American Psychologist

© 2021 American Psychological Association
ISSN: 0003-066X

2022, Vol. 77, No. 2, 196–220
https://doi.org/10.1037/amp0000832

# Active Exploration of Faces in Police Lineups Increases Discrimination Accuracy

Melissa F. Colloff[1], Heather D. Flowe[1], Harriet M. J. Smith[2], Travis M. Seale-Carlisle[3], Christian A. Meissner[4], James C. Rockey[5], Babita Pande[6], Pratibha Kujur[6], Noorshama Parveen[6], Priyanka Chandel[6], Margaret M. Singh[6], Sraddha Pradhan[6], and Arti Parganiha[6]

[1] Centre for Applied Psychology, School of Psychology, University of Birmingham
[2] Division of Psychology, Nottingham Trent University
[3] Wilson Center for Science and Justice, School of Law, Duke University
[4] Department of Psychology, Iowa State University
[5] Department of Economics, Birmingham Business School, University of Birmingham
[6] School of Studies in Life Sciences, Pt. Ravishankar Shukla University

Eyewitness identifications play a key role in the justice system, but eyewitnesses can make errors, often with profound consequences. We used findings from basic science and innovative technologies to develop and test whether a novel interactive lineup procedure, wherein witnesses can rotate and dynamically view the lineup faces from different angles, improves witness discrimination accuracy compared with a widely used procedure in laboratories and police forces around the world—the static frontal-pose photo lineup. No novel procedure has previously been shown to improve witness discrimination accuracy. In Experiment 1, participants ($N$ = 220) identified culprits from sequentially presented interactive lineups or static frontal-pose photo lineups. In Experiment 2, participants ($N$ = 8,507) identified culprits from interactive lineups that were either presented sequentially, simultaneously wherein the faces could be moved independently, or simultaneously wherein the faces moved jointly into the same angle. Sequential interactive lineups enhanced witness discrimination accuracy compared with static photo lineups, and simultaneous interactive lineups enhanced witness discrimination accuracy compared with sequential interactive lineups. These finding were true both when participants viewed suspects who were of the same or different ethnicity/race as themselves. Our findings exemplify how basic science can be used to address the important applied policy issue on how best to conduct a police lineup and reduce eyewitness errors.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This article was published Online First November 18, 2021.
Melissa F. Colloff https://orcid.org/0000-0001-6401-4872
Heather D. Flowe https://orcid.org/0000-0001-5343-5313
Harriet M. J. Smith https://orcid.org/0000-0003-2712-5527
Travis M. Seale-Carlisle https://orcid.org/0000-0002-4522-8549
Christian A. Meissner https://orcid.org/0000-0002-6094-5167
James C. Rockey https://orcid.org/0000-0002-2313-5544
Babita Pande https://orcid.org/0000-0002-0545-6002
Noorshama Parveen https://orcid.org/0000-0002-5366-5052
Arti Parganiha https://orcid.org/0000-0001-9764-5566

This work was supported by the Laura and John Arnold Foundation Grant (to Heather D. Flowe and Christian A. Meissner). Sections of these data were presented by Melissa F. Colloff at the International Meeting of the Psychonomic Society (May 2018), Amsterdam, the Netherlands, and at the Society for Applied Research in Memory and Cognition (June 2019), Cape Cod, Massachusetts, United States. Our data are available at https://osf.io/2x5tg/ (Experiment 1) and https://osf.io/b8tvw/ (Experiment 2).

Melissa F. Colloff served as lead for data curation, formal analysis, resources and writing – original draft, contributed equally to writing – review and editing and served in a supporting role for methodology. Heather D. Flowe served as lead for conceptualization, funding acquisition and methodology, contributed equally to writing – review and editing and served in a supporting role for writing – original draft. Harriet M. J. Smith contributed equally to software and served in a supporting role for writing – review and editing. Travis M. Seale-Carlisle served in a supporting role for writing – review and editing. Christian A. Meissner contributed equally to writing – review and editing and served in a supporting role for methodology. James C. Rockey served in a supporting role for writing – review and editing. Babita Pande served in a supporting role for writing – review and editing. Pratibha Kujur served in a supporting role for writing – review and editing. Noorshama Parveen served in a supporting role for writing – review and editing. Priyanka Chandel served in a supporting role for writing – review and editing. Margaret M. Singh served in a supporting role for writing – review and editing. Sraddha Pradhan served in a supporting role for writing – review and editing. Arti Parganiha served in a supporting role for writing – review and editing. Babita Pande, Pratibha Kujur, Noorsharma Parveen, Priyanka Chandel, Margarent Singh, Sraddha Pradhan, and Arti Parganiha all also served in a supporting role for data curation.

Correspondence concerning this article should be addressed to Melissa F. Colloff, Centre for Applied Psychology, School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom. Email: m.colloff@bham.ac.uk

***Public Significance Statement***
Compared with the photographic identification parade procedure used routinely by police forces worldwide, this article suggests that a novel interactive procedure can increase witness identifications of guilty suspects, while simultaneously minimizing witness identifications of innocent suspects, for both own-race and other-race identifications. This highlights that an interactive procedure, developed using psychological theory and technology, could reduce the significant societal, economic, and psychological costs associated with eyewitness misidentifications.

*Keywords:* eyewitness identification, interactivity, simultaneous lineup, sequential lineup, own race bias

*Supplemental materials:* https://doi.org/10.1037/amp0000832.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Humans can make errors when they attempt to recognize people whom they had previously seen only briefly. In the legal system, such errors have dangerous ramifications. The Innocence Project reports that eyewitness misidentification is the single greatest cause of wrongful conviction in the United States. Since 1989, 365 wrongful convictions have been overturned in the United States on the basis of new DNA evidence; reliance on inaccurate eyewitness testimony has played a role in securing convictions in a large proportion of these cases. Moreover, errors can be compounded further by other factors. Errors are more likely when people try to recognize those of different racial or ethnic groups ("races"), a scientifically robust phenomenon known as Own Race Bias (ORB; Malpass & Kravitz, 1969).[1] Other-race misidentification, in particular, was present in 42% of the wrongful conviction cases. The consequences of wrongful convictions are wide-reaching: those 365 innocent people spent over 5,000 years in prison, guilty culprits were free to commit at least 152 additional violent crimes including sexual assault and murder, and it cost U.S. taxpayers hundreds of millions of dollars (Innocence Project, 2019; Silbert et al., 2015). Because of the importance of eyewitness testimony, the US National Research Council (Policy and Affairs) convened an expert working group and issued a report that reviewed the state of the science (National Research Council, 2014). The report identified ORB as a leading factor in eyewitness misidentification, and called on researchers to develop innovative technologies to improve eyewitness identification accuracy from police lineups. In this article, we heed this call by testing whether a novel interactive lineup procedure, wherein witnesses can rotate and dynamically view the lineup faces, could improve witness discrimination accuracy and also attenuate the ORB.

In a police lineup, the suspect, who may be guilty or innocent, is presented with a number of known-to-be innocent people who physically resemble the suspect, called fillers. In many countries (e.g., United States, Germany, Canada, and Australia), witnesses are presented with one photograph of each lineup member facing the camera, in frontal pose, from the shoulders up (Fitzgerald et al., 2018, 2021). The photos of the lineup members are either presented one at a time (sequentially), or all at once (simultaneously). Other countries use different lineup mediums. In some countries, the images of the lineup members are presented sequentially via video, with each lineup member turning their head to the left and the right, before the next lineup member is shown (e.g., England, Wales, and Scotland). In others, lineup members are presented to the witness in person. Regardless of the lineup medium (photo, video, or live), eyewitnesses can make one of six types of responses. When the real culprit is in the lineup, an eyewitness can correctly identify the culprit, incorrectly identify a filler, or incorrectly say "Not Present." When the real culprit is not in the lineup, they can correctly say "Not Present," incorrectly identify the innocent suspect, or incorrectly identify a filler. Police officers can never be certain that the suspect in the lineup is the real culprit, which makes investigating eyewitness accuracy in the real-world difficult (Horry et al., 2014). For decades, psychological scientists have instead conducted controlled laboratory experiments. Participants in these experiments typically watch a video of a simulated crime and then attempt to recognize the culprit from a lineup. Half of the participants see a lineup that contains the culprit (a target-present lineup), the other half see a lineup that does not (a target-absent lineup). Participants acting as witnesses in these experiments make the wrong identification decision from a photo lineup around half of the time (e.g., Steblay et al., 2011). Moreover, research comparing

[1] Here, we use the term "race" as is commonly used in the psychological literature, to refer to differences in physical appearance that distinguish one group of people from another (e.g., Sporer, 2001). However, it is important to note that biologists have long highlighted how poorly the term race is understood, and questioned whether race refers to meaningful biological categories (e.g., Lewontin, 1972). We refrain from using the term "ethnicity" as an alternative, because ethnicity is concerned with groups defined by a common identity-based ancestry, language, or culture, such as religion or beliefs (e.g., Cornell & Hartmann, 2007).



**Melissa F. Colloff**

the accuracy of participants' identification decisions from photo lineups to video or live lineups has found mixed results, and there is not yet compelling evidence for the benefit of one lineup medium over another (e.g., Clark et al., 2015; Cutler & Fisher, 1990; Rubínová et al., 2021; Seale-Carlisle & Mickes, 2016; Seale-Carlisle et al., 2019; Valentine et al., 2007; for reviews, see Cutler et al., 1994; Fitzgerald et al., 2018).

Despite decades of research, successful attempts to develop innovative procedures to improve eyewitness accuracy based on psychological theory are scant. Proposed lineup procedures often render participants less likely to choose the suspect, and have little effect on people's ability to discriminate between innocent and guilty suspects (e.g., Clark, 2012; Meissner, Tredoux, et al., 2005; Palmer & Brewer, 2012). Put another way, proposed procedures often elicit a more conservative *response bias*, but have not been shown to enhance *discrimination accuracy*. A procedure that elicits a more conservative response bias protects the innocent (by reducing the false identification rate), but this comes at the cost of protecting the guilty (by also reducing the correct identification rate). Implementing lineup procedures that enhance discrimination accuracy should be the goal of policymakers, because procedures that enhance discrimination accuracy can minimize the likelihood of incorrect innocent suspect identifications, while also maximizing the likelihood of correct guilty suspect identifications (Clark, 2012; Gronlund et al., 2015; National Research Council, 2014). To date, no lineup procedure has been shown to improve adult witness discrimination accuracy more than a standard simultaneous frontal pose photo lineup (e.g., Seale-Carlisle et al., 2019).

## Interactive Lineups

We developed and tested whether an interactive lineup improved witness discrimination accuracy compared with a photo lineup. In an interactive lineup, each face can be rotated and held in any position along the vertical axis, from left-profile to right-profile, allowing the witness to dynamically view each face at different orientations (see https://tinyurl.com/t4nc9gp). There are good reasons to predict that an interactive lineup will enhance discrimination accuracy compared with a standard frontal photo lineup. First, face recognition accuracy is improved by allowing observers to actively explore test faces compared with passively viewing movements initiated by another observer. *Active exploration* involves intentional sampling of information, so important facial features can be viewed (Liu et al., 2007). Similarly, recent perspectives in the identification literature hypothesize that discriminability is enhanced by procedures that enable a witness to detect and discount facial features that are shared by all lineup members, so features that are unique to the culprit (i.e., diagnostic features) can be used to identify the culprit (Wixted & Mickes, 2014; Wixted et al., 2018). Theoretically, the opportunity to view and detect more diagnostic or nondiagnostic facial features via active exploration of the faces in an interactive lineup should improve discrimination accuracy compared with passively viewing the front of the faces in a photo lineup.

Second, face recognition accuracy is viewpoint dependent, meaning that accuracy is better when a test face is presented in the same orientation in which it was studied (e.g., Bruce, 1982; Carbon & Leder, 2006). Recently, Colloff et al. (2021) found that participants had better discrimination accuracy in a lineup identification task when they saw or rotated the lineup faces into the same orientation in which they had viewed the culprit in a mock crime video, compared with when they saw the lineup faces in a different orientation. Therefore, the opportunity to rotate lineup faces into the same pose that the witness encoded the culprit—*pose-reinstatement*—during an interactive lineup should improve discrimination accuracy compared with viewing only a frontal pose of faces in a photo lineup.

Third, the *representation enhancement hypothesis* states that facial movement contributes to recognition by facilitating the perception of the three-dimensional structure of a face (O'Toole et al., 2002). Knowing the three-dimensional structure of a face can mitigate viewpoint dependence (Hill et al., 1997; Longmore et al., 2008). Therefore, *motion cues* elicited when witnesses explore the faces during an interactive lineup should improve discrimination accuracy compared with viewing static faces in a photo lineup. In short, psychological theory predicts that interactive lineups will improve eyewitness discrimination accuracy compared with a standard photo lineup procedure for both own- and other-race faces. An outstanding

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.



**Heather D. Flowe**

secondary question is whether the ORB might be attenuated with interactive lineups. We turn to that question, next.

### Own Race Bias

It is well-established that people are less able to accurately perceive and remember faces that are of a different race or group than their own. A meta-analysis found that identifications of other-race compared with own-race faces are 56% more likely to be erroneous (Meissner & Brigham, 2001). While most ORB research has been conducted with Black and White participants and faces, the effect is ubiquitous: it has been found comparing different races and groups around the world (e.g., O'Toole et al., 1994; Sporer et al., 2007). Considering eyewitness identifications, archival studies indicate that about half of witness identifications are other-race (e.g., Behrman & Davey, 2001; Flowe et al., 2018), and experimental studies document the ubiquity of the ORB in photo lineup tasks (e.g., Jackiw et al., 2008; Platz & Hosch, 1988). A number of jurisdictions now recommend that trial judges warn juries about the effect of the ORB on identification inaccuracy (e.g., *People v. Boone*, 2017).

Despite the prevalence of the ORB and the significant potential for miscarriages of justice, there remains considerable debate regarding a theoretical explanation. Causal mechanisms have been attributed to both cognitive and social processes at *encoding*, when the face is learned (e.g., see Hugenberg et al., 2010; Marcon et al., 2009; Meissner, Brigham, & Butz, 2005; Sporer, 2001). Yet, it is currently unknown how information at retrieval influences other-race face recognition. Few studies have examined retrieval, and the results are mixed (Bornstein et al., 2013; Evans et al., 2009; Young et al., 2010). Consistent with an encoding-based account, one study found that only own-race faces, not other-race faces, benefited from representing contextual information, such as a name, at retrieval during a lineup task (Evans et al., 2009); likely attributable to the qualitative encoding of context and other recollective information for own-race faces (Meissner, Brigham, & Butz, 2005; Marcon et al., 2009). Instructions informing witnesses about the challenges of the ORB for identification accuracy are only effective in ameliorating the bias when presented before encoding (Young et al., 2010); not at the time of identification (Bornstein et al., 2013). In experimental research, and in multiple criminal jurisdictions around the world, lineup members are shown facing the camera in a frontal pose. Consequently, our ability to understand the role of information at retrieval in making other-race identifications has been limited by the materials used in previous studies and also the lineup procedures often used in practice. If observers are already relatively proficient at discriminating between faces of their own-race (i.e., near the ceiling of human performance), it is possible that interactive lineups might offer the most improvement to other-race accuracy and therein attenuate the ORB.

In summary, our primary goal was to test whether interactive lineups compared with frontal pose photo lineups enhance discrimination accuracy for both own- and other-race faces. Our secondary goal was to test whether discrimination accuracy on interactive lineups would be particularly enhanced for other-race identifications, since retrieval support and the ORB has not been sufficiently investigated. If interactive lineups improve discriminability, this illustrates that procedures grounded in basic science can improve witness accuracy more than a current procedure used for testing witness memory. If interactive lineups attenuate the ORB, this suggests that encoding-based theories of ORB need to consider factors at retrieval. To test these hypotheses, we used analytical techniques that have been used for decades in basic science, but have only been applied to the eyewitness identification field relatively recently (e.g., Mickes et al., 2012; Wixted & Mickes, 2014).

## Experiment 1

### Method

We preregistered our experiment and analyses (https://osf.io/2x5tg/).

#### *Design*

We used a 2 (Culprit Race: White, South Asian) × 2 (Participant Race: White, South Asian) × 2 (Lineup Procedure: Interactive, Static) × 2 (Target: Present, Absent) mixed

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.



**Harriet M. J. Smith**

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

design.[2] Participant race and lineup procedure were between-participants factors. Culprit race and lineup procedure were within-participant factors; each participant watched eight mock-crime videos (four White and four South Asian culprits) and completed eight lineup tasks (four target-present and four target-absent). We aimed to recruit 110 White and 110 South Asian participants. Receiver Operating Characteristic (ROC) analysis has been used in basic scientific work for decades (e.g., Wixted, 2020; Woodward, 1953) and can be used in lineup experiments to measure the ability of witnesses to discriminate innocent from guilty suspects. Methods for calculating an a priori power analysis for ROC lineup experiments are not well specified. Using the mean difference and standard deviations observed in Dobolyi and Dodson (2013) as a guide, a power analysis indicated that, with 220 participants, power for pairwise comparisons would exceed 80%. The research was reviewed by University of Birmingham Humanities and Social Sciences Ethical Review Committee.

### *Participants*

The participants were 224 students and visitors recruited from the University of Birmingham. Each participant received £7.50, and those who were Psychology students ($n$ = 116) also received course credit. We excluded four participants (2% in total) who reported that they were not White or South Asian. This resulted in a final sample of 220 participants: 110 White, 110 South Asian (age: 18–61 years, $M$ = 21.72, $SD$ = 6.42; sex: 171 female, 48 male, one prefer not to say). We did not advertise that we were recruiting White and South Asian participants. We called the study "Perception and Memory," and advertised it on the University's participant recruitment system to students who had previously self-defined as White or South Asian. We also advertised the study using posters and online advertisements, and potential participants completed a prescreen questionnaire including filler questions and one critical question about their race/ethnicity. Those who self-identified as White or South Asian were invited to participate; those who did not were invited to participate in an unrelated study. Participants completed the experiment in the lab. Participants demonstrated that they attended to the task by answering six attention check questions (91% correctly answered six questions, 9% correctly answered five questions, and one participant correctly answered three questions).

### *Materials*

**Mock-crime Videos.** We developed eight 30 s mock-crime videos, by crossing four crime scenarios (burglary, drink-spiking, laptop theft, and theft from a car) with two culprit races (White and South Asian). In each video, the crime was committed by a single culprit (i.e., eight culprits in total). To create the videos, we filmed each crime scenario committed by one culprit race (either White or South Asian), then we filmed the exact same scenario committed by a culprit of the other race (South Asian or White). The four crime scenarios (burglary, drink-spiking, laptop theft, and theft from a car) differed in a number of ways, such as the location (e.g., house, kitchen, university, and car), distance from the culprit's face, exposure duration of the culprit's face (between 10 and 17 s across the four crime scenarios), sex of the culprit (three male, one female, for each race), and presence of bystanders. Our aim was to capture variability in encoding and test conditions to ensure that any detected effects were reliable and generalizable (e.g., Brewer et al., 2010). In the scenarios, the culprits turned to reveal the left- and right-profile of their face, but were predominantly seen from the front.

**Lineups.** Using a digital camera, we photographed each of the eight culprits. For the lineup fillers, images were captured of over 150 other people. In interactive lineups, faces could be moved from left- to right-profile using a computer mouse (see Figure 1B). To this end, we captured each person's image from multiple angles using video (e.g., moving their head from left-profile, to front, to right-profile) and used our code-based Interactive program to render individual frames from the video into a smooth rotatable object. In photo lineups, faces were shown using the static frontal image from the interactive lineups (see Figure 1A). Each lineup member (for interactive and photo lineups) was photographed using the same camera, background, and focal distance, wearing a cape to cover clothing, and without distinctive nonfacial features (e.g.,

[2] Despite being a poorly defined concept, we use the term "race" to broadly distinguish between the two groups tested (White and South Asian), because within each group there were also variations in self-defined ethnicity (e.g., White British, White American, Indian, Pakistani, etc., see online supplemental materials). Therefore, note that the participants and stimuli are likely to be morphologically diverse, even within the White and South Asian groups.



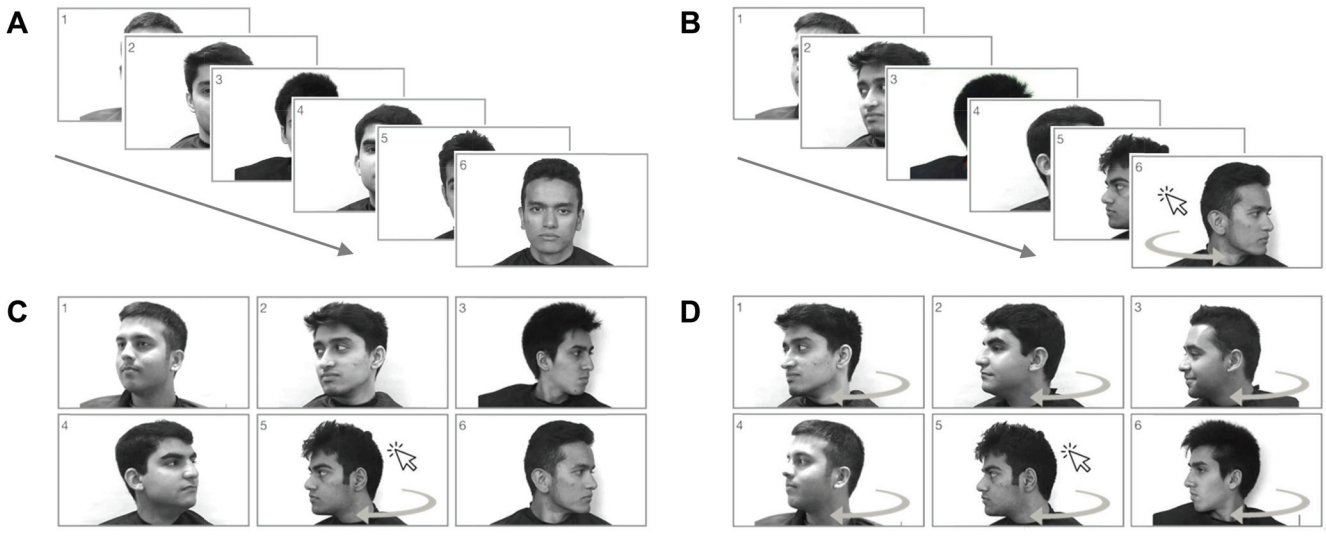
**Travis M. Seale-Carlisle**

removed piercings). Therefore, while both interactive and photo lineups used two-dimensional (2D) images and were captured at the same resolution, the interactive lineup ostensibly provided the witness with three-dimensional (3D) structural information when they rotated a face into different orientations.

To select the fillers for each lineup, we created a modal description of each culprit by asking a group of participants (White $n$ = 10, South Asian $n$ = 10) to watch each mock-crime video and, after each video, answer 10 multiple-choice questions about the culprit's physical appearance (e.g., sex, race/ethnicity). We used these modal descriptions to select 6 other people who matched the description of each culprit. For each culprit, the six selected faces constituted the target-absent lineup and we randomly selected one of the six faces to be substituted with the culprit to create the target-present lineup. A group of participants acting as "mock-witnesses" (White $n$ = 357, South Asian $n$ = 354) read a modal description of one of the culprits, viewed a series of faces, and were asked to decide which lineup member best fit the description of the culprit. The distribution of mock-witness choices confirmed that the members in our lineups were plausible alternatives to the culprits and that the lineups were similarly fair across the experimental conditions. Across the 16 conditions, the mean number of plausible lineup members, as measured by Tredoux's E was 5.35, 95% confidence interval (CI) [5.19, 5.51], and the mean proportion of participants in each lineup who selected the culprit ($M$ = .16, $SD$ = .05) or the culprit's replacement in the target-absent lineup ($M$ = .15, $SD$ = .05) did not differ from chance (.17). None of the measures differed across the experimental conditions (see online supplemental materials). Thus, any differences in discrimination accuracy in the experiment proper are not due to systematic differences in perceptual fairness across the experimental conditions.

**Figure 1**

*A Diagram of the (A) Static Photo Sequential Lineup (Experiment 1), (B) Interactive Sequential Independent-Movement Lineup (Experiment 1 and 2), (C) Interactive Simultaneous Independent-Movement Lineup, and (D) Interactive Simultaneous Joint-Movement Lineup*




*Note.* In A–B, only one face was presented at a time, in C–D all faces were presented together. In B–D, faces were presented in frontal pose and the participant could use the computer mouse to rotate the lineup faces. When the participant clicked on a face to rotate (e.g., face 5), in B–C only that face rotated and in D all of the lineup faces rotated together. Photos printed with permission.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.



**Christian A. Meissner**

### *Procedure*

Participants completed the experiment on a computer. They were told that they would watch a number of videos and answer some questions. Participants first provided their demographic information (age, sex, and race/ethnicity). Following this, participants watched one mock-crime video randomly selected from the pool of videos (featuring a culprit who was either White or South Asian). Next, participants engaged in a distractor task, completing anagrams for 2 min. Participants were then instructed to think back to the video that they had just watched. They were told that they would see a series of people, and that they should attempt to identify the culprit from the video. Following recommended police practice, lineups were administered "double-blind" (i.e., since the experiment was computer-automated, there was not a lineup administrator who knew the identity of the suspect) and participants were told that the culprit may or may not be shown and that not identifying anyone might be the correct answer (e.g., Technical Working Group for Eyewitness Evidence, 1999; Wells et al., 2020). We told participants to write down the number of the person they thought was the culprit, if they believed he was there.

Participants either viewed a target-present or target-absent lineup. Participants in the static lineup condition viewed six static images presented sequentially. Each face was displayed in frontal pose and participants were required to wait for 5 s before the 'next' button appeared and they were allowed to proceed. Participants in the interactive lineup condition viewed six interactive faces presented sequentially. Each interactive face was displayed in frontal pose and participants were told to use the computer mouse to click on each face and drag it to see it from multiple angles. Participants were required to rotate the face by at least 75% to the left and 75% to the right, and to wait for 5 s before the 'next' button appeared and they were allowed to proceed. We recorded how long participants spent viewing a face in a given angle. In both lineup conditions, the presentation order of lineup faces was randomly determined for each participant. Participants could not revisit previously seen faces and there was no stopping rule; once participants had viewed all six faces, they were asked whether the culprit was in the lineup, and, if yes, to enter the number of the person whom they believed to be the culprit. All participants were asked to provide a confidence judgment for their identification decision on an 11-point Likert-type rating scale ranging from 0% *not confident at all* to 100% *absolutely confident* (e.g., 0%, 10%, 20% . . .). Participants answered a multiple-choice attention check about what occurred in the video (e.g., "A female took a laptop from an office," "A male entered a house and took some keys").

Participants repeated this sequence of tasks until they had viewed all eight mock-crime videos and completed eight lineup tasks. The order of the crime videos and target presence in the lineup was randomly determined for every participant, with two constraints: (1) every participant viewed two videos of each race, each followed by a target-present lineup, and two videos of each race, each followed by a target-absent lineup; and (2) crime scenario and target-presence were fixed within participants (e.g., if a participant saw South Asian burglary followed by a target-present lineup, they would, at some point, view the White burglary followed by a target-present lineup).

Finally, participants completed a 9-item interracial contact questionnaire (see online supplemental materials). Participants were asked if the videos played smoothly, if faces in the lineups were displayed clearly, and to describe their ethnicity.

## Results and Discussion

Our data are available (https://osf.io/2x5tg/). Our aim was to determine whether interactive lineups enhance witnesses' ability to discriminate between innocent and guilty suspects and also reduce the size of the ORB, compared with static photo lineups. We first outline the descriptive results and the distribution of identification responses, then we address our research question following our preregistered plan using ROC analysis and fitting a signal-detection model. Finally, we explore the relationship between confidence and accuracy using confidence accuracy characteristic (CAC) analysis.

### *Preliminary Results*

For each participant in the interactive condition, we determined the total length of time that they spent interacting with

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.



**James C. Rockey**

each face, and summed the results across all of the faces in the eight lineups. We excluded from the analysis trials ($n = 1$) on which there was a mouse movement data recording error. The mean total length of time that participants spent interacting with the lineup faces was 153,099 ms ($SD = 38{,}810$, median = 148,272, range = 51,528–297,258 ms). We also analyzed the last region of the lineup faces that participants had viewed, and explored the association between discriminability and the length of time participants spent interacting (see online supplemental materials).

South Asian participants reported more contact with Whites ($M = 36.15$, $SD = 17.12$), than White participants reported contact with South Asians ($M = 13.85$, $SD = 9.95$), $t(175.07) = 11.81$, $p < .001$. Possible scores on the contact questionnaire ranged from 0 to 90, indicating low and high contact with the other-race, respectively. In the analyses that follow, all participants in the final sample were included, regardless of the length of time that they had interacted with the faces, or their other-race contact.

### *Identification Responses*

Table 1 displays the frequency of culprit, filler, and "Not Present" (i.e., reject) identification decisions (IDs) made to own-race and other-race faces in static and interactive lineups. The own-race condition combines data from White participants identifying White culprits, and South Asian participants identifying South Asian culprits. The other-race condition combines data from White participants identifying South Asian culprits and South Asian participants identifying White culprits. On average, there was a higher proportion of correct responses on interactive compared with static lineups, and for own-race compared with other-race decisions. To determine whether the observed trends in proportions reflect changes in discrimination accuracy across the conditions, we used ROC analysis and fit a signal-detection model.

### *ROC Analysis*

ROC analysis has been used for decades in the basic scientific literature and has more recently been applied to the field of eyewitness memory. ROC analysis was recommended to replace the intuitive, yet potentially misleading, measurement methods previously used in the eyewitness field to compare the diagnostic performance of lineup procedures (e.g., Mickes et al., 2012). Although ROC analysis is now a dominant method of analysis in the eyewitness field, some researchers have resisted the use of traditional ROC analysis and advocated their own ROC-like alternative that is not tethered to any formal model of decision-making (e.g., Smith et al., 2020). It has been argued that it is important for eyewitness research to rely on a principled mechanistic understanding of memory, perception, and decision-making (Albright & Rakoff, 2020). Here, we use ROC analysis in the traditional way that it has been used for decades in the basic scientific literature, directly tethered to a longstanding model of decision-making—signal detection theory.

To construct partial ROC curves, we used the 11-point confidence scale (100% to 0% certain) to plot the cumulative hit rate (HR; number of culprit IDs ÷ total number of target-present lineups) against the cumulative false alarm rate (FAR; number of innocent suspect IDs ÷ total number of target-absent lineups) over decreasing levels of confidence (see Gronlund et al., 2014). We estimated the number of innocent suspect IDs by dividing the number of filler IDs by the lineup size (in this case, six). In a fair lineup, estimating innocent suspect IDs in this way leads to exactly the same mean estimate as predesignating an innocent suspect, but it returns an estimate that is statistically more precise (i.e., has a lower standard error) for a given number of participants tested. To statistically compare ROC curves, we computed the partial Area Under the Curve ($p$AUC) using the statistical package $p$ROC (Robin et al., 2011). $p$ROC also calculates $D$, a measure of effect size: $D$ = (AUC1 – AUC2)/s, where $s$ is the standard deviation of the difference between the two AUCs and is estimated using bootstrapping. In all $p$AUC analyses, we defined the specificity (1 – FAR) using the smallest false alarm rate (FAR) range to any filler face in that comparison. Larger $p$AUC values (i.e., higher ROC curves) indicate better empirical discriminability—people's collective ability to discriminate between innocent and guilty suspects.

First, comparing lineup procedures, the partial ROC curves (Figure 2A) and the corresponding $p$AUC values plotted in Figure 2B indicate that ability to discriminate between innocent and guilty suspects was better in interactive than static lineups ($D = 2.40$, 95% CI [.41, 4.37], $p$ =

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.



**Babita Pande**

.02; specificity = .71). For any possible false identification rate of innocent suspects, using interactive lineups instead of static lineups increased the correct identification rate of guilty suspects by 18%. Next, considering how the lineup procedure influenced the ORB, partial ROC curves (Figure 2C) and the corresponding *p*AUC values plotted in Figure 2D indicate that ability to discriminate between innocent and guilty suspects was better for own-race than other-race decisions in both static ($D$ = 5.10, 95% CI [3.27, 7.15], $p < .001$) and interactive lineups ($D$ = 4.39, 95% CI [2.47, 6.39], $p < .001$; specificity = .73; paired ROC tests). A $z$ test on the $D$ values indicated that the size of the ORB was similar in static and interactive lineups, $z = .74$, $p = .23$ (one-tailed).[3]

### *Signal Detection Model*

Sometimes the results of an ROC analysis based on an atheoretical measure like *p*AUC do not agree with the results based on a theoretical measure like $d'$ obtained by fitting a theoretical model to the same data (e.g., Kaesler et al., 2020; Rotello & Chen, 2016; Wilson et al., 2019). Therefore, we fit a signal-detection model to our data, which confirmed our *p*AUC findings. The model accounts for all IDs (culprit, filler, and reject decisions in target-present and target-absent lineups), and measures theoretical discriminability—ability to discriminate between faces that have been seen before (i.e., culprits) and those that have not. It also estimates a set of confidence criteria, which reflect participants' willingness to make a positive ID. Model-fitting details are provided in the online supplemental materials. Table 2 shows a summary of the model-predicted discriminability ($d'$) and positive ID decision criteria ($c_1$) parameters. It is first worth noting that, descriptively speaking, other-race IDs yielded more liberal responding than own-race IDs, consistent with previous findings (Meissner & Brigham, 2001).[4] To answer our main research question: comparison of model-fit statistics illustrated that discriminability was significantly better in interactive than static lineups ($p = .004$); significantly better for own-race than other-race decisions ($p = .001$); and the size of the own-race bias was similar in static and interactive lineups ($p = .94$).[5]

Together, these results indicate that interactive lineups enhanced discrimination accuracy compared with a standard static photo lineup, both for own- and other-race witnesses. However, allowing participants to interact with the faces during an interactive lineup at retrieval did not attenuate the ORB.

### *Confidence Accuracy Characteristic Analysis*

While policymakers should be interested in *discriminability*—which measures the ability of a procedure to sort innocent and guilty suspects into their correct categories—legal decision-makers (e.g., judges, jurors, police officers) are interested in another element of witness performance, called *reliability*. Reliability refers to the probability that an ID made with a certain level of confidence is correct (Mickes, 2015). A procedure that is poorer at sorting innocent and guilty suspects can nevertheless achieve comparable reliability to a procedure that is better at sorting innocent and guilty suspects, because witnesses can often assess the likely accuracy of their memory, assigning high confidence when they are likely to be accurate and low confidence when they are likely to be inaccurate (e.g., Brewer & Wells, 2006; Sporer et al., 1995). Therefore, although witnesses have poorer discrimination accuracy on static compared with interactive lineups, and for other- compared with own-race faces, it is possible that witnesses making IDs from static lineups and other-race IDs are aware of their lower likelihood of accuracy and lower their confidence appropriately. That is, it is possible that there is a good relationship between confidence and accuracy, even for static lineups and other-race faces. Indeed, research indicates that reliability on static photo lineups is impressive (Wixted,

[3] As per our preregistration, we conducted preliminary ROC analyses to examine the ORB in each lineup procedure for White and South Asian participants, and calculated $d'$ for own- and other-race IDs for each crime video and lineup procedure. The results generalise across White and South Asian participants, and across the four mock-crime stimulus sets (see online supplemental materials). This further highlights the benefit of interactive over static lineups.

[4] An exploratory analysis indicated that the difference in the confidence criteria for own- and other-race identifications was not statistically significant (details here https://osf.io/2x5tg/).

[5] Because each participant completed eight lineup tasks and the memory strength distributions were estimated to be of equal variance, $d'$ for static and interactive lineups and for own-race and other-race decisions can also be computed for each participant, using the formula: $d' = z(\text{HR}) - z(\text{FAR})$. This method yields the same pattern of results. Moreover, the size of the ORB (i.e., $d'$ own race - $d'$ other race) did not change systematically across the eight trials.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.



**Pratibha Kujur**

2018; Wixted & Wells, 2017). Moreover, two recent papers examining the ORB have found this pattern of results (Dodson & Dobolyi, 2016; Nguyen et al., 2017). Dodson and Dobolyi (2016) tested participants with 12 lineups and found that confidence was more closely aligned with accuracy for same-race than other-race IDs because other-race IDs were more overconfident; at the same time, generally speaking, same-race and other-race accuracy at each level of confidence was similar and confidence increased with accuracy. Likewise, Nguyen et al. (2017) reanalyzed data from four ORB experiments and found that, in those experiments where performance was above chance levels, accuracy at each level of confidence did not significantly differ for own-race and other-race faces. Given the paucity of research examining the confidence accuracy relationship for own and other-race lineup IDs, and to ensure full understanding of eyewitness performance on interactive versus static lineups, we also plotted confidence accuracy characteristic (CAC) curves, which measure the probability that a suspect who has been identified is guilty at different confidence levels.

To plot CAC curves, we created three confidence bins (0–60, 70–80, 90–100). The 11-point CAC curves were noisy, and three confidence bins resulted in a relatively equal number of IDs each bin, because there were few IDs made with low confidence (e.g., Mickes, 2015). Moreover, the three confidence bins map onto the three confidence criteria parameters that were estimated when fitting the signal-detection model (see online supplemental material). For each level of confidence, we calculated suspect ID accuracy using the formula: HR ÷ (HR + FAR). Figure 3A shows that at each level of confidence, suspect ID accuracy was similar across the static and interactive lineups, indicating that the lineup procedures resulted in suspect IDs that are equally reliable. Next, considering the ORB, Figure 3 shows that in both (B) static and (C) interactive lineups, suspect ID accuracy was generally slightly higher at each level of confidence for own-compared with other-race decisions, but none of these differences were reliable. These results are consistent with previous research and suggest, at least in multiple-trial lineup experiments such as this, the race of the witness and culprit appears to be minimally important when examining the relationship between confidence and accuracy. Of note, however, is that there were slightly more high-confidence hits (correct IDs of guilty suspects made with 90–100% confidence) in the interactive compared with the static lineup, and for own-compared to other-race decisions. This indicates that interactive lineups and own-race decisions yield a greater number of high-confidence correct suspect IDs that are likely to be most influential in the criminal justice system (e.g., Brewer & Burke, 2002).

## Experiment 2

In Experiment 1, we found that sequential interactive lineups enhanced both own-race and other-race witnesses' discriminability compared with sequential static lineups. There are at least three possible mechanisms by which interactive lineups improved discrimination accuracy, namely: active exploration of diagnostic features, pose-reinstatement, and motion cues. In Experiment 2, we focus on the *active exploration* mechanism, specifically drawing on the diagnostic-feature-detection theory, to make predictions about how two additional types of interactive lineups—in which the faces are presented simultaneously—could further boost identification accuracy.

The diagnostic-feature-detection theory suggests that discrimination accuracy is enhanced by procedures that enable witnesses to more easily detect and discount features that are nondiagnostic (i.e., are shared in common by all lineup members, Wixted & Mickes, 2014). The theory was developed to account for the findings that simultaneous lineups, in which all of the faces are displayed at the same time, can yield a higher ROC curve and higher $d'$ than sequential lineups in which the faces are displayed one at a time (see Seale-Carlisle et al., 2019, for a mini meta-analysis, but also Kaesler et al., 2020). One benefit of an interactive lineup is that the faces can be presented simultaneously, harnessing the potential benefit of simultaneous comparison of faces in a lineup, while still allowing an opportunity for retrieval support (e.g., active exploration, pose-reinstatement, and motion cues). As such, we further examined the interactive lineup benefit and the ORB in Experiment 2 by comparing three presentation formats for interactive lineups: sequential-independent (Figure 1B), simultaneous-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.



Noorshama Parveen

independent (Figure 1C), and simultaneous-joint (Figure 1D). Sequential-independent interactive lineups were presented in the same format as Experiment 1. We use the term "independent" to make it clear that when a participant rotated a face, only that face moved and the next sequentially presented face was displayed in frontal pose until it was rotated. In simultaneous-independent interactive lineups, the lineup faces were presented simultaneously and when a participant rotated a face, only that face moved. In simultaneous-joint interactive lineups, the lineup faces were presented simultaneously and when a participant rotated one face, all six faces moved together. Using diagnostic-feature-detection theory, we predicted that discrimination accuracy would be better in both simultaneous-joint and simultaneous-independent lineups compared with sequential lineups, and in simultaneous-joint lineups compared to simultaneous-independent lineups. This is because in the simultaneous-joint lineups, the faces are presented together and maintain a synchronous orientation when rotated, allowing the greatest opportunity for comparison of features across faces to detect and discount nondiagnostic features.

## Method

We preregistered our experiment and analyses before we collected data (https://osf.io/b8tvw/).

### *Design*

We used a 2 (Culprit Race: White, South Asian) × 2 (Participant Race: White, South Asian) × 3 (Lineup Procedure: Sequential-Independent, Simultaneous-Independent, Simultaneous-Joint) × 2 (Target: Present, Absent) between-participants design. We aimed to recruit at least 8,400 participants with usable data, 350 participants in each experimental condition. Using the mean difference and standard deviations observed in Mickes et al. (2012) as a guide, a power analysis indicated that, with 350 participants per between-participants condition, power for this experiment would exceed 80%.

### *Participants*

The participants were 11,435 adults from around the world who completed the task online. We excluded participants who reported that they were not White or South Asian ($n$ = 1,556, 14%), who incorrectly answered an attention check question ($n$ = 797, 7%), who experienced technical issues while watching the video or viewing the lineup ($n$ = 348, 3%), and who completed the experiment more than once ($n$ = 227, 2%; we included only data from the first completion). In total, we excluded 2,928 people, which resulted in a final sample of 8,507: 4,293 White and 4,214 South Asian (age: 16–82 years, $M$ = 32.28, $SD$ = 11.72; sex: 4,252 female, 4,218 male, 37 prefer not to say). Of the final sample, 6,427 participants were recruited from online participation websites, such as Amazon Mechanical Turk and Prolific and received payment in line with local norms (45 cents and 50p, respectively); 1,743 were students recruited from Pt. Ravishankar Shukla University, and 337 were recruited from posts on social-networking and University-based websites and were entered into a prize drawing for one of ten £50 Amazon vouchers. We recruited participants using a variety of methods, because it was not possible to reach our desired sample size of naive participants using one recruitment method. We combined all data for our analysis.[6] Each cell contained between 319 and 405 participants. We did not inform participants that we were recruiting White and South Asian participants. Again, we called the study "Perception and Memory" and advertised it to people in countries that are predominantly White (United States, United Kingdom) or South Asian (Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, and Sri Lanka). On Prolific, we advertised the study to people who had previously self-defined as either White or South Asian.

### *Materials*

**Mock-Crime Videos.** We used six 30 s mock-crime videos, formed by crossing three crime scenarios (mugging, money theft, and theft from a car) with two culprit races (White, South Asian). We used the *theft from a car* scenario from Experiment 1, but created the mugging and money theft scenarios for Experiment 2. We created the new scenarios in the same manner as Experiment 1. Again, the three crime scenarios (mugging, money theft, and theft from a car) differed from each other in a number of ways, such as the location (e.g., park, bedroom, and car), distance from the culprit's

[6] The pattern of findings was the same in all three recruitment methods.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.



**Priyanka Chandel**

face, exposure duration to the culprit (6–19 s across the four videos), sex of culprit, and presence of bystanders.

**Lineups.** We used the same method as Experiment 1 to select and photograph the fillers. Mock-witness testing (White $n$ = 436, South Asian $n$ = 415) confirmed that our lineup members were plausible alternatives to the culprits, and any differences in discrimination accuracy found in the experiment proper are not due to systematic differences in perceptual fairness across the experimental conditions. Across the 24 conditions, the mean number of plausible lineup members, as measured by Tredoux's E, was 4.80 95% CI [4.55, 5.06], and the mean proportion of participants in each lineup who selected the culprit ($M$ = .21, $SD$ = .08) or the culprit's replacement in the target-absent lineup ($M$ = .13, $SD$ = .08), did not differ from chance. None of the measures differed across the experimental conditions (see online supplemental materials).

### *Procedure*

We advertised and conducted the study online in English. We told participants the study involved watching a video and then answering some questions. Participants were first asked to provide their demographic information (age, sex, and ethnicity). Following this, participants watched one mock-crime video randomly selected from the pool of videos (either mugging, money theft, and theft from a car, featuring a culprit who was either White or South Asian). Next, participants engaged in a distractor task in which they watched a cartoon for 2 min. Following this, participants were instructed to think back to the crime video that they had watched and were given the same lineup instructions as Experiment 1.

Next, the lineup was displayed. Again, the lineup was administered "double-blind," via the automated online experiment. Participants either viewed a target-present or a target-absent lineup. Participants in the sequential-independent interactive condition viewed six interactive faces presented sequentially. Each face was displayed in frontal pose and participants were required to rotate the face by at least 75% to the left and 75% to the right, and to wait for 5 s before the 'next' button appeared and they were allowed to proceed. As in Experiment 1, participants could not revisit previously seen faces and there was no stopping rule in the sequential lineup; participants viewed all six faces before making an identification decision. Participants in the simultaneous-independent and simultaneous-joint lineup conditions viewed six interactive faces presented at the same time in two rows of three faces, displayed initially in frontal pose. In the simultaneous-independent condition, participants were required to rotate each face by at least 75% to the left and 75% to the right, and to wait for 30 s before the 'next' button appeared and they were allowed to proceed. When a participant rotated a face, only that face moved. In the simultaneous-joint lineup condition, participants were similarly required to rotate the faces by at least 75% to the left and 75% to the right, and to wait for 30 s before the 'next' button appeared and they were allowed to proceed. When a participant rotated a face, all six faces moved together into the same angle. In all lineups, the order of the faces was randomly determined for each participant and we recorded how long the participant spent viewing a face at a given angle.

Once participants had viewed all six faces, they were asked whether the culprit was in the lineup, and, if yes, were asked to enter the number of the person whom they believed to be the culprit. Participants provided a confidence judgment for their identification decision on an 11-point Likert-type rating scale ranging from 0% *not confident at all* to 100% *absolutely confident*, and answered a multiple-choice attention check about what happened in the video that they had watched (e.g., "A female took some money from a bedroom," "A male threatened a lady and stole her phone"). In the final stage of the experiment, participants completed the same interracial contact questionnaire as in Experiment 1, were asked to describe their ethnicity, and were asked if the video played smoothly and faces in the lineups displayed clearly.

## Results and Discussion

Our data are available (https://osf.io/b8tvw/). Our primary aim was to determine whether interactive presentation formats that might allow for easier detection of nondiagnostic features (i.e., simultaneous-joint lineups compared with simultaneous-independent lineups, and both simultaneous lineups compared with sequential-independent lineups) enhance own- and other-race witnesses' ability to discriminate innocent from guilty suspects. We first

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.



**Margaret M. Singh**

outline descriptive results and present the distribution of identification responses, then we address our research question following our preregistered plan, using ROC analysis and fitting a signal-detection model. Finally, we explore the relationship between confidence and accuracy using CAC analysis.

### *Preliminary Results*

We used the same method as Experiment 1 to analyze how participants interacted. We determined the total length of time that each participant spent interacting with each face, and summed the results across all of the faces in the lineup. Excluding participants ($n = 46$) whose mouse movements did not record properly, the mean length of time participants interacted with the lineup faces was 43,968 ms in the sequential-independent condition (median = 32,800, $SD = 45{,}959$, range = 8,308–1,026,614 ms), 38,617 ms in the simultaneous-joint condition (median = 28,038, $SD = 38{,}122$, range = 993–640,821 ms), and 104,312 ms in the simultaneous-independent condition (median = 79,723, $SD = 90{,}911$, range = 15,686–868,306 ms). Again, we analyzed the last region of the lineup faces that participants had viewed, and explored the association between discriminability and the length of time that participants spent interacting (see online supplemental materials).

South Asian participants reported more contact with Whites ($M = 19.59$, $SD = 20.01$), than White participants reported with South Asians ($M = 6.43$, $SD = 8.29$), $t(5596.4) = 39.48$, $p < .001$. In the results that follow, all of the participants in the final sample were included, regardless of other-race contact, or the length of time they spent interacting with the faces.

### *Identification Responses*

Table 3 displays the frequency of culprit, filler, and reject IDs made to own- and other-race faces in sequential-independent, simultaneous-independent, and simultaneous-joint interactive lineups. For own-race compared with other-race decisions, and for both simultaneous-joint lineups and simultaneous-independent lineups compared to sequential-independent lineups, there was a higher proportion of correct responses. The same pattern was true when comparing simultaneous-joint lineups and simultaneous-independent lineups, though the difference was smaller.

### *ROC Analysis*

First, comparing lineup procedures, the partial ROC curves (Figure 4A) and the corresponding *p*AUC values plotted in Figure 4B indicate that ability to discriminate between innocent and guilty suspects was better in both simultaneous-joint lineups ($D = 4.04$, 95% CI [2.11, 6.07], $p < .001$) and simultaneous-independent lineups ($D = 3.27$, 95% CI [1.26, 5.17], $p = .001$) than sequential-independent lineups, but was similar in simultaneous-joint and simultaneous-independent lineups ($D = .81$, 95% CI [−1.21, 2.73], $p = .42$; specificity = .41). For any possible false identification rate of innocent suspects, using a simultaneous-joint lineup instead of a sequential-independent interactive lineup increased the correct identification rate of guilty suspects by 23%. Next, considering how the lineup procedure influenced the ORB, partial ROC curves (Figure 4C) and the corresponding *p*AUC values plotted in Figure 4D indicate that ability to discriminate between innocent and guilty suspects was better for own-race than other-race decisions in sequential-independent ($D = 5.22$, 95% CI [3.27, 7.16], $p < .001$), simultaneous-independent ($D = 5.71$, 95% CI [3.75, 7.68], $p < .001$), and simultaneous-joint lineups ($D = 6.12$, 95% CI [4.24, 8.10], $p < .001$; specificity = .45). Three *z*-tests (one-tailed) on the *D* values indicated that the size of the ORB was similar in sequential-independent and simultaneous-independent ($z = .51$, $p = .31$), sequential-independent, and simultaneous-joint ($z = .99$, $p = .16$), and simultaneous-independent and simultaneous-joint lineups ($z = .49$, $p = .31$).[7]

### *Signal Detection Model*

When we fit a signal-detection model to our data, the results confirmed our *p*AUC analyses. Model-fitting details are provided in the online supplemental material. Table 4 shows a summary of the model-predicted discriminability

[7] As per our preregistration, we also conducted preliminary ROC analyses to examine the ORB in each lineup procedure for White and South Asian participants. We also calculated $d'$ for own- and other-race IDs for each mock-crime video and lineup procedure. The results generalize across White and South Asian participants, and across the three mock-crime stimulus sets (see online supplemental materials). This further highlights the benefit afforded by simultaneous-joint and simultaneous-independent lineups over sequential-independent interactive lineups.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.



**Sraddha Pradhan**

parameters ($d_a$) and positive identification decision criteria ($c_1$). In all three lineup conditions, other-race identifications yielded more liberal responding than own-race identifications, replicating previous findings (Meissner & Brigham, 2001).[8] To answer our main research question, we conducted three pairwise comparisons of model-fit statistics. Theoretical discriminability was significantly better in both simultaneous-joint and simultaneous-independent lineups than sequential-independent lineups ($p < .001$ and $p = .03$), but was similar in simultaneous-joint and simultaneous-independent lineups ($p = .29$). Theoretical discriminability was better for own-race than other-race decisions (all $p < .001$); and the size of the ORB was similar across all three lineup procedures (all $p > .32$; see online supplemental materials).

Together these results indicate that simultaneous-joint and simultaneous-independent lineups further enhance own- and other-race witnesses' discriminability compared with the sequential-independent interactive lineups that we tested in Experiment 1. We did not find that simultaneous-joint lineups significantly improved discrimination accuracy more than simultaneous-independent lineups, though the results trended in the expected direction.

### *Confidence Accuracy Characteristic Analysis*

First, considering lineup procedure, Figure 5A shows that at each level of confidence, suspect ID accuracy was generally similar across the sequential-independent, simultaneous-independent, and simultaneous-joint interactive procedures. Yet, high-confidence (90–100%) suspect IDs were slightly less accurate on sequential-independent lineups than the simultaneous lineups. In the simultaneous-joint lineup there were also slightly more high-confidence hits (correct IDs of guilty suspects made with 90–100% confidence) than the other two procedures, and fewer high-confidence false alarms (incorrect IDs of innocent suspects made with 90–100% confidence) than the sequential-independent procedure. This indicates that simultaneous-joint lineups yield the greatest number of high-confidence correct IDs and those IDs are more likely to be accurate than the sequential-independent lineup. Next, considering ORB, Figure 5 shows that in (B) sequential-independent, (C) simultaneous-independent, and (D) simultaneous-joint lineups, there was a systematic relationship between confidence and accuracy: as accuracy increased, so did confidence. However, in this one-trial eyewitness identification study, for each lineup procedure, suspect ID accuracy was reliably higher for own-race compared with other-race decisions at the highest-level of confidence. Dodson and Dobolyi (2016) also found that other-race IDs were overconfident compared with own-race IDs. It is possible that high-confidence own-race suspect IDs are more likely to be accurate than other-race IDs, because recollection is superior for own-race faces (Meissner, Brigham, & Butz, 2005) and the same mechanism underlying the ORB on ID accuracy also underlies confidence assessment (Dodson & Dobolyi, 2016; cf. Nguyen et al., 2017). Unlike in Experiment 1, the results from Experiment 2 suggest that the correspondence between the race of the witness and the suspect may be important for legal-decision makers (e.g., police officers, judges, and jurors) determining the likely accuracy of a suspect ID made with high confidence.

## General Discussion

We tested whether a novel interactive lineup procedure, wherein witnesses can rotate and dynamically view lineup faces, could improve witness discrimination accuracy in both own- and other-race witnesses, when compared with a widely used procedure in laboratories and police forces around the world—namely, the static frontal pose photo lineup. In Experiment 1, we found that sequential interactive lineups significantly enhanced discriminability compared with sequential static lineups, for both own- and other-race witnesses. In Experiment 2, we found that simultaneous interactive lineups (in which the faces either moved jointly or moved independently of each other) significantly enhanced discriminability compared with sequential interactive lineups, for both own- and other-race witnesses. In the two experiments, the size of the ORB was similar across the lineup procedures tested. These findings have important implications for testing eyewitness memory.

First, our results illustrate that innovative technologies and psychologically informed procedures can improve eyewitness

[8] An exploratory analysis indicated that the difference in the confidence criteria for own- and other-race identifications was statistically significant (details here https://osf.io/b8tvw/).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.



**Arti Parganiha**

identification accuracy from police lineups. Eyewitness misidentifications are common and reliance on unreliable identifications can have profound implications, such as innocent people being incarcerated while guilty people remain at large to commit further crimes (Innocence Project, 2019; Silbert et al., 2015). In recent years, our understanding of eyewitness memory has undergone a radical revision as the basic science of memory has increasingly been brought to bear on the applied science of memory (e.g., National Research Council, 2014). For example, recent perspectives suggest that eyewitness confidence is informative of accuracy on an initial, uncontaminated, properly administered photo lineup, suggesting that eyewitnesses can be reliable and know if their identification decision is or is not likely to be accurate (e.g., Wixted, 2018). However, eyewitnesses still make incorrect identifications and theoretically informed procedures should be developed to support eyewitness discrimination accuracy (Albright & Rakoff, 2020). Our interactive lineup procedure was developed using basic science evidence that face recognition is aided by active exploration of diagnostic-features (Liu et al., 2007), study-test viewpoint correspondence (e.g., Bruce, 1982; Carbon & Leder, 2006), and motion cues for 3D perception of structure (e.g., O'Toole et al., 2002). Consistent with these effects, we found that interactive lineups improved witness accuracy compared with sequential static photo lineups, which are used in around 5,000 law enforcement agencies in the United States and in other countries worldwide. To highlight the scale of the potential benefit afforded by interactive lineups: for any possible false identification rate of innocent suspects, using sequential-independent interactive lineups instead of sequential static lineups increased the correct identification rate of guilty suspects by 18%. Further, using simultaneous-joint interactive lineups instead of sequential-independent interactive lineups increased the correct identification rate of guilty suspects by a further 23%. In practice, such an increase could markedly improve the effectiveness of police identification procedures. While we call on other researchers to replicate these findings using novel stimulus materials, encoding conditions, and retention intervals, it appears that this approach has the potential to reform the practice of lineup identifications and to foster a more just society.

Like all lineup identification procedures, interactive lineups consist of a "package" of components (e.g., active exploration, pose-reinstatement, movement, multiple viewing angles, viewing a face from ¾ angle), and accuracy differences across lineup procedures may stem from any or a combination or all of these components. Research should continue to isolate the causal mechanisms that underpin enhanced performance in interactive lineups (e.g., Colloff et al., 2021; Smith, Andrews, et al., 2021). In Experiment 2, we began on this path and examined the *active exploration* mechanism, specifically drawing on diagnostic-feature-detection theory, to make predictions about how two additional types of interactive lineups—in which the faces are presented simultaneously—could further boost identification accuracy. According to the diagnostic-feature-detection theory, lineup discrimination accuracy is enhanced by procedures that enable witnesses to more easily detect and discount features that are shared in common by all lineup members (Wixted & Mickes, 2014). Therefore, we predicted that that simultaneous compared with sequential presentation would boost discrimination accuracy, and that participant witnesses who could move the simultaneous interactive lineup faces together (the simultaneous-joint condition) would have better discrimination than those who could only move the simultaneous interactive lineup faces one at a time (the simultaneous-independent condition).

We found significantly higher discrimination accuracy for simultaneous compared with sequential interactive lineups, consistent with the static photo lineup literature (e.g., Dobolyi & Dodson, 2013; Mickes et al., 2012; Seale-Carlisle et al., 2019). Some sequential lineup studies use a stopping rule in which the first lineup member that is identified terminates the lineup procedure. The use of a stopping rule in sequential lineups can impair empirical discriminability as measured by ROC analysis even if $d'$ (the degree to which underlying memory signals generated by innocent and guilty suspects overlap) is not reduced (Wilson et al., 2019; see also Kaesler et al., 2020). The simultaneous superiority effect that we observed in Experiment 2 cannot be explained by the artificial constraint on empirical discriminability imposed by the use of a stopping rule, because we did not use a stopping rule in our sequential procedure. Instead, the data need to be explained by a theory that can account for differences in discrimination accuracy across procedures, such as the diagnostic-feature-detection theory (Wixted & Mickes, 2014).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

**Table 1**

*Frequencies of Culprit, Filler, and "Not Present" (Reject) Identification Decisions Made With Different Confidence Ratings to Own-Race and Other-Race Faces in Static and Interactive Sequential Lineups*

| Lineup procedure and participant confidence rating | Own-race | | | | | Other-race | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Target-present | | | Target-absent | | Target-present | | | Target-absent | |
| | Culprit | Filler | Reject | Filler | Reject | Culprit | Filler | Reject | Filler | Reject |
| Static | | | | | | | | | | |
| 0 | 3 | 0 | 0 | 1 | 2 | 0 | 0 | 4 | 2 | 4 |
| 10 | 2 | 3 | 0 | 3 | 1 | 2 | 1 | 1 | 1 | 2 |
| 20 | 1 | 2 | 0 | 6 | 1 | 1 | 1 | 1 | 4 | 1 |
| 30 | 3 | 2 | 4 | 3 | 3 | 7 | 4 | 2 | 6 | 4 |
| 40 | 7 | 6 | 4 | 9 | 5 | 10 | 6 | 4 | 11 | 7 |
| 50 | 1 | 5 | 5 | 10 | 14 | 14 | 8 | 7 | 13 | 15 |
| 60 | 18 | 5 | 6 | 13 | 16 | 19 | 4 | 10 | 10 | 19 |
| 70 | 24 | 5 | 8 | 9 | 31 | 16 | 3 | 4 | 11 | 17 |
| 80 | 30 | 0 | 2 | 5 | 21 | 26 | 0 | 9 | 5 | 32 |
| 90 | 24 | 0 | 3 | 0 | 24 | 13 | 2 | 4 | 4 | 18 |
| 100 | 38 | 0 | 5 | 3 | 36 | 25 | 2 | 6 | 1 | 29 |
| Total | 151 | 28 | 37 | 62 | 154 | 133 | 31 | 52 | 68 | 148 |
| Proportion | 0.70 | 0.13 | 0.17 | 0.29 | 0.71 | 0.62 | 0.14 | 0.24 | 0.31 | 0.69 |
| Interactive | | | | | | | | | | |
| 0 | 0 | 0 | 2 | 3 | 0 | 0 | 1 | 1 | 1 | 1 |
| 10 | 1 | 0 | 0 | 2 | 1 | 1 | 0 | 1 | 1 | 1 |
| 20 | 4 | 0 | 0 | 2 | 4 | 1 | 1 | 0 | 2 | 1 |
| 30 | 4 | 0 | 4 | 5 | 3 | 3 | 0 | 2 | 5 | 5 |
| 40 | 6 | 2 | 2 | 8 | 5 | 8 | 2 | 3 | 8 | 11 |
| 50 | 14 | 2 | 4 | 11 | 16 | 18 | 0 | 9 | 11 | 22 |
| 60 | 27 | 5 | 1 | 18 | 14 | 17 | 6 | 8 | 20 | 27 |
| 70 | 27 | 3 | 7 | 7 | 27 | 16 | 4 | 9 | 12 | 26 |
| 80 | 25 | 3 | 5 | 3 | 30 | 29 | 3 | 5 | 5 | 21 |
| 90 | 28 | 1 | 0 | 1 | 18 | 26 | 1 | 4 | 2 | 10 |
| 100 | 44 | 1 | 2 | 1 | 45 | 40 | 1 | 4 | 3 | 29 |
| Total | 180 | 17 | 27 | 61 | 163 | 159 | 19 | 46 | 70 | 154 |
| Proportion | 0.80 | 0.08 | 0.12 | 0.27 | 0.73 | 0.71 | 0.08 | 0.21 | 0.31 | 0.69 |

*Note.* Target-present lineups contained the culprit from the mock-crime video and five fillers, and target-absent lineups contained six fillers. The total row shows the frequency of each identification decision, collapsed over participants' postidentification confidence. The proportion row shows the total number of identification decisions, divided by the number of lineups in that condition. For example, the proportion of own-race culprit IDs in static lineups is calculated by taking the total number of culprit IDs and dividing that by the total number of target-present own-race static lineups, that is, $151/(151 + 28 + 37)$.

Yet, the ability to move and view the faces in the same angle (simultaneous-joint condition) did not significantly increase discrimination accuracy compared with when the faces moved independently (simultaneous-independent condition). We forced all participants to rotate each of the lineup faces to the left and right to have the greatest amount of experimental control possible and to ensure that any differences in accuracy could be isolated as due to the presentation format. One possibility is that by forcing participants to rotate the lineups faces, participants in the simultaneous-independent lineup rotated the lineup faces to see them from the same orientation and could easily compare across faces, mimicking the simultaneous-joint condition and reducing the size of the predicted joint-movement advantage. Indeed, we found that participants in the simultaneous-independent condition had rotated and were viewing the majority of the lineup faces from the same angle at the end of the trial, before they made an identification decision (see online supplemental materials). It is possible that the predicted difference between the simultaneous-joint and simultaneous-independent conditions would be larger if we let participants inspect the faces as they naturally would, rather than forcing rotation. Future studies should examine this further.

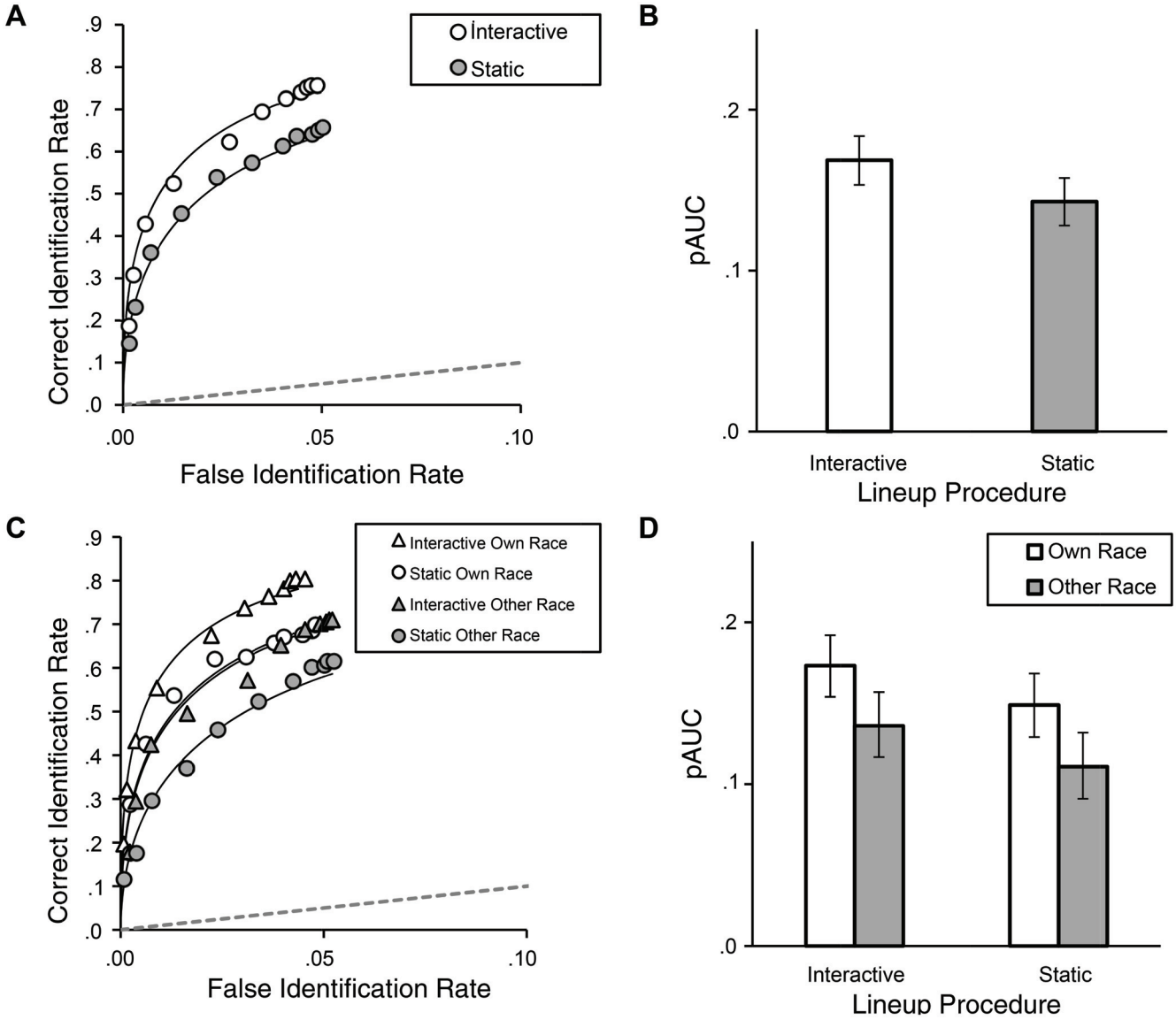
It is also perhaps interesting that while we did not find a significant difference in discriminability between the simultaneous-joint and simultaneous-independent lineups, the simultaneous-joint lineup might ultimately be preferred by legal policymakers. In the simultaneous-joint lineup, all of the lineup members are shown moving in the same manner, whereas an identification from a simultaneous-independent lineup could potentially be discredited in court, for example, on the grounds that the witness did not sufficiently examine the suspect compared with other lineup members, leading to a less reliable ID decision. It is clear that a partnership between science and law is necessary to develop and use successful identification procedures that foster accurate identifications and are legally compliant (Albright, 2017).

Future research should compare the interactive lineup to other lineup procedures that are used globally. Considering photo lineups, recent research suggests that static photo simultaneous lineups improve discriminability compared with a static photo sequential lineups (Dobolyi &

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

**Figure 2**

*Partial Receiver Operating Characteristic (ROC) Curves and Partial Area Under the Curve (pAUC) Statistics for (A, B) Interactive Lineups and Static Lineups, Collapsed Over Own- and Other-Race Decisions, and (C, D) Own- and Other-Race Decisions in Interactive and Static Lineups*




*Note.* In A and C, ROC lines of best fit were drawn using parameters estimated by an equal-variance signal-detection model, reported in the online supplemental materials, and the dashed lines represent chance-level performance. In B and D, error bars are 95% confidence intervals (CIs).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Dodson, 2013; Mickes et al., 2012; Seale-Carlisle et al., 2019; also see Kaesler et al., 2020). We do not know yet whether enabling witnesses to actively explore faces and view them from multiple angles in interactive lineups can increase witness discrimination accuracy compared with static photo simultaneous lineups that are frequently used in practice. We would predict that simultaneous interactive lineups also improve upon simultaneous static lineups, for the same reasons that we predicted sequential interactive lineups improve upon sequential static lineups (Experiment 1). Nevertheless, this is a major gap in the literature and the next step of research is to test this directly.

Future research should also compare interactive lineups to live and video lineups. One might intuit that live and video lineups, which ostensibly provide more facial information, would improve discrimination accuracy compared with photo lineups. Research to date, however, has not provided convincing evidence to support that intuition (e.g., Cutler et al., 1994; Fitzgerald et al., 2018; Rubínová et al., 2021). Indeed, some research has shown that simultaneous photo

**Table 2**
*Discriminability (d') and Positive Identification Decision Criteria ($c_1$) Parameters Estimated by the Equal Variance Signal-Detection Model*

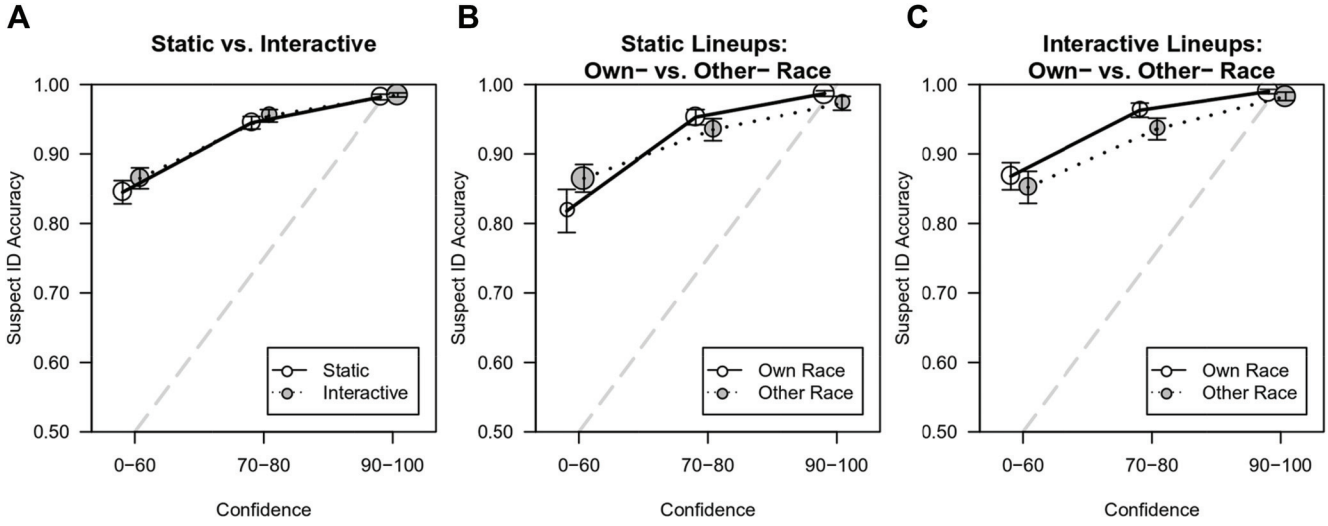
| | Interactive lineups | | Static lineups | |
|---|---|---|---|---|
| Race | $d'$ | $c_1$ | $d'$ | $c_1$ |
| Own | 2.54 | 1.61 | 2.23 | 1.55 |
| Other | 2.21 | 1.58 | 1.89 | 1.52 |

*Note.* $d'$ and $c_1$ are model-predicted discriminability and positive ID decision criteria parameters, respectively. Larger values of $d'$ indicate better ability to discriminate between faces that have and have not been seen before. Larger values of $c_1$ indicate more conservative responding.

lineups boost discrimination compared to sequential video lineups, because of the benefit afforded by simultaneous presentation (Seale-Carlisle et al., 2019) and other research has shown that sequential lineup performance can be impaired when the suspect appears later in the lineup (e.g., Kaesler et al., 2020; Wilson et al., 2019). Video lineups require sequential presentation because lineup members move their heads at different rates and so cannot easily be presented simultaneously. Therefore, interactive lineups could improve discrimination performance compared with video lineups because the lineup members can be explored from multiple angles simultaneously, but research should also test this.

Moreover, policymakers need to consider the practicality of lineup procedures. Legal guidelines in many countries express a preference for live lineups, though this preference is not always translated into practice (e.g., Australia; see Fitzgerald et al., 2018) because live lineups are inconvenient and impractical. Live lineups rely on the attendance of multiple parties at the same time (e.g., witnesses, fillers, and suspects) and it can be difficult and time-intensive to locate suitable fillers to attend in-person (Pike et al., 2002). In photo and video procedures, databases of images can be searched and lineups can be constructed instantly. Therefore, even if live lineups do turn out to yield a memory advantage, policymakers may decide that live lineups are not be worth the impracticalities they impose (Fitzgerald et al., 2018). In practice, the interactive lineup system would be relatively simple and inexpensive to implement, because the interactive technology that enables witnesses to rotate faces is readily available, and lineups could be conducted electronically (like photo and video lineups) and created using existing video libraries in the United Kingdom that capture each lineup member from multiple angles. However, in the United States and other countries that currently only capture static frontal images, a move to interactive lineups would require an initial start-up cost as the lineup member images would need to be captured from multiple angles. There would also be costs involved in training law enforcement agencies and implementing the interactive lineup. Such costs might be worth the long-term investment, though, if research replicates our findings and extends our work to illustrate that interactive lineups improve discrimination accuracy compared with existing procedures used worldwide. On the other hand, depending on the existence and size of the interactive advantage over other commonly used procedures that have not yet been tested (e.g., the simultaneous static photo lineups), the costs might not be considered to be worth the investment. Ultimately, psychological scientists can provide the relevant empirical data

**Figure 3**
*Confidence Accuracy Characteristic Curves for (A) Static and Interactive Lineups, Collapsed Over Own-Race and Other-Race Decisions; and Own-Race and Other-Race Decisions in (B) Static and (C) Interactive Lineups*




*Note.* The dashed lines represent chance-level performance at the lowest confidence bin (0–60) and perfect performance at the highest confidence bin (90–100). The size of the symbols represents the number of suspect IDs at a given level of confidence relative to the number of suspect IDs collapsed over confidence (Seale-Carlisle et al., 2019). Error bars indicate ±1 *SE* estimated by bootstrapping (Seale-Carlisle & Mickes, 2016).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

**Table 3**

*Frequencies of Culprit, Filler, and "Not Present" (Reject) Identification Decisions Made With Different Confidence Ratings to Own-Race and Other-Race Faces in Sequential-Independent-Movement, Simultaneous-Independent-Movement, and Simultaneous-Joint-Movement Interactive Lineups*

| Lineup procedure and participant confidence rating | Own-race | | | | | Other-race | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Target-present | | | Target-absent | | Target-present | | | Target-absent | |
| | Culprit | Filler | Reject | Filler | Reject | Culprit | Filler | Reject | Filler | Reject |
| Sequential-independent | | | | | | | | | | |
| 0 | 2 | 2 | 1 | 4 | 1 | 2 | 1 | 2 | 1 | 4 |
| 10 | 0 | 3 | 2 | 5 | 3 | 0 | 4 | 2 | 7 | 2 |
| 20 | 4 | 2 | 4 | 6 | 2 | 5 | 5 | 3 | 9 | 6 |
| 30 | 1 | 7 | 3 | 12 | 7 | 4 | 9 | 7 | 16 | 9 |
| 40 | 9 | 15 | 1 | 20 | 10 | 6 | 16 | 1 | 21 | 6 |
| 50 | 25 | 29 | 13 | 55 | 18 | 23 | 47 | 15 | 59 | 13 |
| 60 | 27 | 30 | 12 | 62 | 24 | 25 | 39 | 10 | 63 | 15 |
| 70 | 44 | 43 | 28 | 99 | 29 | 48 | 52 | 14 | 103 | 36 |
| 80 | 62 | 32 | 21 | 82 | 50 | 51 | 58 | 19 | 93 | 42 |
| 90 | 64 | 20 | 10 | 50 | 46 | 59 | 35 | 14 | 78 | 31 |
| 100 | 116 | 27 | 28 | 38 | 68 | 76 | 33 | 18 | 64 | 33 |
| Total | 354 | 210 | 123 | 433 | 258 | 299 | 299 | 105 | 514 | 197 |
| Proportion | 0.52 | 0.31 | 0.18 | 0.63 | 0.37 | 0.43 | 0.43 | 0.15 | 0.72 | 0.28 |
| Simultaneous-independent | | | | | | | | | | |
| 0 | 1 | 2 | 7 | 3 | 5 | 2 | 4 | 4 | 2 | 6 |
| 10 | 0 | 3 | 1 | 5 | 4 | 3 | 1 | 4 | 4 | 1 |
| 20 | 1 | 5 | 1 | 5 | 4 | 4 | 3 | 3 | 10 | 2 |
| 30 | 8 | 7 | 6 | 13 | 9 | 3 | 14 | 5 | 16 | 7 |
| 40 | 2 | 5 | 2 | 23 | 9 | 3 | 11 | 4 | 18 | 10 |
| 50 | 23 | 30 | 7 | 41 | 12 | 15 | 34 | 12 | 59 | 16 |
| 60 | 26 | 26 | 11 | 48 | 19 | 15 | 47 | 18 | 67 | 25 |
| 70 | 41 | 41 | 25 | 94 | 44 | 38 | 41 | 19 | 93 | 33 |
| 80 | 67 | 43 | 15 | 80 | 49 | 65 | 53 | 22 | 76 | 44 |
| 90 | 83 | 18 | 21 | 47 | 65 | 55 | 22 | 25 | 63 | 42 |
| 100 | 141 | 14 | 24 | 33 | 81 | 84 | 34 | 25 | 45 | 54 |
| Total | 393 | 194 | 120 | 392 | 301 | 287 | 264 | 141 | 453 | 240 |
| Proportion | 0.56 | 0.27 | 0.17 | 0.57 | 0.43 | 0.41 | 0.38 | 0.20 | 0.65 | 0.35 |
| Simultaneous-joint | | | | | | | | | | |
| 0 | 0 | 0 | 3 | 0 | 3 | 1 | 3 | 3 | 5 | 3 |
| 10 | 0 | 1 | 0 | 8 | 4 | 1 | 6 | 1 | 2 | 3 |
| 20 | 1 | 4 | 2 | 8 | 3 | 2 | 4 | 3 | 8 | 4 |
| 30 | 2 | 4 | 3 | 9 | 14 | 8 | 14 | 5 | 16 | 13 |
| 40 | 8 | 12 | 9 | 23 | 8 | 3 | 17 | 8 | 17 | 9 |
| 50 | 23 | 31 | 16 | 51 | 20 | 16 | 24 | 14 | 63 | 21 |
| 60 | 23 | 31 | 10 | 70 | 23 | 24 | 45 | 8 | 55 | 27 |
| 70 | 55 | 44 | 18 | 79 | 44 | 41 | 45 | 16 | 81 | 47 |
| 80 | 60 | 25 | 29 | 67 | 61 | 77 | 54 | 35 | 91 | 48 |
| 90 | 87 | 15 | 15 | 45 | 76 | 63 | 28 | 18 | 67 | 48 |
| 100 | 147 | 18 | 19 | 38 | 68 | 99 | 25 | 25 | 71 | 58 |
| Total | 406 | 185 | 124 | 398 | 324 | 335 | 265 | 136 | 476 | 281 |
| Proportion | 0.57 | 0.26 | 0.17 | 0.55 | 0.45 | 0.46 | 0.36 | 0.18 | 0.63 | 0.37 |

*Note.* Target-present lineups contained the culprit from the mock-crime video and five fillers, and target-absent lineups contained six fillers. The total row shows the number of each identification decision, collapsed over participants' identification confidence. The proportion row shows the total number of identification decisions, divided by the number of lineups in that condition. For example, the proportion of own-race culprit identifications in sequential-independent lineups is calculated by taking the total number of culprit IDs and dividing that by the total number of target-present own-race sequential-independent lineups, that is 354 / (354 + 210 + 123).

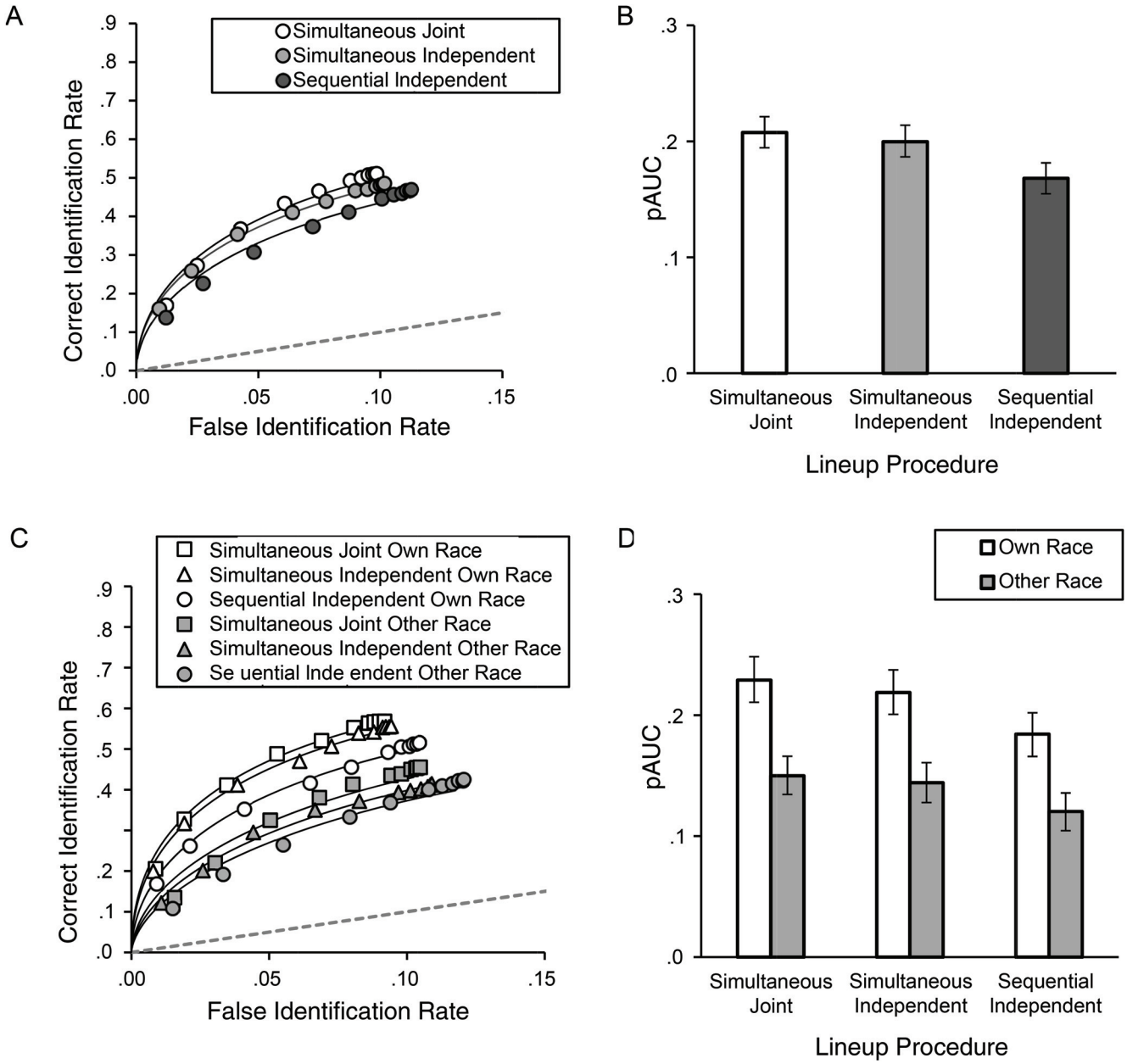
on eyewitness discriminability, but it will up to policymakers consider costs and normative and legal factors beyond the empirical data to determine which lineup procedure to use (Clark, 2012).

While policymakers should be interested in which procedures improve eyewitness discrimination accuracy so that eyewitnesses are best able to distinguish innocent from guilty suspects, legal decision-makers (e.g., judges, jurors, and police officers) are interested in the reliability of identification evidence (Mickes, 2015). Namely, legal decision-makers need to know "if a witness expresses high certainty in their suspect ID, how likely is this ID to be accurate?" In both experiments, the procedures that yielded poorer discriminability (i.e., sequential static in Experiment 1 and sequential interactive in Experiment 2), nevertheless achieved generally comparable reliability to the procedures that yielded better discriminability (i.e., sequential interactive in Experiment 1 and the simultaneous interactive

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

**Figure 4**

*Partial Receiver Operating Characteristic (ROC) Curves and Partial Area Under the Curve Statistics for (A, B) Sequential-Independent Movement, Simultaneous-Independent Movement, and Simultaneous-Joint Movement Lineups, Collapsed Over Own-Race and Other-Race Decisions, and (C, D) Own-Race and Other-Race Decisions in Sequential-Independent Movement, Simultaneous-Independent Movement, and Simultaneous-Joint Movement Lineups in Experiment 2*




*Note.* In A and C, ROC lines of best fit were drawn using parameters estimated by an unequal-variance signal-detection model, reported in the online supplemental materials, and the dashed lines represent chance-level performance. In B and D, error bars are 95% confidence intervals (CIs).

procedures in Experiment 2). This underscores the important distinction between discriminability versus reliability measures. Other research testing inventive novel lineup procedures—such as requiring witnesses to rate their confidence that each lineup member is the culprit (Brewer et al., 2020), or asking witnesses to eliminate members who they do not believe to be the culprit (e.g., Pozzulo & Lindsay, 1999)—seem promising when considering the reliability of identification evidence, but the effect of such procedures on discriminability is currently unknown and requires testing. As noted previously, implementing lineup procedures that enhance discrimination accuracy should be the goal of

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

**Table 4**
*Summary of Model-Predicted Discriminability ($d_a$) and Positive Identification Decision Criteria ($c_1$) in Experiment 2*

| | Simultaneous-joint | | Simultaneous-independent | | Sequential-independent | |
|---|---|---|---|---|---|---|
| Race | $d_a$ | $c_1$ | $d_a$ | $c_1$ | $d_a$ | $c_1$ |
| Own | 1.51 | 1.15 | 1.46 | 1.13 | 1.28 | 1.04 |
| Other | 1.08 | 1.01 | 1.00 | 0.99 | 0.93 | 0.86 |

*Note.* $d_a$ and $c_1$ are model-predicted discriminability and positive ID decision criteria parameters, respectively. Larger values of $d_a$ indicate better ability to discriminate between faces that have and have not been seen before. Larger values of $c_1$ indicate more conservative responding.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

policymakers, because procedures that enhance discrimination accuracy minimize the likelihood of incorrect innocent suspect identifications, while simultaneously maximizing the likelihood of correct guilty suspect identifications.

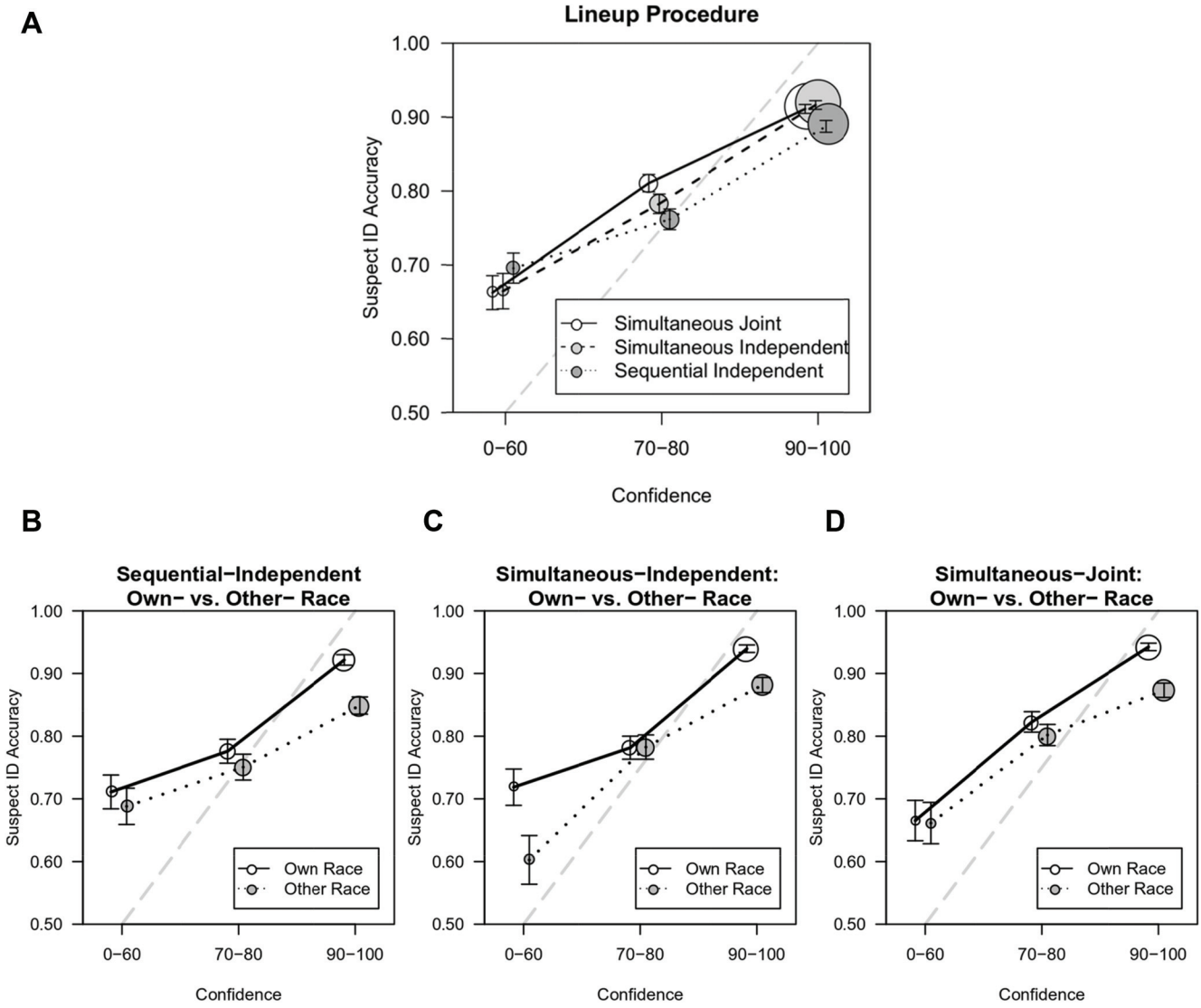
Our results also have implications for understanding of ORB. Studies suggest that the ORB is an encoding-based phenomenon, with multiple sociocognitive mechanisms likely contributing to people having a decreased ability to encode other-race faces (e.g., Meissner & Brigham, 2001; Sporer, 2001). Yet, nearly all of the experiments conducted to date have presented participants with static frontal pose faces at study and test (see Meissner & Brigham, 2001, for a meta-analysis, but also see Evans et al., 2009; Jackiw et al., 2008). Consequently, our ability to understand the role of information at retrieval in making other-race identifications was limited. The current experiments tested if retrieval support afforded by an interactive lineup could attenuate the ORB. We found that enabling participant witnesses to interact with the lineup faces and view them from multiple angles improved performance to the same extent for own- and other-race witnesses, further supporting the hypothesis that the ORB is an encoding-based phenomenon. Again, it is noteworthy that the simultaneous interactive lineup improved witness discrimination accuracy for both own- and other-race faces, given the ubiquity of the ORB and its prominence in wrongful conviction cases (e.g., Flowe et al., 2018; National Research Council, 2014).

Considering the reliability of own- and other-race identification decisions, in Experiment 1, we found that confidence increased with accuracy, and that the relationship between confidence and accuracy was the same for own- and other-race identifications. This finding suggests that the race of the witness and suspect is minimally important when legal decision-makers are estimating the likely accuracy of identification evidence made with a particular level of confidence. In Experiment 2, confidence also increased with accuracy for own- and other-race identifications, but high confidence other-race IDs were less likely to be accurate than high confidence own-race IDs across all three lineup presentation methods, suggesting that the race of the witness and suspect may be important for legal decision-makers to consider. On one hand, these differing findings might have occurred because discrimination accuracy was higher in Experiment 1 ($d'$ ranged from 1.89 to 2.54) compared with Experiment 2 ($d_a$ ranged from 0.92 to 1.54). In their analysis of studies that used a yes/no face recognition task, Nguyen et al. (2017) found a good confidence-accuracy relationship for own- and other-race participants when discrimination exceeded chance levels, but found a poorer confidence-accuracy relationship for other-race compared to own-race participants when discrimination accuracy did not exceed chance. Participants' performance in our Experiment 2 exceeded chance levels, but one possibility is that this relationship may not differ for own- and other-race identifications when memory strength is high, but it is weakened for other-race identifications when memory strength is relatively poor (see also Dodson & Dobolyi, 2016).

On the other hand, in Experiment 1, we tested participants over multiple trials, whereas in Experiment 2, we tested them with a single lineup. Perhaps this methodological difference is important to consider when assessing the confidence-accuracy relationship. Specifically, other-race participants in Experiment 1 compared with Experiment 2 may have been better able to judge the likely accuracy of their suspect identification due to practice effects accrued over multiple trials (Nguyen et al., 2017). Recollection is thought to be better for own-race faces, and own-race faces yield a lower proportion of high-confidence incorrect identifications, possibly because individuals qualitatively encode more information about own-race faces (Meissner, Brigham, & Butz, 2005). Previous research assessing the confidence-accuracy for own- and other-race faces has used multiple-trial designs (e.g., Dodson & Dobolyi, 2016; Nguyen et al., 2017). Future research should examine the impact of underlying memory strength and multiple-trial study design on the relationship between confidence and accuracy. Nevertheless, it is worth noting that for both own- and other-race identifications, increased confidence was associated with increased suspect ID accuracy, highlighting that witness confidence can be a useful (but imperfect) indicator of likely accuracy. This finding runs contrary to what was previously concluded in the eyewitness literature because misleading statistics, such as the point biserial correlation coefficient, were used and underestimated the relationship between witness confidence and accuracy (e.g., see Brewer & Wells, 2006; Juslin et al., 1996). Instead, the finding is consistent with positive assessments of

**Figure 5**

*Confidence Accuracy Characteristic Curves for (A) Sequential-Independent, Simultaneous-Independent, Simultaneous-Joint Lineups, Collapsed Over Own-Race and Other-Race Decisions; and Own-Race and Other-Race Decisions in (B) Sequential-Independent, (C) Simultaneous-Independent, and (D) Simultaneous-Joint Interactive Lineups*




*Note.* The dashed lines represent chance-level performance for the lowest confidence bin (0–60) and perfect performance at the highest confidence bin (90–100). The size of the symbols represents the number of suspect IDs at a given level of confidence relative to the number of suspect IDs collapsed over confidence (Seale-Carlisle et al., 2019). Error bars indicate ±1 *SE*, estimated by bootstrapping (Seale-Carlisle & Mickes, 2016).

the relationship between confidence and accuracy that have been observed more recently using appropriate statistical techniques, such CAC analysis and calibration approaches (e.g., Mickes, 2015; see Wixted & Wells, 2017, for a review).

In closing, the current research demonstrates that enabling witnesses to actively explore faces and view them from multiple angles in an interactive lineup can increase the discrimination accuracy of own- and other-race identifications. Interactive lineup procedures, especially if administered simultaneously, can increase the identification of guilty suspects, and simultaneously minimize the identification of innocent suspects and potentially prevent miscarriages of justice. Our findings exemplify how the basic science of face recognition and the analytical techniques tethered to signal-detection theory can be used to address the important applied policy issue on how best to conduct a police lineup.

## References

Albright, T. D. (2017). Why eyewitnesses fail. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(30), 7758–7764. https://doi.org/10.1073/pnas.1706891114

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Albright, T. D., & Rakoff, J. S. (2020). A clearer view: The impact of the National Academy of Sciences Report on Eyewitness Identification. *Judicature*, *104*(1), 21–29.

Behrman, B. W., & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior*, *25*(5), 475–491. https://doi.org/10.1023/A:1012840831846

Bornstein, B. H., Laub, C. E., Meissner, C. A., & Susa, K. J. (2013). The cross-race effect: Resistant to instructions. *Journal of Criminology*. Article ID, 745836. https://doi.org/10.1155/2013/745836

Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Behavior*, *26*, 353–364. https://doi.org/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. https://doi.org/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. https://doi.org/10.1037/1076-898X.12.1.11

Brewer, N., Weber, N., & Guerin, N. (2020). Police lineups of the future? *American Psychologist*, *75*, 76–91. https://doi.org/10.1037/amp0000465

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology*, *73*, 105–116. https://doi.org/10.1111/j.2044-8295.1982.tb01795.x

Carbon, C. C., & Leder, H. (2006). When faces are heads: View-dependent recognition of faces altered relationally or componentially. *Swiss Journal of Psychology*, *65*(4), 245–252. https://doi.org/10.1024/1421-0185.65.4.245

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

Clark, S. E., Moreland, M. B., & Rush, R. A. (2015). Lineup composition and lineup fairness. In T. Valentine & J. P. Davis (Eds.), *Forensic facial identification: Theory and practice of identification from eyewitnesses, composites and CCTV* (pp. 129–157). Wiley Blackwell.

Cochran, W. G. (1952). The $\chi^2$ test of goodness of fit. *Annals of Mathematical Statistics*, *25*, 315–345. https://doi.org/10.1214/aoms/1177729380

Colloff, M. F., Seale-Carlisle, T. M., Karoğlu, N., Rockey, J. C., Smith, H. M. J., Smith, L., Maltby, J., Yaremenko, S., & Flowe, H. D. (2021). Perpetrator pose reinstatement during a lineup test increases discrimination accuracy. *Scientific Reports*, *11*, Article 13830. https://doi.org/10.1038/s41598-021-92509-0

Colloff, M. F., Wade, K. A., & Strange, D. (2016). Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychological Science*, *27*(9), 1227–1239. https://doi.org/10.1177/0956797616655789

Cornell, S., & Hartmann, D. (2007). *Ethnicity and race: Making identities in a changing world*. Pine Forge Press.

Cumming, G. (2008). Inference by eye: Reading the overlap of independent confidence intervals. *Statistics in Medicine*, *28*, 205–220. https://doi.org/10.1002/sim.3471

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, *60*(2), 170–180. https://doi.org/10.1037/0003-066X.60.2.170

Cutler, B. L., & Fisher, R. P. (1990). Live lineups, videotaped lineups, and photoarrays. *Forensic Reports*, *3*(4), 439–448.

Cutler, B. L., Berman, G. L., Penrod, S., & Fisher, R. P. (1994). Conceptual, practical, and empirical issues associated with eyewitness identification test media. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 163–181). Cambridge University Press. https://doi.org/10.1017/CBO9780511759192.009

Dobolyi, D. G., & Dodson, C. S. (2013). Eyewitness confidence in simultaneous and sequential lineups: A criterion shift account for sequential mistaken identification overconfidence. *Journal of Experimental Psychology: Applied*, *19*, 345–357. https://doi.org/10.1037/a0034596

Dodson, C. S., & Dobolyi, D. G. (2016). Confidence and eyewitness identifications: The cross-race effect, decision time and accuracy. *Applied Cognitive Psychology*, *30*(1), 113–125. https://doi.org/10.1002/acp.3178

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

Duncan, M. (2006). *A signal detection model of compound decision tasks*. Defence Research and Development Canada.

Evans, J. R., Marcon, J. L., & Meissner, C. A. (2009). Cross-racial lineup identification: Assessing the potential benefits of context reinstatement. *Psychology, Crime & Law*, *15*(1), 19–28. https://doi.org/10.1080/10683160802047030

Fitzgerald, R. J., Price, H. L., & Valentine, T. (2018). Eyewitness identification: Live, photo, and video lineups. *Psychology, Public Policy, and Law*, *24*(3), 307–325. https://doi.org/10.1037/law0000164

Fitzgerald, R. J., Rubínová, E., & Juncu, S. (2021). Eyewitness identification around the world. In A. M. Smith, M. P. Toglia, & J. M. Lampinen (Eds.), *Methods, measures, and theories in eyewitness identification tasks*. Taylor and Francis.

Flowe, H. D., Carline, A., & Karoğlu, N. (2018). Testing the reflection assumption: A comparison of eyewitness ecology in the laboratory and criminal cases. *The International Journal of Evidence & Proof*, *22*(3), 239–261. https://doi.org/10.1177/1365712718782996

Gronlund, S. D., Mickes, L., Wixted, J. T., & Clark, S. E. (2015). Conducting an eyewitness lineup: How the research got it wrong. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 63, pp. 1–43). Academic Press.

Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using Receiver Operating Characteristic analysis. *Current Directions in Psychological Science*, *23*(1), 3–10. https://doi.org/10.1177/0963721413498891

Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition*, *62*, 201–222. https://doi.org/10.1016/S0010-0277(96)00785-8

Horry, R., Halford, P., Brewer, N., Milne, R., & Bull, R. (2014). Archival analyses of eyewitness identification test outcomes: What can they tell us about eyewitness memory? *Law and Human Behavior*, *38*, 94–108. https://doi.org/10.1037/lhb0000060

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychological Review*, *117*(4), 1168–1187. https://doi.org/10.1037/a0020463

Innocence Project. (2019). *DNA exonerations in the United States*. https://www.innocenceproject.org/dna-exonerations-in-the-united-states/

Jackiw, L. B., Arbuthnott, K. D., Pfeifer, J. E., Marcon, J. L., & Meissner, C. A. (2008). Examining the cross-race effect in lineup identification using Caucasian and First Nations samples. *Canadian Journal of Behavioural Science*, *40*(1), 52–57. https://doi.org/10.1037/0008-400x.40.1.52

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

Kaesler, M., Dunn, J. C., Ransom, K., & Semmler, C. (2020). Do sequential lineups impair underlying discriminability? *Cognitive Research: Principles and Implications*, *5*, 35. https://doi.org/10.1186/s41235-020-00234-5

Lewontin, R. C. (1972). The apportionment of human diversity. In T. Dobzhansky, M. K. Hecht, & W. C. Steere (Eds.), *Evolutionary biology* (pp. 381–398). Springer. https://doi.org/10.1007/978-1-4684-9063-3_14

Liu, C. H., Ward, J., & Markall, H. (2007). The role of active exploration of 3D face stimuli on recognition memory of facial information. *Journal*

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

*of Experimental Psychology: Human Perception and Performance*, *33*(4), 895–904. https://doi.org/10.1037/0096-1523.33.4.895

Longmore, C. A., Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(1), 77–100. https://doi.org/10.1037/0096-1523.34.1.77

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Erlbaum.

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology*, *13*, 330–334. https://doi.org/10.1037/h0028434

Mansour, J. K., Beaudry, J. L., Kalmet, N., Bertrand, M. I., & Lindsay, R. C. (2017). Evaluating lineup fairness: Variations across methods and measures. *Law and Human Behavior*, *41*(1), 103–115. https://doi.org/10.1037/lhb0000203

Marcon, J. L., Susa, K. J., & Meissner, C. A. (2009). Assessing the influence of recollection and familiarity in memory for own- versus other-race faces. *Psychonomic Bulletin & Review*, *16*, 99–103. https://doi.org/10.3758/PBR.16.1.99

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*(1), 3–35. https://doi.org/10.1037/1076-8971.7.1.3

Meissner, C. A., Brigham, J. C., & Butz, D. A. (2005). Memory for own- and other-race faces: A dual-process approach. *Applied Cognitive Psychology*, *19*(5), 545–567. https://doi.org/10.1002/acp.1097

Meissner, C. A., Tredoux, C. G., Parker, J. F., & MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. *Memory & Cognition*, *33*, 783–792. https://doi.org/10.3758/BF03193074

Mickes, L. (2015). Receiver operating characteristic analysis and confidence–accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory & Cognition*, *4*(2), 93–102. https://doi.org/10.1016/j.jarmac.2015.01.003

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

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

Nguyen, T. B., Pezdek, K., & Wixted, J. T. (2017). Evidence for a confidence-accuracy relationship in memory for same- and cross-race faces. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *70*(12), 2518–2534. https://doi.org/10.1080/17470218.2016.1246578

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition*, *22*, 208–224. https://doi.org/10.3758/BF03208892

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences*, *6*, 261–266. https://doi.org/10.1016/S1364-6613(02)01908-3

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

People v. Boone, 30 N.Y.3d 521 (2017).

Pike, G., Brace, N., & Kynan, S. (2002). *The visual identification of suspects: Procedures and practice. Home Office Briefing Note*. Home Office Research, Development and Statistics Directorate. https://pdfs.semanticscholar.org/dea9/e198cb32585a2aa33e7b4e26175a26d58599.pdf

Platz, S. J., & Hosch, H. M. (1988). Cross-racial/ethnic eyewitness identification: A field study. *Journal of Applied Social Psychology*, *18*(11), 972–984. https://doi.org/10.1111/j.1559-1816.1988.tb01187.x

Pozzulo, J. D., & Lindsay, R. C. L. (1999). Elimination lineups: An improved identification procedure for child eyewitnesses. *Journal of Applied Psychology*, *84*(2), 167–176. https://doi.org/10.1037/0021-9010.84.2.167

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

Rotello, C. M., & Chen, T. (2016). ROC curve analyses of eyewitness identification decisions: An analysis of the recent debate. *Cognitive Research: Principles and Implications*, *1*, 10. https://doi.org/10.1186/s41235-016-0006-7

Rubínová, E., Fitzgerald, R. J., Juncu, S., Ribbers, E., Hope, L., & Sauer, J. D. (2021). Live presentation for eyewitness identification is not superior to photo or video presentation. *Journal of Applied Research in Memory & Cognition*, *10*(1), 167–176. https://doi.org/10.1016/j.jarmac.2020.08.009

Sauerland, M., & Sporer, S. L. (2009). Fast and confident: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology: Applied*, *15*, 46–62. https://doi.org/10.1037/a0014560

Seale-Carlisle, T. M., & Mickes, L. (2016). U.S. line-ups outperform U.K. line-ups. *Royal Society Open Science*, *3*(9), 160300. https://doi.org/10.1098/rsos.160300

Seale-Carlisle, T. M., Colloff, M. F., Flowe, H. D., Wells, W., Wixted, J. T., & Mickes, L. (2019). Confidence and response time as indicators of eyewitness identification accuracy in the lab and in the real world. *Journal of Applied Research in Memory & Cognition*, *8*(4), 420–428. https://doi.org/10.1016/j.jarmac.2019.09.003

Seale-Carlisle, T. M., Wetmore, S. A., Flowe, H. D., & Mickes, L. (2019). Designing police lineups to maximize memory performance. *Journal of Experimental Psychology: Applied*, *25*, 410–430. https://doi.org/10.1037/xap0000222

Silbert, R., Hollway, J., & Larizadeh, D. (2015). *Criminal (in)justice: A cost analysis of wrongful convictions, errors, and failed prosecutions in California's criminal justice system*. Warren Institute on Law and Social Policy.

Smith, A. M., Yang, Y., & Wells, G. L. (2020). Distinguishing between investigator discriminability and eyewitness discriminability: A method for creating full receiver operating characteristic curves of lineup identification performance. *Perspectives on Psychological Science*, *15*, 589–607. https://doi.org/10.1177/1745691620902426

Smith, H. M. J., Andrews, S., Baguley, T., Colloff, M. F., Davis, J. P., White, D., Rockey, J. C., & Flowe, H. D. (2021). Performance of typical and superior face recognisers on a novel interactive face matching procedure. *British Journal of Psychology*. Advance online publication. https://doi.org/10.1111/bjop.12499

Sporer, S. L. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law*, *7*(1), 36–97. https://doi.org/10.1037/1076-8971.7.1.36

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. https://doi.org/10.1037/0033-2909.118.3.315

Sporer, S. L., Trinkl, B., & Guberova, E. (2007). Matching faces: Differences in processing speed of out-group faces by different ethnic groups. *Journal of Cross-Cultural Psychology*, *38*(4), 398–412. https://doi.org/10.1177/0022022107302310

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*(1), 99–139. https://doi.org/10.1037/a0021650

Susa, K. J., Meissner, C. A., & de Heer, H. (2010). Modeling the role of social-cognitive processes in the recognition of own- and other-race

faces. *Social Cognition*, *28*(4), 523–537. https://doi.org/10.1521/soco.2010.28.4.523

Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement*. National Institute of Justice.

Tredoux, C. G. (1998). Statistical inference on lineup measures. *Law and Human Behavior*, *22*(2), 217–237. https://doi.org/10.1023/A:1025746220886

Valentine, T., Darling, S., & Memon, A. (2007). Do strict rules and moving images increase the reliability of sequential identification procedures? *Applied Cognitive Psychology*, *21*(7), 933–949. https://doi.org/10.1002/acp.1306

Wells, G. L., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C. A., & Wixted, J. T. (2020). Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior*, *44*(1), 3–36. https://doi.org/10.1037/lhb0000359

Wilson, B. M., Donnelly, K., Christenfeld, N., & Wixted, J. T. (2019). Making sense of sequential lineups: An experimental and theoretical analysis of position effects. *Journal of Memory and Language*, *104*, 108–125. https://doi.org/10.1016/j.jml.2018.10.002

Winsor, A., Flowe, H. D., Seale-Carlisle, T. M., Killeen, I. M., Hett, D., Jores, T., Ingham, M., Lee, B. P., Stevens, L. M., & Colloff, M. F. (in press). Children's expressions of certainty are informative. *Journal of Experimental Psychology: General*.

Wixted, J. T. (2018). Time to exonerate eyewitness memory. *Forensic Science International*, *292*, e13–e15. https://doi.org/10.1016/j.forsciint.2018.08.018

Wixted, J. T. (2020). The forgotten history of signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(2), 201–233. https://doi.org/10.1037/xlm0000732

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

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, *18*(1), 10–65. https://doi.org/10.1177/1529100616686966

Wixted, J. T., Vul, E., Mickes, L., & Wilson, B. M. (2018). Models of lineup memory. *Cognitive Psychology*, *105*, 81–114. https://doi.org/10.1016/j.cogpsych.2018.06.001

Woodward, P. M. (1953). *Probability and information theory with applications to radar*. Pergamon Press.

Young, S. G., Bernstein, M. J., & Hugenberg, K. (2010). When do own-group biases in face recognition occur? Encoding versus post-encoding. *Social Cognition*, *28*(2), 240–250. https://doi.org/10.1521/soco.2010.28.2.240

Received July 9, 2020
Revision received January 25, 2021
Accepted January 27, 2021 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.