Age Differences (or Lack Thereof) in Discriminability for Lineups and Showups

Kylie N. Key[1], Daniella K. Cash[2], Jeffrey S. Neuschatz[1], Jodi L. Price[1], Stacy A. Wetmore[3], & Scott D. Gronlund[3]

[1]The University of Alabama in Huntsville

[2]Louisiana State University

[3]University of Oklahoma

Author Notes

Kylie N. Key, Jeffrey S. Neuschatz, and Jodi L. Price, Department of Psychology, The University of Alabama in Huntsville; Daniella K. Cash, Department of Psychology, Louisiana State University; Scott D. Gronlund and Stacy Wetmore, Department of Psychology, University of Oklahoma. This work was supported by the National Science Foundation grant SES-1060921 to JSN, SES-1060902 to SDG, and SES-1060913 to Charles Goodsell. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of the NSF.

Abstract

Previous research reveals that showups are an inferior eyewitness identification procedure to lineups, but no single study has compared younger and older adults' identification decisions for both of these procedures. We had witnesses watch a mock crime video and then make an identification decision from a fair lineup, a biased lineup, or a showup that contained the perpetrator or a designated innocent suspect. ROC analysis showed that identification accuracy was higher from a lineup than from a showup for both age groups, even if the lineup was biased. In addition, calibration curves revealed that witnesses were underconfident when choosing from a fair lineup but overconfident when choosing from a showup. These results reinforce prior research asserting the superiority of lineups over showups.

Keywords: calibration, ROC analysis, lineups, showups, eyewitness identification

Age Differences (or Lack Thereof) in Discriminability for Lineups and Showups

Eyewitness identification is a frequently employed but fallible source of evidence used by the judicial system (Wells, 1993). The fallibility is obvious given that mistaken eyewitness identifications played a role in 75% of false convictions overturned by DNA exoneration (innocenceproject.org, last viewed April 21, 2014). Mistaken identifications might be even more common when older adults are eyewitnesses. Considerable research suggests that older adults experience a general memory decline (Bartlett, 2014), and do not perform as well as younger adults on an abundance of memory tasks, including eyewitness identification (see Price, Mueller, Wetmore, & Neuschatz, 2014). This is even more troubling given that middle-aged and older adults are more likely to be a witness to a crime than younger adults (Bornstein, 1995). One explanation for the higher proportion of older adult witnesses is that older adults make up a large percentage of the population and, according to the U.S. Bureau of Statistics (1997, as cited in Searcy, Bartlett, Memon & Swanson, 2001), are also likely to have had a personal encounter with the perpetrator (Price et al., 2014). Additionally, these older adults are frequently asked to participate in an identification procedure (Price et al., 2014; Rothmans, Dunlop, & Entzel, 2000).

The memory decline that older adults experience, coupled with the frequency with which they are involved in eyewitness events, warrants further research examining older adult performance in identification procedures. Previous research has shown that older adults make as many correct identifications of the perpetrator as younger adults, but more false identifications of innocent suspects from lineups (see Bartlett, 2014, or Searcy, Bartlett & Memon, 1999 for a review of this literature). When younger adults are compared to older adults in terms of accuracy (identifying the perpetrator when he is present) or correct decisions (rejecting the lineup when he

is absent), younger adults are consistently superior, suggesting an age-related deficit in memory performance on identification tasks.

Research on younger adults demonstrates that showups, a one-person identification procedure, consistently result in lower performance than lineups (Neuschatz et al., in press). But to our knowledge, no study has compared the accuracy of older and younger adult witnesses participating in showup identifications. This gap in the literature is surprising given how often showups are conducted in real criminal cases in the United States. Garrett (2011) reviewed trial transcripts from 160 DNA exoneration cases and found that 34% (53/160) involved misidentifications from showups. Other researchers estimate that as many as 77% of eyewitness cases involve showups rather than lineups (Dysart & Lindsay, 2007).

Studies that have examined witness accuracy in showup identifications generally have involved younger adult samples, rather than older adults (e.g., Gronlund et al., 2012; Wetmore et al., in press). These researchers have concluded that showups result in worse discriminability (ability to distinguish between the perpetrator and innocent suspect) than lineups. As noted by Clark and Godfrey (2009), showups put innocent suspects at greater risk of being falsely identified, a point that has been echoed by the U. S. Supreme Court (Stovall v. Denno, 1967). Although a meta-analysis conducted by Steblay, Dysart, Fulero and Lindsay (2003) found a higher false identification rate (choosing the innocent suspect) from showups than lineups, Clark and Godfrey (2009) argued that correct and false identification rates from lineups are affected by the filler identification rate, and thus are not appropriate for comparing to showups (which contain no fillers). They instead argued for a measure of conditional probability, which they termed “innocence risk”, and found that showups put innocent suspects at greater risk of being falsely identified.

Recently, it has been argued that Receiver Operating Characteristic (ROC) analysis is the proper analytic technique for evaluating eyewitness identification evidence (Wixted & Mickes, 2012). This technique disregards filler identifications and assesses whether differences in identification procedures are the result of discriminability differences (ability to distinguish the perpetrator from the innocent suspect) or a response bias (willingness to choose). A study conducted by Gronlund et al. (2012) using ROC analysis found that lineups yielded greater discriminability than showups. In their recent meta-analysis, Neuschatz et al. (in press) used ROC analysis to show that showups consistently result in poorer discriminability than lineups, even when the lineup is conducted after a retention interval (but the showup is conducted immediately) or when the lineup is biased (contains fillers that do not match the description of the perpetrator). The conclusion from these studies seems clear: showups put innocent suspects at greater risk than lineups.

Why is eyewitness performance worse when there is only one option at test in a showup compared to (typically) six in a lineup? Gonzales, Ellsworth, and Pembroke (1993) argued that a showup involves a different type of processing than a lineup. In addition, Gonzales and colleagues argued that the presence of fillers forces witnesses to have a higher criterion for choosing, making them more careful about making a false identification of a suspect. Lineup fillers also protect against witnesses who are choosing with low confidence. Because showups do not contain fillers, they do not offer these protections to potentially innocent suspects.

Alternatively, the poor witness accuracy arising from showups can be explained in terms of Wixted and Mickes' (2014) diagnostic feature-detection hypothesis. This hypothesis states that simultaneous presentation of faces in a lineup results in better discriminability because the witness can discern that some features (i.e., those shared by all suspects) are not diagnostic of

identifying the perpetrator. For example, witnesses viewing a simultaneous lineup in which all members have dark hair and a beard should discern that they will not be able to pick the perpetrator by relying on these features. Thus, they must look at the features that are not shared by all lineup members in order to determine who, if anyone, is the perpetrator. This capability to make comparisons between faces is not available in showups, making it difficult for witnesses to determine which features are diagnostic of the perpetrator. Pairing the lack of alternatives with the more liberal choosing results in little protection for innocent suspects in showups.

Making an accurate identification from a showup may be particularly difficult for older adults. Bartlett (2014) recently outlined several explanations for this, including a reliance on gist memories, a reliance on familiarity, and less differentiation in the areas of the brain associated with processing faces. All are cognitive declines that may put older adults at a particular disadvantage when viewing showups. We review each of these explanations in turn.

Older adults tend to remember general information about an event (gist memory) rather than specific details (verbatim memory) (Brainerd & Reyna, 2001; Rabinowitz, Craik, & Ackerman, 1982; Reyna & Brainerd, 2011). When applied to facial recognition, it is easy to see why older adults may struggle with identification procedures. Specifically, older adult witnesses may not remember specific facial features or other cues, and thus be inclined to identify the suspect if he matches the general description of the perpetrator. Furthermore, while a lineup is designed to offer a safeguard to innocent suspects by surrounding them with similar fillers, the showup does not offer the same protection. This would increase the likelihood that older witnesses may incorrectly identify innocent suspects in a showup.

Greater reliance on familiarity may also impair older adults' identification accuracy. Bartlett (2014) found that older adults are more likely to make an identification based on reliance

on the suspect's familiarity as opposed to directly remembering the encounter. Furthermore, Bartlett argued that older adults have viewed more faces than younger adults, which may increase the likelihood that a face feels familiar to them. Alternatively, older adults may experience difficulties with encoding new faces, which may impair their ability to distinguish differences between a familiar and an unfamiliar face. The presentation of multiple suspects (i.e., in a lineup) may discourage older eyewitnesses from making an identification if two or more faces feel familiar to them. In the case of a showup, however, the suspect is placed in the showup based on the description given by the witness, which could easily elicit feelings of familiarity and lead to more false identifications.

There is also a neurological basis for why older adults may be more likely to make inaccurate identifications from showups. Neuroimaging evidence has revealed that older adults experience deficiencies in face processing. The use of functional magnetic resonance imaging (fMRI) has revealed more activation in the Fusiform Face Area (FFA) when viewing faces as opposed to other visual stimuli (see Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). However, research also has demonstrated an age-related decline in activation of this brain region (Park et al., 2012). This decline suggests that older adults exhibit less distinctive encoding of faces. Consequently, older adults likely are disadvantaged from a neurological standpoint. If an older adult is unable to recall specific details about the perpetrator, it increases the chance of a false identification. Lineups enable older eyewitnesses to compare and contrast lineup members having features that they may have trouble recalling, something that is not afforded to them when viewing a single face in a showup. Thus, false identifications could be expected to be more prevalent in showups than in lineups.

From a forensic standpoint, it is important to examine not only witness accuracy, but also how confident they are in their identification decisions. Confident witnesses are more likely to testify in court, and the confidence of the witness is the most powerful predictor of court verdicts (Cutler, Penrod & Dexter, 1990). This is because the jury assumes that a confident witness must have a good memory for the perpetrator and of the details surrounding the crime. A false identification made with high confidence would put an innocent suspect at great risk of being convicted; thus, it is important to know not only how often older and younger adults make false identifications but also how confident they are in those identifications. Recently, researchers have argued that there is a moderately strong confidence-accuracy relationship for lineups when the data are analyzed using confidence calibration (e.g., Brewer & Wells, 2006; Palmer, Brewer, Weber and Nagesh, 2013). However, only Neuschatz et al. (in press) has compared lineups and showups using calibration, and they found that witnesses viewing lineups were better calibrated than those viewing showups. However, showup witnesses were overconfident, which puts innocent suspects at risk because confident witnesses are more likely to testify.

The goal of the current study is to compare older and younger adults' identification performance in showups and lineups, which no previous study has done. Based on previous findings involving younger adults (e.g., Gronlund et al., 2012; Wetmore et al., in press), we expected our witnesses, both older and younger adults, to show greater discriminability in lineups than showups. This difference in accuracy should also be reflected in witness' confidence judgments, such that lineups should yield better confidence calibration than showups (as suggested by Neuschatz et al., in press). Additionally, based on previous research showing age-related declines in memory, we expected younger adults to show superior discriminability to older adults for the lineups and showups. Finally, we predict an interaction between age and

identification procedure, such that older adults making identifications from showups should have the worst performance.

We tested these hypotheses by having witnesses watch a mock crime video and then make an identification from a showup or a lineup. The lineup was either fair or biased. A fair lineup has many viable options (i.e., fillers who match the description of the perpetrator), whereas a biased lineup includes fillers who do not match the description of the perpetrator. Wells and Quinlivan (2009) argued that a biased lineup should be worse than a showup because the lack of viable options may entice witnesses to choose the perpetrator or innocent suspect more often than they would in a showup. We included the biased lineup in an effort to test this prediction. After making an identification, each participant rated confidence in his or her identification decision, and answered follow up questions about willingness to testify, certainty, and other court-relevant questions.

Method

Participants

Young adults (*n*=342) ages 18-25, middle-aged adults (*n*=528) ages 26 to 59, and older adults (*n*=1541) ages 60 and older (age range over 60 years) were recruited from a regional southern university in the United States and from SurveyMonkey[1] (final *N*=2411). Demographic characteristics of these witnesses are presented in Table 1. All participants received a small monetary reward in exchange for participation. All participants were treated in accordance with the ethical guidelines of the American Psychological Association.

[1] SurveyMonkey is an online recruitment website used for data collection. Participants are recruited nationally and the only demographic information provided is what we asked participants for in the survey. We specifically did not ask for any identifying information such as email addresses or names to ensure confidentiality and anonymity. Participants are encouraged to contact the experimenters via email if they have any questions or concerns regarding the survey.

Design

The experiment conformed to a 2 (Target: present, absent) x 3 (Identification procedure: fair lineup, biased lineup, showup) x 3 (Age: young adult, middle-aged adult, older adult) between-participants design. The dependent variables were participants' identification decisions and their confidence in those decisions.

Materials

Video. The study utilized the video from Gronlund et al. (2009). The participants viewed a mock crime video lasting about 1 min 45 s. The video depicts a male and female couple walking down a sidewalk until the male actor says goodbye and enters a building. The female actor, who serves as the victim, continues walking to her car. The perpetrator subsequently jumps out from behind a bush, steals the victim's purse, and runs away. The last image participants see is a train crossing the perpetrator's path, which participants were asked about as the manipulation check question. The perpetrator is best described as a White male between the ages of 20 and 25, with dark brown short hair, brown eyes, 5'8" to 6'0" tall, weighing 160 to 185 lbs. The face of the perpetrator is visible for about 15 s.

Showups and Lineups. There was one perpetrator and one innocent suspect for each identification procedure, resulting in a total of two showups and four lineups. The two showups included a photograph of either the perpetrator or the innocent suspect. There were two lineups (one fair, one biased) for both the perpetrator and innocent suspect, taken from Gronlund et al. (2009). Gronlund et al. had research assistants search through a criminal database to find fillers that would either match the description of the perpetrator (fair lineup) or not (biased lineup). After the fillers were selected and lineups created, naive participants read the description of the perpetrator and chose the person who most closely matched that description. The resulting choice probabilities were used to compute Tredoux's *E'* (Tredoux, 1998), which quantifies lineup

fairness, with higher values reflecting greater fairness. The target present fair lineup had a Tredoux's *E'* value of 4.51, the innocent suspect fair lineup had a value of 3.88. The target present biased lineup had a value of 1.29, and the innocent suspect lineup had a value of 1.85. This indicates that in the biased lineups, there were 1.29 viable options (people who match the description of the perpetrator) for target present, and 1.85 viable options for the innocent suspect lineup. The target present fair and biased lineups are displayed in Figure 1, with the perpetrator presented in position 5.

Procedure

The procedure was similar to Gronlund et al. (2009). The study was presented to all participants via an online survey collection site (SurveyMonkey). Informed consent was obtained and participants were told they would view a video where they would be asked to make judgments of the people in the video based on their nonverbal behavior. Participants were also advised to pay close attention. The participants were then shown the mock crime video. After watching, participants were asked to select the last thing that they saw in the video (the train) to ensure that video played, and that they watched and paid attention to the video. Participants who answered this question incorrectly [young adults n = 15 (4% of total sample), middle-aged adults n = 56 (6% of total sample), older adults n = 103 (6% of total sample)] or failed to complete the survey (n = 261, ages unknown) were excluded from all data analyses. This left the final sample of N = 2411 (342 young adults, 528 middle-aged adults, and 1541 older adults) to be included in data analyses. Participants then completed a distractor task intended to take approximately 5 min. The distractor was a word scramble of various states (e.g., LBAAMAA). Participants were told to spend up to 20 s on each letter string.

Upon completion of the distractor task, participants were informed that they had just witnessed a crime and that they were to identify the perpetrator. Participants were informed that the perpetrator may or may not be present in the identification procedure. Participants were randomly assigned to receive one of the identification procedures. Those in the showup condition were shown a single photograph and asked to indicate whether the individual shown was the perpetrator; those who received the lineup condition saw an array of six photographs and were asked to identify the perpetrator or to indicate that he was not there.

After making their identification choice, participants rated their confidence in the identification decision on a 7-point Likert scale where "1" indicated "not confident at all" and "7" indicated "extremely confident". Follow-up questions evaluated participants' certainty in their identifications, view, difficulty of the identification procedure, willingness to testify, whether they believed that a witness with the same view of the crime should be trusted, and demographic information. Participants were then debriefed regarding the nature of the experiment and thanked for their participation.

Results

The goal of this study was to compare the identification accuracy of older and younger adults in lineup and showup identification procedures. To assess this, we calculated the correct and false identification rates for each condition, along with the corresponding probative values. We also conducted logistic regression analyses with identification procedure and age as predictors of suspect identifications. We then conducted ROC analyses to ascertain if discriminability—the ability to distinguish between the perpetrator and innocent suspect—was better for lineups or showups, and whether it differed for older or younger adults. Finally, we

assessed the confidence-accuracy relationship using point-biserial correlations and confidence calibration curves.

Descriptive Statistics

Table 2 displays the number and percentage of suspect identifications and rejections made in each procedure. The correct identification rates are the percentages of suspect identifications made from the target present procedures. The false identification rates are the percentages of designated innocent suspect identifications made in target absent procedures. The rejection rates are the percentages of witnesses who indicated the perpetrator was not present in the identification procedure. The showup had a much higher percentage of false identifications than the fair lineup, but a similar percentage of correct identifications. The biased lineup resulted in increased choosing of the suspect (guilty or innocent), which can be expected because the fillers do not match the description of the perpetrator. Th5ese data replicate previous studies indicating that the showup is an inferior identification procedure (Clark & Godfrey, 2009: Gronlund et al., 2012; Wetmore et al., in press).

Logistic Regression

Traditionally, logistic regression analyses are used to assess differences in performance on identification procedures. We conducted binary logistic regressions with age (young adult, middle-aged adult, older adult) and identification procedure (fair lineup, biased lineup, showup) as predictors of identification type (suspect identification, other) separately for target present and target absent procedures.

Target Present. The overall model was significant, χ^2 (8, N = 972) = 35.66, p < .001, indicating that age and identification procedure significantly predicted correct identifications. There was no main effect for age, Wald χ^2 (2, N = 972) = 4.61, p = .10 and no significant

interaction between age and identification procedure, Wald χ^2 (4, N = 972) = 1.91, p = .75. However, there was a significant main effect for identification procedure, Wald χ^2 (2, N = 972) = 19.80, p <.001, so we next conducted individual Chi-Square Tests to break down these effects. There were significantly more correct identifications in the biased lineup than the fair lineup in young adults, χ^2 (1, N = 32) = 4.57, p = .033, V = .38, middle-aged adults, χ^2 (1, N = 173) = 5.56, p = .018, V = .18, and older adults, χ^2 (1, N = 339) = 11.29, p = .001, V = .18. The biased lineup also had more correct identifications than the showup in the middle-aged adults, χ^2 (1, N = 179) = 8.12, p = .004, V = .21, and older adults, χ^2 (1, N = 395) = 18.87, p < .001, V = .22. No other comparisons were significant.

Target Absent. The overall model was significant, χ^2 (8, N = 1439) = 42.73, p < .001, indicating that age and identification procedure significantly predicted false identifications. There was no significant main effect of age, Wald χ^2 (2, N = 1439) = 1.73, p = .42 and no significant interaction, Wald χ^2 (4, N = 1439) = 3.20, p = .53. There was a main effect of identification procedure, Wald χ^2 (2, N = 1439) = 23.58, p <.001. As in the target present analyses, we conducted individual Chi-Square Tests to break down these effects. There were significantly more false identifications in the biased lineup than the fair lineup in middle-aged adults, χ^2 (1, N = 177) = 10.20, p = .001, V = .24, and older adults, χ^2 (1, N = 717) = 18.70, p <.001, V = .16. The showup also had more false identifications than the fair lineup in older adults, χ^2 (1, N = 625) = 20.31, p <.001, V = .18. No other comparisons were significant.

However, separately assessing correct and false identification rates through logistic regression (Gronlund & Neuschatz, 2014), or combining these quantities in a ratio (Wixted &

Mickes, 2012), arguably conflates accuracy and response bias. Given these problems with logistic regression analyses, we also conducted ROC analysis.

ROC Analysis

ROC analysis plots the correct identification rate versus the false identification rate at each level of witness confidence for each identification procedure. The identification procedure that has the highest ROC curve (the one closest to the upper left corner of the y-axis, or furthest from the chance diagonal) exhibits the best discriminability. In order to test for a significant difference between identification procedures, the area under the curve (AUC) is calculated. When the curves do not extend across the x-axis from 0 to 1, we compute a restricted portion of the area under the curve (i.e., partial area under the curve or – pAUC) (for details see Gronlund, Wixted, & Mickes, 2014). We constructed ROC graphs for each relevant comparison, and provide the pAUC values and results of the significance testing below. Unfortunately, the amount of data needed to conduct ROC analyses made it ill-advised to conduct the analyses on the three age groups separately. ROC analyses require about 100 responses per cell in order for the curves to be stable. Moreover, only suspect identifications are included in the analysis; participants who do not make a suspect identification are not included. In order to maintain stable curves, the data were broken into two age subsets: younger adults (age 18-59) and older adults (age 60 or older). Given that there were no age differences using the logistic regression analyses, we felt that collapsing the younger and middle-aged adults is acceptable.

An ROC comparison of the performance on lineup and showup procedures in older adults is displayed in the top panel of Figure 2. The pAUC for the fair lineup (.14) was significantly greater than that of the showup (.09), $D = 3.38$, $p < .001$. The biased lineup pAUC (.15) also was greater than the showup, $D = 3.80$, $p < .001$. The fair lineup and the biased lineup were not

significantly different from one another, D =-0.62, p =.54. Thus, regardless of lineup fairness, older witnesses were better able to discriminate between the guilty and innocent suspect in lineups than in showups.

An ROC comparison of the performance on lineup and showup procedures in younger adults is displayed in the bottom panel of Figure 2. The pAUC for the fair lineup (.14) was significantly greater than that of the showup (.09), D = 2.05, p = .03. The biased lineup pAUC (.13) was not significantly different from the showup, D = 1.64, p = .10 or the fair lineup, D = .28, p = .78. Once again, regardless of lineup fairness, younger witnesses were better able to discriminate between the guilty and innocent suspect in lineups than in showups.

We next compared older and younger adults' performance for each identification procedure. There were no significant differences between the older and younger adult pAUC's for the fair lineup, D = -.08, p = .93, biased lineup, D = -.88, p = .37, or showup, D = .48, p = .63. The only significant differences arose when we compared the younger adult lineups to the older adult showups, and vice versa. The young adult fair (and biased) lineup was significantly better than the older adult showup, D = 2.55, p = .01 (D = 2.10, p = .04). The older adult fair (and biased) lineup was significantly better than the young adult showup, D = 2.69, p < .001 (D = 3.16, p = .001).

Confidence

We first calculated point-biserial correlations on witness confidence and accuracy to assess the confidence-accuracy relationship. Overall, the confidence-accuracy correlations were low at r (1541) = .081, p = .001 for older adults, r (528) = .113, p = .009 for middle-aged adults, and r (342) = .103, p = .058 for young adults. This was not surprising, as past researchers have found a low point-biserial correlation between confidence and accuracy (e.g., Sporer, Penrod,

Read & Cutler, 1995; Wells & Murray, 1984). When separated by identification procedure, the data revealed no strong confidence-accuracy correlations for any age group or procedure (see Table 3); although lineups yielded higher accuracy than showups, according to the point-biserial correlations this did not translate into witness confidence judgments.

However, confidence calibration provides a better depiction of the confidence-accuracy relationship than does the point biserial correlation (Juslin Olsson, & Winman, 1996; Roediger, Wixted & Desoto, 2012). Calibration plots accuracy as a function of subjective confidence. Perfect calibration occurs when witnesses that express 100% confidence are 100% accurate, witnesses that express 90% confidence are 90% accurate, and so on. In order to assess which identification procedure yields the best calibration, we plotted the calibration curves and calculated the calibration index and over/underconfidence (O/U) statistic for each identification procedure. The calibration index (CI) ranges from 0 (perfect calibration) to 1 (no calibration), and reflects the weighted average of the squared difference between confidence and accuracy for each confidence level. The O/U statistic ranges from -1 to +1, with negative numbers representing that witnesses are on average underconfident; positive numbers represent that witnesses are on average overconfident.

We converted participant confidence ratings into proportions, so that the first confidence level was 1/7 (.14), the next was 2/7 (.29), and so on. Note that we present the calibration results only for the witnesses who made a suspect identification from a target present or target absent procedure, as this has been argued to be the most forensically relevant (Wixted, Mickes, Clark, Gronlund, & Roediger, under review; but see Sporer et al., 1995 who found a moderate CA correlation for choosers only). The best identification procedure will have a CI and O/U closest to zero. The calibration curves, much like ROC analyses, require about 100 data points per

identification procedure (as stated by Juslin et al., 1996); therefore, we could only plot calibration curves for younger adults (age 18-59) and older adults (age 60+).

As can be seen in the top panel of Figure 3, the best calibration for older adults resulted from the fair lineup (CI = .01, OU= -.22), and the worst calibration occurred in the showup (CI = .03, OU = -.17) and the biased lineup (CI = .02, OU = -.17) conditions. For younger adults, the fair lineup (CI = .01, OU = -.19) was better calibrated than the showup (CI = .04, OU = -.18) and the biased lineup (CI = .03, OU = -.17) conditions (refer to the bottom panel of Figure 3).

It may not be best to rely on the CI and OU because they are averages; therefore, if witnesses are overconfident at some confidence levels but underconfident at others, this would average out to reveal a very good CI and OU. We argue that visual inspection of the calibration curves is as important, if not more so, to assess. Visual inspection of the curves shows that showup witnesses are underconfident at the lowest levels of confidence (1 and 2) and overconfident at the highest levels of confidence. A similar pattern is true for biased lineups. However, fair lineup witnesses are closest to perfect calibration, but nevertheless are underconfident across almost the entire range of the confidence scale.

Discussion

The purpose of this study was to determine if there was a difference in identification accuracy between a lineup and showup in older adult witnesses and to compare the identification performance of older, middle-aged, and younger adults for all procedures. ROC analysis demonstrated that witnesses were better able to discriminate between the guilty and innocent suspects when choosing from a lineup. In hindsight, perhaps this finding is not surprising given recent research showing that identification accuracy from showups is poor despite other factors

thought to benefit showup identifications (reduced retention interval, a clothing match at encoding and test; see Neuschatz et al., in press; Wetmore et al., in press; Wetmore, Neuschatz, Gronlund, Key, & Goodsell, under review). As Wixted and Mickes (2014) suggest, the showup does not allow for a comparison among lineup members, and thus witnesses have difficulty determining which features are diagnostic of guilt. However, it is important to note that using a relative judgment strategy results in more false positive identifications, and thus could have negative implications for the legal system.

Note that, consistent with Wells and Quinlivan's (2009) contention that the biased lineup performance should be comparable to the showup, we found similar false identification rates in the biased lineup (31.3% young adults, 32.6% middle adults, 23.9% older adults) and the showup (29.7% young adults, 22.5% middle adults, and 25.5% older adults). Even though biased lineups and showups had similar false alarm rates, ROC analysis revealed that a biased lineup yielded better discriminability than showups in older adults; this was also the trend for young adults, although not to a significant degree. This is because the biased lineup had a higher correct identification rate, which resulted in superior discriminability from the biased lineup compared to the showup. Moreover, it would be incorrect to claim that our manipulation of lineup fairness was too weak, or to conclude that lineup fairness had no effect on choosing. On the contrary, the ROC curves for fair and biased lineups extend over very different ranges. The greater range over which the biased lineup ROCs extend signals that more participants viewing a biased lineup are willing to make choices at lower levels of confidence. Some might view this as bad, but these lower confidence identifications are just as likely to discriminate guilty from innocent suspects as the identifications made from the fair lineups. Since biased lineups are likely not used frequently (Wells & Quinlivan, 2009) and low confidence judgments usually do not proceed to

trial (Cutler et al., 1990), the practical import of this finding is simply that it shows that even a biased lineup is better than a showup.

An argument for the lineup advantage is that, because lineups include fillers, the false identification rate is lower than that of showups. In other words, the decision error is spread around in a lineup rather than being focused on one person (the innocent suspect) in a showup. There are several reasons this argument is incorrect. First, if this were the case, one also should see a concomitant decrease in correct identifications in lineups, but that was not the case (compare the fair lineup correct identification rates of 37.5% for young adults, 44.6% for middle-aged adults, and 44.8% for older adults to the showup correct identification rates of 52.5% for younger adults, 41.3% for middle-aged adults, and 40.9% for older adults). In contrast to these similar correct identification rates, the showup had more than double the percentage of false identifications obtained in lineups. This point alone demonstrates that simply having more options in the lineup is not producing the higher performance of the lineup. In addition, filler identifications are forensically unimportant. Because fillers are known to be innocent, a filler identification made in the real world is evaluated by law enforcement as being incorrect and serves to eliminate witnesses who are guessing or who have a poor memory for the perpetrator[2]. These witnesses are then likely to be excluded from further investigation and would not testify. Filler identifications do not lead to wrongful convictions, hence the focus on the identifications of suspects in showups and lineups.

Calibration curves revealed that showups yielded a poorer confidence-accuracy relationship than fair lineups. Not only do witnesses viewing a showup exhibit poorer

[2] Police might use a filler identification as evidence that they have an innocent suspect (Clark, 2012).

discriminability, they express overconfidence when doing so. Specifically, these witnesses are overconfident at high levels of confidence, which is problematic for the legal system because confident witnesses are perceived as more believable by jurors (Cutler et al., 1990).

The lack of age differences in our study does not confirm previous research and is, at first glance, unexpected. However, one difference between the present research and that of Searcy et al. (1999) is that we focused only on identifications of the suspect (e.g., the perpetrator and designated innocent suspect), whereas Searcy et al. focused on correct *decisions.* When we calculated percentages of correct decisions (perpetrator identifications from target present procedures and correct rejections from target absent procedures) and incorrect decisions (filler choices from target present procedures and any choices from target absent procedures), we obtained the same pattern as Searcy et al. in both fair lineups and showups. Specifically, the lineup correct decision rate was higher for younger adults (*M* = .50) than for older adults (*M* = .41) and the incorrect decision rate was higher for older adults (M = .59) than for younger adults (*M* = .50). The same pattern emerged for showups, more correct decisions (*M* = .62) and fewer incorrect decisions (*M* = .38) for younger adults than for older adults (correct decision rate *M* = .59, incorrect decision rate *M* = .41). In sum, our data are entirely consistent with what Searcy et al. found, and younger adults do show superior performance when examining correct decisions or accuracy. However, the move to analyzing suspect identifications is important for two reasons (1) suspect identifications are more forensically relevant (2) this allows us to conduct ROC analyses, which disentangle discriminability from response bias and are a superior way to compare different identification procedures.

Limitations and Implications

There are limitations to the current study that should be acknowledged. The difference between a simulated crime video and a real life crime lies in the fact that a witness feels more stress or arousal during the crime that cannot be replicated in the laboratory or with an online sample of participants. Furthermore, witnesses in real life identification tasks experience stronger demand characteristics (e.g., police pressure to choose, Dysart & Lindsay, 2007) than mock witnesses. This might particularly be the case for showup identifications that take place at the scene of the crime, shortly after the crime occurs. Because the witnesses in this study did not experience the stress of witnessing a live event, nor the demand characteristics associated with the identification procedure, it is likely that the results of the current study understate what happens in the real world. But one thing that may improve the performance of real witnesses is the ability to access more cues while viewing the crime and making an identification, especially from a live showup. For example, perpetrator height, gait, and other characteristics, are more salient in real crimes (see Valentine et al., 2012). These extra cues may help witness performance when identifying the perpetrator: Future research is needed comparing the identification accuracy of live showups to photo showups and photo lineups.

Another limitation of the study is that we did not test for vision, and it is possible that visual impairment may be confounded with age (i.e., older adults may suffer more visual impairment). We believe this may only be a small concern for three reasons. First, the manipulation check question asked all participants to indicate the last thing they saw in the video. Any participants who answered this incorrectly were eliminated from our final data set. In analyzing the data, the number of exclusions based on this question did not vary with age. Second, since there was no main effect of age in the ROC analyses, it seems unlikely that the

results could be attributed to poorer vision in the older adult sample. Finally, since all older adult data were collected online, it is likely that our sample wass comfortable with using a computer, which stands to reason that they do not have problems seeing text or images on the computer.

A final concern with this study may be the use of one suspect/perpetrator pair in this study. It is the case that this will indeed raise concerns about generalizability; however, we have replicated the result that lineups are superior to showups with many different studies (Gronlund et al., 2012; Wetmore et al., in press). Future research should replicate these findings with different materials, procedures, and laboratories.

Given that showups are such a commonly used identification procedure in the United States, the poor performance of this procedure in both older and younger adults is problematic and has important implications for the legal system. The use of showups rather than lineups is putting innocent suspects at greater risk, but the use of showups also increases the likelihood that guilty suspects will be missed. As it is likely that showup use will continue, future research should investigate ways to possibly strengthen the showup as an identification procedure. It is important for police officers to be able to conduct field identifications quickly and efficiently. As has been cited elsewhere, the reasons for doing field identifications are (1) this gets potentially dangerous people off the streets quickly and (2) exonerates innocent people quickly. There is no question that a good field procedure would have great value to law enforcement and society. However, at the present time, showups are not a reliable method, even though they can be conducted quickly and efficiently (Neuschatz et al., in press; Wetmore et al., in press). It is incumbent on researchers to examine better ways to conduct these important field identifications. But while we await this research, it behooves legal practitioners to use lineups instead of showups. Given that lineups can now be constructed in police vehicles (Wells, Steblay, &

Dysart, in press), little need be lost in terms of time or efficiency by constructing a lineup. And whatever the cost encumbered by constructing a lineup, it surely is outweighed by the benefit of using a more reliable identification procedure.

References

Bartlett, J. C. (2014). True and false recognition of faces by young and old adults. In M. P. Toglia, D. F., Ross, J. A. & Pozzulo, E. Pica (Eds.), *The elderly witness in court*. London: Francis & Taylor.

Bornstein, B.H. (1995). Memory processes in elderly eyewitnesses: What we know and what we don't know. *Behavioral Sciences and the Law, 13* (3), 337-348. doi: 10.1002/bsl.2370130303

Brainerd, C. J., & Reyna, V. F. (2001). Fuzzy-trace theory: Dual processes in reasoning, memory, and cognitive neuroscience. *Advances in Child Development and Behavior, 28,* 49-100.

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* (1), 11-30.

Clark, S. E. & Godfrey, R. D. (2009). Eyewitness identification evidence and innocence risk. *Psychonomic Bulletin & Review, 16* (1), 22-42, doi: 10.3758/PBR.16.1.22

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform psychological science and public policy. *Perspectives on Psychological Science, 7*(3), 238-259, doi: 10.1177/1745691612439584

Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1990). Juror sensitivity to eyewitness identification evidence. *Law and Human Behavior, 14*(2), 185-191.

Dysart, J. E., & Lindsay, R. C. L. (2007). Show-up identifications: Suggestive technique or reliable method? In Rod C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.),

The handbook of eyewitness psychology, Vol II: Memory for people. (pp. 137-153). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Garrett, B.L. (2011). *Convicting the Innocent: Where Criminal Prosecutions Go Wrong. Cambridge*: Harvard University Press.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nature Neuroscience, 2*, 568-573. doi: 10.1038/9224

Gonzales, R., Ellsworth, P. C., & Pembroke, M. (1993). Response biases in lineups and showups. *Attitudes and Social Cognition, 64* (4), 525-537. doi: 10.1037/0022-3514.64.4.525

Gronlund, S. D., Carlson, C. A., Neuschatz, J. S., Goodsell, C. A., Wetmore, S. A., Wooten, A., & Graham, M. (2012) Showups versus lineups: An evaluation using ROC analysis. *Journal of Applied Research in Memory and Cognition, 1*, 221-228. doi:10.1016/j.jarmac.2012.09.003

Gronlund, S. D., Carlson, C. A., Dailey, S. B., & Goodsell, C. A. (2009). Robustness of the sequential lineup advantage. *Journal of Experimental Psychology: Applied, 15* (2), 140-152. doi: 10.1037/a0015082

Gronlund, S. D., & Neuschatz, J. S. (2014). Eyewitness identification discriminability: ROC analysis versus logistic regression. *Journal of Applied Research in Memory and Cognition, 3*, 54-57.

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), 3-10. doi: 10.1177/0963721413498891

Innocence Project (2014). Understand the causes: eyewitness misidentification. Retrieved from http://www.innocenceproject.org/understand/Eyewitness-Misidentification.php on April 21, 2014

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22* (5), 1304-1316. doi: 10.1037/0278-7393.22.5.1304

Neuschatz, J. S., Wetmore, S. A., Key, K., Cash, D., Gronlund, S. D., & Goodsell, C. A. (in press). Comprehensive evaluation of showups. In M. Miller & B. Bornstein (Eds.), *Advances in Psychology and Law*. New York: Springer.

Palmer, M., Brewer, N., Weber, N. & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied, 19*, 55-71.

Park, J., Carp, J., Kennedy, K. M., Rodrigue, K. M., Bischof, G. N., Huang, C., Rieck, J. R., Polk, T. A., & Park, D. C. (2012). Neural broadening or neutral attenuation? Investigating age-related dedifferentiation in the face network in a large lifespan sample. *The Journal of Neuroscience, 32* (6), 2154-2158. doi:10.1523/JNEUROSCI.4494-11.2012

Price, J., Mueller, N., Wetmore, S. A., & Neuschatz, J.S. (2014). Eyewitness Memory and Metamemory in Older Adults. In M. P. Toglia, D. F. Ross, J. A. & Pozzulo, & E. Pica (Eds.), *The elderly witness in court*. London: Francis & Taylor.

Rabinowitz, J. C., Craik, F. I. M., & Ackerman, B. P. (1982). A processing resource account of age differences in recall. *Canadian Journal of Psychology, 36,* 325-344. doi: 10.1037/h0080643

Reyna, V. F. & Brainerd, C. J. (2011). Dual processes in decision making and developmental neuroscience: A fuzzy-trace model. *Developmental Review, 31,*180-206. doi: 10.1016/j.dr.2011.07.004

Roediger, H. L., Wixted, J. H., & Desoto, K. A. (2012). The curious complexity between confidence and accuracy in reports from memory. In L. Nadel & W. Sinnott-Armstrong (Eds.), Memory and Law (pp. 84-118). New York, NY: Oxford University Press.

Rothmans, M. B., Dunlop, B. D., & Entzel, P. (2000). *Elders, crime and the criminal justice system: Myth, perceptions and reality in the 21st century*. Springer Series on Life Styles and Issues in Aging. New York: Springer.

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. doi: 10.3758/BF03211547

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118* (3), 315-327. doi: 10.1037/0033-2909.118.3.315

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2003). Eyewitness accuracy rates in police showup and lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 27* (5), 523-540. doi: 10.1023/A:1025438223608

Stovall v. Denno, (1967). 388 U.S. 293.

SurveyMonkey Inc. (2013). [Online computer software]. Palo Alto, California.

Tredoux, C. G. (1998). Statistical inference on lineup measures. *Law & Human Behavior, 22,* 217 – 237. doi: 10.1023/A:1025746220886

Valentine, T., Davis, J. P., Memon, A. & Roberts, A. (2012). Live showups and their influence on a subsequent video line-up. *Applied Cognitive Psychology, 26,* 1-23. doi: 10.1002/acp.1796

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist, 48* (5), 553-571. doi:10.1037/0003-066X.48.5.553

Wells, G. L., & Murray, D. M. (1984). Eyewitness confidence. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 155–170). New York, NY: Cambridge University Press.

Wells, G. L. & Quinlivan, D. S. (2009). Suggestive eyewitness identification procedures and the Supreme Court's reliability test in light of eyewitness science: 30 years later. *Law & Human Behavior, 33,* 1-24. doi: 10.1007/s10979-008-9130-3

Wells, G. L, Steblay, N. K, & Dysart, J. E., (in press). Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior*

Wetmore, S. A., Neuschatz, J. S., Gronlund, S. D., Goodsell, C. A., Wooten, A., & Carlson, C. A. (2014). Effect of retention interval on showup and lineup performance. *Journal of Applied Research in Memory and Cognition*, http://dx.doi.org/10.1016/j.jarmac.2014.07.003

Wetmore, S. A., Neuschatz, J. S., Gronlund, S. D., Key, K. N., & Goodsell, C. A. (2015). Do the clothes make the criminal in showups? The influence of clothing match on identification accuracy in showups. *Journal of Applied Research in Memory and Cognition, 4* (1), 36-42. doi:10.1016/j.jarmac.2014.12.003

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*(3), 275-278.

Wixted, J. T. & Mickes, L (2014). A signal-detection based diagnostic-feature model of eyewitness identification. *Psychological Review*, *121* (2), 262-276. doi: 10.1037/a0035940

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., Roediger, H. L. (under review). Confidence judgments are useful in eyewitness identifications: A new perspective. *American Psychologist*

Table 1

Demographic Information Separated by Age

		Young Adult	Middle Adult	Older Adult
Gender	Male	133 (38.9)	245 (46.4)	766 (49.7)
	Female	209 (61.1)	283 (53.6)	775 (50.3)
Mean Age		19.9 (*SD*=2.1)	43.9 (*SD*=10.0)	66.4 (*SD*=5.2)
Ethnicity	Hawaiian/Pacific Islander	2 (.58)	0 (0)	1 (.06)
	American Indian	10 (2.9)	4 (.76)	3 (.84)
	African American	25 (7.3)	39 (7.4)	49 (3.2)
	Hispanic	34 (10.0)	32 (6.1)	15 (.97)
	Asian	24 (7.0)	14 (2.7)	9 (.58)
	Caucasian	235 (68.7)	421 (79.7)	1428 (92.7)
	Other	12 (3.5)	18 (3.4)	26 (1.7)
Education	High School Diploma/GED	259 (75.7)	175 (33.1)	412 (26.7)
	Associates Degree	30 (8.8)	94 (17.8)	245 (15.9)
	Bachelor’s Degree	42 (12.3)	155 (29.4)	400 (26.0)
	Master’s Degree	6 (1.75)	59 (11.2)	324 (21.0)
	Doctoral Degree	0 (0)	22 (4.2)	107 (7.0)
	Other	5 (1.5)	23 (4.4)	53 (3.4)

Note: Numbers in parentheses are percentages.

Table 2

Identification Decisions by Age

		Younger Adults	Middle Adults	Older Adults
Fair	TP Suspect ID	37.5%	44.6%	44.8%
	Rejection	25%	28.4%	33.9%
	N	16	74	174
	TA Suspect ID	21.4%	12.5%	11.6%
	Rejection	57.1%	42.1%	39.7%
	N	28	88	370
Biased	TP Suspect ID	75.0%	62.6%	63.0%
	Rejection	18.8%	24.2%	26.7%
	N	16	99	165
	TA Suspect ID	31.3%	32.6%	23.9%
	Rejection	43.8%	38.2%	42.1%
	N	16	89	347
Showup	TP Suspect ID	52.5%	41.3%	40.9%
	Rejection	47.5%	58.8%	59.1%
	N	118	80	230
	TA Suspect ID	29.7%	22.5%	25.5%
	Rejection	70.3%	77.6%	74.5%
	N	148	98	255

Table 3

Confidence-Accuracy Point-Biserial Correlations

		N	*r*	*p*
Young Adult	Fair	44	.060	.701
	Biased	32	.211	.247
	Showup	266	.095	.122
Middle Adult	Fair	162	.162	.040
	Biased	188	.127	.083
	Showup	178	.021	.785
Older Adult	Fair	544	.038	.372
	Biased	512	.073	.100
	Showup	485	.079	.083

Figure 1. The top panel shows the target present fair lineup used in this study. The bottom panel shows the target present biased lineup. The perpetrator is in position 5 in both lineups.

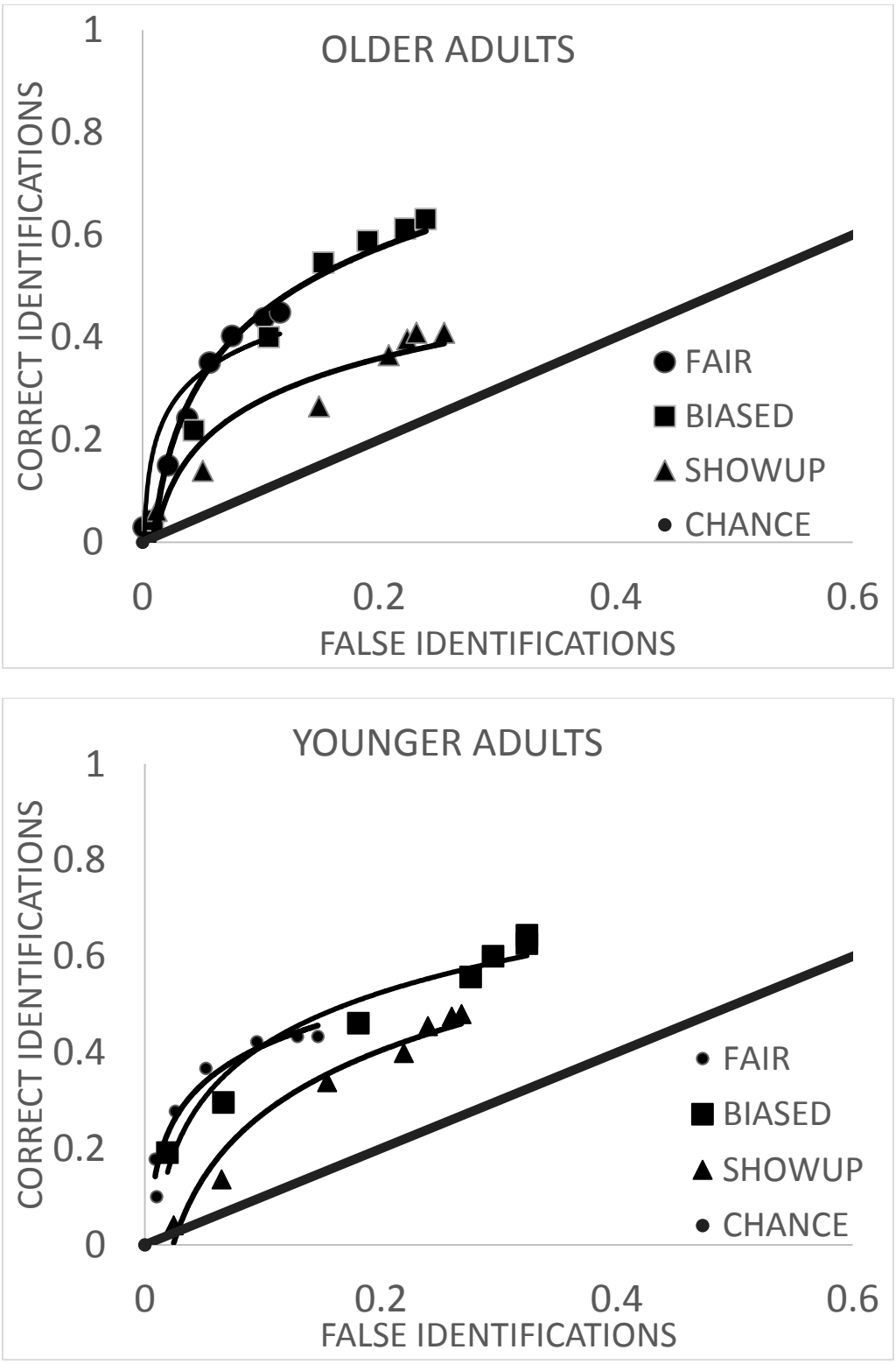

Figure 2. The top panel displays the ROC curves for older adult performance in the fair lineup, biased lineup, and showup; the diagonal line represents chance performance. The bottom panel displays the ROC curves for young adult performance on the same procedures.

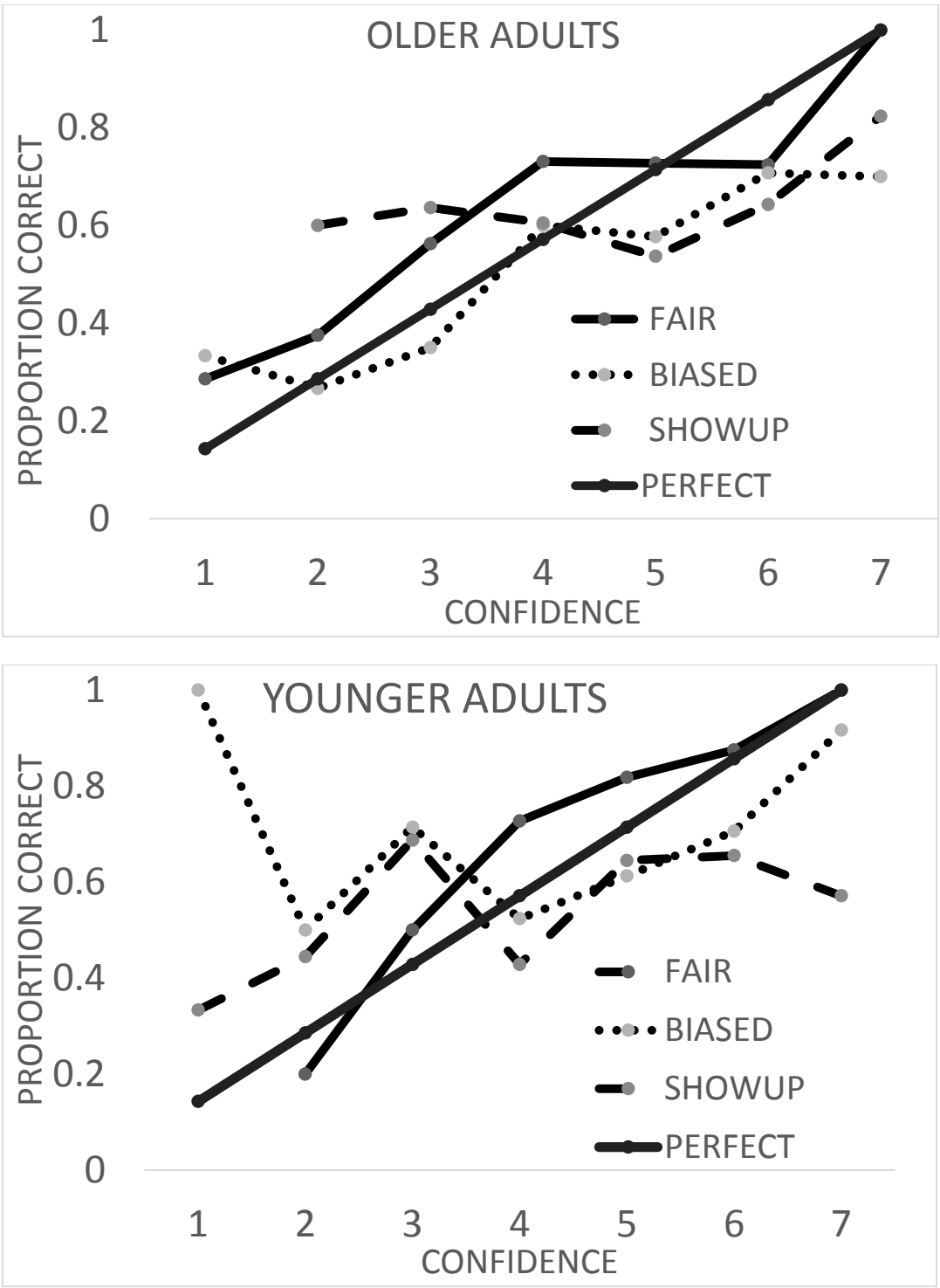

Figure 3. Calibration plots for each identification procedure for older adults (top panel) and younger adults (bottom panel). The diagonal line is the perfect calibration line; the fair lineup is the solid line, the showup is the dashed line, and the biased lineup is the dotted line.