However, if the increase in erroneous identifications with age is due to more liberal responding, then there should be a marked leftward shift of the decision criteria (i.e., c_1 through c_5 should decline) with age. The data contained 15 degrees of freedom, corresponding to the 5 levels of confidence for guilty suspect identifications and foil identifications in target-present lineups, and the 5 levels of confidence for foil identifications in target-absent lineups. Once these response frequencies were known, the number of rejections made in target-present and target-absent lineups was fixed. The model had 6 free parameters (μ_{guilty}, c_1, c_2, c_3, c_4, c_5) because we fixed $\mu_{innocent}$ to 0 and set the standard deviations for each distribution to 1, for simplicity. Thus, the fit had 15 – 6 = 9 degrees of freedom.

We fit the model to our data by minimizing the chi-square goodness-of-fit statistic. Table 2 shows our observed data and the values predicted by the best-fitting model, whereas Table 3 shows best-fitting parameters and the chi-square goodness-of-fit statistics. Table 2 shows that the model proficiently captured the trends in our data, and this is reflected in the (nonsignificant) chi-square goodness-of-fit statistics in the left-hand column (full model) of Table 3.[4] Figure 5 displays the best-fitting parameters for all age groups. The overlap in the guilty and innocent distributions increases (i.e., d' declines) with age. Interestingly, the response criteria also spread out on the decision axis from young to older subjects—this trend is more easily observed by considering the confidence parameter estimates for the young and older adults displayed in the left-hand column (full model) of Table 3. Larger estimates correspond to more conservative confidence criterion settings, whereas smaller estimates correspond to more liberal confidence criterion settings. Compared to young adults, older adults set their high-confidence criteria (i.e., c_4 and c_5) in a slightly more conservative position, but place their remaining criteria (i.e., c_1, c_2, c_3) in a more liberal position. A similar pattern has been observed when memory strength is manipulated in studies of younger subjects and is a natural consequence of a decline in d' (Stretch & Wixted, 1998).

Discriminability. To test whether the decline in d' with age was statistically significant, we performed three pairwise comparisons: young versus middle-aged, young versus older, and middle-aged versus older. We fit the same model, allowing the confidence criteria to differ, but we constrained d' to be equal in the two age groups being compared. The overall χ^2, *df*, and *p* rows in Table 3 show the full (unconstrained) and constrained model fit statistics. Compared to the full model, the constrained model did not provide a significantly worse fit of the data for the young and middle-aged comparison, $\chi^2(1) = 1.87$, $p = .17$, but it did provide a significantly worse fit of the data for the young and older, $\chi^2(1) = 22.43$, $p < .001$, and middle-aged and older comparisons, $\chi^2(1) = 11.31$, $p < .001$. These results indicate that aging was accompanied by a decline in theoretical discriminability, but the decline from young to middle age was not statistically significant.

Decision criteria. To examine how the decision criteria changed with age, we compared the criteria settings in the young

[4] Nonsignificant chi-square goodness-of-fit statistics (i.e., $p > .05$) indicate that the data do not significantly deviate from the model-predicted values, that is, they indicate that the model fits the data well.

Table 2
Observed and Predicted Identification Responses in Each Confidence Bin in the Fair Lineups for the Young, Middle-Aged, and Older Adults

	Target present			Target absent	
Confidence	Guilty suspect	Foil	Incorrect rejection	Foil	Correct rejection
Young					
0–20					
Observed	11.00	19.00	—	40.00	—
Predicted	13.26	17.88	—	38.66	—
30–40					
Observed	22.00	28.00	—	42.00	—
Predicted	20.58	24.22	—	46.93	—
50–60					
Observed	38.00	25.00	—	57.00	—
Predicted	33.45	31.72	—	53.97	—
70–80					
Observed	40.00	21.00	—	50.00	—
Predicted	42.57	27.94	—	41.29	—
90–100					
Observed	39.00	10.00	—	15.00	—
Predicted	36.72	11.61	—	15.23	—
Total					
Observed	—	—	87.00	—	144.00
Predicted	—	—	80.04	—	151.92
Middle-aged					
0–20					
Observed	15.00	21.00	—	26.00	—
Predicted	11.38	17.40	—	34.28	—
30–40					
Observed	16.00	13.00	—	38.00	—
Predicted	13.56	18.86	—	34.45	—
50–60					
Observed	40.00	35.00	—	74.00	—
Predicted	36.40	41.86	—	68.42	—
70–80					
Observed	45.00	21.00	—	48.00	—
Predicted	37.96	30.75	—	44.01	—
90–100					
Observed	24.00	15.00	—	21.00	—
Predicted	30.94	13.14	—	17.01	—
Total					
Observed	—	—	95.00	—	141.00
Predicted	—	—	87.73	—	149.83
Older					
0–20					
Observed	11.00	23.00	—	35.00	—
Predicted	10.22	22.20	—	36.12	—
30–40					
Observed	14.00	25.00	—	42.00	—
Predicted	13.11	26.42	—	40.80	—
50–60					
Observed	31.00	61.00	—	89.00	—
Predicted	35.04	59.86	—	85.31	—
70–80					
Observed	26.00	25.00	—	40.00	—
Predicted	22.93	29.49	—	38.70	—
90–100					
Observed	15.00	6.00	—	18.00	—
Predicted	13.99	11.62	—	14.56	—
Total					
Observed	—	—	103.00	—	124.00
Predicted	—	—	95.12	—	132.49

Note. The total row displays all reject identification decisions because the model does not account for the confidence level with which lineup rejections are made.

versus older adults. Figure 6 shows the best-fitting confidence criteria parameters. The confidence criteria for the young and older adults were linearly related; therefore we fit the same model, but we replaced the 5 confidence parameters (c_1, c_2, c_3, c_4, c_5) for the older adults with a linear transformation of the 5 confidence parameters for the young adults. For instance, $c_{1old} = a * c_{1young} + b$, where a and b are free parameters. We allowed d' to differ across the young and older groups. The overall χ^2, *df*, and *p* rows in Table 4 show the full (unconstrained confidence parameters) and reduced (linear transformation of c_1 - c_5) model fit statistics. The model fit statistic indicates that the reduced (linear transformation of c_1 - c_5) model fit the data well, but, surprisingly, it provided a significantly worse fit of the data than the full model, $\chi^2(3) = 12.70$, $p = .01$. Looking back at Figure 6, it is clear that c_3 falls slightly away from the line of best fit. Therefore, this one criterion could explain why the fit of the reduced model was significantly worse than the fit for the full model. To address this, we fit the same linear transformation model, but this time we allowed c_3 to vary across the young and older groups. The model

Table 3
Full and Constrained (d′) Model Fits for the Young Versus Middle-Aged, Young Versus Older, and Middle-Aged Versus Older Fair Lineup Comparisons

	Full model		Constrained model	
Estimate	Young	Middle-aged	Young	Middle-aged
μ_{guilty} (d')	1.21	1.07	1.14	1.14
c_1	1.13	1.12	1.12	1.13
c_2	1.31	1.28	1.30	1.29
c_3	1.54	1.44	1.53	1.46
c_4	1.89	1.86	1.88	1.87
c_5	2.44	2.39	2.42	2.41
Overall χ^2	26.12		27.99	
Overall *df*	18		19	
Overall *p*	.10		.08	
	Young	Older	Young	Older
μ_{guilty} (d')	1.21	0.72	0.99	0.99
c_1	1.13	1.04	1.09	1.08
c_2	1.31	1.21	1.27	1.25
c_3	1.54	1.40	1.49	1.44
c_4	1.89	1.92	1.84	1.97
c_5	2.44	2.45	2.36	2.50
Overall χ^2	15.90		38.33	
Overall *df*	18		19	
Overall *p*	.60		.005	
	Middle-aged	Older	Middle-aged	Older
μ_{guilty} (d')	1.07	0.72	0.91	0.91
c_1	1.12	1.04	1.07	1.09
c_2	1.28	1.21	1.23	1.25
c_3	1.44	1.40	1.42	1.41
c_4	1.86	1.92	1.96	1.83
c_5	2.39	2.45	2.49	2.37
Overall χ^2	23.98		35.29	
Overall *df*	18		19	
Overall *p*	.16		.01	

Note. The full model allows d' to differ between the two age groups being compared. The constrained model holds d' constant across the two age groups being compared. Overall χ^2, *df*, and *p* rows represent goodness-of-fit statistics when the model was fit to the two age groups together.

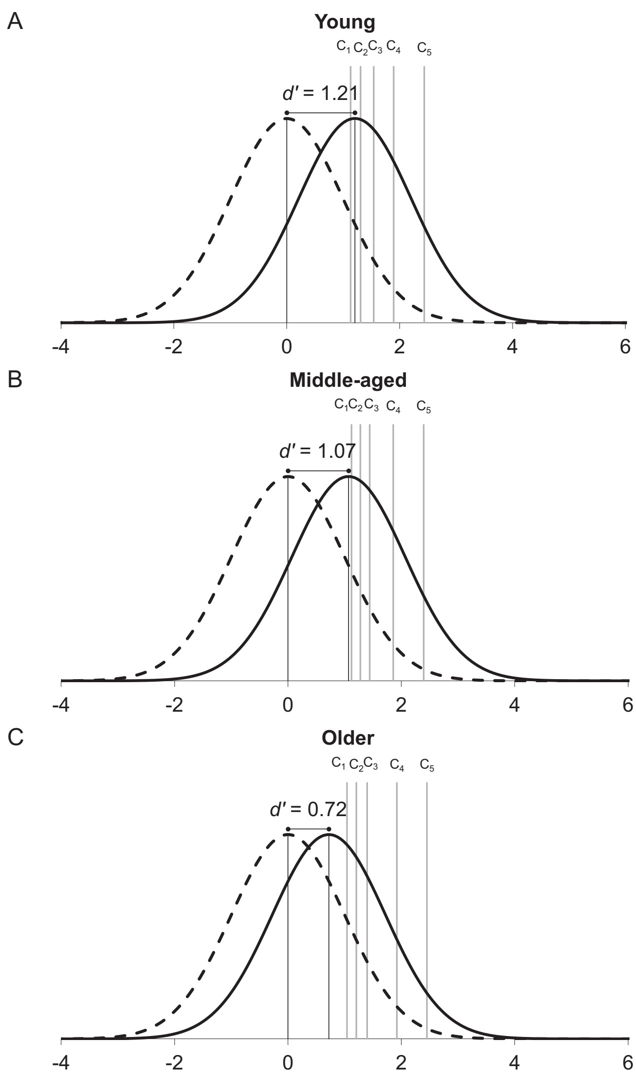

Figure 5. Innocent and guilty distributions for the (A) young, (B) middle-aged, and (C) older adults using the best-fitting signal-detection model parameters. d' measures subjects' ability to discriminate between innocent and guilty faces, with lower values indicating poorer discriminability. c_1, c_2, c_3, c_4 and c_5 are a set of response criteria that reflect different levels of confidence.

fit statistic in Table 4 indicates that the reduced (linear transformation of c_1, c_2, c_4, c_5) model fit the data well, and it did not provide a significantly worse fit of the data than the full model, $\chi^2(2) = 3.02$, $p = .22$. This suggests that a linear transformation, while allowing c_3 to vary, adequately characterized the confidence criteria in the young versus older groups.

Next, we fit the same model, but this time we equated the confidence parameters in the young and older groups, setting $a = 1$ and $b = 0$. Again, we allowed d' to differ across the young and older groups. The overall χ^2, *df*, and *p* rows in Table 4 show the reduced (linear transformation of c_1, c_2, c_4, c_5) and constrained (equated confidence parameters) model fit statistics. Compared to the reduced model, the constrained model provided a significantly worse fit of the data, $\chi^2(3) = 15.52$, $p = .001$. This indicates that aging is accompanied by a statistically significant change in criteria settings. This change is, generally speaking, linear, suggesting that the older adults tend to spread out their decision criteria more than the young adults. Setting the high-confidence criteria to more conservative positions, while spreading the remaining decision criteria to more liberal positions in this way at least approximates an optimal strategy because it means that identifications made with high confidence are likely to remain highly accurate, even though there is a general decline in d' (Stretch & Wixted, 1998). Thus, this provides preliminary evidence that older adults adjust their criteria in a way that maintains a good confidence-accuracy relationship.

Confidence and Accuracy

So far, our analyses have illustrated that discriminability declines with age on fair lineups, but older adults spread out their decision criteria in a more-or-less optimal manner. This suggests that older adults are aware that their memory accuracy is poor and that they make adjustments accordingly. Here, we tested this idea more concretely. If middle-aged and older subjects realize that their memory is error-prone, they should lower their confidence judgments to reflect their poorer performance and the proportion of correct identifications should be similar in all three age groups at each level of confidence.

To test this, we calculated suspect identification accuracy (guilty suspect identifications ÷ [guilty suspect identifications + innocent suspect identifications]) separately for each level of confidence (100, 90, 80 and so forth, as per Mickes, 2015). We calculated the number of innocent suspect identifications in the same way that we did in the ROC analysis. We then binned confidence into five categories (0–20, 30–40, 50–60, 70–80, 90–100) to provide more stable estimates (e.g., Brewer & Wells, 2006).

Figure 7A shows the confidence-accuracy curves for fair lineups in the young, middle-aged, and older subjects. The error bars largely overlap, which indicates that the differences in suspect

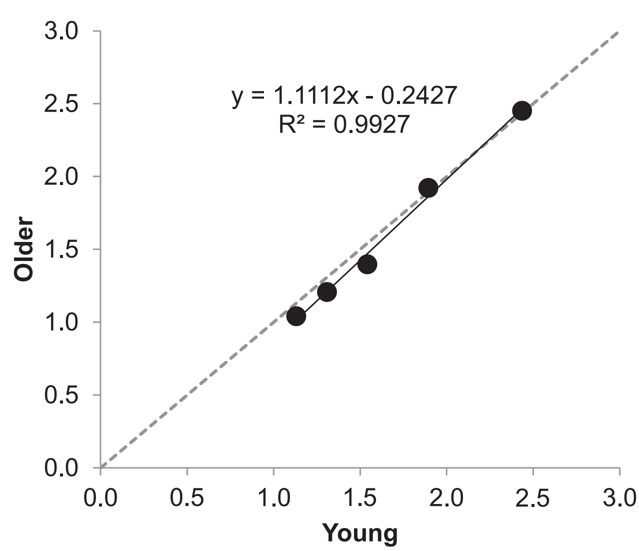

Figure 6. The best-fitting model confidence criteria parameters (c_1, c_2, c_3, c_4, c_5) for the young versus older adults. The dashed line is y = x.

Table 4
Full, Reduced, and Constrained (Confidence Criteria) Model Fits for the Young Versus Older Fair Lineup Comparisons

	Full model		Reduced (linear, c_1-c_5) model		Reduced (linear, c_1, c_2, c_4, c_5) model		Constrained model	
Estimate	Young	Older	Young	Older	Young	Older	Young	Older
μ_{guilty} (d')	1.21	0.72	1.20	0.71	1.20	0.72	1.17	0.76
c_1	1.13	1.04	1.13	1.03	1.13	1.04	1.08	1.08
c_2	1.31	1.21	1.30	1.21	1.29	1.22	1.25	1.25
c_3	1.54	1.40	1.50	1.43	1.54	1.40	1.46	1.46
c_4	1.89	1.92	1.92	1.89	1.91	1.90	1.91	1.91
c_5	2.44	2.45	2.44	2.45	2.42	2.47	2.44	2.44
a	—		1.09		1.11		1	
b	—		−.21		−.21		0	
Overall χ^2	15.90		28.60		18.92		34.44	
Overall df	18		21		20		23	
Overall p	.60		.12		.53		.06	

Note. The full model allows the confidence criteria (c_1-c_5) to differ between the young and older groups. The reduced (linear c1-c5) model allows the confidence criteria to differ between the young and older groups by a linear transformation. The reduced (linear, c_1, c_2, c_4, c_5) model allows the confidence criteria c_1, c_2, c_4, and c_5 to differ between the young and older groups by a linear transformation, and leaves c_3 free to vary. The constrained model holds the confidence criteria constant across the young and older groups. Overall χ^2, df, and p rows represent goodness-of-fit statistics when the model was fit to the two age groups together.

identification accuracy between the three age groups at each level of confidence were not, on the whole, statistically reliable (e.g., Sauer et al., 2010). Despite being significantly poorer at distinguishing between who is guilty and who is innocent than the young and middle-aged adults, older adults seem to be reasonably effective at regulating their confidence judgments to reflect the likely accuracy of their suspect identification decisions. Nevertheless, descriptively speaking, Figure 7A shows that the older adults are slightly less accurate at every level of confidence than the young and middle-aged adults. This suggests that older adults, while they do adjust their confidence criteria in the appropriate direction, do not quite adjust their confidence criteria enough, given their decline in memory ability. For example, if we look back at Figure 5, older adults would need to set c_5 to a more conservative position if they were to be as accurate as the young and middle-aged adults at the highest level of confidence (90–100).

Finally, comparing the fair and unfair lineups in Figure 7B–D, suspect identification accuracy was reduced in the unfair lineups in young, middle-aged, and older adults. Specifically, within each age group, high-confidence suspect identifications made from unfair lineups were substantially less trustworthy than high-confidence suspect identifications made from fair lineups. This suggests that subjects were not aware that their accuracy was poor in the unfair lineups and did not adjust their confidence accordingly.

Discussion

We asked (a) whether the age-related decline in accurate identification decisions is due to an increased willingness to make an identification, a decline in discriminability, or both, and (b) whether older and middle-aged adults are able to gauge the likely accuracy of their suspect identification decisions and assign appropriate confidence judgments to the same extent as young adults. Our findings suggest that aging is associated with a genuine decline in ability to discriminate between who is innocent and who is guilty. Remarkably, despite a substantial decline in memory ability, older adults were able to gauge the accuracy of their suspect identifications, and were, generally speaking, as accurate as the young and middle-aged adults at each level of confidence.

At first glance, our results are perhaps unsurprising. Many previous studies have shown that older adults make more mistakes on lineups than younger adults (see Bartlett & Memon, 2007; Sporer & Martschuk, 2014, for reviews). Indeed, the distribution of identification responses indicated that the number of erroneous identifications increased with age. But our analyses show that this pattern is not simply due to older adults being more willing to make an identification. Instead, our data suggest that the errors are due to a genuine decline in ability to discriminate between those who are innocent and guilty.

Why might aging be associated with a decline in recognition performance? One explanation is that our ability to recollect source-specific information declines over the life span, which results in a greater reliance on familiarity processes with age (Healy et al., 2005; Searcy et al., 1999). Older adults were more likely to make erroneous identifications than young adults. Presumably this is because the faces in the lineups were very similar and so even the new faces evoked signals of perceived familiarity (Bartlett et al., 1984; Edmonds et al., 2012; Young et al., 1985). Further support for this theoretical account comes from our modeling. If older adults are more reliant on a general feeling of familiarity, then the strength of the memory signal from new faces in the lineup (i.e., those in the innocent distribution) should be closer to the strength of the memory signal from the real culprit (i.e., those in the guilty distribution). Indeed, we found a statistically significant increase in the overlap of the innocent and guilty distributions with age.

Our finding that ability to discriminate between innocent and guilty suspects declines with age is concordant with face recognition studies in the broader literature (e.g., Lamont et al., 2005). Face recognition is dependent on processing the spatial distances between facial features (configural processing) and processing the face as a whole (holistic processing; see Tanaka & Gordon, 2011,

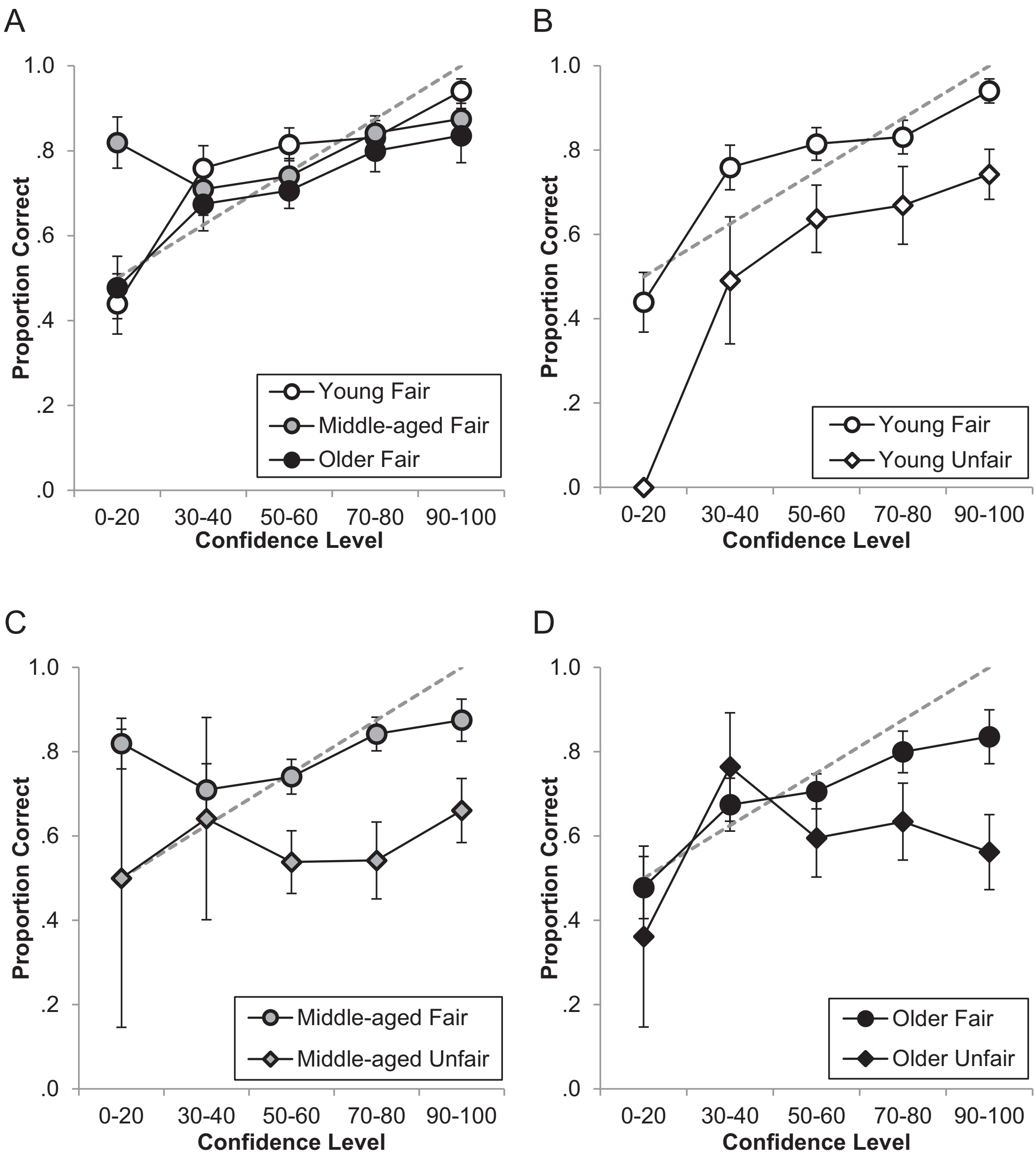

Figure 7. Confidence-accuracy curves for suspect identifications in (A) fair lineups made by each age group, (B) fair and unfair lineups made by young adults, (C) fair and unfair lineups made by middle-aged adults, and (D) fair and unfair lineups made by older adults. Error bars indicate ±1 *SE*. The dashed diagonal line signifies chance-level accuracy at the lowest confidence bin (0–20) and perfect accuracy at the highest confidence bin (90–100).

for a review). Therefore, our finding also fits with the idea that aging might be associated with a decline in configural or holistic processing (see Boutet, Taler, & Collin, 2015, for a review). More specifically, considering the eyewitness identification literature, our finding that discriminability declines with age is consistent with a recent meta-analysis (Fitzgerald & Price, 2015). Key et al. (2015), by contrast, found equivalent performance in their young and older subjects using ROC analysis. One possible reason for these contradictory findings is that Key et al.'s young and older groups consisted of subjects aged 18–59 and over 60, respectively. Our results suggest that discriminability begins to decline from early adulthood (aged 18–30) to middle-age (aged 31–59). Performance in Key et al.'s young group may have been artificially low because of its wide age range. Therefore, it is possible that the nonsignificant difference in discriminability between the young and older adults reflected how their young and older age groups were defined.

So, why is all this important? Greater theoretical understanding of how memory changes with healthy aging can be used to advance appropriate procedures to help aid identification accuracy. Many studies have attempted to reduce older adults' false identification rate by reducing their proclivity to choose (e.g., Memon & Gabbert, 2003; Rose et al., 2005; Wilcock et al., 2005). But our data suggest that encouraging older adults to be more conservative when they make a decision will not reduce the age-related deficit in performance. Instead, our results indicate that procedures need to target middle-aged and older adults' ability to discriminate between who is innocent and who is guilty.

One might argue that our older adult sample made more identification errors simply because their eyesight was poorer than the young and middle-aged subjects. However, there are at least three reasons why this counter explanation is unlikely to explain our results. First, the older adults, like the young and middle-aged adults, were more willing to identify the suspect in the unfair lineups than in the fair lineups. This suggests that older subjects saw the distinctive feature in the video because they subsequently picked the only person with a distinctive feature in the lineup task. Second, we asked a separate group of young ($n = 20$, aged 18–30) and older ($n = 29$, aged 60–85) adults to watch the mock crime video and then describe the culprit's appearance. The proportion of young and older adults who correctly described the distinctive feature did not differ for either the mugging or the graffiti video ($ps > .19$). This suggests that the vision of both young and older adults was good enough to see and encode the face of the culprit. Finally, the findings from our identification responses analyses are consistent with many laboratory-based studies that likely had greater control over whether subjects were wearing glasses, if necessary (e.g., Badham, Wade, Watts, Woods, & Maylor, 2013). Therefore, it seems that recognition memory ability on lineup tasks declines with age.

Perhaps most strikingly, our study has shown that despite older adults' poorer recognition memory ability, suspect identifications made by older adults can be almost as accurate as those made by young and middle-aged adults, when the confidence judgment expressed immediately after the identification decision is taken into account. In practice, this finding is important for legal decision-makers because it means that a suspect identification made with a particular level of confidence is likely to be similarly accurate regardless of whether it is made by a young, middle-aged or older adult. Recall that in our modeling (which accounted for all identification decisions) we found that the confidence criteria naturally spread out along the decision axis as d' declined with age. The fact that there were no significant differences in suspect identification accuracy between the age groups at each level of confidence indicates that the extent of spreading was generally appropriate to account for the decline in d'. Theoretically, this illustrates that older adults are, on the whole, able to assess the likely accuracy of their memories.

Recall also, however, that there was a trend for the older adults to be slightly (but not significantly) less accurate at each confidence level than the young and middle-aged adults in our confidence-accuracy plot. To investigate this further, we separated our older adults into young-old (aged 60–70) and old-old (aged 71+) groups, and we saw the same numerical pattern: old-old adults were slightly (but not significantly) less accurate at each confidence level than the young-old adults (see Supplemental Figure S3 in the online supplement). This trend accords with other research that shows that older adults can have reduced metacognitive monitoring of recently encountered information (e.g., Dodson & Krueger, 2006), can experience high-confidence false memories (e.g., Dodson, Bawa, & Krueger, 2007), and sometimes have a tendency for less flexible criterion placement in difficult memory tasks (e.g., Koutstaal, 2006). Thus, there is some basis for the idea that aging may be associated with a difference in adjusting criteria to account for poorer memory ability. One theory suggests that people are usually adept at assigning appropriate confidence judgments because they have learned through error feedback training the situations in which their memory is and is not likely to be accurate (Mickes, Hwe, Wais, & Wixted, 2011; Stretch & Wixted, 1998; also see Lindsay, Read, & Sharma, 1998). Therefore, it is possible that, as we age, memory ability declines quicker than we are able to learn about the degree of our memory impairment though error feedback training. This might explain why our older adults failed to adjust their confidence criteria to the extent required for them to be just as accurate as the young and middle-aged adults. Notably, this idea is based on trends, and not statistically significant differences, in our data. Therefore, our main conclusion still stands: suspect identifications made by older adults are as accurate as those made by young and middle-aged adults when their confidence judgment is taken into account. Nevertheless, examining the role of error feedback training in older adults could be a fruitful avenue for further research.

Finally, our comparison between performance on fair and unfair lineups is also important. We found that subjects of all ages were more willing to identify the suspect, but, critically, were also less able to discriminate between innocent and guilty suspects on unfair lineups compared to fair lineups. Indeed, ability to discriminate on unfair lineups was remarkably poor in all age groups. Suspect identification accuracy was also reduced at almost every confidence level on unfair lineups, compared to fair lineups. This suggests that subjects were not aware that their accuracy was poor on the unfair lineups and did not adjust their confidence judgments accordingly. These results replicate the findings of Colloff et al. (2016) and reiterate the need for fair lineups for witnesses of all ages.[5] Interestingly, these results are predicted by the diagnostic-feature-detection model, which suggests that witnesses are less able to distinguish between innocent and guilty suspects when they rely on features that both innocent and guilty suspects share (Wixted & Mickes, 2014). Presumably, our fair lineups prevented subjects from relying on the distinctive feature to make their identification decision because the feature was either concealed (pixelation and block) or appeared on every lineup member (replication). Our unfair lineups, however, did not provide this protection because only one lineup member—the suspect—had the dis-

[5] In the current study, the suspect identification accuracy of the young adults at the lowest level of confidence (i.e., 0–20) on the unfair lineups was considerably poorer than the suspect identification accuracy at the lowest level of confidence as reported by Colloff et al. (2016). The accuracy of the young adults on unfair lineups in the 0-20 confidence bin in the current study should be interpreted with caution because there were only three subjects in this age group who made a suspect identification decision with this level of confidence.

tinctive feature, and so subjects relied on the feature to make their identification decision. Theoretically, then, our research lends support to the idea that fair lineups may enhance people's ability to discriminate between innocent and guilty suspects because fair lineups prevent reliance on facial features that are nondiagnostic of guilt, whereas unfair lineups do not.

To conclude, we have shown that errors made by older individuals on lineup tasks are likely attributable to a genuine decline in ability to tell the difference between who is innocent and who is guilty, rather than an increased willingness to make an identification. Although further research is required before practical recommendations are made to the Criminal Justice System, our results add to the growing literature that suggests that if you were a police officer you should always use fair lineups to enhance your witness's accuracy. But, crucially, our results provide new, preliminary evidence that if you were a judge considering an identification made at a particular level of confidence, you should impart the same amount of trust in the identification regardless of whether it was made by a young, middle-aged, or older eyewitness.

References

Acierno, R., Hernandez, M. A., Amstadter, A. B., Resnick, H. S., Steve, K., Muzzy, W., & Kilpatrick, D. G. (2010). Prevalence and correlates of emotional, physical, sexual, and financial abuse and potential neglect in the United States: The National Elder Mistreatment Study. *American Journal of Public Health, 100,* 292–297. http://dx.doi.org/10.2105/AJPH.2009.163089

Adams-Price, C. (1992). Eyewitness memory and aging: Predictors of accuracy in recall and person recognition. *Psychology and Aging, 7,* 602–608. http://dx.doi.org/10.1037/0882-7974.7.4.602

Badham, S. P., Wade, K. A., Watts, H. J. E., Woods, N. G., & Maylor, E. A. (2013). Replicating distinctive facial features in lineups: Identification performance in young versus older adults. *Psychonomic Bulletin & Review, 20,* 289–295. http://dx.doi.org/10.3758/s13423-012-0339-2

Bartlett, J. C. (2014). The older eyewitness. In T. J. Perfect & D. S. Lindsay (Eds.), *The SAGE handbook of applied memory* (pp. 654–674). London, UK: SAGE. http://dx.doi.org/10.4135/9781446294703.n36

Bartlett, J. C., & Fulton, A. (1991). Familiarity and recognition of faces in old age. *Memory & Cognition, 19,* 229–238. http://dx.doi.org/10.3758/BF03211147

Bartlett, J. C., Hurry, S., & Thorley, W. (1984). Typicality and familiarity of faces. *Memory & Cognition, 12,* 219–228. http://dx.doi.org/10.3758/BF03197669

Bartlett, J. C., & Memon, A. (2007). Eyewitness memory in young and older eyewitnesses. 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. 309–338). Mahwah, NJ: Erlbaum Inc.

Bartlett, J. C., Strater, L., & Fulton, A. (1991). False recency and false fame of faces in young adulthood and old age. *Memory & Cognition, 19,* 177–188. http://dx.doi.org/10.3758/BF03197115

Boutet, I., Taler, V., & Collin, C. A. (2015). On the particular vulnerability of face recognition to aging: A review of three hypotheses. *Frontiers in Psychology, 6,* 1139. http://dx.doi.org/10.3389/fpsyg.2015.01139

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. http://dx.doi.org/10.1037/1076-898X.8.1.44

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. http://dx.doi.org/10.1037/1076-898X.12.1.11

Cochran, W. G. (1952). The χ^2 test of goodness of fit. *Annals of Mathematical Statistics, 23,* 315–345. http://dx.doi.org/10.1214/aoms/1177729380

Colloff, M. F., Wade, K. A., & Strange, D. (2016). Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychological Science, 27,* 1227–1239. http://dx.doi.org/10.1177/0956797616655789

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115,* 107–117. http://dx.doi.org/10.1037/0096-3445.115.2.107

Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory, and high-confidence errors: A misrecollection account. *Psychology and Aging, 22,* 122–133. http://dx.doi.org/10.1037/0882-7974.22.1.122

Dodson, C. S., Bawa, S., & Slotnick, S. D. (2007). Aging, source memory, and misrecollections. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 169–181. http://dx.doi.org/10.1037/0278-7393.33.1.169

Dodson, C. S., & Krueger, L. E. (2006). I misremember it well: Why older adults are unreliable eyewitnesses. *Psychonomic Bulletin & Review, 13,* 770–775. http://dx.doi.org/10.3758/BF03193995

Dywan, J., & Jacoby, L. (1990). Effects of aging on source monitoring: Differences in susceptibility to false fame. *Psychology and Aging, 5,* 379–387. http://dx.doi.org/10.1037/0882-7974.5.3.379

Edmonds, E. C., Glisky, E. L., Bartlett, J. C., & Rapcsak, S. Z. (2012). Cognitive mechanisms of false facial recognition in older adults. *Psychology and Aging, 27,* 54–60. http://dx.doi.org/10.1037/a0024582

Erickson, W. B., Lampinen, J. M., & Moore, K. N. (2016). Eyewitness identifications by older and younger adults: A meta-analysis and discussion. *Journal of Police and Criminal Psychology, 31,* 108–121. http://dx.doi.org/10.1007/s11896-015-9176-3

Field, A. P., Miles, J. N. V., & Field, Z. C. (2012). *Discovering statistics using R: And sex and drugs and rock 'n' roll*. London, UK: Sage Publications.

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

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. http://dx.doi.org/10.1037/0882-7974.6.4.623

Goodsell, C. A., Neuschatz, J. S., & Gronlund, S. D. (2009). Effects of mugshot commitment on lineup performance in young and older adults. *Applied Cognitive Psychology, 23,* 788–803. http://dx.doi.org/10.1002/acp.1512

Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using Receiver Operating Characteristic analysis. *Current Directions in Psychological Science, 23,* 3–10. http://dx.doi.org/10.1177/0963721413498891

Havard, C., & Memon, A. (2009). The influence of face age on identification from a video line-up: A comparison between older and younger adults. *Memory, 17,* 847–859. http://dx.doi.org/10.1080/09658210903277318

Healy, M. R., Light, L. L., & Chung, C. (2005). Dual-process models of associative recognition in young and older adults: Evidence from receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 768–788. http://dx.doi.org/10.1037/0278-7393.31.4.768

Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 3–22. http://dx.doi.org/10.1037/0278-7393.25.1.3

Jennings, J. M., & Jacoby, L. L. (1997). An opposition procedure for detecting age-related deficits in recollection: Telling effects of repetition. *Psychology and Aging, 12,* 352–361. http://dx.doi.org/10.1037/0882-7974.12.2.352

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. http://dx.doi.org/10.1037/0278-7393.22.5.1304

Key, K. N., Cash, D. K., Neuschatz, J. S., Price, J. L., Wetmore, S. A., & Gronlund, S. D. (2015). Age differences (or lack thereof) in discriminability for lineups and showups. *Psychology, Crime & Law, 21,* 871–889. http://dx.doi.org/10.1080/1068316X.2015.1054387

Koutstaal, W. (2006). Flexible remembering. *Psychonomic Bulletin & Review, 13,* 84–91. http://dx.doi.org/10.3758/BF03193817

Lamont, A. C., Stewart-Williams, S., & Podd, J. (2005). Face recognition and aging: Effects of target age and memory load. *Memory & Cognition, 33,* 1017–1024. http://dx.doi.org/10.3758/BF03193209

Lampinen, J. M. (2016). ROC analyses in eyewitness identification research. *Journal of Applied Research in Memory & Cognition, 5,* 21–33. http://dx.doi.org/10.1016/j.jarmac.2015.08.006

Lindsay, D. S., Read, J. D., & Sharma, K. (1998). Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. *Psychological Science, 9,* 215–218. http://dx.doi.org/10.1111/1467-9280.00041

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252–271. http://dx.doi.org/10.1037/0033-295X.87.3.252

Mansour, J. K., Beaudry, J. L., Kalmet, N., Bertrand, M. I., & Lindsay, R. C. L. (2017). Evaluating lineup fairness: Variations across methods and measures. *Law and Human Behavior, 41,* 103–115. http://dx.doi.org/10.1037/lhb0000203

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6,* 255–260. http://dx.doi.org/10.1016/S1364-6613(02)01903-4

Memon, A., & Gabbert, F. (2003). Improving the identification accuracy of senior witnesses: Do prelineup questions and sequential testing help? *Journal of Applied Psychology, 88,* 341–347. http://dx.doi.org/10.1037/0021-9010.88.2.341

Memon, A., Hope, L., Bartlett, J., & Bull, R. (2002). Eyewitness recognition errors: The effects of mugshot viewing and choosing in young and old adults. *Memory & Cognition, 30,* 1219–1227. http://dx.doi.org/10.3758/BF03213404

Mickes, L. (2015). Receiver operating characteristic analysis and confidence–accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory & Cognition, 4,* 93–102. http://dx.doi.org/10.1016/j.jarmac.2015.01.003

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus sequential lineups. *Journal of Experimental Psychology: Applied, 18,* 361–376. http://dx.doi.org/10.1037/a0030609

Mickes, L., Hwe, V., Wais, P. E., & Wixted, J. T. (2011). Strong memories are hard to scale. *Journal of Experimental Psychology: General, 140,* 239–257. http://dx.doi.org/10.1037/a0023007

National Research Council. (2014). *Identifying the culprit: Assessing eyewitness identification*. Washington, DC: The National Academies Press.

Neuschatz, J. S., Preston, E. L., Burkett, A. D., Toglia, M. P., Lampinen, J. M., Neuschatz, J. S., . . . Goodsell, C. A. (2005). The effects of post-identification feedback and age on retrospective eyewitness memory. *Applied Cognitive Psychology, 19,* 435–453. http://dx.doi.org/10.1002/acp.1084

Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and Human Behavior, 36,* 247–255. http://dx.doi.org/10.1037/h0093923

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics, 12,* 77–84. http://dx.doi.org/10.1186/1471-2105-12-77

Rose, R., Bull, R., & Vrij, A. (2005). Non-biased lineup instructions do matter - a problem for older witnesses. *Psychology, Crime & Law, 11,* 147–159. http://dx.doi.org/10.1080/10683160512331316307

Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence-accuracy relationship for eyewitness identification. *Law and Human Behavior, 34,* 337–347. http://dx.doi.org/10.1007/s10979-009-9192-x

Searcy, J. H., Bartlett, J. C., & Memon, A. (1999). Age differences in accuracy and choosing in eyewitness identification and face recognition. *Memory & Cognition, 27,* 538–552. http://dx.doi.org/10.3758/BF03211547

Searcy, J., Bartlett, J. C., & Memon, A. (2000). Influence of post-event narratives, line-up conditions and individual differences on false identification by young and older eyewitnesses. *Legal and Criminological Psychology, 5,* 219–235. http://dx.doi.org/10.1348/135532500168100

Searcy, J. H., Bartlett, J. C., Memon, A., & Swanson, K. (2001). Aging and lineup performance at long retention intervals: Effects of metamemory and context reinstatement. *Journal of Applied Psychology, 86,* 207–214. http://dx.doi.org/10.1037/0021-9010.86.2.207

Sporer, S. L., & Martschuk, N. (2014). The reliability of eyewitness identifications by the elderly: An evidence-based review. In M. P. Toglia, D. F. Ross, J. Pozzulo, & E. Pica (Eds.), *The elderly eyewitness in court* (pp. 3–37). New York, NY: Psychology Press.

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1379–1396. http://dx.doi.org/10.1037/0278-7393.24.6.1379

Tanaka, J. W., & Gordon, I. (2011). Features, configuration, and holistic face processing. In G. Rhodes, A. Calder, M. Johnson, & J. V. Haxby (Eds.), *Oxford handbook of face perception* (pp. 149–176). http://dx.doi.org/10.1093/oxfordhb/9780199559053.013.0010

Wilcock, R., Bull, R., & Vrij, A. (2005). Aiding the performance of older eyewitnesses: Enhanced non-biased line-up instructions and line-up presentation. *Psychiatry, Psychology and Law, 12,* 129–140. http://dx.doi.org/10.1375/pplt.2005.12.1.129

Willoughby, M. (2015). *Witnessing crime—Findings from the crime survey for England and Wales 2013/14*. London, UK: Ministry of Justice. Retrieved March 5, 2016, from https://www.gov.uk/government/publications/witnessing-crime-findings-from-the-crime-survey-2013-to-2014

Wixted, J. T., & Mickes, L. (2012). The field of eyewitness memory should abandon probative value and embrace receiver operating characteristic analysis. *Perspectives on Psychological Science, 7,* 275–278. http://dx.doi.org/10.1177/1745691612442906

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review, 121,* 262–276. http://dx.doi.org/10.1037/a0035940

Wixted, J. T., & Wells, G. L. (2016). *The relationship between eyewitness confidence and identification accuracy: A new synthesis*. [Manuscript submitted for publication].

World Health Organization. (2015, September). *Ageing and health*. Retrieved January 29, 2016, from http://www.who.int/topics/ageing/en/

Wylie, L. E., Bergt, S., Haby, J., Brank, E. M., & Bornstein, B. H. (2015). Age and lineup type differences in the own-race bias. *Psychology, Crime & Law, 21,* 490–506. http://dx.doi.org/10.1080/1068316X.2014.989173

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441–517. http://dx.doi.org/10.1006/jmla.2002.2864

Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin, 133,* 800–832. http://dx.doi.org/10.1037/0033-2909.133.5.800

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception, 14,* 737–746. http://dx.doi.org/10.1068/p140737

Zarkadi, T., Wade, K. A., & Stewart, N. (2009). Creating fair lineups for suspects with distinctive features. *Psychological Science, 20,* 1448–1453. http://dx.doi.org/10.1111/j.1467-9280.2009.02463.x

Received June 17, 2016
Revision received February 2, 2017
Accepted February 13, 2017 ■