Psychology, Public Policy, and Law

ISSN: 1076-8971

2023, Vol. 29, No. 1, 19–31
https://doi.org/10.1037/law0000373

The Rule Out Procedure: A Signal-Detection-Informed Approach to the Collection of Eyewitness Identification Evidence

Andrew M. Smith, Nydia T. Ayala, and Rebecca C. Ying
Department of Psychology, Iowa State University

Visual recognition memory has a remarkable capacity to discriminate between previously seen and novel items. Yet, research on eyewitness lineups suggests that memory is useful for detecting culprit presence, but less useful for detecting culprit absence. We show that this asymmetry is predicted by the equal-variance signal-detection model. When witnesses reject lineups, they provide a global confidence rating that none of the lineup members is the culprit. These ratings do not scale match-to-memory for the suspect and are low in diagnostic value. Consequently, the equal-variance signal-detection model predicts that a one-person showup will have better discriminability than a six-person lineup. A large-scale experiment (N = 3281) supported that prediction. However, a modified lineup in which participants were asked to follow categorical identification decisions by assigning a confidence rating to each lineup member had better discriminability than both the showup and the standard simultaneous lineup. We call this modified lineup the rule out procedure. Results also revealed a relatively weak confidence-accuracy relation for global rejections of lineups, but a much stronger confidence-accuracy relation for rejections of individual faces. Past failures to detect suspect innocence with lineups should be attributed to flawed design, not to limitations of visual recognition memory.

Keywords: eyewitness memory, eyewitness lineups, memory, signal detection theory, ROC analysis

Visual recognition memory has a remarkable capacity to discriminate between what has been seen before and what has not been seen before (Brady et al., 2008). Hence, it is unsurprising that in the process of solving crimes, the criminal justice system routinely presents eyewitnesses with photos of suspects. What is surprising is the criminal justice system's asymmetric treatment of different identification outcomes. Whereas the affirmative identification of a suspect is attributed to suspect guilt, the rejection of a lineup is not attributed to suspect innocence, but rather to a failure of the visual recognition system (Brandon & Davies, 1973; Clark & Wells, 2008; Clifford & Bull, 1978; Wells & Lindsay, 1980). But why should a system as remarkable as visual recognition memory be limited to detecting the presence of a face? Should a system this remarkable not also be useful for detecting the absence of a face? On one hand, we believe that visual recognition memory should be useful both for detecting culprit presence and for detecting culprit absence. On the other hand, 45 years of research on eyewitness identification provides compelling evidence that suspect identifications are far more diagnostic of guilt than rejections are of innocence (Wells et al., 2015). Moreover, whereas eyewitness confidence is strongly related to the accuracy of suspect identifications, it bears a more modest relation to the accuracy of rejection decisions (Brewer & Wells, 2006; Sporer et al., 1995; Wixted & Wells, 2017).

Why suspect identifications are more diagnostic of guilt than rejections are of innocence has baffled psychological scientists for decades. We offer a novel signal-detection-based account for this fundamental asymmetry. We then use this same model to imagine what a better eyewitness identification procedure might look like. To that end, we introduce the simultaneous lineup plus rule out procedure (Ayala et al., 2022). We demonstrate that the rule out procedure is superior to the two identification procedures most used in real cases: simultaneous lineups and showups.

The Purpose of an Eyewitness Lineup Procedure

An eyewitness lineup is a procedure in which a single suspect is surrounded by fillers (known-innocent persons) and presented to an eyewitness for an identification attempt. Lineup members can

This article was published Online First December 22, 2022.
Andrew M. Smith https://orcid.org/0000-0002-4184-9364
Nydia T. Ayala https://orcid.org/0000-0003-4073-8141
Rebecca C. Ying https://orcid.org/0000-0002-5666-7790
Our preregistration, model predictions, de-identified data, and analysis code are available on the following Open Science Framework Page: https://osf.io/k95py/?view_only=30fdc6b76a6f4a7f8a7e69e6f3dbd3a0.

A preprint of a draft of this article was published on ResearchGate on May 10, 2022: https://www.researchgate.net/publication/360504290_The_Rule_Out_Procedure_A_Signal-Detection-Informed_Approach_to_the_Collection_of_Eyewitness_Identification_Evidence.

Andrew M. Smith served as lead for conceptualization, formal analysis, visualization, and writing–original draft. Nydia T. Ayala served as lead for data curation and writing–review and editing and served in a supporting role for conceptualization. Rebecca C. Ying served in a supporting role for conceptualization, data curation, and writing–review and editing.

Correspondence concerning this article should be addressed to Andrew M. Smith, Department of Psychology, Iowa State University, Lagomarcino Hall, 1347 Stange Road, Ames, IA 50011, United States. Email: amsmith@iastate.edu

be presented to the witness either simultaneously or sequentially (Lindsay & Wells, 1985). The reason for surrounding a suspect with fillers is because many suspects are innocent. If a suspect is innocent, fillers offer that person protection from mistaken identification as many witnesses who would have identified the innocent suspect identify a filler instead (e.g., Greathouse & Kovera, 2009; Lee & Penrod, 2019; Smith et al., 2018, 2022). Although a filler identification is still a mistaken identification on the part of the eyewitness, the identification of a filler is a far better forensic outcome than is an innocent-suspect identification. Whereas the mistaken identification of an innocent suspect may lead to arrest and wrongful conviction, fillers are not at risk of arrest or wrongful conviction. Indeed, filler identifications are known errors and because police investigators know these identifications are errors, they discount them.

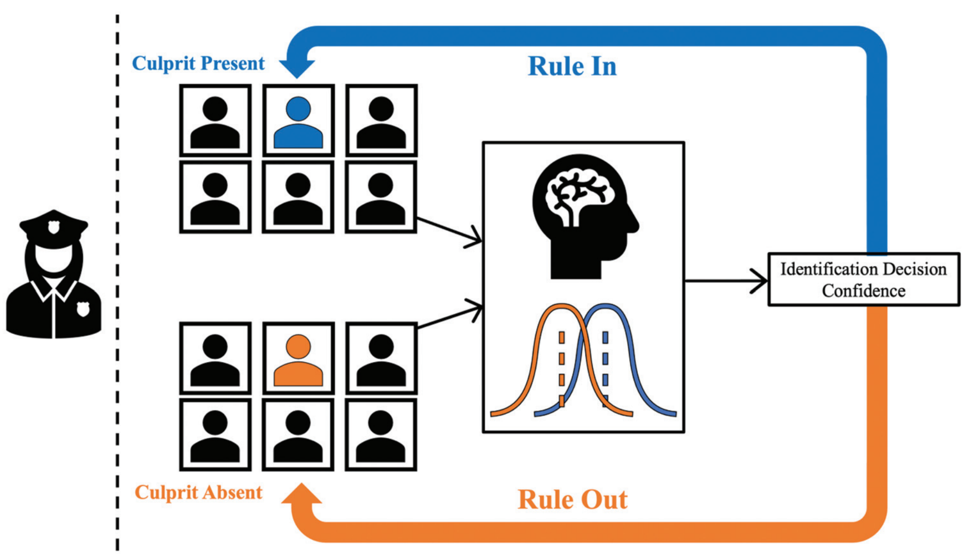
At a more conceptual level an eyewitness lineup is a psychometric instrument (see Figure 1; e.g., Gepshtein et al., 2020; Smith et al., 2020; Smith & Ayala, 2021; Starns et al., 2021). The investigator is tasked with determining whether the suspect in the lineup is guilty or innocent—in other words, the task of the investigator is to determine whether she has a culprit-present lineup or a culprit-absent lineup. Because the investigator cannot directly measure suspect guilt, she presents a lineup to the witness to measure how strongly the suspect matches the witness' memory for the culprit and uses this match value as a proxy for suspect guilt. The rationale is that the witness has a memory trace for the culprit from the initial crime scene and therefore a guilty suspect should provide a relatively strong match to memory and an innocent suspect a relatively weak match to memory. Because match to memory is a latent construct, the eyewitness is asked to make an identification decision and to qualify that decision with an expression of confidence. If the witness' behavior suggests that the suspect provides a strong match to memory—for example, she identifies the suspect with high confidence—the investigator might infer guilt. In other words, lineups can be used to "rule in" guilty suspects. But Figure 1 demonstrates another essential property of lineups: At least in theory, lineups can be used to "rule out" innocent suspects. For example, if a witness were to reject a lineup with high confidence, the investigator might infer that the suspect is innocent.

Eyewitness Lineups Are Flawed by Design

As we likened lineups to psychometric instruments, attentive readers might have questioned, what if the witness identifies a filler? Filler identifications are problematic. When a witness identifies the suspect, an expression of confidence provides nuanced

Figure 1

An Eyewitness Lineup Is a Psychometric Instrument for Measuring the Probability of Suspect Guilt



Note. The task of the investigator is to determine whether she has a culprit-present lineup (suspect is guilty) or a culprit-absent lineup (suspect is innocent). The silhouette on the far left of the figure represents the police investigator. In both lineups, the black-filled silhouettes represent known-innocent fillers. The blue-filled silhouette in the culprit-present lineup represents the culprit and the orange-filled silhouette in the culprit-absent lineup represents the innocent suspect. The black-filled silhouette in the middle of the figure represents the eyewitness and the Gaussian distributions represent the range of match-to-memory values for innocent suspects (orange filled) and guilty suspects (blue filled), respectively. Following convention, we assume that the guilty-suspect distribution is shifted to the right of the innocent-suspect distribution, reflecting that, on average, guilty suspects provide a stronger match-to-memory than do innocent suspects. The arrows extending from the witness' behaviors back to the lineups represent the hypothetical inferences the investigator might make about which lineup she has based on the witness' behavior. See the online article for the color version of this figure.

information about how well the suspect matches the witness' memory for the culprit, which in turn, informs on the probability that the suspect is guilty (e.g., Brewer & Wells, 2006; Palmer et al., 2013; Sauer et al., 2010; Wixted & Wells, 2017). But when a witness identifies a filler, confidence scales match-to-memory for that individual (the filler) and does not directly inform on the probability that the suspect is guilty. To put the magnitude of this problem in perspective, 28% of lineups conducted in laboratory settings and 24% of lineups conducted *with witnesses to actual crimes* result in filler identifications (Wells et al., 2015, 2020). It is tempting to infer from these figures that for approximately one quarter of all eyewitnesses, lineups do not measure what they are intended to measure—how strongly the suspect matches the witness' memory for the culprit. Unfortunately, that drastically underestimates the scope of the problem. As we show next, even when a witness rejects a lineup, confidence will often scale match to memory for a filler rather than for the police suspect.

We assume that identification decisions and expressions of confidence are driven, primarily, by the strength of match between the best-matching lineup member and the witness' memory for the culprit. If the best-matching lineup member exceeds some internal decision criterion the witness identifies that person, otherwise the witness rejects the lineup. Expressions of confidence can be thought of as reflecting the extent to which the best-matching lineup member exceeds or falls short of the witness' decision criterion (e.g., Clark et al., 2011; Smith & Ayala, 2021; Starns et al., 2021). Henceforth, we refer to this as the MAX decision rule and to the best-matching lineup member as the MAX lineup member (e.g., Starns et al., 2021; Yang & Burke, 2022).

For a fair six-person culprit-absent lineup, an innocent suspect will provide the MAX memory signal only 1/6th of the time. In other words, five in every six rejection decisions and associated expressions of confidence reflect match to memory for a filler rather than for the suspect. The staunchest of critics might argue that this is not problematic because confidence still indicates the witness' minimum degree of certainty that the suspect is innocent. That reasoning is deeply problematic. It is not difficult to envision a situation where a witness might tentatively reject a lineup because one of the fillers provides a relatively strong match to memory, but if the witness had been asked about the suspect directly she would have rejected him with absolute certainty. In fact, it is also easy to envision situations where witnesses who identify fillers would nevertheless be able to reject the police suspect with high confidence. In both scenarios the failure of the lineup to directly measure how strongly the suspect matches the witness' memory for the culprit means a loss in diagnostic information.

That eyewitness lineups often fail to measure how strongly the suspect matches the witness' memory for the culprit is a major design flaw. But merely establishing that lineups are flawed does not explain why rejection decisions are less diagnostic of innocence than suspect identifications are of guilt (Wells et al., 2015) and why confidence provides a good index of accuracy for suspect identifications but only a modest index of accuracy for rejection decisions (e.g., Brewer & Wells, 2006; Sporer et al., 1995; Wixted & Wells, 2017). Something more is needed.

Why Are Lineups More Effective at Demonstrating Guilt Than Innocence?

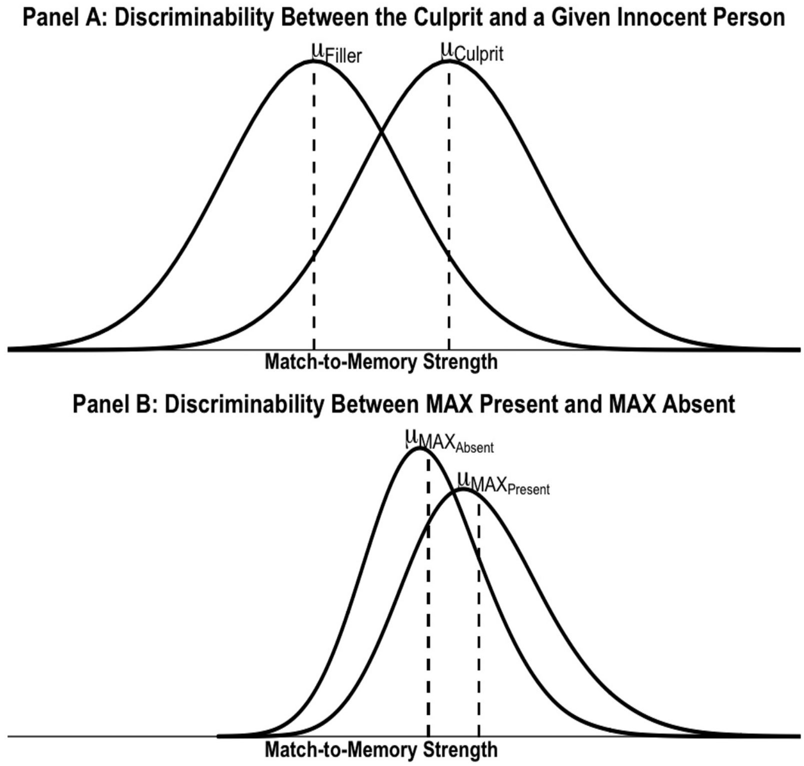
We use the equal-variance signal detection model with a MAX decision rule to explain why lineups are more effective at demonstrating guilt than they are at demonstrating innocence (Egan, 1958; Green & Swets, 1966; Macmillan & Creelman, 2005; Swets, 1973). For brevity, we refer to this as the MAX model. This model is depicted in Figure 2, Panel A. The *X*-axis is a random variable reflecting how strongly a lineup member matches the witness' memory for the culprit. We assume that match strength increases from left to right, that a culprit-absent lineup can be represented by six draws from the filler distribution ($\mu = 0.00, \sigma = 1.00$), and that a culprit-present lineup can be represented by one draw from the culprit distribution ($\mu = 1.50, \ \sigma = 1.00$) and five draws from the filler distribution. As the overlap between the culprit distribution and filler distribution decreases, the witness' ability to discriminate the guilty suspect from a given innocent person increases.

Figure 2, Panel A, also reflects the ability of an investigator to discriminate between guilty-suspect identifications and innocent-suspect identifications. Surprisingly, however, Figure 2, Panel A, does not reflect the ability of an investigator to discriminate between culprit-present rejections and culprit-absent rejections. As it turns out, this simple model predicts that discriminability will be far worse for rejection decisions relative to suspect identifications. Fortunately, there is a straightforward explanation for this asymmetry. There is only one suspect in the culprit-absent lineup and one suspect in the culprit-present lineup, and their status as suspects is determined before the witness ever sees the lineup. Therefore, the innocent suspect reflects one representative draw from the filler distribution and the guilty suspect reflects one representative draw from the culprit distribution. Hence, when a witness identifies the suspect, the ability of an investigator to discriminate between the guilty and the innocent is as depicted in Figure 2, Panel A.

But lineup rejections are not based on a priori representative draws. Lineup rejections are based on the strength of the MAX lineup member who, by definition, provides an atypically strong match-to-memory (i.e., the MAX lineup member provides a stronger match than the other five lineup members). Consequently, the distribution of match-to-memory values for the MAX lineup member is shifted to the right of the filler distribution from which each individual lineup member was drawn (Yang & Burke, 2022). Critically, discriminability for rejection decisions is defined by the overlap between the MAX distribution for the culprit-absent condition and the MAX distribution for the culprit-present condition (Figure 2, Panel B). Given the parameters in Figure 2, Panel A, the MAX culprit-absent distribution has a mean of 1.27 and a standard deviation of .65. The MAX culprit-present distribution has a mean of 1.83 and a standard deviation of .76. Hence, whereas discriminability between guilty-suspect identifications and innocent-suspect identifications was equal to 1.50 standard units (d' = 1.50), discriminability between culprit-present rejections and culprit-absent rejections is equal to only .79 standard units (d_a = .79). Hence, lineups as currently conducted are far better equipped to rule in guilty suspects than they are to rule out innocent suspects. The code for our instantiation of the MAX model that was used to produce these predictions is available on the Open Science Framework (Smith et al., 2022).

Figure 2

A Signal Detection-Based Explanation of Why Lineups Are More Effective at Demonstrating Guilt Than They Are at Demonstrating Innocence



Note. The ability of an investigator to discriminate guilty-suspect identifications from innocent-suspect identifications is depicted in Panel A. Given the parameters in Panel A, the ability of an investigator to discriminate culprit-present rejections from culprit-absent rejections is depicted in Panel B. By design, lineups undermine the potential for eyewitness memory to rule out innocent suspects.

This discovery is closely related to the "filler curtain effect" (Starns et al., 2021). When a witness identifies the suspect, the investigator learns that the suspect provides a stronger match to the witness' memory for the culprit than do the remaining lineup members and confidence provides a direct index of how strongly the suspect matches the witness' memory for the culprit. But lineup rejections are devoid of rank-order information—they do not tell the investigator whether the suspect provides a stronger match to memory than the remaining lineup members, a weaker match to memory than the remaining lineup members, or a match-to-memory value that falls somewhere in between. Further, confidence in a rejection decision will often measure match to memory for a filler rather than for the police suspect. This undermines the potential for eyewitness memory to rule out innocent suspects from police suspicion (Smith & Ayala, 2021).

There are numerous scenarios in which a witness who rejected a lineup with low confidence or even a witness who mistakenly identified a filler might be able to reject the suspect with high confidence. Maybe the witness rejected the lineup with low confidence because one of the fillers seemed sort of familiar and that gave the witness pause. But if the investigator had asked for more information about each of the individual lineup members, the witness might have been able to reject the innocent suspect with high confidence. Likewise, even a witness who mistakenly identified a filler might be able to reject an innocent suspect with high confidence (Ayala et al., 2022). There are countless scenarios where a witness might be able to provide more diagnostic information if asked to specifically speak to the degree of match between the suspect and her memory for the culprit. This information could help investigators to rule out innocent suspects. What is more, the sooner law enforcement personnel can clear the wrong person, the sooner they can start reallocating resources to finding the right person.

A Signal Detection-Informed Approach to Increasing Discriminability for Rejections

If the filler curtain effect explains why rejections are less diagnostic of innocence than suspect identifications are of guilt, then it should be possible to increase the diagnostic value of rejections by ensuring that rejections always reflect match to memory for the suspect. One procedure for which this situation holds is the showup. A showup involves presenting a lone suspect to the eyewitness for an identification attempt. However, both psychological

scientists (e.g., Steblay et al., 2003; Wells et al., 1998, 2020) and the legal system (e.g., Manson v. Braithwaite, 1977; Neil v. Biggers, 1972) have cautioned against the use of showups. Showups lead to higher rates of innocent-suspect identification and worse discriminability between guilty-suspect identifications and innocent-suspect identifications than do lineups (Smith et al., 2017; Wetmore et al., 2015). That is to say that showups are less effective at ruling in guilty suspects (they have a worse trade-off for culprit IDs and innocent-suspect IDs). Yet, more recent research focusing on both the potential to rule in guilty suspects and on the potential to rule out innocent suspects finds that showups are better at ruling out innocent suspects than are lineups (they have a better trade-off for correct rejections and misses). In fact, showups are so much more effective at ruling out innocent suspects that they have better overall discriminability than do lineups (Eisen et al., 2022; Starns et al., 2021). This creates a serious dilemma. In actual criminal cases, police investigators do not know when the suspect is guilty or innocent and fillers offer a safeguard for catching unreliable witnesses who otherwise might have identified an innocent suspect. Lineups do a better job than do showups of protecting innocent suspects and ruling in the guilty. However, showups are better at ruling out innocent suspects, so reasonable people can disagree about whether lineups or showups should be used in practice. To the extent one subjectively values culprit identifications more than correct rejections and views false alarms as more costly than misses, that person could make a reasonable argument that lineups should be preferred to showups despite their inferior discriminability.

Ideally, there would be a single procedure that could maximize the potential for eyewitness memory to both (a) rule in guilty suspects and (b) rule out innocent suspects. Such a procedure would be *objectively* superior to the alternatives. One procedure that might be able to achieve this lofty goal is the simultaneous lineup plus rule out procedure. Henceforth, we refer to this as the rule out procedure. The rule out procedure works as follows. The witness is first presented with a simultaneous lineup. She makes an identification decision and expresses her confidence in that decision. This is the point at which the standard simultaneous lineup concludes. But with the rule out procedure, the witness is presented with the same lineup for a second viewing and is asked to indicate for each individual she did not identify, how confident she is that individual is *not* the culprit. Hence, even when the witness rejects the lineup or identifies a filler, the rule out procedure ensures that there is always a confidence rating that directly scales the match between the suspect and the witness' memory for the culprit. We recently demonstrated that the rule out procedure does a better job of ruling out innocent suspects and has better overall discriminability than does the standard simultaneous lineup procedure. Moreover, rule-out ratings discriminated guilty suspects from innocent suspects both when witnesses rejected the lineup and even when witnesses mistakenly identified fillers (Ayala et al., 2022). The rule out procedure bears some resemblance to a ratings-based identification procedure that was developed over the past 15 years (e.g., Brewer et al., 2012; Sauer et al., 2008). We reserve a more in-depth comparison between the rule out and ratings-based procedures for our discussion section.

What remains unclear is how the discriminability of the rule out procedure compares to that of the showup. Because the rule out procedure starts with a simultaneous lineup, by definition, it must be as effective at ruling in the guilty as is the simultaneous lineup. Past research demonstrates that the simultaneous lineup is more effective at ruling in the guilty than is the showup procedure (e.g., Smith et al., 2017; Wetmore et al., 2015). Hence, the rule out procedure should be more effective at ruling in the guilty than is the showup procedure. Moreover, the signal detection model depicted in Figure 2 suggests that the reason lineups are less effective at ruling out innocent suspects than are showups is because of the filler curtain effect (see also Smith & Ayala, 2021; Starns et al., 2021). Because the rule out procedure eliminates the filler curtain effect, the simultaneous lineup procedure should be as effective at ruling out innocent suspects as is the showup. Putting everything together, we predict that the rule out procedure will have better overall discriminability than both the standard simultaneous lineup and showup procedures.

Method

Our preregistration, data analysis plan, de-identified data, and code can be accessed on the Open Science Framework (Smith et al., 2022). The Institutional Review Board at Iowa State University approved this experiment. We predicted that the rule out procedure would have better discriminability than the showup, which in turn, would have better discriminability than the standard simultaneous lineup. An a priori power analysis revealed that a sample size of 3142 participants was required to achieve 80% power (two-tailed) to detect a standardized effect of $d = .10$ using an independent-samples t test (Faul et al., 2007). We expected that some participants would provide unusable data due to failed attention checks, technological issues, or because they self-reported prior exposure to these stimuli in a past experiment and so we recruited 3,500 participants.

Participants

In total, 3443 Mechanical Turk workers participated in this study in exchange for $1.50 USD. We used the Cloud Research platform to expedite the data collection process (Litman et al., 2017). Participants completed informed consent and provided demographic information on a Qualtrics survey page. From there, participants were directed to a PsychoPy experiment (Peirce et al., 2019) where they completed the study and received a remuneration code. Fifty-seven participants completed the study on PsychoPy, but either did not complete the Qualtrics portion of the study or entered incorrect information, making it impossible for us to establish they had provided informed consent. Hence, our final dataset includes data from 3443 participants rather than 3500 participants. We excluded 162 participants who failed the attention check ($n = 83$), encountered a technological issue with the encoding video ($n = 18$), encountered a technological issue with the identification task ($n = 8$), or recognized the stimuli from a previous experiment ($n = 54$). One additional participant disclosed that she accidentally pushed the wrong key and made a response that she did not intend to make. Of the 3281 participants who provided valid data, 102 participants in the lineup condition skipped one or more of the confidence ratings associated with the rule out procedure. We analyzed our data both with these 102 participants included and excluded and reached identical conclusions in both

cases. Hence, we include these 102 participants with partially missing data in all analyses.

Of the 3281 participants who provided valid data, 73% (*n* = 2382) were White or Caucasian, 10% (*n* = 329) were Black or African American, 7% (*n* = 233) were Hispanic or Latino/Latina, 7% (*n* = 230) were Asian or Asian American, 3% (*n* = 106) indicated that they belonged to another race or ethnicity, and one person did not report. Fifty-nine percent (*n* = 1951) of participants were Female, 40% of participants were Male (*n* = 1304), and slightly less than 1% of participants belonged to another gender identity (*n* = 26). On average, participants were 39.25 years of age (*SD* = 12.36).

Design

We randomly assigned participants to a 2 (culprit: present, absent) × 2 (identification procedure: simultaneous lineup plus rule out, showup) between-participants design. Participants in the lineup condition completed the standard simultaneous lineup followed by the rule out procedure. Hence, we were able to use the data from this condition both to estimate performance for the standard simultaneous lineup and to estimate performance for the rule out procedure. Participants were also randomly assigned to view one of two different encoding videos. The sole purpose of this manipulation was to create some degree of stimulus sampling (Wells & Windschitl, 1999).

Material

Encoding Videos

Participants watched one of two versions of a video of a man committing a crime at an airport. The videos used different actors but depicted the same general event. The video shows several people standing in a check-in line, each with a small piece of luggage. At one point, the second person in line begins searching for his airline ticket and, not finding it, allows two people to move in front of him. He then switches his similar-looking suitcase with the suitcase belonging to the individual immediately in front of him. He then leaves the line, walking toward the camera for a close-up of his face. The video lasts approximately 90 seconds and the culprit is in view for 45 seconds. The culprit's face is in close-up view for approximately five seconds.

The first target was a White male of average build in his early 20s, with brown eyes and shaggy brown hair. The second target was a White male of average build, in his early 20s, with close-cut blonde hair and blue eyes. Both targets were clean-shaven and had no distinctive features (e.g., tattoos, piercings, scars).

Identification Procedures

All simultaneous lineups included photographs of six faces presented to participants in a 2 (rows) × 3 (columns) array. Culprit-present lineups included a photo of the culprit along with the photographs of five fillers who matched the general appearance of the culprit. Culprit-absent lineups included the photographs of six fillers. We counterbalanced which of the six culprit-absent fillers were included in the culprit-present lineup. The positions of lineup members were also randomized for both culprit-present and culprit-absent lineups.

Culprit-present showups included a photograph of the culprit and no photographs of fillers. Culprit-absent showups included the photograph of one filler. For both targets, we counterbalanced which of the six lineup fillers served as the innocent suspect for the showup.

Procedure

We used a similar procedure to that of Ayala et al. (2022). We instructed participants that they would watch a video and not to advance to the next page until they were ready to watch it. After viewing the video, participants completed a six-minute filler task (solving anagrams). Participants were instructed that they would now view an identification procedure that might or might not include the culprit from the initial encoding video and that they should identify the culprit if present and otherwise reject the lineup. After making an identification decision, participants were asked to express their level of confidence: 0%, 20%, 40%, 60%, 80%, or 100%. At this point the standard simultaneous lineup and showup procedures were complete. Participants in the lineup condition were now asked to complete the rule out procedure. Participants were asked to indicate for any person they did not initially identify, how confident they were that person *was not* the culprit. If a participant had initially rejected the lineup, a rating scale (0%, 20%, 40%, 60%, 80%, 100%) appeared above each lineup member in the top row of the lineup and below each lineup member in the bottom row of the lineup. If a participant had initially made an affirmative identification, the identified person's photo was present during the rule out procedure, but there was no rating scale attached to his photograph.

Results

We first describe the categorical outcomes for standard simultaneous lineups and showups. We then present the results of a suspect-identification Receiver Operating Characteristic (ROC) analysis to show that we replicated the typical finding that lineups better discriminate guilty-suspect identifications from innocent-suspect identifications than do showups. Next, we present the results of a full ROC analysis, that includes all lineup outcomes, to compare the overall discriminabilities of the standard simultaneous lineup, the showup, and the rule out procedure. Finally, we present supplemental analyses that help to clarify our primary results. In the process of analyzing the data, we used RStudio (RStudio Team, 2018), tidyverse (Wickham et al., 2019), pROC (Robin et al., 2011), boot (Canty & Ripley, 2021), gridExtra (Baptiste, 2017), and MBESS (Kelley, 2007).

Table 1 provides an overview of the categorical identification outcomes from both the simultaneous lineup and showup conditions. Because the rule out procedure results in the same categorical outcomes as the standard simultaneous lineup, it is not depicted in Table 1. The showup procedure led to a culprit identification rate of 72% and an innocent-suspect identification rate of 24%. The lineup led to a culprit identification rate of 48% and an innocent-suspect identification rate of 5%.

Table 1
Identification Decisions as a Function of Culprit Presence and Lineup Type

	Culprit presence					
	Culprit present			Culprit absent		
Procedure	Suspect ID	Filler ID	Rejection	Suspect ID	Filler ID	Rejection
Showup	72% (581)	—	28% (224)	24% (191)	—	76% (616)
Lineup	48% (396)	14% (118)	37% (306)	5%	26% (265)	69% (584)

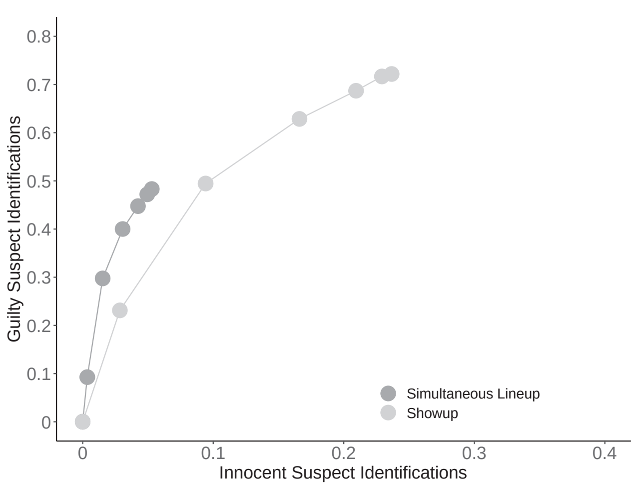
Note. The values in parentheses are frequencies. Showups do not include fillers and so there are no filler identifications for the showup conditions. We estimated the innocent-suspect identification rate for the culprit-absent lineup member by dividing the total culprit-absent false alarm rate (31%) by the number of lineup members (6).

Suspect-Identification ROC Curves

Numerous previous studies have demonstrated that lineups better discriminate guilty-suspect identifications from innocent-suspect identifications than do showups (e.g., Clark, 2012; Steblay et al., 2003; Wetmore et al., 2015). In other words, lineups have better potential to rule in the guilty and protect innocent suspects than do showups. Before comparing the overall discriminability of the procedures under investigation, we first examine whether the present experiment replicated the typical finding that lineups are more effective at ruling in the guilty than are showups.

Figure 3 depicts suspect-identification ROC curves for both the standard simultaneous lineup and showup procedures. In generating ROC curves, the goal is to order operating points from the strongest evidence of guilt (weakest evidence of innocence) to the weakest evidence of guilt (strongest evidence of innocence). The eyewitness' expressed level of confidence serves as a good proxy for evidence of guilt—high-confidence suspect identifications provide stronger evidence of guilt than do low-confidence suspect identifications. Accordingly, for both procedures, the leftmost operating point represents the origin, and the second-to-leftmost operating points represent the guilty- and innocent-suspect identification rates made with 100% confidence. The third-to-leftmost points represent the proportions of guilty- and innocent-suspect identifications made with *at least* 80% confidence. In other words, the third-to-leftmost operating point includes both suspect identifications made with 100% confidence and suspect identifications made with 80% confidence. One continues plotting points in this fashion until all the guilty-suspect identifications and all the innocent-suspect identifications are reflected in a single operating point. That is the operating point at which the suspect-identification ROC curve terminates. Because we did not have a designated innocent-suspect in the culprit-absent lineup, we divided the overall false alarm rate by the total number of lineup members (6) to estimate the innocent-suspect identification rate.

Figure 3
Suspect-Identification ROC Curves for Lineups and Showups



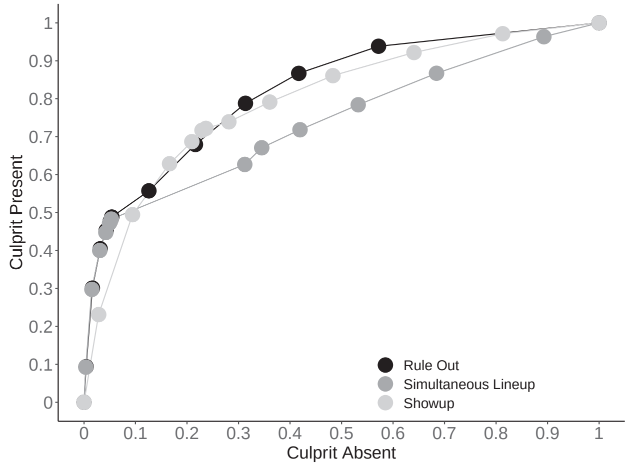
Note. In the instance that a witness identifies the suspect, there is no distinction between the standard simultaneous lineup and the rule out procedure. Therefore, the suspect identification ROC curves for the standard simultaneous lineup and rule out procedures are identical, and we omit the rule out curve to avoid redundancy.

Two things are immediately apparent from Figure 3 First, the showup curve spans a greater range on both the X (Innocent-Suspect Identifications) and Y (Guilty-Suspect Identifications) axes. This tells us that the showup results in both higher rates of guilty-suspect identification and innocent-suspect identification than does the standard simultaneous lineup. Second, the lineup curve bows above the showup curve, which suggests that the lineup might be more effective at discriminating guilty-suspect identifications from innocent-suspect identifications than is the showup procedure. To test this statistically, we compared the partial Area Under the ROC curves (pAUC) for the range of the *X*-axis that is covered by the standard simultaneous lineup (.00–.06). Replicating past research, the simultaneous lineup was more effective at discriminating guilty-suspect identifications from innocent-suspect identifications (pAUC = .021) than was the showup procedure (pAUC = .013), $D = 3.80$, $p < .001$.

Full ROC Curves Comparing Overall Discriminability

Figure 3 reflects only what happens when the witness identifies the suspect. Assessing the overall discriminability of a procedure, however, requires an examination of all possible lineup outcomes. This is because the overall discriminability of a lineup is determined both by its potential to rule in guilty suspects and by its potential to rule out innocent suspects. As previously mentioned, when generating ROC curves, operating points should be ordered from strongest evidence of guilt (weakest evidence of innocence) to weakest evidence of guilt (strongest evidence of innocence).

Figure 4
Full ROC Curves Comparing the Relative Discriminability of the Simultaneous Lineup, Showup, and Rule Out Procedures



Because suspect identifications provide stronger evidence of guilt than do rejections or filler identifications, the ROC curves depicted in Figure 4 include, and extend, the suspect-identification ROC curves that were depicted in Figure 3 For example, on a showup procedure, the strongest evidence of guilt is a 100%-confidence suspect identification, and the strongest evidence of innocence is a 100%-confidence rejection. Accordingly, we extended the suspect-identification showup ROC curve depicted in Figure 3 by adding a point for each level of rejection confidence and ordered these points in ascending order. The showup ROC curve depicted in Figure 4 now spans the entire width of the ROC space.

For the rule out procedure we generated an ROC curve based on the confidence ratings participants assigned to the suspect and then ordered those operating points from strongest evidence of guilt (100% confidence suspect identification) to strongest evidence of innocence (100% confidence rule out rating). In other words, to create the rule out ROC curve, we extended the suspect-identification lineup ROC curve depicted in Figure 3 by adding a point for each level of rule out confidence (0%, 20%, 40%, 60%, 80%, 100%) assigned to the suspect and ordered these points in ascending order. Because we did not have a designated innocent-suspect in the culprit-absent condition, we used all six confidence ratings that each participant provided to produce the rule out ROC curve. The result is that each operating point is equivalent to having divided the total number of culprit-absent outcomes at that interval by six (the total number of lineup members; e.g., Ayala et al., 2022).

For the standard simultaneous lineup ROC curve, we used the same suspect identification data that we used for the rule out procedure (once again assuming a perfectly fair lineup). However, rather than extending this curve with rule out ratings, we extended the standard simultaneous lineup curve by using filler identifications and confidence in global (or initial) rejection decisions. We first added a point to reflect the overall suspect identification rate plus the overall filler identification rate. We then added an additional point for each level of global rejection confidence (0%, 20%, 40%, 60%, 80%, 100%) and ordered these points in ascending order. The decision to include only one operating point for fillers collapsed over all levels of confidence and to embed that point between the suspect identifications and rejection decisions was based on the predictions of an equal-variance signal detection model with a MAX decision rule. According to that model, filler identifications should be low in diagnostic value and confidence in a filler identification should bear little relation to culprit presence (Smith & Ayala, 2021; Yang & Smith, 2022). Readers who are interested in more nuanced details about the computations are directed to our publicly available analysis code.

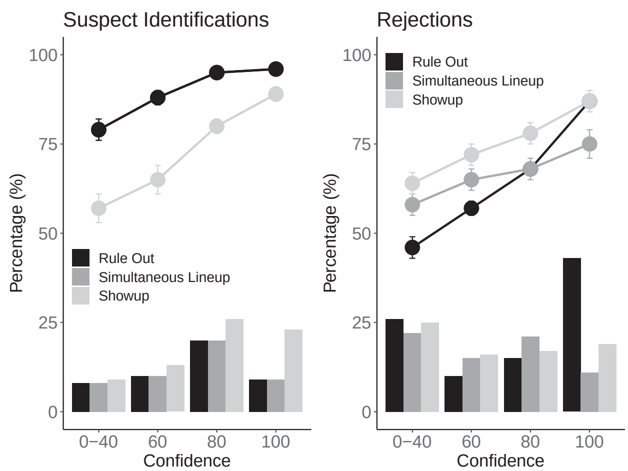
As we predicted, the showup (AUC = .7976) better discriminated between guilty suspects and innocent suspects than did the standard simultaneous lineup (AUC = .7409), $D = 3.49$, $p < .001$. Likewise, the rule out procedure (AUC = .8232) also better discriminated between guilty suspects and innocent suspects than did the standard simultaneous lineup (AUC = .7409), $D = 5.79$, $p < .001$. Finally, we found that the rule out procedure (AUC = .8232) led to a slight increase in discriminability when compared to the showup (AUC = .7976), $D = 1.92$, $p = .027$ (one-tailed). The benefit of the rule out procedure over the showup was slight even though the rule out curve visually dominated the showup curve over almost the entire width of the ROC space (see Figure 4).

Confidence-Accuracy Characteristic Analysis

Figure 5 depicts confidence-accuracy characteristic analyses for all identification procedures separated by suspect identifications and rejections (Mickes, 2015). The points and lines provide information about the confidence-accuracy relation and the bars provide information about the yield. For lineups, suspect identifications made with 100% confidence were correct 96% of the time and suspect identifications made with 80% confidence were correct 95% of the time. In short, high confidence implied high accuracy (Wixted & Wells, 2017). For showups, suspect identifications made with 100% confidence were correct 89% of the time, but identifications made with lower levels of confidence were far less accurate. These patterns largely replicate past research (e.g., Eisen et al., 2022; Sauerland et al., 2018; Wixted & Wells, 2017). For suspect identifications, the yield reflects the total proportion of culprit-present conditions leading to a particular outcome. For example, only 9% of witnesses who viewed a culprit-present lineup were able to identify the culprit with 100% confidence. That might be reason for concern if it were not for the fact that 20% of witnesses who viewed a culprit-present lineup were able to identify the culprit with 80% confidence and their accuracy was nearly as high as the 100%-confidence group (95% versus 96%).

Our primary objective was to examine why rejections from lineups are typically low in diagnostic value and to test a theoretically informed approach to increasing the diagnostic value of those decisions. In that respect, there are several noteworthy points that can be drawn from Figure 5. First, lineup rejections were relatively low in accuracy and as is evident from the relatively flat slope of the confidence-accuracy function, confidence was only a modestly related to accuracy. For example, rejections made with 100% confidence from lineups were accurate only 75% of the time. Second, and in contrast to lineups, for showups the confidence-accuracy function for rejections mirrored that of the confidence-accuracy

Figure 5
Confidence-Accuracy Characteristic Functions for Suspect Identifications and Rejections From Simultaneous Lineups, Showups, and Rule Out Procedures



Note. In the left panel, points represent suspect-identification accuracy at each level of confidence. In the right panel, points represent the proportion of correct rejections at each level of confidence. In both panels, bars represent the yield (or probability mass function). For suspect identifications (left panel), the yield is the proportion of culprit-present trials that resulted in a suspect identification at some expressed level of confidence. For rejections (right panel), the yield represents the proportion of culprit-absent trials that resulted in a correct rejection at some expressed level of confidence. For suspect identifications, the standard simultaneous lineup and rule out procedures have identical confidence-accuracy functions and the confidence-accuracy function for the standard simultaneous lineup is completely covered by the confidence-accuracy function for the rule out procedure.

function for suspect identifications—low confidence was associated with relatively low accuracy and high confidence was associated with relatively higher accuracy. In fact, rejections made with 100% confidence from showups were accurate 87% of the time. This asymmetry between lineups and showups is consistent with the model presented in Figure 2 and with the filler curtain effect more generally. In the absence of fillers, confidence should be a relatively good predictor of accuracy for both suspect identifications and rejection decisions and the showup data are consistent with that expected pattern. In the presence of fillers, rejection accuracy should be much lower than suspect-identification accuracy and the lineup data show that pattern.

Most importantly the rule out ratings showed a relatively strong confidence-accuracy relation. Rule out ratings made with 100% confidence were correct 87% of the time, which is the same 100%-confidence accuracy rate that was achieved by the showup. At lower levels of confidence, accuracy was lower for the rule out procedure than it was for the showup. This does not mean that, at the aggregate, the rule out procedure had worse rejection accuracy than did the showup. Inferences about rejection accuracy must also consider yield and the rule out procedure yielded a much higher rate of 100%-confidence correct rejections (43%) than did the showup procedure (19%). Because rejections made with 100%-confidence were much higher in accuracy than rejections made with lower levels of accuracy, the rule out procedure, if anything, had better rejection accuracy than did the showup (see the right side of Figure 4 where the rule out ROC curve dominates the showup ROC curve).

Can Witnesses Who Mistakenly Identify Fillers Be Used to Discriminate Between Guilty Suspects and Innocent Suspects?

We conclude our results section with an examination of whether rule out ratings are useful both for witnesses who initially rejected the lineup and for witnesses who initially identified a filler. A witness who identifies a filler has demonstrated that her memory is less than perfectly reliable. Yet, past research suggests that even these demonstrably unreliable eyewitnesses possess memorial information that can be used to discriminate between guilty suspects and innocent suspects (e.g., Ayala et al., 2022; Brewer et al., 2020; McAdoo & Gronlund, 2016). In our initial test of the rule out procedure, we found that rule out ratings discriminated between guilty suspects and innocent suspects even for those witnesses who mistakenly identified a filler (Ayala et al., 2022).

We computed an average rule-out rating for each culprit-absent witness. For witnesses who identified a filler, we averaged rule out ratings for the five lineup members the witness did not identify. For witnesses who rejected the lineup, we averaged rule out ratings for all six lineup members. We then compared these averages to the ratings that culprit-present witnesses assigned to the culprit. We used Welch independent-samples t-tests for these comparisons. Culprit-absent witnesses who rejected the lineup assigned higher rule out ratings to innocent persons (M = 74.30, SD = 22.41) than culprit-present witnesses who rejected the lineup assigned to the culprit (M = 46.03, SD = 31.85), $t(459.81) = 13.76$, $p < .001$, $d = .97$, (95% CI: .83, 1.12). Interestingly, culprit-absent witnesses who initially identified a filler assigned higher rule out ratings to innocent persons (M = 61.05, SD = 29.93) than culprit-present witnesses who identified a filler assigned to the culprit (M = 44.96, SD = 31.11), $t(204.92) = 4.65$, $p < .001$, $d = .51$, (95% CI: .29, .73). In summary, rule out ratings discriminated between guilty suspects and innocent suspects both when they came from witnesses who initially rejected the lineup and even when they came from witnesses who mistakenly identified fillers.

Discussion

We examined why lineup rejections are less diagnostic of innocence than suspect identifications are of guilt and tested a signal detection-informed approach for increasing the diagnostic value of lineup rejections. According to the MAX signal detection model (see Figure 2), lineup rejections often scale match to memory for a filler rather than for the suspect, which undermines the potential for lineups to rule out innocent suspects. Fillers were so detrimental to the rule out potential of lineups that lineups had worse discriminability than did showups (see Figure 4; Eisen et al., 2022; Starns et al., 2021). This pattern demonstrates that lineups are not objectively superior to showups and reinforces recent calls to consider all identification outcomes when making inferences about which of two identification procedures is superior (Smith et al., 2020; Starns et al., 2021). Whereas lineups are better at ruling in the guilty, showups are better at ruling out the innocent. Given this

trade-off, reasonable people can disagree about which procedure is a better investigative tool. Yet, it is important to appreciate that because showups lack fillers, they have no mechanism for catching mistaken identifications and might often lead to drastically high rates of innocent-suspect identification (e.g., Eisen et al., 2017; Smith et al., 2013, 2020; Wetmore et al., 2015). Fortunately, the criminal justice system need not be forced to choose between these two procedures. The rule out procedure was as effective at ruling in guilty suspects as was the standard simultaneous lineup and was as effective at ruling out innocent suspects as was the showup procedure. In other words, the rule out procedure was objectively superior to both the standard lineup and to the showup procedure.

We also found that the confidence-accuracy relation for rejection decisions was stronger when witnesses made decisions about individual faces compared to when witnesses made decisions about sets of faces. Others have inferred from this pattern that the reason the confidence-accuracy relation is weaker for lineup rejections than for suspect identifications must have to do with differences in how witnesses generate confidence judgments (e.g., Brewer & Wells, 2006; Lindsay et al., 2013; Sauerland et al., 2012; Yilmaz et al., 2022). According to this dual-process account, suspect-identification confidence is based on the degree of match between the suspect and the witness' memory for the culprit, but lineup-rejection confidence is based on integrating the individual match values for each lineup member into one global confidence judgment. This integration process is assumed to be suboptimal, attenuating the strength of the confidence-accuracy relation for lineup rejections. Our theoretical account shows that such elaborate theorizing is not needed to explain the confidence-accuracy asymmetry. Consider again the MAX signal detection model in Figure 2. For both suspect identifications and lineup rejections the MAX model assumes that confidence reflects the degree of match between the MAX lineup member and the witness' memory for the culprit. Even so, the MAX model still predicts a weaker confidence-accuracy relation for lineup rejections compared to suspect identifications. This is because confidence in a lineup rejection will often reflect match-to-memory for a filler who provides a better match to the witness' memory for the culprit than does the innocent suspect. The result is that witnesses who view culprit-absent lineups tend to be less confident that the culprit is absent from the lineup than they are that the suspect is innocent. Indeed, only 11% of culprit-absent participants were able to reject the entire set of lineup members with 100% confidence, but 43% of these same witnesses were able to reject the innocent suspect with 100% confidence (see Figure 5).

A major limitation of the standard simultaneous lineup is that it often results in filler identifications. In fact, a recent descriptive summary of 6,734 actual police lineups estimates that eyewitnesses identify a filler on 24% of lineups (Wells et al., 2020). Although fillers are essential for protecting innocent suspects from mistaken identification, when a standard simultaneous lineup culminates in a filler identification there is a huge loss of potential information. This is because when a witness identifies a filler, the investigator learns that at least one person provides a stronger match to memory than does the suspect (viz., the identified filler), but little else. Worse yet, confidence in a filler identification bears almost no relation to suspect guilt (Smith & Ayala, 2021). Yet, as we have shown here, even witnesses who mistakenly identify fillers possess additional memorial information that is useful for discriminating between guilty suspects and innocent suspects (Ayala et al., 2022; McAdoo & Gronlund, 2016). By not following up on filler identifications by directly probing the match between the suspect and the witness' memory for the culprit, the standard simultaneous lineup is failing to extract potentially diagnostic information from the witness' memory.

The rule out procedure led to slightly better discriminability than did the showup procedure (see Figure 4). As predicted, the rule out procedure was more effective at ruling in guilty suspects than was the showup procedure (see Figure 3). However, it also appears as though the rule out procedure may be slightly more effective at ruling out innocent suspects than is the showup (see the right side of Figure 4). For both the showup and rule out procedures rejections made with 100% confidence were correct 87% of the time. But the showup procedure led to far fewer 100%-confidence correct rejections (19%) than did the rule out procedure (43%). We can only speculate, but maybe this reflects a benefit of being able to compare faces (e.g., Meissner et al., 2005; Sauer et al., 2008; Wixted & Mickes, 2014). For example, if one or more of the fillers provided a stronger match to memory than did the suspect, that might have increased a witness' confidence that the suspect was innocent. Of course, witnesses do not know which lineup member is the suspect and which lineup members are fillers, but investigators are privy to this information and if relative signal strength tends to increase rule out ratings for innocent suspects, that could help investigators to discriminate between guilty suspects and innocent suspects.

Of course, maximizing discriminability is also desirable (and more desirable than adopting a more conservative test in our opinion). Fortunately, the simultaneous lineup plus rule out procedure affords the same protections to innocent suspects as does the simultaneous lineup, while also being as effective at ruling out the innocent suspect as the showup. Hence, the rule out procedure has better discriminability and is objectively superior to both the standard lineup and showup procedures. Given the observed benefits of the rule out procedure (see also Ayala et al., 2022), which were predicted a priori by formal quantitative theory, policy recommendations seem straightforward. The rule out procedure should be endorsed by the scientific community and implemented by law enforcement personnel. The rule out procedure simultaneously maximizes the potential for eyewitness memory to rule in guilty suspects and to rule out innocent suspects.

Finally, it is important to note the similarities between the rule out procedure and the ratings-based identification procedure (Brewer et al., 2012, 2020; Brewer & Doyle, 2021; Sauer et al., 2008, 2012; Sauer & Brewer, 2021). With the ratings-based identification procedure the witness is never makes a categorical identification decision and instead assigns a confidence rating to each lineup member. These confidence ratings are then used by an outside observer, such as the police investigator, to discriminate between guilty suspects and innocent suspects. The ratings-based identification procedure has much in common with the rule out procedure, but the overall discriminability of the ratings-based procedure has never been compared to either the standard simultaneous lineup or showup procedures. In theory, we would be surprised to find anything more than trivial differences between the ratings-based and rule out procedures. In fact, when we first discovered that the standard simultaneous lineup undermined the potential for eyewitness memory to rule out innocent suspects, we

theorized that the ratings-based procedure would prove superior (Smith & Ayala, 2021). However, in a subsequent review of the literature, we noted some points of resistance to the ratings-based procedure. For example, an identification procedure is the point in an investigation when a witness either formally accuses the suspect or does not (Dodson & Garrett, 2021; Steblay & Brooks, 2021). Because this is an important property of a lineup from a legal standpoint, the criminal justice system would likely be resistant to any identification procedure that does not secure a categorical identification decision. Hence, we developed the rule out procedure, which we view as hybrid of the standard simultaneous lineup and ratings-based identification procedures.

Conclusion

It seems clearer now than ever that eyewitness memory has a remarkable potential to detect when the culprit is absent from a lineup. The asymmetric value of suspect identifications and lineup rejections is attributable to a discrepancy between what the standard simultaneous lineup measures when a witness identifies a suspect compared to when a witness rejects a lineup (see Figure 2). When a witness identifies a suspect, police investigators learn that the suspect provides a stronger match to memory than do the remaining lineup members and confidence indexes the strength of this match. But when a witness rejects a lineup, investigators learn nothing about whether the suspect provided a stronger or weaker match to memory than did the other lineup members and confidence typically indexes match to memory for a filler rather than for the suspect. These global lineup rejections provide only weak evidence of innocence (Wells et al., 2015) and confidence in these decisions bears a relatively weak relation to accuracy (Brewer & Wells, 2006; Sporer et al., 1995; Wixted & Wells, 2017). In contrast, the present work demonstrates that the rejection of a specific face is relatively diagnostic of innocence and confidence in these specific rejections bears a relatively strong relation to accuracy. In fact, even for witnesses who mistakenly identified fillers rule out ratings discriminated between guilty suspects and innocent suspects.

It is difficult not to be excited about the potential applied implications of this work. The standard simultaneous lineup is flawed by design. When witnesses identify fillers or reject lineups, the standard simultaneous lineup typically does not measure match to memory for the suspect and thus, does not measure the probability that the suspect is guilty. This undermines the potential for eyewitness memory to rule out innocent suspects and the potential for police investigators to discriminate between guilty suspects and innocent suspects. We have now demonstrated these limitations theoretically (see Figure 2), computationally (Smith & Ayala, 2021), and empirically (Ayala et al., 2022). Clearly it is time to move on from this deeply flawed identification procedure. In fact, contrary to consensus among psychological scientists and legal scholars, we also demonstrated that the showup procedure better discriminates guilty suspects from innocent suspects than does the standard simultaneous lineup (Eisen et al., 2022; Starns et al., 2021). Given the inherent risk that a showup poses to a potentially innocent suspect, many will find that evidence unsettling. We too would be unsettled if it were not for the fact that there is an easy way to maintain the protective features of lineups, while also maximizing discriminability. After making categorical identification decisions, witnesses should be asked to provide a series of confidence ratings—one for each face depicted in the lineup.

References

Ayala, N. T., Smith, A. M., & Ying, R. C. (2022). The rule out procedure: Increasing the potential for police investigators to detect suspect innocence from eyewitness lineup procedures. *Journal of Applied Research in Memory & Cognition*. Advance online publication. https://doi.org/10.1037/mac0000018

Baptiste, A. (2017). gridExtra: Miscellaneous functions for “Grid” graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(38), 14325–14329. https://doi.org/10.1073/pnas.0803390105

Brandon, R., & Davies, C. (1973). *Wrongful imprisonment*. Allen & Unwin.

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, *12*(1), 11–30. https://doi.org/10.1037/1076-898X.12.1.11

Brewer, N., & Doyle, J. (2021). Changing the face of police lineups: Delivering more information from witnesses. *Journal of Applied Research in Memory & Cognition*, *10*(2), 180–195. https://doi.org/10.1016/j.jarmac.2020.12.004

Brewer, N., Weber, N., & Guerin, N. (2020). Police lineups of the future? *American Psychologist*, *75*(1), 76–91. https://doi.org/10.1037/amp0000465

Brewer, N., Weber, N., Wootton, D., & Lindsay, D. S. (2012). Identifying the bad guy in a lineup using confidence judgments under deadline pressure. *Psychological Science*, *23*(10), 1208–1214. https://doi.org/10.1177/0956797612441217

Canty, A., & Ripley, B. D. (2021). boot: Bootstrap R (S-Plus) Functions. *R Package version 1*.

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., & Wells, G. L. (2008). On the diagnosticity of multiple-witness identifications. *Law and Human Behavior*, *32*(5), 406–422. https://doi.org/10.1007/s10979-007-9115-7

Clark, S. E., Erickson, M. A., & Breneman, J. (2011). Probative value of absolute and relative judgments in eyewitness identification. *Law and Human Behavior*, *35*(5), 364–380. https://doi.org/10.1007/s10979-010-9245-1

Clifford, B. R., & Bull, R. (1978). *The psychology of person identification*. Routledge.

Dodson, C. S., & Garrett, B. L. (2021). Can we get over IDs? *Journal of Applied Research in Memory & Cognition*, *10*(2), 212–214. https://doi.org/10.1016/j.jarmac.2021.03.008

Egan, J. P. (1958). Recognition memory and the operating characteristic. *USAF Operational Applications Laboratory Technical Note*, 58–51, ii, 32.

Eisen, M. L., Smith, A. M., Olaguez, A. P., & Skerritt-Perta, A. S. (2017). An examination of showups conducted by law enforcement using a field-simulation paradigm. *Psychology, Public Policy, and Law*, *23*(1), 1–22. https://doi.org/10.1037/law0000115

Eisen, M. L., Ying, R. C., Chui, C., & Swaby, M. A. (2022). Comparing witness performance in the field versus the lab: How real-world conditions affect eyewitness decision-making. *Law and Human Behavior*, *46*(3), 175–188. https://doi.org/10.1037/lhb0000485

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Gepshtein, S., Wang, Y., He, F., Diep, D., & Albright, T. D. (2020). A perceptual scaling approach to eyewitness identification. *Nature Communications*, *11*(1), Article 3380. https://doi.org/10.1038/s41467-020-17194-5

Greathouse, S. M., & Kovera, M. B. (2009). Instruction bias and lineup presentation moderate the effects of administrator knowledge on eyewitness identification. *Law and Human Behavior*, *33*(1), 70–82. https://doi.org/10.1007/s10979-008-9136-x

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley.

Kelley, K. (2007). Methods for the behavioral, educational, and social sciences: An R package. *Behavior Research Methods*, *39*(4), 979–984. https://doi.org/10.3758/BF03192993

Lee, J., & Penrod, S. D. (2019). New signal detection theory-based framework for eyewitness performance in lineups. *Law and Human Behavior*, *43*(5), 436–454. https://doi.org/10.1037/lhb0000343

Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups. Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, *70*(3), 556–564. https://doi.org/10.1037/0021-9010.70.3.556

Lindsay, R. C. L., Kalmet, N., Leung, J., Bertrand, M. I., Sauer, J. D., & Sauerland, M. (2013). Confidence and accuracy of lineup selections and rejections: Postdicting rejection accuracy with confidence. *Journal of Applied Research in Memory & Cognition*, *2*(3), 179–184. https://doi.org/10.1016/j.jarmac.2013.06.002

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

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

Manson v. Braithwaite. (1977). 432 U.S. 98

McAdoo, R. M., & Gronlund, S. D. (2016). Relative judgment theory and the mediation of facial recognition: Implications for theories of eyewitness identification. *Cognitive Research: Principles and Implications*, *1*(1), Article 11. https://doi.org/10.1186/s41235-016-0014-7

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*(5), 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

Neil v. Biggers. (1972). 409 U.S. 188

Palmer, M. A., Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied*, *19*(1), 55–71. https://doi.org/10.1037/a0031602

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*(1), Article 77. https://doi.org/10.1186/1471-2105-12-77

RStudio Team. (2018). RStudio: Integrated development for R. RStudio, Inc. http://www.rstudio.com/

Sauer, J. D., & Brewer, N. (2021). Ratings-based identification procedures. In A. M. Smith, M. P. Toglia, & J. M. Lampinen (Eds.), *Methods, measures, and theories in eyewitness identification tasks* (pp. 192–210). Taylor & Francis. https://doi.org/10.4324/9781003138105-12

Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General*, *137*(3), 528–547. https://doi.org/10.1037/a0012712

Sauer, J. D., Weber, N., & Brewer, N. (2012). Using ecphoric confidence ratings to discriminate seen from unseen faces: The effects of retention interval and distinctiveness. *Psychonomic Bulletin & Review*, *19*(3), 490–498. https://doi.org/10.3758/s13423-012-0239-5

Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence-accuracy relationship for eyewitness identification. *Law and Human Behavior*, *34*(4), 337–347. https://doi.org/10.1007/s10979-009-9192-x

Sauerland, M., Sagana, A., & Sporer, S. L. (2012). Assessing nonchoosers' eyewitness identification accuracy from photographic showups by using confidence and response times. *Law and Human Behavior*, *36*(5), 394–403. https://doi.org/10.1037/h0093926

Sauerland, M., Sagana, A., Sporer, S. L., & Wixted, J. T. (2018). Decision time and confidence predict choosers' identification performance in photographic showups. *PLoS ONE*, *13*(1), Article e0190416. https://doi.org/10.1371/journal.pone.0190416

Smith, A. M., & Ayala, N. T. (2021). Do traditional lineups undermine the capacity for eyewitness memory to rule out innocent suspects? *Journal of Applied Research in Memory & Cognition*, *10*(2), 215–220. https://doi.org/10.1016/j.jarmac.2021.03.003

Smith, A. M., Ayala, N. T., & Ying, R. C. (2022). The rule out procedure: A signal-detection-infomed approach to the collection of eyewitness identification evidence. Open Science Framework. https://osf.io/k95py/?view_only=30fdc6b76a6f4a7f8a7e69e6f3dbd3a0

Smith, A. M., Leach, A.-M., & Cutler, B. L. (2013). Facilitating accuracy in showup identification procedures: The effects of the presence of stolen property. *Applied Cognitive Psychology*, *27*(2), 216–221. https://doi.org/10.1002/acp.2898

Smith, A. M., Mackovichova, S., Jalava, S. T., & Pozzulo, J. (2020). Fair forensic-object lineups are superior to forensic-object showups. *Journal of Applied Research in Memory & Cognition*, *9*(1), 68–82. https://doi.org/10.1037/h0101843

Smith, A. M., Smalarz, L., Wells, G. L., Lampinen, J. M., & Mackovichova, S. (2022). Fair lineups improve outside observers' discriminability, not eyewitness' discriminability: Evidence for differential filler-siphoning using empirical data and the WITNESS computer-simulation architecture. *Journal of Applied Research in Memory & Cognition*. Advance online publication. https://doi.org/10.1037/mac0000021

Smith, A. M., Wells, G. L., Lindsay, R. C. L., & Penrod, S. D. (2017). Fair lineups are better than biased lineups and showups, but not because they increase underlying discriminability. *Law and Human Behavior*, *41*(2), 127–145. https://doi.org/10.1037/lhb0000219

Smith, A. M., Wells, G. L., Smalarz, L., & Lampinen, J. M. (2018). Increasing the similarity of lineup fillers to the suspect improves applied value of lineups without improving memory performance. *Psychological Science*, *29*(9), 1548–1551. https://doi.org/10.1177/0956797617698528

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 procedures. *Perspectives on Psychological Science*, *15*(3), 589–607. https://doi.org/10.1177/1745691620902426

Sporer, S. L., Penrod, S. D., 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

Starns, J. J., Cohen, A., & Rotello, C. M. (2021). A complete method for assessing the effectiveness of eyewitness identification procedures: Expected information gain. *Psychological Review*. Advance online publication. https://doi.org/10.1037/rev0000332

Steblay, N. K., & Brooks, W. G. (2021). Practical concerns for investigations and Courtroom: A commentary on Brewer and Doyle (2021).

Journal of Applied Research in Memory & Cognition, *10*(2), 208–211. https://doi.org/10.1016/j.jarmac.2021.03.005

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

Swets, J. A. (1973). The relative operating characteristic in psychology: A technique for isolating effects of response bias finds wide use in the study of perception and cognition. *Science*, *182*(4116), 990–1000. https://doi.org/10.1126/science.182.4116.990

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentificaitons. *Psychological Bulletin*, *88*(3), 776–784. https://doi.org/10.1037/0033-2909.88.3.776

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, *25*(9), 1115–1125. https://doi.org/10.1177/01461672992512005

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

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, *22*(6), 603–647. https://doi.org/10.1023/A:1025750605807

Wells, G. L., Yang, Y., & Smalarz, L. (2015). Eyewitness identification: Bayesian information gain, base-rate effect equivalency curves, and reasonable suspicion. *Law and Human Behavior*, *39*(2), 99–122. https://doi.org/10.1037/lhb0000125

Wetmore, S. A., Neuschatz, J. S., Gronlund, S. D., Wooten, A., Goodsell, C. A., & Carlson, C. A. (2015). Effect of retention interval on showup and lineup performance. *Journal of Applied Research in Memory & Cognition*, *4*(1), 8–14. https://doi.org/10.1016/j.jarmac.2014.07.003

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., . . . Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), Article 1686. https://doi.org/10.21105/joss.01686

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

Yang, Y., & Smith, A. M. (2022). fullROC: An R package for generating and analyzing eyewitness-lineup ROC curves. https://doi.org/10.13140/RG.2.2.20415.94885/1

Yang, Y., & Burke, J. (2022). *A multidimensional signal detection model for eyewitness identification* [unpublished manuscript]. Department of Psychology, University of Nevada.

Yilmaz, A. S., Lebensfeld, T. C., & Wilson, B. M. (2022). The reveal procedure: A way to enhance evidence of innocence from police lineups. *Law and Human Behavior*, *46*(2), 164–173. https://doi.org/10.1037/lhb0000478

Received August 18, 2022
Revision received September 23, 2022
Accepted September 24, 2022 ■