Research Article

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
2016, Vol. 27(9) 1227–1239

DOI: 10.1177/0956797616655789
pss.sagepub.com
SAGE

Unfair Lineups Make Witnesses More Likely to Confuse Innocent and Guilty Suspects

Melissa F. Colloff[1], Kimberley A. Wade[1], and Deryn Strange[2]
[1]Department of Psychology, University of Warwick, and [2]Department of Psychology, John Jay College of Criminal Justice, City University of New York

Abstract

Eyewitness-identification studies have focused on the idea that unfair lineups (i.e., ones in which the police suspect stands out) make witnesses more willing to identify the police suspect. We examined whether unfair lineups also influence subjects' ability to distinguish between innocent and guilty suspects and their ability to judge the accuracy of their identification. In a single experiment (N = 8,925), we compared three fair-lineup techniques used by the police with unfair lineups in which we did nothing to prevent distinctive suspects from standing out. Compared with the fair lineups, doing nothing not only increased subjects' willingness to identify the suspect but also markedly impaired subjects' ability to distinguish between innocent and guilty suspects. Accuracy was also reduced at every level of confidence. These results advance theory on witnesses' identification performance and have important practical implications for how police should construct lineups when suspects have distinctive features.

Keywords

eyewitness memory, lineup fairness, distinctive features, diagnostic feature detection, open data

Received 7/30/15; Revision accepted 5/27/16

In 1986, a woman viewed a lineup and identified the police suspect, Leonard Callace, as her attacker. She had described the attacker as a White male with reddish-blonde Afro-style hair and a full beard. But Callace—who had a full beard and straight hair—appeared in the lineup with 5 men who had only moustaches. After Callace served 6 years in prison, DNA evidence revealed he was not the attacker. Callace's case and many others highlight the importance of preventing suspects with distinctive features from standing out in lineups (for more examples of wrongful convictions, see http://www.innocenceproject.org). But why do unfair lineups make eyewitnesses prone to making erroneous identification decisions? Is it because unfair lineups make witnesses more willing to identify the police suspect, regardless of whether that suspect is guilty or innocent? Or is it because unfair lineups make it more difficult for witnesses to determine whether the lineup contains the actual culprit? We aimed to answer these questions.

A lineup typically consists of one police suspect, who is either guilty (i.e., the culprit) or innocent, and a number of other lineup members (foils) who are not connected to the case and are therefore known to be innocent. Research has shown that suspects who stand out are prone to be selected, but not for the right reasons (i.e., they are not selected because they are an exact match to the witness's memory of the culprit; Wells, Rydell, & Seelau, 1993). Why? The long-standing explanation is that witnesses tend to select the person who looks most like the culprit, in much the same way that a student answering a multiple-choice question tends to select the option that looks most like the right answer (Wells, 1984). Indeed, it is well established that when the police suspect is the only person who matches the witness's description of the culprit, the witness tends to select the suspect instead of another member of the lineup (Doob & Kirshenbaum, 1973; Wells, Leippe, & Ostrom, 1979). Reviews and meta-analyses have

Corresponding Author:
Kimberley A. Wade, Department of Psychology, University of Warwick, Coventry, CV4 7AL United Kingdom
E-mail: k.a.wade@warwick.ac.uk

also shown that suspects who look less like the other members of a lineup are more often identified by witnesses (Clark, 2012; Fitzgerald, Price, Oriet, & Charman, 2013). Two problems arise from this tendency. First, if the police suspect is the culprit, the identification is correct, but not for the right reasons—much like the student who gets the correct answer but does not actually know the right answer. Second, if the police suspect is not the culprit, the misidentification might send an innocent person to prison. The observation that witnesses are more willing to identify a suspect—that is, correctly identify a guilty suspect when he or she is present in the lineup but incorrectly identify an innocent suspect when the real culprit is absent from the lineup—can help us to understand why unfair lineups often result in misidentifications.

However, a new approach, the diagnostic-feature-detection model, supports an additional prediction: Unfair lineups may also impair witnesses' ability to differentiate between the actual culprit and an innocent suspect (Wixted & Mickes, 2014). Consider what happens when a witness views a lineup, whether fair or unfair. The idea is that for each lineup member's face, features combine to create a memory signal (a sense of familiarity and recollection), and the witness uses that signal to make an identification decision. Some features differ between the culprit and an innocent suspect, and such features can help the witness make a better decision. For instance, Leonard Callace had straight hair, whereas the culprit had an Afro. But the culprit and an innocent suspect may share other facial features, and such shared features cannot help the witness. For instance, Callace and the culprit each had a full beard. If witnesses give weight to these shared features, their ability to distinguish between culprits and innocent suspects will suffer.

How, then, do witnesses make identifications in an unfair lineup, in which only the police suspect possesses the distinctive facial feature (e.g., a full beard) that the witnesses remember? To the extent that witnesses do not realize that the distinctive feature is unhelpful, they might erroneously weight that feature. Giving weight to an unhelpful feature will impair their ability to discriminate between real culprits and innocent suspects. Consistent with this idea, the results of one study showed that witnesses were better able to distinguish between guilty and innocent suspects when all lineup members, including the suspect, had the same emotional expression. But witnesses found it harder to distinguish between innocent and guilty suspects when the suspect was the only lineup member with that expression (Flowe, Klatt, & Colloff, 2014). Presumably, witnesses who saw the matched-expression lineup discounted the emotional expression and used other useful information to make an identification. By contrast, witnesses who saw the unmatched-expression lineup weighted the emotional expression, even though it was objectively unhelpful because it was something that both the innocent and guilty suspect shared. Other studies have found that people are better able to distinguish between innocent and guilty suspects when they are presented with a fair lineup rather than a single photo of a suspect (Key et al., 2015; Wetmore et al., 2015). Again, the fair lineup might permit subjects to discount unhelpful features, but a single photo might not.

In the real world, guidelines for constructing lineups often state that the police should prevent suspects with distinctive features from unduly standing out. In the United States, England, and Wales, for instance, police sometimes artificially replicate a suspect's distinctive feature across the lineup members (*replication*; see Fig. 1a); other times, they conceal the feature on the suspect and conceal a similar area on the other members (Police and Criminal Evidence Act 1984, Code D, 2011; Technical Working Group for Eyewitness Evidence, 1999). Concealing involves either pixelating the area of the feature (*pixelation*; Fig. 1b) or covering the area with a solid black rectangle (*block*; Fig. 1c). These techniques represent a heartening translation of science into practice. Nonetheless, many efforts to make lineups fair are unsuccessful, and police officers still often do nothing and allow distinctive suspects to stand out (e.g., MacLin, MacLin, & Albrechtsen, 2006; Valentine & Heaton, 1999; Wogalter, Malpass, & Mcquiston, 2004).

How, then, might replication, pixelation, or block lineups affect witness performance? First, because the suspect does not unduly stand out, witnesses should be less willing to identify the suspect. Second, because the distinctive feature appears either on every lineup member (replication lineup) or on none of the lineup members (pixelation lineup and block lineup), witnesses should be more likely to weight something other than the distinctive feature. Therefore, they should also be better able to distinguish between the culprit and an innocent suspect. By contrast, if the police do nothing, and a suspect is allowed to stand out (i.e., a *do-nothing* lineup; see Fig. 1d), witnesses should be more willing to choose the suspect, and they should find it harder to distinguish between the culprit and an innocent suspect. In the current research, we tested these hypotheses.

Method

Design

We used a 4 (lineup type: replication, pixelation, block, do-nothing) × 2 (target: present, absent) between-subjects design. We recruited as many subjects as possible before the end of spring term, aiming for at least 1,000 subjects with usable data in each of the eight conditions.

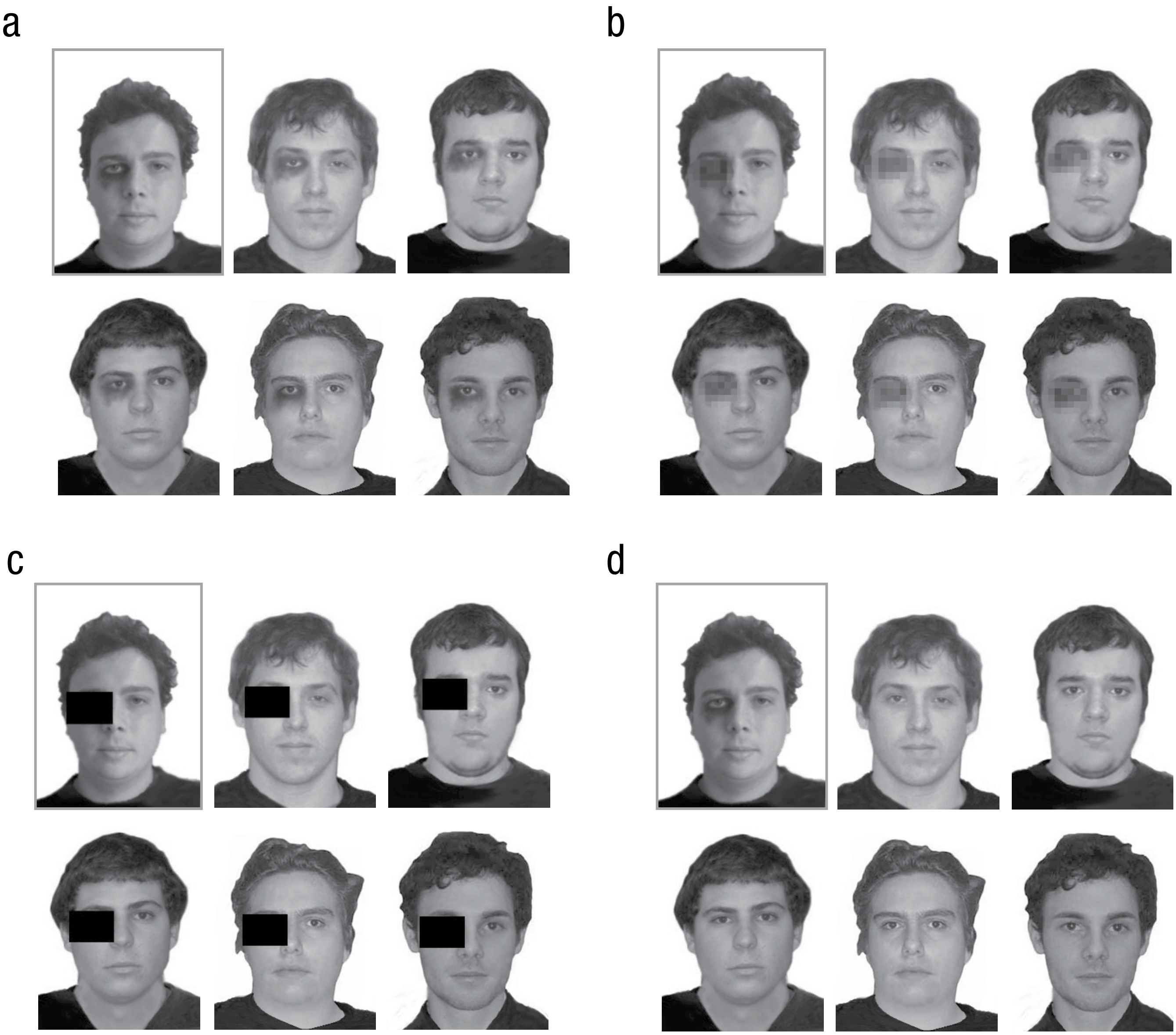

Fig. 1. Examples of lineup types. A suspect's distinctive feature can be replicated (a) or concealed either by pixelation (b) or with a block (c). These are considered fair lineups. Doing nothing about the distinctive feature (d; a do-nothing lineup) constitutes an unfair lineup. The boxed image in each lineup indicates the suspect with the distinctive facial feature.

Subjects

The subjects were 9,841 adults from around the world who completed the task online. We excluded subjects who experienced technical difficulties while watching the video (n = 689, 7%), experienced programming errors while viewing the lineup (n = 128, 1%), or incorrectly answered an attention-check question on the content of the video (n = 99, 1%). In total, we excluded 916 people (9%; between 89 and 218 in each of the eight conditions), which resulted in a final sample size of 8,925. Of these, 5,495 subjects were recruited from social-networking sites and were entered into a prize drawing for four £50 Amazon vouchers, 2,405 subjects were recruited via Amazon Mechanical Turk and received $0.60, 871 students were recruited from John Jay College of Criminal Justice and received extra credit in a course, and 154 students were recruited from a sixth form (final year of high school) in the United Kingdom and completed the study as part of a research-methods course. Because the pattern of results was the same among the Internet and student samples, we combined their data for our analyses. Each cell contained between 1,017 and 1,145 subjects. We also checked for multiple responses by the same

Table 1. Distribution of the Samples by Sex, Age, and Race or Ethnicity

	Sample			
Characteristic	Social media	Mechanical Turk	University	High school
Sex				
Male	1,498	1,091	265	40
Female	3,960	1,309	599	114
Prefer not to say	37	5	7	0
Age (years)				
16–20	1,606	79	593	149
21–30	1,693	997	252	0
31–40	870	675	18	0
41–50	649	326	4	0
51–60	395	224	0	0
61–70	161	86	0	0
≥ 71	46	13	0	0
Prefer not to say	75	5	4	5
Race or ethnicity				
White or European	4,633	1,494	195	72
Latin or Hispanic	52	102	339	0
Black, African, or Caribbean	72	178	140	6
South Asian	156	399	41	5
East Asian	175	90	42	6
Middle Eastern	25	7	13	2
Mixed	136	71	37	11
Other	147	41	27	39
Prefer not to say	99	23	37	13

individual by examining internet-protocol addresses and e-mail addresses. These checks revealed 26 possible duplicates (i.e., 0.3% of subjects). Our results were the same regardless of whether we included or excluded data from these subjects. Table 1 shows a demographic breakdown of the sample.

Materials

Videos. It is widely documented that variability in encoding and test conditions is crucial when trying to detect reliable and generalizable effects (Brewer, Keast, & Sauer, 2010; D. S. Lindsay, Read, & Sharma, 1998). Accordingly, we created four 30-s videos depicting four different nonviolent crimes, so that encoding conditions varied on several dimensions, including (a) the appearance of the target (each video featured a different White male culprit), (b) the distinctive feature on the target (each culprit had a unique distinctive feature), (c) the crime committed (carjacking, graffiti attack, mugging, theft), and (d) the exposure duration of the target in each video (5–16 s across the four videos). At testing, there was some variation between the encoding stimuli (the target in the crime video) and the test stimuli (the target's photographic image), simply because videos and photos of people can vary to different extents. Targets also varied in their similarity to the foils (for a more complete description of each crime, see the Supplemental Material available online).

Lineups. Each lineup contained six photos. The lineup consisted of either one culprit and five foils (a target-present lineup) or six foils (a target-absent lineup). We created a pool of 40 foils for each culprit, so that we could randomly generate lineups from these pools. To create the pools of foils, we created *modal descriptions* of the culprits by first asking a group of 18 subjects to watch each crime video and then answer 16 questions about the culprit's physical attributes, including questions about his sex, eye color, hair color, height, weight, and ethnicity. Some characteristics required a categorical option choice (e.g., sex), whereas others required free-text responses (e.g., height and weight). As researchers have done in other studies (Carlson, Gronlund, & Clark, 2008; Zarkadi, Wade, & Stewart, 2009), we then entered the modal descriptions into the Florida Department of Corrections Inmate Database (http://www.dc.state.fl.us/AppCommon/) to retrieve 40 photos of men who matched the modal description of each of the four culprits (i.e., 160 photos in total). This approach fits with the widely accepted recommendation that foils should match the witness's description of the culprit (Technical Working Group for Eyewitness Evidence, 1999; Wells, 1993).

The photos we selected from the database depicted men directly facing the camera. To control for the influence of emotional display, we selected men with neutral facial expressions (Flowe et al., 2014). We used Adobe Photoshop CS5 to transform the images to gray scale and to remove any background color or pattern. We removed any distinctive facial features. To prevent biases attributable to clothing, we also digitally altered each photo so that all foils appeared to be wearing a plain black T-shirt (R. C. L. Lindsay, Wallbridge, & Drennan, 1987). We took similar-looking mug shots of the culprits on the day we filmed the mock crimes. We edited these mug shots in the same way as the foil photos, including adjusting the resolution to match that of the foil photos.

Next, we edited the four pools of 40 images (160 in total) to create foils for the replication, pixelation, and block lineups (see Fig. 1). For the replication lineups, we digitally added the culprit's distinctive feature to each face in the pool of 40 foil photos. To reflect current police practice in several jurisdictions including England, Wales, New Zealand, Canada, and Germany, we made the size, appearance, and location of this distinctive feature very similar to—but not identical to—the culprit's distinctive feature. For pixelation lineups, we concealed the culprit's distinctive feature by pixelating it, and we pixelated the same region on each of the faces in the 40 foil photos in the corresponding pool. For block lineups, we concealed the culprit's distinctive feature with a solid black rectangle; we overlaid the same shape in the same region on each of the faces in the 40 foil photos in the corresponding pool. For target-present do-nothing lineups, we left the culprit's distinctive feature uncovered and did nothing to the faces in the foil photos. For each target-absent do-nothing lineup, we needed one foil-photo face that had a distinctive feature similar to the culprit's; accordingly, we used one replication foil photo to which the culprit's distinctive feature had been added. The other 5 foil photos in each do-nothing target-absent lineup remained undoctored. Note that a target-absent do-nothing lineup mirrors the real-world situation in which a witness reports the culprit's distinctive feature to the police, but the police apprehend an innocent person with a similar distinctive feature and place him in the lineup. Ultimately, this process resulted in a total of 640 foil photos across all four culprits and all four lineup types.

To check that we had doctored our foil photos in the same way that police do, we gathered evidence of ecological validity by consulting with a Detective Inspector from a local police force in the United Kingdom who sat on the National Committee for Identification Evidence. We asked her to evaluate 18 randomly selected foil photos to which we had applied the replication, pixelation, and block manipulations. The officer agreed that the images were concordant with police practice in England and Wales.

To make sure our replication foil photos did not look doctored, we then asked 5 new subjects to view all four replication-foil pools (one pool for each culprit, 160 photos total) and to identify any images that either did not match the modal description of the culprit or appeared to have been digitally altered. These subjects said that all the faces in the foil photos matched the descriptions of the culprits, but they identified 14 photos that they believed to have been digitally altered. We then reedited the distinctive features on these 14 photos until all 5 subjects were satisfied. Next, we asked a new group of 39 subjects to evaluate four target-present replication lineups (one for each culprit) in which the foils were selected at random. We asked them to identify which photo had not been digitally altered; they performed no better than chance at this task (all ps > .20). Taken together, these findings suggest that our replication photos did not look manipulated and that our procedure for generating lineups did not bias subjects toward or against the suspect.

Procedure

Subjects were told that the study was about personality and perception. They were randomly assigned to one of the eight experimental conditions and to one of the four crime videos (with the constraint that subject numbers were relatively equal in each condition).

There were three phases in the experiment. In the first phase, subjects watched a video of a crime. They were instructed to pay close attention because they would be asked questions about it later. After the video ended, we asked subjects whether they had encountered any technical problems while viewing the video. The second phase, a filler phase, then began. In this phase, subjects worked on three questionnaires and an anagram puzzle for a total of 8 min. The questionnaires were the Autism Spectrum Quotient (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001), the Six-Item Short-Form State scale of the Spielberger State-Trait Anxiety Inventory (Marteau & Bekker, 1992), and the Ten-Item Personality Inventory (Gosling, Rentfrow, & Swann, 2003). We do not discuss subjects' performance on these scales because they served as a filler task. In the third phase, we asked subjects to indicate their confidence that they would be able to recognize the culprit. Subjects responded on a 100-point Likert-type scale ranging from 1 (*completely uncertain*) to 100 (*completely certain*). Immediately after this task, subjects saw a lineup composed of two rows of three photos.

In target-present lineups, the culprit was presented alongside five foils who were selected at random from the corresponding pool. The position of the culprit in the lineup was determined randomly for each subject. In do-nothing target-absent lineups, one foil with the culprit's distinctive feature and five foils without it were selected at random; the position of the innocent suspect (i.e., the foil who had the culprit's distinctive feature) was randomly determined for each subject. In replication, pixelation, and block target-absent lineups, six foils were selected at random (i.e., there was no designated innocent suspect). We chose this method of generating lineups to increase the generalizability of our results and to avoid the problems associated with using a small number of pairs of culprits and innocent suspects. By randomly generating lineups, we also avoided using lineup fairness and bias measures, which are not always stable (Mansour, Beaudry, Kalmet, Bertrand, & Lindsay, in press).

All subjects were told that the culprit "may or may not be present" and were then asked to make a single identification by clicking on either the person they believed to be the culprit or on an option labeled "Not Present." Next, subjects used a 100-point Likert-type scale (1 = *completely uncertain*, 100 = *completely certain*) to rate their confidence in their decision. Finally, subjects answered an attention-check question ("What happened in the video that you watched?"), and they also answered a number of demographic questions.

Results

Recall that our primary aim was to determine the extent to which unfair lineups affect witnesses' (a) willingness to identify the suspect and (b) ability to distinguish between real culprits and innocent suspects. We addressed these questions by using receiver-operating-characteristic (ROC) analysis and gathered further information by examining both the distribution of subjects' identification responses and the subjects' ability to judge the accuracy of their identification decisions.

ROC analysis

Overview. Because the ROC approach is relatively new in the field of eyewitness memory, a brief overview should prove helpful. In ROC analysis, the first step is to construct an ROC curve for each lineup technique. Each curve plots the correct-identification rate, or hit rate (HR), of guilty suspects in target-present lineups against the false-identification rate, or false alarm rate (FAR), of innocent suspects in target-absent lineups. In many ways, ROC analysis is similar to analysis of the traditional

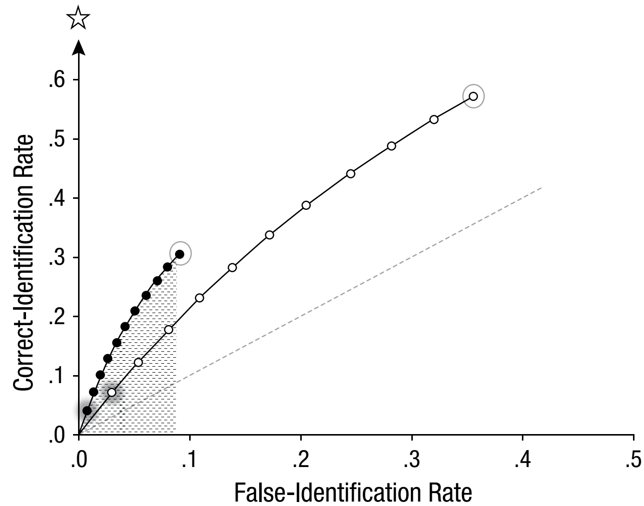

Fig. 2. Two hypothetical receiver-operating-characteristic (ROC) curves. The lowest points of the curves, on the left (with gray shading), show the correct- and false-identification rates at the highest level of confidence (100% certain); the second points on the curves show correct- and false-identification rates at the highest level of confidence (100% certain) and the second-highest level of confidence (i.e., 100% certain and 90% certain), and so forth. The points farthest to the right (circled in gray) show the rates for all subjects who made an identification. The star at the top of the *y*-axis represents perfect accuracy (i.e., hit rate = 1, false alarm rate = 0); the dashed line represents chance-level performance. To calculate the shaded area under the curve with the solid black circles, one would set the specificity (i.e., 1 – false alarm rate at the right-most edge of the shaded area) to .91.

diagnosticity ratio, determined by the HR-to-FAR ratio (Steblay, Dysart, & Wells, 2011). But instead of calculating a single diagnosticity ratio (one HR-FAR pair), several HR-FAR pairs are plotted over decreasing levels of confidence. Confidence serves as a proxy for willingness to choose; decreasing levels of confidence equate to more liberal responding (Wixted & Mickes, 2014). Therefore, by plotting these HR-FAR pairs over the full range of confidence, one can determine how different lineup types affect subjects' ability to distinguish between real culprits and innocent suspects, independently of their willingness to identify the suspect (Gronlund, Wixted, & Mickes, 2014; National Research Council, 2014).

Figure 2 displays this idea in a concrete way, depicting two hypothetical ROC curves. The left-most points of the curves show the HR and FAR at the highest level of confidence (100% certain); the second points on the curves show the HR and FAR at the highest level of confidence and the second-highest level of confidence (i.e., 100% certain and 90% certain), and so forth. The right-most points of the curves show the rates for all subjects who made an identification. A key idea is that for any point on the lower ROC, there is an achievable point on the higher ROC that is associated with both a higher HR and a lower FAR. Therefore, the ROC curve that falls closest to the upper left corner of the plot—closest to the star and farthest from the dashed chance line—is the objectively superior procedure because it maximizes identifications of culprits while minimizing identifications of innocent suspects. Put simply, this procedure allows witnesses to most accurately discriminate between culprits and innocent suspects.

To compare ROC curves, one uses the partial area under the curve (pAUC) because the FAR for innocent suspects is less than 1. In pAUC analysis, one defines the specificity (1 – FAR) for calculating the AUC. For example, if we were interested in calculating the shaded area under the curve with the solid black circles in Figure 2, we would define the specificity as .91 (i.e., 1 – .09). When comparing ROC curves, the specificity must be set to the same value in every pAUC calculation. Thus, in the current example, when calculating the area under the curve with the white circles, we would also set the specificity as .91.

Current analysis. Our ROC analysis measured people's ability to discriminate between guilty suspects and innocent suspects, setting aside choices of known-to-be innocent foils. This is different from an absolute notion of memory discriminability—which would be the ability to discriminate between guilty suspects and anyone else in the lineup (i.e., innocent suspects and foils; for a discussion, see Wixted & Mickes, 2015). From a practical standpoint, discriminating between guilty suspects and innocent suspects is arguably the key discriminability to measure because false identifications of foils do not result in any legal action against the foil that is selected. Nevertheless, our signal detection modeling accounts for foil choices and thus also estimates (a) people's ability to discriminate between guilty suspects and foils and (b) people's ability to discriminate between innocent suspects and foils in unfair lineups (for more information, see the Supplemental Material).

To calculate pAUC, we used the statistical package pROC (Version 1.8; Robin et al., 2011) with RStudio (Version 0.98.1103; RStudio Team, 2015) and the R software environment (Version 3.2.0; R Development Core Team, 2015); pROC also calculated a measure of effect size, D, using the formula: $D = (\text{AUC1} - \text{AUC2})/s$. In this formula, s is the standard error of the difference between the two AUCs and is estimated using bootstrapping. To construct our ROC curves, we collapsed the data across the four crime videos. We rounded subjects' confidence ratings (made on the 100-point Likert scale) to the nearest 10 so that each curve would have 11 operating points of decreasing confidence (i.e., 100, 90, 80, and so forth). We then calculated the HRs and FARs over the decreasing confidence levels. HR was calculated as the number of identifications of guilty suspects divided by the number of target-present lineups. FAR was calculated as the number of identifications of innocent suspects divided by number of target-absent lineups.

We calculated identifications of innocent suspects differently for the unfair and fair lineups. In the unfair (do-nothing) lineups, subjects made identifications of innocent suspects when they identified the single lineup member with the distinctive feature. In the fair (replication, pixelation, and block) lineups, recall that there was no designated innocent suspect—thus, we estimated the number of identifications of innocent suspects in these conditions using a common approach. We divided the number of false identifications made in target-absent lineups by the total number of people in the lineup—here, six (Brewer & Wells, 2006; Mickes, 2015). In general, this procedure works on the assumption that the lineup member that best matches the subject's memory of the culprit is the innocent suspect (Palmer, Brewer, Weber, & Nagesh, 2013). One particular benefit of estimating false identifications in this way is that it leads to a more conservative measure of false identifications. Because the innocent suspect may not always be the most similar in appearance to the actual culprit, this method of estimation can only overestimate, not underestimate, the number of false identifications in target-absent lineups. Thus, in the current study, using this estimation method in replication, pixelation, and block lineups provided a conservative test of how well these (fair) techniques enhanced witness identification performance compared with the (unfair) do-nothing lineups.

Figure 3 shows the ROC curves for the fair and unfair lineups. When calculating pAUC statistics, we set the specificity to 0.91—which corresponded to the FAR range (0–.09) covered by the least extensive curve (block)—for two main reasons. First, by setting the FAR range from 0 to .09, we prevented the pROC program from having to extrapolate the three fair-lineup curves over a vast range (FAR of .09–.40). The pROC program uses a crude method of extrapolation, so doing so over large ranges can reduce statistical accuracy. Second, the lower FAR range (0–.09) may have greater practical relevance, because the legal system (a) is interested in knowing which conditions increase witnesses' ability to distinguish between innocent and guilty suspects when the FAR is low, and (b) may take these high-confidence identifications more seriously than low-confidence identifications (see Gronlund et al., 2012). We are confident that limiting the pAUC analysis to a small subset of the do-nothing curve did not affect our findings. When we fit a signal detection process model of lineup performance to our data, we found the same pattern of results (see the Supplemental Material). This modeling technique uses the largest FAR range that a target-absent lineup can support.

To what extent did our lineup types affect witnesses' performance? More specifically, did the unfair lineups increase witnesses' willingness to identify the (guilty or innocent) suspect—or did those lineups impair witnesses' ability to distinguish between guilty and innocent suspects?

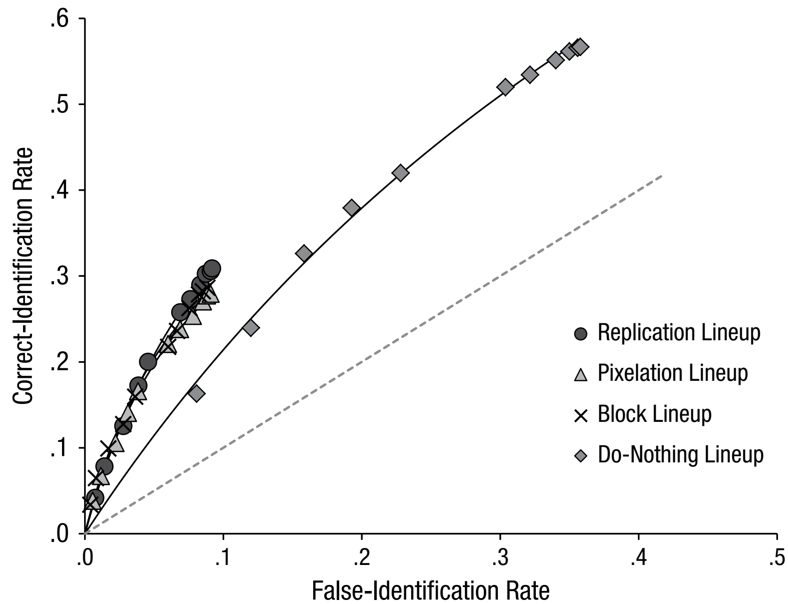

Fig. 3. Receiver-operating-characteristic curves for the fair (replication, pixelation, block) and unfair (do-nothing) lineups. For an explanation of how the curves relate to confidence ratings, see Fig. 2. The dashed line represents chance-level performance.

As Figure 3 shows, compared with the replication, pixelation, and block (i.e., fair) lineup techniques, doing nothing increased subjects' willingness to identify the suspect and also markedly impaired subjects' ability to discriminate between real culprits and innocent suspects. We can see that the do-nothing ROC points have shifted more to the right than any of the fair-lineup ROC points. This rightward shift shows that there was an increase in both correct and false identifications. That is, subjects' willingness to identify the suspect increased in the do-nothing lineups compared with replication, pixelation, and block lineups.

A more striking finding, though, was that do-nothing lineups made it more difficult for subjects to distinguish between innocent and guilty suspects. The pAUC for do-nothing lineups (0.008, 95% CI = [0.006, 0.010]) was significantly smaller than the pAUC for replication lineups (0.016, 95% CI = [0.013, 0.019]), $D = 4.11$, $p < .001$, block lineups (0.016, 95% CI = [0.013, 0.019]), $D = 4.35$, $p < .001$, and pixelation lineups (0.015, 95% CI = [0.012, 0.018]), $D = 4.17$, $p < .001$. Finally, the three fair lineups led to similar levels of identification performance—the pAUCs did not differ significantly between replication and block lineups ($D = 0.08$, $p > .250$), between replication and pixelation lineups ($D = 0.32$, $p > .250$), or between block and pixelation lineups ($D = 0.24$, $p > .250$). We also fit a signal detection process model of lineup performance to our data to further confirm these findings (see Lampinen, 2016; Wixted & Mickes, 2014). The modeling procedure and results are presented in the Supplemental Material. Note that the model-fitting exercise and our pAUC analysis led to the same results. Taken together, these findings fit with the additional prediction of the diagnostic-feature-detection model—that doing nothing to stop distinctive suspects from standing out not only increases witnesses' willingness to choose the suspect but also markedly impairs their ability to discriminate between guilty and innocent suspects.

Identification responses

To further understand the effect of unfair lineups on subjects' identification performance, we calculated the proportion of identifications of suspects, identifications of foils, and lineup rejections (i.e., "Not Present" responses) for each lineup type. Table 2 shows the frequencies and percentages of identification responses for each lineup type. Note that we knew from the ROC analysis that unfair lineups led to more identifications of both guilty and innocent suspects than did fair lineups. The data in Table 2 indicate that this overall increase in identifications of suspects was accompanied by a decrease in both identifications of foils and lineup rejections in target-present lineups but only by a decrease in identifications of foils in target-absent lineups.

Target-present lineups. A 4 (lineup type: replication, pixelation, block, do-nothing) × 3 (identification response: guilty suspect, foil, incorrect rejection) analysis showed

Table 2. Frequencies and Percentages of Identification Responses in the Replication, Pixelation, Block, and Do-Nothing Lineups

Target presence and identification response	Replication lineup		Pixelation lineup		Block lineup		Do-nothing lineup	
	n	%	*n*	%	*n*	%	*n*	%
Target present								
Guilty suspect identified	347.00	30.84	320.00	27.95	323.00	28.66	629.00	56.67
Foil identified	382.00	33.96	411.00	35.90	390.00	34.61	206.00	18.56
Incorrect rejection	396.00	35.20	414.00	36.16	414.00	36.73	275.00	24.77
Target absent								
Innocent suspect identified	104.50	9.17	102.33	9.10	100.50	8.84	364.00	35.79
Foil identified	522.50	45.83	511.67	45.52	502.50	44.20	219.00	21.53
Correct rejection	513.00	45.00	510.00	45.37	534.00	46.97	434.00	42.67

that lineup type influenced identification responses, $\chi^2(6, N = 4{,}507) = 282.70$, $p < .001$, Cramer's $V = .18$. Specifically, fair lineups led to fewer identifications of guilty suspects (replication: $z = -2.84$, $p < .01$; pixelation: $z = -4.50$, $p < .001$; block: $z = -4.07$, $p < .001$) but more identifications of foils (replication: $z = 1.90$, $p > .05$; pixelation: $z = 3.09$, $p < .01$; block: $z = 2.29$, $p < .05$) and more rejections (replication: $z = 1.13$, $p > .05$; pixelation: $z = 1.70$, $p > .05$; block: $z = 2.02$, $p < .05$), than expected. Conversely, unfair lineups led to more identifications of guilty suspects ($z = 11.53$, $p < .001$), but fewer identifications of foils ($z = -7.36$, $p < .001$) and fewer rejections ($z = -4.90$, $p < .001$), than expected. In short, when the suspect stood out in target-present lineups, there was an increase in identifications of guilty suspects along with a reduction in both identifications of foils and incorrect rejections.

Target-absent lineups. Recall that in replication, pixelation, and block target-absent lineups, there was no designated innocent suspect. We therefore estimated the number of identifications of innocent suspects by dividing the total number of false identifications by six (the number of faces in the lineup). Likewise, we estimated the number of identifications of foils by dividing the total number of false identifications by six (the number of faces in the lineup) and then multiplying by five (the number of faces that were not the innocent suspect). A 4 (lineup type: replication, pixelation, block, do-nothing) × 3 (identification response: innocent suspect, foil, correct rejection) analysis using these estimates showed that lineup technique influenced identification responses, $\chi^2(6, N = 4{,}418) = 481.70$, $p < .001$, Cramer's $V = .23$. Fair lineups led to fewer identifications of innocent suspects (replication: $z = -5.22$, $p < .001$; pixelation: $z = -5.24$, $p < .001$; block: $z = -5.50$, $p < .001$) but more identifications of foils (replication: $z = 3.26$, $p < .001$; pixelation: $z = 3.08$, $p < .001$; block: $z = 2.38$, $p < .001$), than expected. Conversely, unfair lineups led to more identifications of innocent suspects ($z = 16.85$, $p < .001$), but fewer identifications of foils ($z = -9.21$, $p < .001$), than expected. The proportion of correct rejections in all four lineup types was similar (replication: $z = -0.03$, $p > .05$; pixelation: $z = 0.15$, $p > .05$; block: $z = 0.95$, $p > .05$; do-nothing: $z = -1.14$, $p > .05$). This analysis indicates that when the suspect stood out in target-absent lineups, subjects shifted their identifications from the other lineup members onto the innocent suspect.

Confidence and accuracy

Recall that the diagnostic-feature-detection model suggests that unfair lineups impair a witness's ability to distinguish between innocent and guilty suspects because it is not obvious to the witness that the suspect's distinctive feature is unhelpful. If witnesses fail to realize that the distinctive feature is unhelpful, they may not lower their confidence judgment to compensate for their poorer performance. If this account is correct, then subjects who viewed the unfair do-nothing lineups should be less accurate, at every level of confidence, than subjects who viewed the fair replication, pixelation, and block lineups.

To test this prediction, we plotted suspect-identification accuracy (correct identifications of guilty suspects in target-present lineups divided by correct identifications of guilty suspects in target-present lineups plus false identifications of innocent suspects in target-absent lineups) separately for each level of confidence (100, 90, 80, and so forth, according to the method of Mickes, 2015). This method of calculating suspect-identification accuracy reflects the probability of guilt, given that the suspect was identified (i.e., the posterior probability of guilt). We estimated the number of identifications of innocent suspects in the replication, block, and pixelation lineups in the same way that we did for the ROC analysis. To provide more stable estimates, confidence level was binned into five categories (0–20, 30–40, 50–60, 70–80, 90–100; see

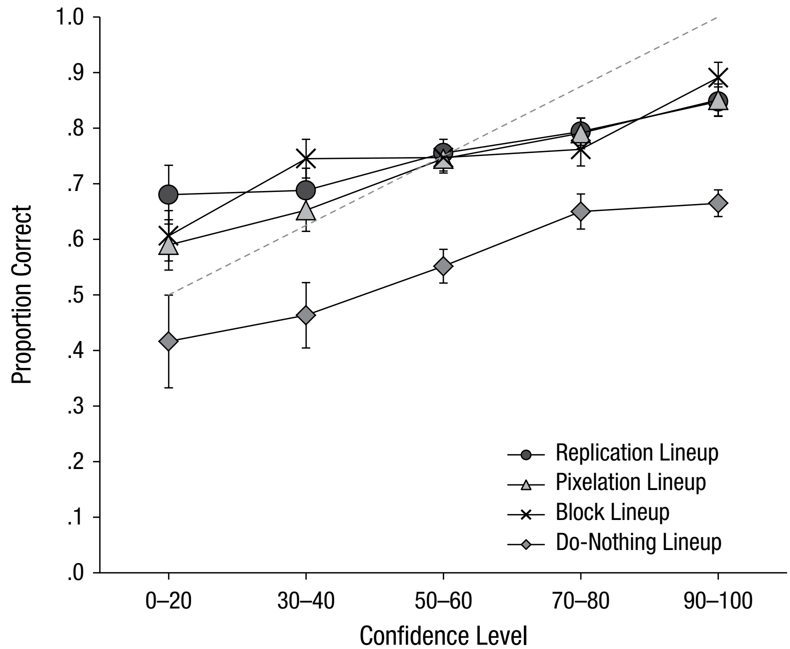

Fig. 4. Confidence-accuracy curves for the fair (replication, pixelation, block) and unfair (do-nothing) lineups. Error bars indicate ±1 *SE*. The dashed diagonal line represents chance accuracy at the lowest confidence bin (i.e., 0–20) and perfect accuracy at the highest confidence bin (i.e., 90–100).

Brewer & Wells, 2006). The frequencies of identification responses in each confidence bin in replication, pixelation, block, and do-nothing lineups are shown in Table S1 in the Supplemental Material available online.

Figure 4 shows the confidence-accuracy curves for each lineup type. Nonoverlapping error bars denote reliable differences between the lineup techniques (e.g., Sauer, Brewer, Zweck, & Weber, 2010). As predicted, subjects who viewed the unfair, do-nothing lineups showed lower levels of accuracy at every level of confidence than subjects who viewed the fair lineups. Put another way, an identification made at any level of confidence from an unfair lineup was less trustworthy than an identification made with the same level of confidence from a fair lineup. These data align with the diagnostic-feature-detection model, which suggests that when nothing was done to stop the distinctive suspect from standing out, subjects may have been unaware that their memory accuracy was worse than if the subject did not stand out and therefore failed to adjust their confidence accordingly.

Discussion

In this study, we explored why unfair lineups promote mistaken identifications. Our findings suggest that unfair lineups—compared with fair lineups—make people more likely to identify the suspect; worse still, unfair lineups impair people's ability to distinguish between guilty and innocent suspects and distort people's ability to judge the trustworthiness of their identification decision.

It is arguably unsurprising that our unfair lineups, in which the suspect stood out, increased subjects' willingness to identify that suspect. Many eyewitness identification studies have demonstrated this already (Clark, 2012; Doob & Kirshenbaum, 1973; Fitzgerald et al., 2013; Wells, Leippe, & Ostrom, 1979; Wells et al., 1993). The fascinating finding is that unfair lineups also dramatically hindered subjects' ability to sort innocent and guilty suspects into their appropriate categories. This mechanism has not been discussed until now, yet it is important. Procedures that simply make witnesses less willing to choose the police suspect decrease identifications of innocent suspects but also come at a cost: They stifle identifications of culprits (Clark, 2012). Procedures that enhance a witness's ability to distinguish between innocent and guilty suspects minimize identifications of innocent suspects and maximize identifications of culprits, regardless of the witness's willingness to choose. Arguably then, this is the critical mechanism to investigate (Gronlund et al., 2014; National Research Council, 2014).

So, why might unfair lineups harm people's ability to distinguish between the real culprit and an innocent suspect? One explanation is that witnesses fail to appreciate that the suspect's distinctive feature is not useful in an unfair lineup, and so they rely heavily on it to make their identification. By contrast, when lineups are fair and the suspect does not stand out, witnesses can appropriately discount the distinctive feature and give more weight to other, more informative cues (Wixted & Mickes, 2014). Support for this theoretical account comes from the finding that, in the unfair lineups, subjects failed to compensate by setting a more conservative confidence criterion when making an identification. This fits with a mechanism in which subjects do not realize that their accuracy is impaired.

A growing body of research suggests that subjects acting as witnesses in studies are generally good at judging the likely accuracy of their memories even when their memory accuracy is impaired (e.g., Brewer & Wells, 2006; Mickes, 2015, Experiment 1; Palmer et al., 2013; Sauer et al., 2010). Palmer et al., for instance, showed that divided attention significantly impaired people's memory ability, yet when the authors plotted accuracy at each level of confidence, it did not matter whether subjects had full or divided attention at encoding—their accuracy at each level of confidence was generally the same (Palmer et al., 2013, Experiment 2, Figs. 3 and 4). Palmer et al. concluded that their experimental manipulations did not undermine the usefulness of confidence as an indicator of accuracy. This study, along with many others, shows that people typically recognize when their memories are poor and adjust their confidence appropriately (Mickes, 2015, Experiment 1; Palmer et al., 2013, Experiment 1; Sauer et al., 2010). There are some instances, however, in which confidence is uninformative of accuracy (e.g., Chandler, 1994; Mickes, 2015, Experiment 2). Indeed, our findings show that unfair lineups can systematically distort confidence.

One consequence of the finding that identifications from unfair lineups were less accurate at every level of confidence is that subjects in the do-nothing condition made high-confidence identifications of suspects (certainty of 90–100) when accuracy was moderate (.6). This finding has serious implications for criminal justice because legal decision makers are strongly influenced by highly confident witnesses (Brewer & Burke, 2002; Wells, Lindsay, & Ferguson, 1979). Although subjects in the fair-lineup conditions (i.e., replication, pixelation, or block) were more accurate at the lower end of the confidence scale than their confidence level would indicate (i.e., they were underconfident), the critical point is that their identifications were consistently and substantially more trustworthy than the identifications made by subjects in the unfair-lineup condition. Moreover, subjects who viewed the fair lineups identified the suspect with high confidence (certainty of 90–100) only when they were very likely to be accurate (> .8). Therefore, highly confident identifications of suspects made from replication, pixelation, and block lineups are likely to be very informative for triers of fact (for further discussion on subjects' confidence ratings, see the Supplemental Material).

Two face-recognition studies have suggested that replicating distinctive features is better than removing them (Badham, Wade, Watts, Woods, & Maylor, 2013; Zarkadi et al., 2009); at first glance, our findings appear to conflict with these results. Zarkadi et al., for example, found that replication increased correct identifications by approximately 20% in target-present lineups, whereas we found that replication and concealment techniques were equally effective. There is, however, a crucial methodological difference to consider. The previous research compared replication lineups with removal lineups (i.e., those in which the target's distinctive feature was simply removed). Subjects made more incorrect rejections in target-present removal lineups, possibly because the person they believed to be the culprit was now missing a prominent distinctive feature that they remembered (Wixted & Mickes, 2014). Subjects in our study were unlikely to use this strategy because we tested pixelation and block lineups, both of which indicate the possibility of a distinctive feature underneath the concealed area. Therefore, in our findings, unlike those of the previous research, we did not observe a relatively high number of incorrect rejections in pixelation and block lineups compared with replication lineups. Instead, we observed similar performance in all three fair conditions.

On a practical level, our research suggests that law enforcement officers should take steps to prevent distinctive suspects from standing out. If unfair lineups only increased witnesses' willingness to choose the suspect (and did not affect their ability to distinguish between innocent and guilty suspects), then officers could remedy this by inducing more conservative responding. For instance, urging witnesses to be cautious (e.g., "Be certain before making a decision") should increase the amount of memory information that witnesses demand before choosing and result in fewer positive identifications and therefore fewer identifications of suspects (Clark, 2005). Our data, however, suggest that law enforcement officers need to apply fair lineup techniques to improve identification accuracy, and that replication, pixelation, or block techniques are equally effective.

In sum, our data fit the predictions of a new model: the diagnostic-feature-detection model. Testing theoretical models is important; once refined, theories can be used to develop procedures that further enhance eyewitness accuracy. More specifically, our findings shed light on the processes underlying the harmful effects of unfair lineups and suggest that when suspects are unduly distinctive, witnesses are not only more willing to choose the suspect but also struggle to distinguish between guilty and innocent suspects. Perhaps if Leonard Callace had been placed in a fair lineup, alongside foils who also had full beards or whose chins had been concealed, he would not have spent 6 years in prison for a crime he did not commit.

Action Editor

Colleen M. Kelley served as action editor for this article.

Author Contributions

K. A. Wade and M. F. Colloff developed the study concept and design. Data collection was performed by all the authors, and M. F. Colloff performed the data analysis and interpretation under the supervision of K. A. Wade. M. F. Colloff and K. A. Wade drafted the manuscript, and D. Strange provided critical revisions. All the authors approved the final version of the manuscript for submission.

Acknowledgments

We thank John Wixted and Laura Mickes for their insightful discussions and indispensable practical advice. We also thank Neil Stewart for valuable programming assistance and Maryanne Garry and Elizabeth Maylor for their helpful comments on an earlier draft.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Supplemental Material

Additional supporting information can be found at http://pss.sagepub.com/content/by/supplemental-data

Open Practices

All data have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/axphc/. The complete Open Practices Disclosure for this article can be found at http://pss.sagepub.com/content/by/supplemental-data. This article has received the badge for Open Data. More information about the Open Practices badges can be found at https://osf.io/tvyxz/wiki/1.%20View%20the%20Badges/ and http://pss.sagepub.com/content/25/1/3.full.

References

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. doi:10.3758/s13423-012-0339-2

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism Spectrum Quotient (AQ): Evidence from Asperger syndrome/high functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5–17. doi:10.1023/A:1005653411471

Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Behavior*, *26*, 353–364. doi:10.1023/A:1015380522722

Brewer, N., Keast, A., & Sauer, J. D. (2010). Children's eyewitness identification performance: Effects of a *Not Sure* response option and accuracy motivation. *Legal and Criminological Psychology*, *15*, 261–277. doi:10.1348/135532509X474822

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. doi:10.1037/1076-898X.12.1.11

Carlson, C. A., Gronlund, S. D., & Clark, S. E. (2008). Lineup composition, suspect position, and the sequential lineup advantage. *Journal of Experimental Psychology: Applied*, *14*, 118–128. doi:10.1037/1076-898X.14.2.118

Chandler, C. C. (1994). Studying related pictures can reduce accuracy, but increase confidence, in a modified recognition test. *Memory & Cognition*, *22*, 273–280. doi:10.3758/BF03200854

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior*, *29*, 575–604. doi:10.1007/s10979-005-7121-1

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

Doob, A. N., & Kirshenbaum, H. M. (1973). Bias in police lineups—partial remembering. *Journal of Police Science and Administration*, *1*, 287–293.

Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. D. (2013). The effect of suspect-filler similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law*, *19*, 151–164. doi:10.1037/a0030618

Flowe, H. D., Klatt, T., & Colloff, M. F. (2014). Selecting fillers on emotional appearance improves lineup identification accuracy. *Law and Human Behavior*, *38*, 509–519. doi:10.1037/lhb0000101

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*, 504–528. doi:10.1016/S0092-6566(03)00046-1

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., Wixted, J. T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using receiver operating characteristic analysis. *Current Directions in Psychological Science*, *23*, 3–10. doi:10.1177/0963721413498891

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. doi:10.1080/1068316X.2015.1054387

Lampinen, J. M. (2016). ROC analyses in eyewitness identification research. *Journal of Applied Research in Memory and Cognition*. Advance online publication. doi: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. doi:10.1111/1467-9280.00041

Lindsay, R. C. L., Wallbridge, H., & Drennan, D. (1987). Do the clothes make the man? An exploration of the effect of lineup attire on eyewitness identification accuracy. *Canadian Journal of Behavioural Science*, *19*, 463–478. Abstract retrieved from https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=116390

MacLin, M. K., MacLin, O. H., & Albrechtsen, J. S. (2006). Using image manipulation to construct fair lineups: The case of the Buddy Holly glasses. *Canadian Journal of Police and Security Services*, *4*, 1–16.

Mansour, J. K., Beaudry, J. L., Kalmet, N., Bertrand, M. I., & Lindsay, R. C. L. (in press). Evaluating lineup fairness: Variations across methods and measures. *Law and Human Behavior*.

Marteau, T., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State–Trait Anxiety Inventory (STAI). *British Journal of Psychology*, *31*, 301–306. doi:10.1111/j.2044-8260.1992.tb00997.x

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 and Cognition*, *4*, 93–102. doi:10.1016/j.jarmac.2015.01.003

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

Palmer, M. A., 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. doi:10.1037/a0031602

Police and Criminal Evidence Act 1984, Code D. (2011). Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/253831/pace-code-d-2011.pdf

R Development Core Team. (2015). R: A language and environment for statistical computing (Version 3.2.0) [Computer software]. Retrieved from https://www.r-project.org/index.html

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. doi:10.1186/1471-2105-12-77

RStudio Team. (2015). RStudio: Integrated development for R (Version 0.98.1103) [Computer software]. Retrieved from http://www.rstudio.com/

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. doi:10.1007/s10979-009-9192-x

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law*, *17*, 99–139. doi:10.1037/a0021650

Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement*. Washington, DC: U.S. Department of Justice, Office of Justice Programs. Retrieved from https://www.ncjrs.gov/pdffiles1/nij/178240.pdf

Valentine, T., & Heaton, P. (1999). An evaluation of the fairness of police line-ups and video identifications. *Applied Cognitive Psychology*, *13*, S59–S72. doi:10.1002/(SICI)1099-0720(199911)13:1+<S59::AID-ACP679>3.0.CO;2-Y

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology*, *14*, 89–103. doi:10.1111/j.1559-1816.1984.tb02223.x

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, *48*, 553–571. doi:10.1037/0003-066X.48.5.577

Wells, G. L., Leippe, M. R., & Ostrom, T. M. (1979). Guidelines for empirically assessing the fairness of a lineup. *Law and Human Behavior*, *3*, 285–293. doi:10.1007/BF01039807

Wells, G. L., Lindsay, R. C., & Ferguson, T. J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification. *Journal of Applied Psychology*, *64*, 440–448. doi:10.1037/0021-9010.64.4.440

Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). The selection of distractors for eyewitness lineups. *Journal of Applied Psychology*, *78*, 835–844. doi:10.1037/0021-9010.78.5.835

Wetmore, S. A., Neuschatz, J. S., Gronlund, S. D., Wooten, A., Goodsell, C. A., & Carlson, C. A. (2015). Effect of retention interval on showup and lineup performance. *Journal of Applied Research in Memory and Cognition*, *4*, 8–14. doi:10.1016/j.jarmac.2014.07.003

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

Wixted, J. T., & Mickes, L. (2015). ROC analysis measures objective discriminability for any eyewitness identification procedure. *Journal of Applied Research in Memory and Cognition*, *4*, 329–334. doi:10.1016/j.jarmac.2015.08.007

Wogalter, M. S., Malpass, R. S., & Mcquiston, D. E. (2004). A national survey of US police on preparation and conduct of identification lineups. *Psychology, Crime & Law*, *10*, 69–82. doi:10.1080/10683160410001641873

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