A Comprehensive Evaluation of Showups

Chapter · June 2016

DOI: 10.1007/978-3-319-29406-3_2

Jeffrey S Neuschatz
University of Alabama in Huntsville

Stacy A. Wetmore
Roanoke College

Scott D Gronlund
University of Oklahoma

Charles Goodsell
Canisius College

A Comprehensive Evaluation of Showups

Jeffrey S. Neuschatz, Stacy A. Wetmore, Kylie N. Key, Daniella K. Cash, Scott D. Gronlund, and Charles A. Goodsell

A showup is a procedure in which police present an eyewitness with a single person, either live or in a photograph, for the purpose of identification (Valentine, Davis, Memon, & Roberts, 2012). Typically, these one-on-one confrontations occur in the field, and in close spatial and temporal proximity to the crime (Behrman & Davey, 2001). Showup identifications are an alternative to lineup identifications (which involve multiple individuals being presented to the witness); however, the U.S. Supreme Court, state courts, and social science researchers have stated repeatedly that showup identifications are less reliable than lineup identifications (Gronlund et al., 2012; *Stovall v. Denno*, 1967; Wetmore et al., 2015a). A study by Gronlund et al. (2012) found that showups led to worse identification performance than lineups. This finding has been confirmed by several other researchers (Clark & Godfrey, 2009; Steblay, Dysart, Fulero, & Lindsay, 2003; Wetmore et al., 2015a). Garrett (2011) reviewed trial transcripts from 160 DNA exoneration cases and found that 34 % (53/160) involved misidentifications from showups. Nevertheless, showups remain one of the most widely employed identification procedures (see Goodsell, Wetmore, Gronlund, and Neuschatz, 2013). The estimates of showup identifications as a percentage of all identifications range from 30 to 77 % (Gonzalez, Ellsworth, & Pembroke, 1993; McQuiston & Malpass, 2001). Therefore, showups are a significant percentage of all identifications and warrant further investigation.

This chapter examines showups and lineups, particularly the comparative claim that showups are inherently more suggestive than lineups. The chapter also examines

J.S. Neuschatz (✉) • K.N. Key • D.K. Cash
The University of Alabama in Huntsville, Huntsville, AL, USA
e-mail: neuschaj@uah.edu

S.A. Wetmore • S.D. Gronlund
University of Oklahoma, Norman, OK, USA

C.A. Goodsell
Canisius College, Buffalo, NY, USA

M.K. Miller, B.H. Bornstein (eds.), *Advances in Psychology and Law*,
Advances in Psychology and Law 1, DOI 10.1007/978-3-319-29406-3_2

if there are situations in which a showup is comparable in performance to a lineup. In the sections that follow, we operationally define terms important to the comparison of these identification procedures, as well as outline advantages and disadvantages of showups relative to lineups. The extant literature is meta-analyzed comparing showups and lineups, using traditional measures (e.g., correct identification rates, false identification rates, and ratio-based probative value measures). These analyses are followed by Receiver Operating Characteristic (ROC) analyses, where feasible. Finally, the confidence–accuracy relationship yielded by each identification procedure is assessed using both point-biserial correlations and confidence calibration. We conclude by addressing proposed theoretical explanations, real-world implications of the findings, and ways in which showup procedures might be improved to yield more accurate identifications.

Advantages and Disadvantages of Showups

In order to understand the advantages and disadvantages of showups, it is important to define some terms that will be used throughout the chapter. Choices made from both perpetrator present (PP) and perpetrator absent (PA) identification procedures are compared when trying to determine which procedure is superior. Perpetrator present lineups (or showups) contain the actual perpetrator; perpetrator absent lineups (or showups) replace the perpetrator with an innocent suspect designated by the researcher. This is only true in the lab (in the field, the suspect is not known to be innocent). There are two correct choices that can be made: correct identifications (i.e., identifications of the perpetrator) in a perpetrator present identification procedure or correctly indicating that the perpetrator is not there from a perpetrator absent procedure. There are three possible errors that can occur: false identifications, incorrect rejections, and filler identifications (lineups only). False identifications refer to identifications of innocent suspects; incorrect rejections arise when a rejection is made from a perpetrator present lineup (or showup); filler identifications refer to identifications of people in the lineup who are known to be innocent. Because fillers are known to be innocent, filler identifications are not considered harmful errors. If a filler is chosen, the lineup administrator knows that the witness is unsure of the perpetrator's identity and should not be trusted as an accurate eyewitness. We focus on three identification responses for purposes of evaluating identification performance: correct identifications from perpetrator present identification procedures, false identifications from perpetrator absent identification procedures, and the confidence with which those decisions are made.

Showup identifications may confer performance advantages over traditional lineups. First, showups have the potential to be conducted faster than lineups, and therefore implicate criminals or absolve innocent people of suspicion quickly. Second, as has been well documented, memory performance decreases with time, so it is better to test memory (e.g., administer an eyewitness identification test) at a short rather than long delay (e.g., Light, 1996). Consistent with basic memory research, research on eyewitness identifications has shown that witnesses tend to perform more poorly at identifying guilty suspects as time progresses (Clark & Godfrey, 2009). Consistent with the psychological research, the U.S. Supreme Court and state courts have held that one of the circumstances under which it is acceptable to conduct a showup is when it occurs close in time to the incident in question (see Gardner & Anderson, 2004; *Stovall v. Denno*, 1967). Thus, a potential advantage of showups is that they may result in better memory performance than lineups due to a shorter delay—retention interval—between the incident and the administration of the identification procedure. Given the shorter delay, the perpetrator may not have time to alter distinctive features noticed by the witness (e.g., clothing, facial hair). Thus, the speed at which a showup can be conducted should afford extra cues like clothing match that are likely not available in a lineup conducted after a delay. These extra cues could aid recognition memory.

Despite these potential advantages, the U.S. Supreme Court repeatedly has argued against the use of showups (*Stovall v. Denno*, 1967; *United States v. Wade*, 1967). In *Stovall*, the court argued that showing suspects singly for the purpose of identification was a widely condemned practice. Consequently, showups are widely thought to be suggestive and should, with few exceptions, not be used by law enforcement. From an applied research perspective, the disadvantages arise from three factors: administration, lack of fillers, and choosing rates. We discuss these next.

First, showups are difficult to administer in a double-blind manner. Double-blind administration of a lineup refers to a situation where both the lineup administrator and the eyewitness are unaware of the identity of the suspect before the identification is conducted. This is problematic because an investigator who knows who the suspect is can intentionally or inadvertently convey the “correct” answer to the eyewitness and therefore influence the identification decision (see Haw & Fisher, 2004), without the witness even being aware of the influence (Garrioch & Brimacombe, 2001). When evaluating a positive identification under these circumstances, it is impossible to know whether the identification is the product of the witness’s memory for the perpetrator, influenced by the investigator, or some combination of the two. Given that the purpose of an identification is to test the hypothesis that the suspect is the perpetrator, ruling out alternative explanations for positive identifications is critical. Haw and Fisher (2004) minimized witness contact with the lineup administrator, which resulted in decreased false identifications without reducing hits. It is important to note that blind administration has never, to our knowledge, been tested with showups; however, given the long history demonstrating experimenter bias in the psychology literature, it is reasonable to assume that the same concerns apply to showups (see Rosenthal, 1966; Rosenthal & Jacobson, 1968).

Second, there are no fillers in a showup, which means there is no protection against witnesses who are guessing (i.e., willing to respond despite very low confidence). The problem is compounded by the fact that the likelihood of choosing the suspect in showups is greater than in lineups ($1/n$ where n is greater than 2 in a lineup). Luus and Wells (1991) outlined some of the important functions of fillers in a well-constructed, fair lineup. A selection by the witness of a filler is a known error, which gives the lineup administrator information regarding the accuracy of the eyewitness. The filler also serves as a control for guessing—the suspect should not be chosen more often than each of the fillers if a witness has no memory of the perpetrator (Malpass & Lindsay, 1999).

Finally, a witness may be more likely to make an identification from a showup regardless of whether the perpetrator is present or not (see Dysart & Lindsay, 2007; Goodsell et al., 2013). It is commonly assumed that showups will produce higher choosing rates because they include only one individual—witnesses are more likely to choose if there is only one option (Kassin, Tubb, Hosch, & Memon, 2001). According to signal detection theory (Green & Swets, 1966), participants use a more liberal criterion when there is only one option. The higher choosing rate will lead to more correct identifications but also more false identifications of suspects. This translates into a greater potential of innocent people being identified, indicted, and potentially incarcerated from showups than lineups. In fact, in the Steblay et al. (2003) meta-analysis comparing showups and lineups, participants did have higher choosing rates in showups than lineups. Consistent with this meta-analysis, Meissner, Tredoux, Parker, and MacLin (2005) examined face memory utilizing identification procedures ranging in size from 1 to 12 options and found that criterion became more liberal as lineup size decreased, until lineup size equaled 1 (showup).

Regarding choosing rates, laboratory studies probably underestimate the actual choosing rates in showups because researchers cannot simulate the social pressure that a real eyewitness feels (Dysart, Lindsay, & Dupuis, 2006). Indeed, in their archival analysis of real-world cases, Behrman and Davey (2001) reported that the suspect identification rate in showups was 76 % as opposed to 48 % in lineups. According to the Best Practice Guidelines outlined by the Technical Working Group for Eyewitness Evidence (1999), using unbiased instructions, or indicating that the perpetrator may or may not be present in the identification procedure, may alleviate this pressure to choose. Typically, lineup administrators employ unbiased instructions, but this likely is less true for showups. Given the different circumstances surrounding a showup, unbiased instructions may be less effective. Although the use of unbiased instructions has never been examined using showup identification procedures, research has demonstrated that the expectation that a perpetrator is present in a lineup eliminates the protection of unbiased instructions (Quinlivan et al., 2011). There are several reasons to believe that the expectation of perpetrator presence are even greater in a showup than a lineup due to circumstances surrounding showups (proximity, time, one person).

Showups Versus Lineups

The remainder of the chapter is organized around the central question: Are showups always a less reliable identification procedure than lineups? To tackle this question, in the subsections that follow, we compare showups and lineups across a range of factors. However, we begin by collapsing over all conditions and then break down the data into conditions that are forensically or theoretically important to the question.

First, we compare showups to simultaneous (all photographs shown at the same time) and sequential (photographs shown one at a time) lineups. There has been a vigorous debate in the literature about which lineup yields better accuracy (for a review, see Clark, 2012; Gronlund, Andersen, & Perry, 2013; Steblay, Dysart, & Wells, 2011); therefore, we felt it is necessary to compare both simultaneous and sequential lineups to showups. Next, we compare showups to simultaneous and sequential lineups with suspects in different lineup positions (2 or 5). Previous research has identified suspect position as a factor that moderates the effectiveness of sequential lineups (Carlson, Gronlund, & Clark, 2008; Gronlund, Carlson, Dailey, & Goodsell, 2009). Gronlund et al. (2012) found that when the suspect appears early in the sequential lineup (position 2), performance was no better than a showup, but when the suspect appears late in the sequential lineup (position 5), sequential performance was as good as simultaneous performance. Finally, we examine variables that should supposedly provide an advantage for showups (i.e., short retention interval, similar cues at encoding and test). We also compared showups to fair and biased lineups; fair lineups contain fillers that all match the general description of a suspect, whereas biased lineups contain only one or two plausible options. Because of the lack of viable options in a biased lineup, a showup might be preferred to a biased lineup. We examined these comparisons using traditional meta-analysis, ROC analysis, point-biserial correlations, and confidence calibration.

Meta-analysis

Steblay et al. (2003) conducted the first meta-analytic comparison of showups and lineups. They analyzed eight published articles that included 12 tests of identification performance in showups and lineups, and found that the choosing rate, collapsed over perpetrator present and perpetrator absent conditions, was significantly higher in lineups (54 %) than in showups (27 %). Despite choosing less often, the laboratory data indicated that showup choices were more accurate: Correct decisions (perpetrator identifications from perpetrator present conditions + correct rejections from perpetrator absent conditions) were significantly higher in showups (69 %) than in lineups (51 %). In addition, the number of incorrect identifications from perpetrator absent conditions was significantly lower in showups (15 %) than in lineups (43 %). However, Steblay et al. (2003) argued that this was a misleading comparison because filler identifications from lineups were not "dangerous errors." Instead, they argued that it made more sense to compare "dangerous errors" involving only the identifications of an innocent suspect. This is important because researchers must compare perpetrator present and perpetrator absent conditions to get an understanding of the utility of an identification procedure—comparing the benefits of identifying the guilty to the costs of accusing the innocent.

Only five of the 12 tests reviewed by Steblay et al. (2003) included a designated innocent suspect. When they focused on just those studies (Dekle, Beal, Elliott, & Huneycutt, 1996; Yarmey, Yarmey, & Yarmey, 1994, 1996), they found that the false identification rate from showups (23 %) was higher than for lineups (10 %). But unfortunately, Steblay et al. failed to consider correct identifications. In the aforementioned studies the correct identification rate for showups (59 %) was greater than for lineups (39 %). Thus, the benefit of fewer false identifications was offset by the cost of fewer correct identifications. This complicates the decision about which identification procedure is best (see Clark (2012) for a discussion of the cost-benefit tradeoff regarding showups and lineups).

A second meta-analysis was conducted by Clark and Godfrey (2009). They included the five studies from Steblay et al. (2003), plus three additional studies, which resulted in a total of 15 showup–lineup comparisons (all with adult participants). Contrary to Steblay et al., Clark and Godfrey found that correct identification rates in perpetrator present conditions and innocent suspect identifications from perpetrator absent conditions were not significantly different between lineups and showups. Showups also were not found to significantly increase choosing rates relative to lineups.

Clark and Godfrey (2009), however, argued that focusing on correct identifications in perpetrator present and correct rejections in perpetrator absent lineups places lineups at a disadvantage. Witnesses can choose a filler in a lineup but not a showup, and every filler choice reduces the correct rejection rate of a lineup. Clark and Godfrey argued that conditional probability (CP = correct identification/[correct identifications + false identifications]) was a better measure of identification performance. Clark and Godfrey actually reported 1-CP or what they termed the innocence risk. The innocence risk for lineups (.21) was significantly less than for showups (.31), indicating that showups put an innocent suspect at greater risk of being falsely identified. However, this result was mitigated by retention interval. There was no difference at short retention intervals, but at 2- and 24-h delays, the innocence risk for showups was much higher than for lineups.

Updated Meta-analysis

We used the studies examined by the Clark and Godfrey (2009) meta-analysis, and studies published since. To be included, studies had to compare showups and lineups and use adult participants. When no designated innocent suspect was indicated, the false identification rates were estimated by using the probability of selecting any lineup member (1/number of lineup members).[1] There were a total of 21 showup–lineup comparisons, and we utilized the method of analysis in Clark and Godfrey (2009): *t*-tests to compare identification procedures and Cohen's *h* as a measure of effect size (see Cohen, 1988). The experiment, not the participants, is the unit of

[1] Although Lawson and Dysart (2014) included showups and lineups in their paper, they did not include correct identification rates. Attempts to get that information were unsuccessful.

analysis. The addition of these extra studies (compared to 13 in Clark and Godfrey and 8 in Steblay et al., 2003) also allowed us to conduct ROC analyses, which will be reported in the next section. Table 1 contains the correct identification rates, false identifications rates, and probative values for all the studies included in the analyses.[2]

As can be seen in Table 1, the correct and false identification rates for showups and lineups were similar and not significantly different. Consistent with Clark and Godfrey (2009), the only significant differences were observed for the probative value measure. This indicates that, given that the witness makes an identification, it is more likely that the witness will choose the perpetrator from a lineup than from a showup. Stated more simply, showups put innocent suspects at greater risk than lineups. In addition to overall correct and false identification rates, we compared showups conducted immediately to lineups conducted after a delay (defined as a retention interval greater than 24 h).[3] Because there are very few delayed lineups ($n=5$), we only present the descriptive statistics. Surprisingly, even with a retention interval greater than 24 h, there are more correct identifications ($M=.52$) and fewer false identifications ($M=.13$) from lineups than showups conducted at shorter retention intervals (Ms = .45 and .18 for the correct and false identification rates, respectively). Thus, although common sense might lead one to expect witnesses to display a showup advantage at a short retention interval, the data contradict this supposition.

However, before we accept any of these conclusions, we have to acknowledge the problems with using correct/false identification rates and ratio-based probative value measures to determine which identification procedure is superior. These measures of performance are confounded by response bias (different choosing rates across procedures). Thus, a measure like ROC analysis that can disentangle response bias from discriminability must be used (see Gronlund & Neuschatz, 2014; Wixted & Mickes, 2012).

ROC Analysis

As Gronlund and colleagues argue (Gronlund & Neuschatz, 2014; Gronlund, Wixted, & Mickes, 2014), ROC analysis can determine if an identification procedure results in a performance benefit (i.e., better discriminability of the perpetrator from the innocent suspect), a difference in response bias, or both. ROC analysis is grounded in signal-detection theory (Macmillan & Creelman, 2005) and is a well-established analytic technique used for evaluating diagnostic decisions in

[2] One measure of probative value (PV) is the probability of choosing the guilty suspect given that a suspect (innocent or guilty) was chosen (=[(guilty suspect identifications from perpetrator present lineups)/(guilty suspect identifications from perpetrator-absent lineups + innocent suspect identifications from perpetrator-absent lineups)]).

[3] Given the way the data were reported, it was impossible to analyze other variables, but we will examine those in the next section with ROC analyses.

Table 1 Proportion of correct and false identifications and conditional probabilities for studies included in the meta-analysis

Study	Lineup size	Delay	Lineup			Showup			Cohen's H		
			CID	FID	PV	CID	FID	PV	CID	FID	PV
Wagenaar and Veefkind (1992)[a]	2	0 h	.56	.12	.82	.35	.11	.76	.42	.03	.15
Wagenaar and Veefkind (1992)[a]	6	0 h	.5	.07	.88	.35	.11	.76	.30	−.14	.28
Wagenaar and Veefkind (1992)[a]	10	0 h	.42	.05	.89	.35	.11	.76	.14	−.23	.38
Wagenaar and Veefkind (1992) Exp. 2[a]	6	7 days	.75	.05	.94	.5	.15	.77	.52	−.34	.51
Gonzalez et al. (1993) **Exp. 1**	6	10 min	.64	.5	.56	.21	.07	.75	.90	1.04	.20
Gonzalez et al. (1993) **Exp. 2**	6	0 h	.25	.2	.56	0	.03	0.00	1.05	.61	2.09
Dekle et al. (1996)[a]	6	0 h	.3	.06	.83	.28	.04	.88	.04	.09	−.12
Dekle (1997)	6	2–3 days	.3	.06	.83	.35	.05	.88	−.11	.04	−.14
Dekle (1997)	6	7 days	.31	.1	.76	.3	.13	.70	.02	−.09	.16
Yarmey et al. (1994)[a]	6	5 min	.46	.05	.90	.57	.12	.83	−.22	−.26	.21
Yarmey et al. (1996)	6	0 h	.49	.16	.75	.7	.18	.80	−.43	−.05	−.12
Yarmey et al. (1996)	6	5 h	.39	.33	.54	.64	.44	.59	−.51	−.23	−.12
Yarmey et al. (1996)	6	2 h	.36	.14	.72	.54	.58	.48	−.36	−.96	.50
Yarmey et al. (1996)	6	1 day	.32	.14	.70	.55	.53	.51	−.47	−.86	.39
Lindsay, Pozzulo, Craig, Lee, and Corber (1997) **simultaneous**	6	0 h	.55	.06	.91	.5	.07	.88	.10	−.04	.10
Lindsay et al. (1997) **sequential**	6	0 h	.62	.04	.94	.5	.07	.88	.24	−.13	.21
Gronlund et al. (2012)	6	10 min	.65	.18	.78	.49	.24	.67	.32	.15	.32

(continued)

Table 1 (continued)

Study	Lineup size	Delay	Lineup CID	Lineup FID	Lineup PV	Showup CID	Showup FID	Showup PV	Cohen's H CID	Cohen's H FID	Cohen's H PV
Valentine et al. (2012)	6	15 min	.68	.12	.85	.65	.05	.93	.06	.26	−.25
Wetmore et al. (2015a)	6	0 h	.74	.28	.73	.62	.44	.58	.26	−.34	.30
Wetmore et al. (2015b)	6	2 days	.72	.3	.71	.56	.41	.58	.34	−.23	.27
Key et al. (2015)	6	0 h	.65	.18	.78	.44	.26	.63	.42	−.20	.34
M			.51	.15	.78	.45	.20	.70	.14	−.09	.27
SD									.41	.42	.47
t(20)									1.89	−.97	2.62

Note. Bold studies were included in Clark and Godfrey (2009)
[a]Studies were included in Steblay et al. (2003). Wetmore et al. (2015b) was included in the data analyses but not in the table because they only used showups. PV = CID/(CID+FID)

many different domains (Swets, Dawes, & Monahan, 2000). An eyewitness ROC curve is created by plotting the correct identification rate against false identification rate at each level of witness confidence (see Fig. 1 for an example). Filler identifications are excluded, just as they are from probative value calculations, because the identification of a filler is not a harmful error. The lower left-hand point on the ROC curve depicts the most confident correct and false identification responses (confidence level of 7 or the most conservative response bias). The next point on the ROC curve depicts the most and second-most confident correct and false identifications (confidence of 7 and 6), and so forth for the remainder of the confidence scale. Thus, ROCs depict a cumulative record of suspect identification decisions across the entire confidence scale. The data are summarized by a trend-line to better depict each curve.

The identification procedure whose ROC curve is closest to the upper left corner of the space exhibits better discriminability than ROCs that fall below it. Discriminability can be assessed by computing the area under the curve (AUC) for each identification procedure; the greater the AUC, the better the discriminability of that procedure. Lineup ROCs are truncated and do not trace out over the entire probability space (0–1) because the curve only extends to the highest false identification rate (recall that filler identifications are excluded). Therefore, the partial area under the curve (pAUC)[4] must be computed.

[4]pAUC values are computed using a false ID rate range from 0 to q, where q is set to a value slightly greater than the maximum false ID rate for the ROCs used in a comparison (see Wixted & Mickes, 2012). To evaluate different identification procedures, statistically compare the pAUCs and report the pAUCs, D, and p values. D is defined as (AUC1 – AUC2)/s, where s is the standard error of the difference between the two pAUCs (Robin et al., 2011). The standard error is estimated by the bootstrap method using 10,000 bootstraps (see Mickes, Flowe, & Wixted, 2012; for a tutorial, see Gronlund et al., 2014).

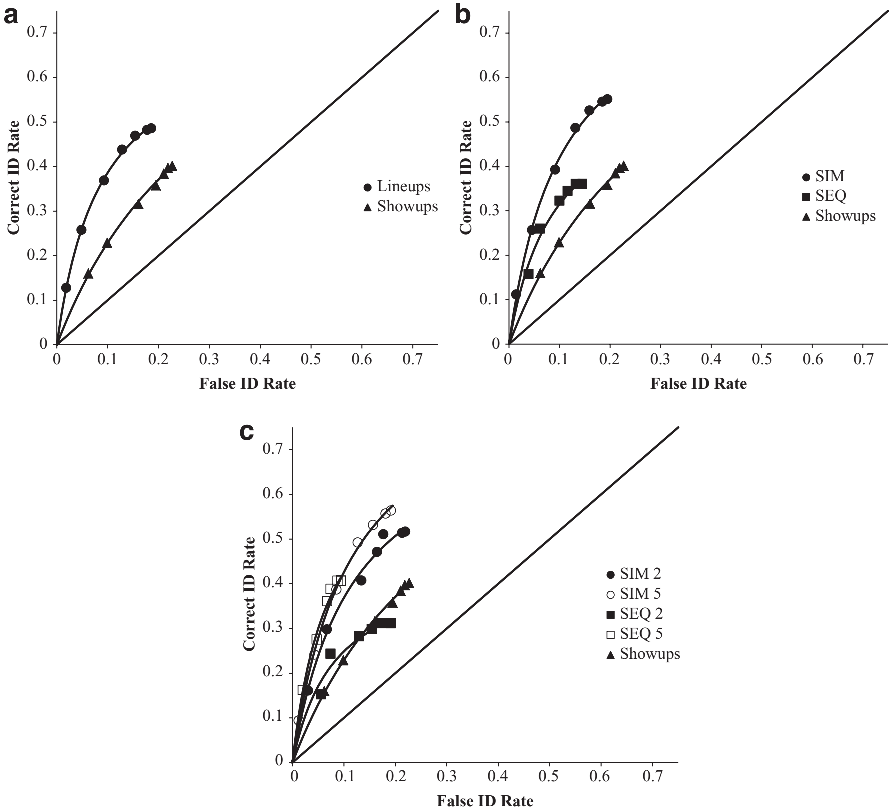

Fig. 1 ROC curves comparing lineups and showups. (**a**) Compares lineup and showups overall, (**b**) compares showups versus simultaneous and sequential lineups. (**c**) Compares showups versus lineups separated by lineup type and suspect position

In order to conduct ROC analyses, one needs to have confidence ratings associated with each identification response. This limited the sample of available studies because most researchers do not report confidence ratings in this way.[5] There were six such studies, which included a total of 7877 participants.[6] These studies are italicized in Table 1.

As we did for the evaluation of the data meta-analytically, we first review the ROC data by evaluating showups versus lineups overall, and then consider other variables pertinent to discriminability (e.g., lineup type, suspect position). ROC curves comparing showups and lineups, with showups and lineups broken down as

[5] Gronlund et al. (2012) utilized the same data reported in Gronlund et al. (2009). Wetmore et al. (2015a) had different data but utilized the same stimuli as Gronlund et al. (2009).

[6] A total of 1050 participants in the innocent strong condition were removed from the analyses because performance was poor. Clark (2012) also excluded these data from his meta-analysis.

a function of lineup type, and as a function of suspect position, are all presented in Fig. 1. Panel a shows that participants are better able to discriminate perpetrators from innocent suspects in lineups (pAUC = .08) than showups (pAUC = .05), $D = 7.59$, $p < .001$. Panel b shows that lineups, irrespective of presentation method, result in better performance than showups. The pAUC for the simultaneous lineup (.09) was significantly greater than that of the showup (.05), $D = 8.32$, $p < .001$. Similarly, the pAUC was greater for the sequential lineup (.07) than the showup, $D = 2.24$, $p = .02$. These findings are consistent with expert opinion that showups are less reliable than lineups (Kassin et al., 2001) and confirm what has been reported in previous studies (Gronlund et al., 2012; Wetmore et al., 2015a). Additionally, there was a significant difference between the pAUCs for the simultaneous versus sequential lineup, $D = 3.57$, $p < .001$. This finding is consistent with a number of studies that have recently found better discriminability from simultaneous than sequential lineups (Carlson & Carlson, 2014; Dobolyi & Dodson, 2013; Mickes et al., 2012).

Panel c of Fig. 1 shows the ROC curves for showups and lineups broken down by suspect position. Consistent with Gronlund et al. (2012), we found no difference in discriminability between showups and sequential lineups when the suspect was in position 2. The pAUCs for sequential position 2 (.05) and showups (.05) were not different, $D = .08$, $p = .93$. By contrast, the simultaneous lineup position 2 (pAUC = .08) and the showup were significantly different, $D = 2.80$, $p = .01$. Moreover, the remaining pAUCs were greater than that of the sequential position 2, and therefore significantly superior to showups, Ds > 3.03, $p < .001$, but not different from one another. In sum, simultaneous lineups are better than showups, as are sequential lineups if the suspect occurs late in the lineup.

Thus far, the analyses portray a dim view of showup identifications. However, they fail to consider two situations in which showups might be more reliable than lineups. First, as discussed previously, an immediate showup should confer a memorial advantage relative to a delayed lineup. Second, a showup may induce better discriminability when the clothing of the suspect matches what the perpetrator was wearing. According to the principle of encoding specificity (Tulving & Thomson, 1973), similar cues at study and test (same clothing) should enhance discriminability relative to different cues (different clothing). Moreover, a clothing match should be more likely in a showup because it generally takes place shortly after the crime and in close temporal and spatial proximity (see Valentine et al., 2012), making it less likely that a perpetrator could change clothing before the identification. Although some authors have reported that a clothing match increases false identifications and has no effect on correct identifications (Dysart Lindsay, & Dupuis, 2006; Yarmey et al., 1996), Wetmore, Neuschatz, Gronlund, Key, and Goodsell (2015b) recently found, using ROC analysis, that clothing match enhanced discriminability from showups relative to non-matching clothing. Therefore, it is possible that a clothing match might be beneficial to a showup identification.

To test the retention interval predictions, immediate showups were compared to simultaneous lineups conducted after a 2-day retention interval. This dataset did not include any sequential lineups conducted at a delay. Surprisingly, as can be seen

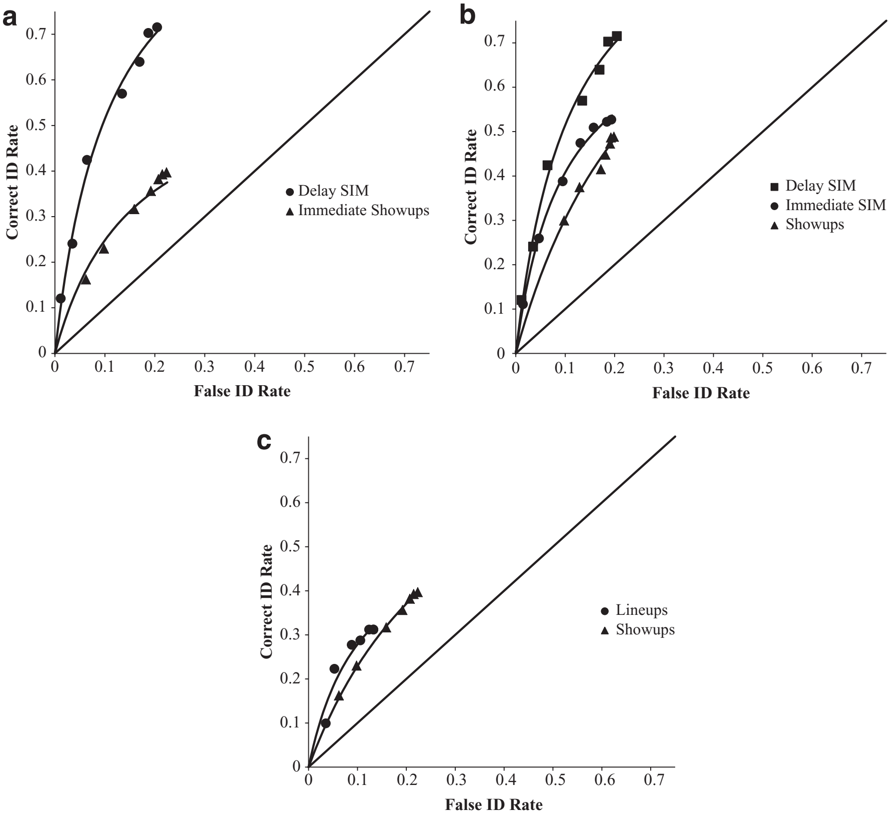

Fig. 2 ROC curves comparing simultaneous lineups and showups. (**a**) Depicts ROCs comparing immediate showups versus delayed simultaneous lineups. (**b**) Depicts immediate showups, match condition versus immediate and delay simultaneous lineups, no match. (**c**) Compares fair and biased lineups to showup conditions

in Panel a of Fig. 2, discrimination was significantly worse in showups conducted immediately (pAUC = .05) compared to simultaneous lineups after a 2-day delay (pAUC = .11), $D = 4.68$, $p < .001$.

To test the clothing match predictions, we compared immediate showups in which the clothing was the same during the crime and the identification (the match conditions from Wetmore et al., 2015b) to lineups that were conducted immediately or after a delay with no clothing match. As can be seen in Panel b of Fig. 2, discriminability was significantly better for simultaneous lineups conducted immediately (pAUC = .09) than for showups conducted immediately, (pAUC = .07), $D = 2.72$, $p < .001$, even when the showup suspects wore the same clothing during the crime and identification. Even the delayed simultaneous lineup (pAUC = .11) outperformed the immediate showup with clothing match, $D = 3.15$, $p < .001$.

It is important to note that there were other differences between the clothing match showups (Wetmore et al., 2015b) and the lineups in Wetmore et al. (2015a), including the use of video showups that showed the suspect's entire body, compared to lineup conditions that utilized photo headshots. However, because the showup revealed the suspect's entire body, it likely provided additional cues to help participants discriminate the guilty from the innocent suspect. Despite this, participants' discrimination ability was still inferior in the showup conditions.

Lineup fairness is another factor that could differentially affect eyewitness performance as a function of lineup procedure (Carlson et al., 2008; Gronlund et al., 2009). Although Wells and Quinlivan (2009) suggested that showups may put an innocent suspect at greater risk than a fair lineup, the showup may be better than placing an innocent suspect in a biased lineup. Lineup fairness refers to the degree to which the fillers in the lineup match the perpetrator. Traditionally, lineup fairness is assessed using Tredoux's *E,* which assesses the nominal size of the lineup—how many viable lineup members there are in the lineup. If a six-person lineup has a Tredoux's *E* of 4, it indicates that there are four viable alternatives from which the witness might choose. For the studies that manipulated lineup fairness (Gronlund et al., 2009; Wetmore et al., 2015a), a fair lineup required a Tredoux's *E* value greater than 4 and a biased lineup required a value near 1. Examples of the fair and biased lineup for the innocent suspect can be seen in Fig. 3.

The ROC curves for the fair lineup, biased lineup, and showup are shown in Fig. 2c. The fair lineup pAUC (.11) and biased lineup pAUC (.12) were significantly greater than the showup, pAUC (.07), $D=6.59$, $p<.001$ and $D=7.13$, $p<.001$, respectively. Surprisingly, there was no significant difference between the fair and biased lineup pAUCs, $D=-.95$, $p=.34$, indicating a similar ability to discriminate guilty from innocent suspects irrespective of lineup fairness. However, it would be wrong to conclude that the lack of a discriminability difference signals that our fairness manipulation was too weak because lineup fairness greatly impacted the range over which the curves extend.

The reason the ROC curve for the biased lineup extended over a greater range than the ROC curve for the fair lineup is apparent upon examination of Fig. 4. Figure 4 depicts possible underlying memory strength distributions portraying biased and fair lineups. The distribution in blue represents the range of different memory strengths for the perpetrator that different participants possess. That is, a few participants have an excellent memory for the perpetrator, most have a moderate memory, and a few have a poor memory for the perpetrator. Likewise, the distribution in green represents the varying degrees to which an innocent suspect matches participants' memory for the perpetrator. If a lineup is fair, the fillers should not stand out from an innocent suspect and therefore we can assume that the green distribution also represents the range of memory strengths of these fillers. However, if the lineup is biased it means that fillers are worse matches to the perpetrator. This poor filler distribution (in red) is shifted lower (has lower memory strength) than the distribution of good fillers (in green). Because ROC discriminability is a function of the overlap between the perpetrator and innocent suspect distributions, it is little affected by the distribution of biased fillers. Hence, there is no difference in discriminability as a function of lineup fairness. But the extent of the ROC curve arises from the placement of the confidence criteria (the vertical lines in the figure). The greater range over which the ROC curve for the biased lineup extends is due to

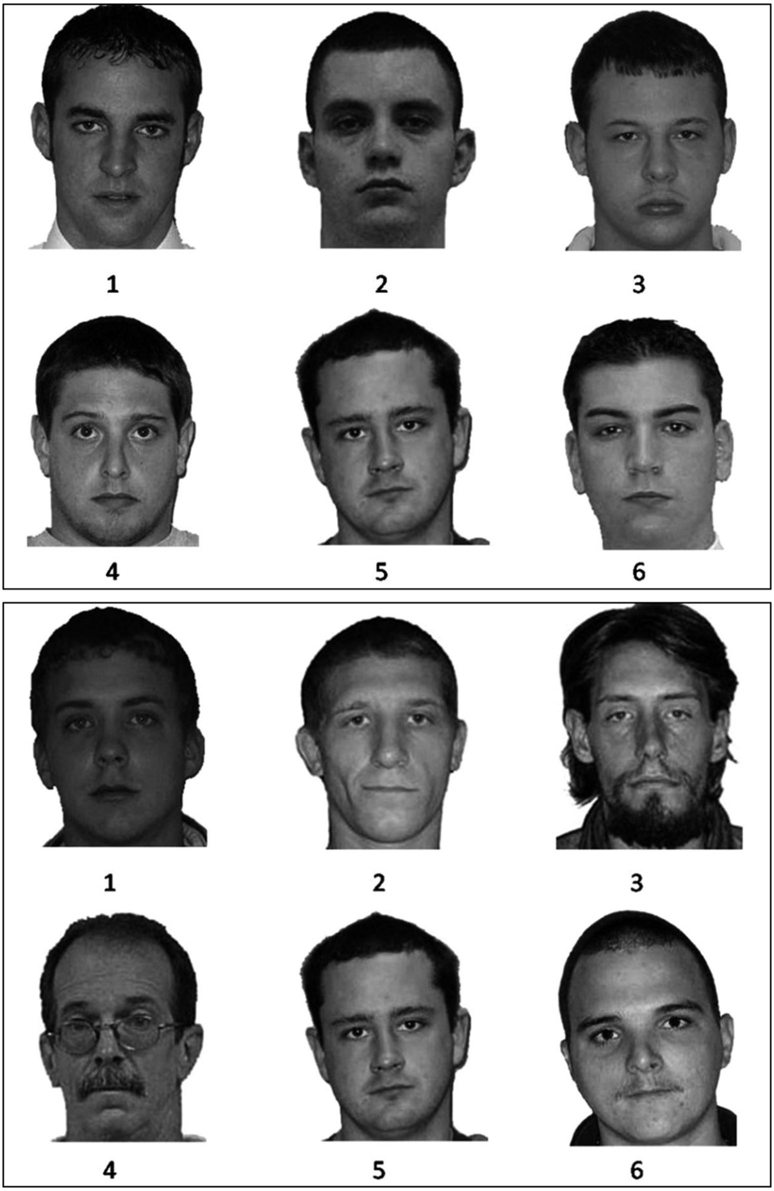

Fig. 3 Examples of a fair (*top*) and biased lineup (*bottom*). The innocent suspect is in position 5 in both lineups

the greater range over which the confidence criteria are spread (the dashed lines versus the solid vertical lines).

In summary, the data from both the meta-analysis of traditional measures and the ROC analyses paint a grim picture of showups as an identification procedure.

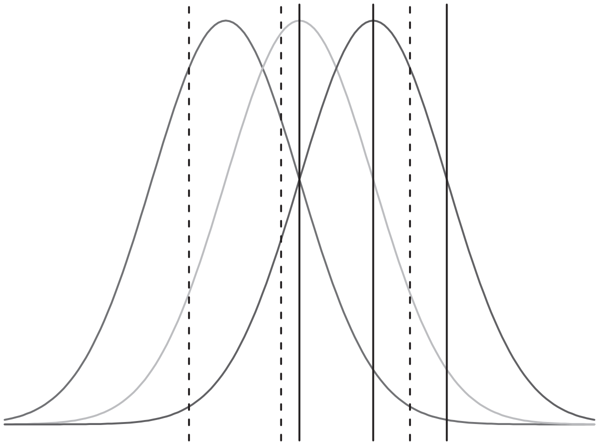

Fig. 4 Memory strength distributions depicting situations involving biased and fair lineups. The *blue* distribution represents the range of memory strengths for the perpetrator. The *green* distribution represents the varying degrees to which an innocent suspect matches participants' memory for the perpetrator. If a lineup is fair, the fillers do not stand out from an innocent suspect, akin to being selected from the green distribution. If the lineup is biased, the fillers are selected from the red distribution, which depicts a poorer match (on average) to the perpetrator. The extent of the ROC curve arises from the placement of the confidence criteria (the *vertical lines*). The greater range over which the ROC curve extends is due to the greater expanse over which the confidence criteria (the *dashed lines*) are spread if the fillers are poor (the blue distribution vs. the red distribution) than if the fillers are good matches (the blue distribution vs. the green distribution, the *solid vertical lines*)

This was true even when showups were compared to biased lineups. The only instance in which showup performance was equivalent to a lineup condition involved a sequential lineup with the suspect early in the lineup. Furthermore, even in situations in which memory theory would predict better performance in showups (i.e., immediate testing, similar cues at study and test), the performance in lineups still exceeded that of showups. Furthermore, one explanation of the superior performance in lineups to showups has been that in lineups there are filler choices, so that false identifications are spread amongst fillers rather than just to the innocent suspect, as is in the case with showups. This explanation seems faulty for two reasons. First, if there were fewer false identifications in the target absent conditions because of filler identifications, there should also be a concomitant decrease in correct identifications from target present conditions. We did not see this in our data (see Table 1), as the correct identification rate for showups was very similar to lineups (*$M = .45$ and $M = .51$, respectively*). Second, the ROC analysis clearly shows that discrimination, not bias (i.e., choosing fillers), is superior in lineups than showups (see Fig. 1, panel a).

One caveat to these conclusions is that one research team collected all of the data in the ROC analyses, and most studies utilized only one set of stimuli. Nevertheless, we believe that it is unlikely, given the large number and variety of participants, the range of independent and dependent variables used, and the supporting role provided by theory, that the results are stimulus specific. Given the public policy implications, it is important for independent research teams to replicate the results using different materials, samples, and methodologies, before accepting the conclusions as definitive.

Confidence

The poorer performance yielded by showups might be less problematic if witnesses were not confident in their identifications. That is, perhaps the lower confidence identifications arising from a procedure deemed "suggestive," even by the U.S. Supreme Court, would not proceed to trial due to witness uncertainty. However, the data reported above signal that participants are willing to make identifications from showups with high confidence. This necessitates further investigation of how confidence relates to identification accuracy by assessing the confidence–accuracy relationship for showups and lineups.

The confidence–accuracy correlation is one of the most widely studied topics in forensic psychology (see Roediger, Wixted, & DeSoto, 2012). There are likely two major reasons for the amount of attention and research devoted to this relationship. First, under the *Biggers* criteria, jurors are instructed to use witness confidence as one index of identification accuracy (see *Neil v. Biggers*, 1972; Wells & Quinlivan, 2009). Moreover, jurors not only rely heavily on witness confidence as a proxy for accuracy, they also are more likely to excuse inconsistencies in testimony of witnesses who are highly confident in their identifications (Brewer & Burke, 2002; Krug, 2007). Second, past research has consistently shown that the correlation between confidence and accuracy for eyewitnesses is low to moderate (see Sporer, Penrod, Read, & Cutler, 1995). This is counterintuitive, as one would logically think that witnesses who display high confidence would be more accurate in their identification decisions. In fact, other research domains, such as general and text-based knowledge, show a strong positive correlation between confidence and accuracy (Stephenson, 1984; Stephenson, Clark, & Wade, 1986).

Recently, several researchers (Juslin, Olsson, & Winman, 1996; Roediger et al., 2012) have argued that the inconsistent results across research domains are likely due to the statistic used to measure the correlation. Psycho-legal researchers typically have employed the point-biserial correlation to measure the relationship between confidence and accuracy (see Roediger et al., 2012). Juslin et al. (1996) demonstrated that this measure can be misleading, as it compares confidence with the unreasonable standard of perfect discrimination. That is, all correct responses should be in one confidence category and all incorrect responses should be in the other confidence category. Furthermore, they argued that the point-biserial correlation would be low in situations where participants are presented with one stimulus and make one choice amongst similar alternatives. This is, of course, exactly the scenario used in the typical eyewitness experiment. Juslin et al. (1996) and Roediger et al. (2012) argued in favor of confidence calibration as a more

Table 2 Point-Biserial correlations for confidence and accuracy

Identification type	Non-choosers and choosers		Choosers		Suspect identifications	
	r	*N*	*r*	*N*	*r*	*N*
Showups	0.093^{A}	4388	0.143^{C}	1349	0.143^{E}	1349
Lineup (all)	0.152^{B}	3569	0.328^{D}	2254	0.307^{F}	1221
Simultaneous	0.144^{B}	2624	0.318^{D}	1790	0.292^{F}	947
Sequential	0.193^{B}	945	0.352^{D}	514	0.271^{F}	274
Fair lineup	0.102	1483	0.269	948	0.319^{F}	410
Bias lineup	0.162^{B}	1466	0.381^{D}	917	0.309^{F}	660
Fair simultaneous	0.93	1168	0.259	776	0.332^{F}	332
Fair sequential	0.165^{B}	315	0.310	172	0.240	78
Biased simultaneous	0.149^{B}	1148	0.352	752	0.281^{F}	524
Biased sequential	0.190	318	0.307	166	0.193	136
Overall	0.145^{B}	7877	0.291^{A}	3603	0.179^{B}	2570

Note. All identification types are compared within and overall correlations are compared between columns; values with different *superscripts* are significantly different

appropriate measure of the relationship between confidence and accuracy. Specifically, they argue that this measure allows researchers to answer the forensically relevant question: Are higher ratings of confidence associated with more accurate identifications?

With these considerations in mind, we examined the confidence–accuracy relationship for simultaneous and sequential lineups, and showups, using point-biserial correlation and calibration curves. Is the confidence–accuracy relationship weaker for showups than for lineups? To our knowledge, there have been no comparisons of this relationship across the different identification procedures.

Point-Biserial Correlation

Sporer et al. (1995) argued that it makes the most sense from a forensic standpoint to evaluate the confidence–accuracy relationship for only those people that make a choice from the identification procedure because these are the witnesses most likely to be called to testify at trial. These individuals are referred to as "choosers." In a showup, a witness can either choose the suspect or not; in lineups, witnesses also can choose fillers. The point-biserial correlations in Table 2 include the data from the same studies that contributed to the ROC analyses, as these were the only studies for which we had, or could get, confidence data. The right panel of Table 2 displays the correlations for choosers. Collapsing over identification procedure yielded a higher correlation for choosers ($r = .29$) than when choosers and non-choosers were combined ($r = .14$, $Z = 7.64$, $p < .01$). Our data mirror the results of Sporer et al. (1995), who also found that the confidence–accuracy correlation for choosers was higher ($r = .40$) than when choosers and non-choosers were combined ($r = .29$). The confidence–accuracy correlation was significantly higher for simultaneous and sequential lineups combined than for showups, $Z = 5.71$, $p < .01$. Separately, simultaneous and sequential lineups each had significantly higher correlations than showups (Zs > 4.31, ps $< .01$). Thus, for choosers, the confidence–accuracy correlation is significantly greater for lineups than showups, as assessed by the point-biserial correlation.

The problem with the previous analysis is that it takes into account all choosers, including filler identifications from lineups. Filler identifications distort the calibration for lineups and puts them at a disadvantage in the same way that it complicates interpretation of identification rates (Clark & Godfrey, 2009). To address this bias, we calculated the correlations for only suspect identifications. Wixted, Mickes, Clark, Gronlund, and Roediger (2015) argued that the most forensically relevant choices are those of the perpetrator and the innocent suspect. Analyzing only suspect identifications yielded a confidence–accuracy correlation ($r = .18$) similar to that of the inclusion of all choices ($r = .14$, $Z = 1.54$, $p = .12$). Like the analysis of choosers, the confidence–accuracy correlation was significantly higher for simultaneous and sequential lineups combined than for showups, $Z = 4.38$, $p > .01$. Separately, simultaneous and sequential lineups each had higher confidence–accuracy correlations than showups ($Z = 3.69$, $p < .01$ and $Z = 2.01$, $p < .04$, respectively). Thus, for those who made suspect identifications, the confidence–accuracy correlation, albeit still low, was significantly better for lineups than showups.

Confidence Calibration

Calibration refers to how well a witness's confidence reflects his or her accuracy. Point-biserial correlations can be misleading; Juslin et al. (1996) showed that point-biserial correlations ranging from 0 to 1 could exhibit perfect calibration. The important point, as noted by Wixted et al. (2015), is that low point-biserial correlations do not necessarily imply poor confidence–accuracy relationships. The true magnitude of the relationship between confidence and accuracy is revealed by calibration curves.

Calibration curves are created by calculating the proportion of accurate decisions at each level of confidence. Perfect calibration occurs when all decisions made with 100 % confidence are accurate, decisions made with 90 % confidence are accurate 90 % of the time, and so forth. Typically, one reports several different statistics that accompany the calibration curves, which summarizes, in one form or another, how confidence relates to accuracy (Sauer, Brewer, Zweck, & Weber, 2010). The calibration index (CI) is the weighted average of the squared difference between confidence and accuracy at each confidence level. It can range from 0, representing perfect calibration, to 1, which represents no calibration. The over/under confidence (O/U) statistic measures participants' tendency to respond with greater or lesser confidence than their accuracy. For example, people who give confidence estimates of

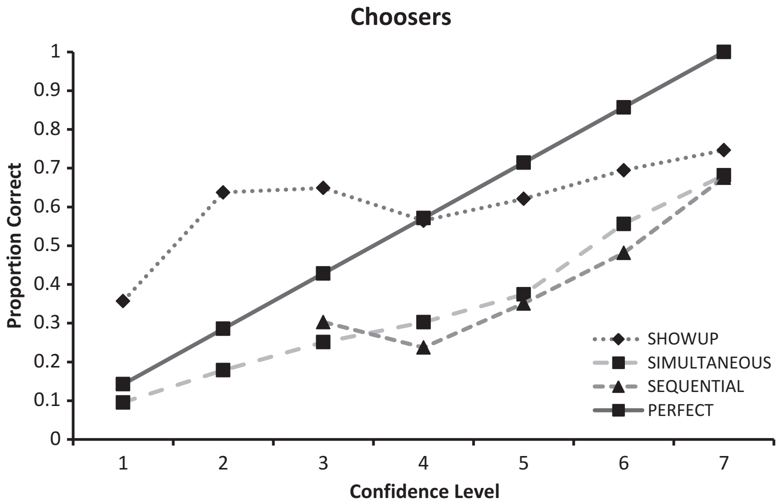

Fig. 5 Confidence calibration curves for choosers comparing simultaneous lineups, sequential lineups, and showups

30 % but are only accurate 10 % of the time are overestimating their confidence. O/U ranges from +1 to −1, with positive numbers representing overconfidence and negative numbers representing under confidence.

Before presenting the curves, it is important to discuss how the curves were generated. All of the lineups used a confidence scale from 1 (least confident) to 7 (most confident). The scale was converted to proportions by dividing each scale value by 7, so that the first entry is 1/7 (.14), the second is 2/7 (.29), and so on. We used the same scale for all the showup studies except Wetmore et al. (2015b), which had confidence ranging from 1 (least confident) to 10 (most confident). For this study, we combined confidence categories 2 and 3, 5 and 6, and 8 and 9 to put confidence on a 7-point scale.[7] The perfect calibration line starts at .14 to reflect the proportions on the scale. We calculated CI and O/U separately for the showups on the different scales (1–7 and 1–10), and then calculated a weighted average. All of our calibrations are between subjects because we only had participants make one identification decision/confidence judgment.

Figure 5 displays the relationship between confidence and accuracy for choosers only. Simultaneous and sequential lineup witnesses were overconfident at every level of confidence; showup witnesses, on the other hand, were underconfident at low levels of confidence and overconfident at high levels of confidence. But more to the point, note that the proportion correct from showups varies little from a confidence level of 2 (.64) to a confidence level of 7 (.75). In contrast, the proportion

[7] We reanalyzed the data excluding the Wetmore et al. (2015b) study, which was the only study using a 10-point scale. Doing this did not change the pattern of results. Consequently, we elected to leave the Wetmore et al. (2015b) in the analyses in the text.

Table 3 Calibration Index and over/under confidence for identification tasks

	Choosers		Suspect	
Identification task	CI	O/U	CI	O/U
Simultaneous	.078	−0.10	.008	−0.18
Sequential	.11	−0.06	.011	−0.12
Showup	.027	0.07	.027	0.07
Fair lineup	.03	−0.21	.017	−0.21
Biased lineup	.03	−0.14	.009	−0.14

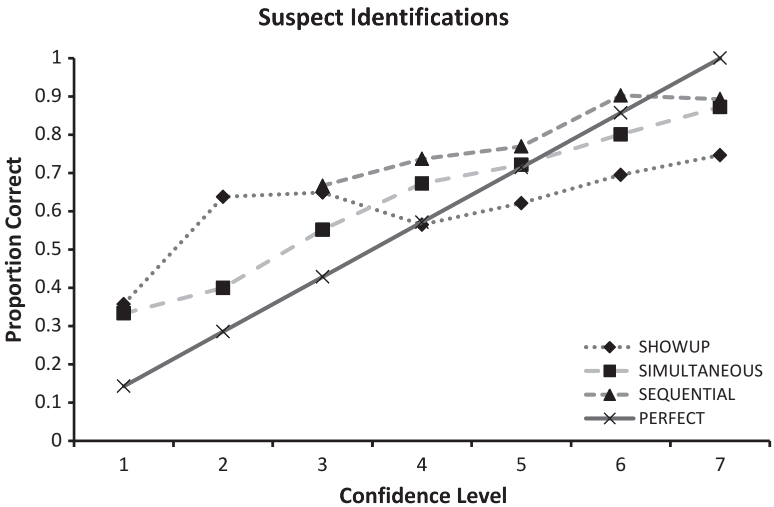

Fig. 6 Confidence calibration curves for suspect identifications comparing simultaneous lineups, sequential lineups, and showups

correct from simultaneous lineups over this same range, varies from .18 to .68. Table 3 shows that showups are better calibrated, and also indicate a penchant for overconfidence (O/U) from showups, but underconfidence from lineups. This could be problematic because the witnesses who report the highest level of confidence in their identification decisions are more likely to be called to testify in court, but from showups, these individuals were less accurate than lineup respondents.

A better comparison of showups and lineups excludes fillers and focuses on suspect identifications (see Fig. 6). Of course, the showup calibration curve does not change because all showup choosers have made a suspect identification. Table 3 now reveals the calibration of lineups to be superior to showups. We also assessed the calibration of suspect identifications from showups to fair and biased lineups (see Fig. 7). Now both biased and fair lineups were calibrated to the same degree as showups, but both still exhibited general underconfidence compared to showups.

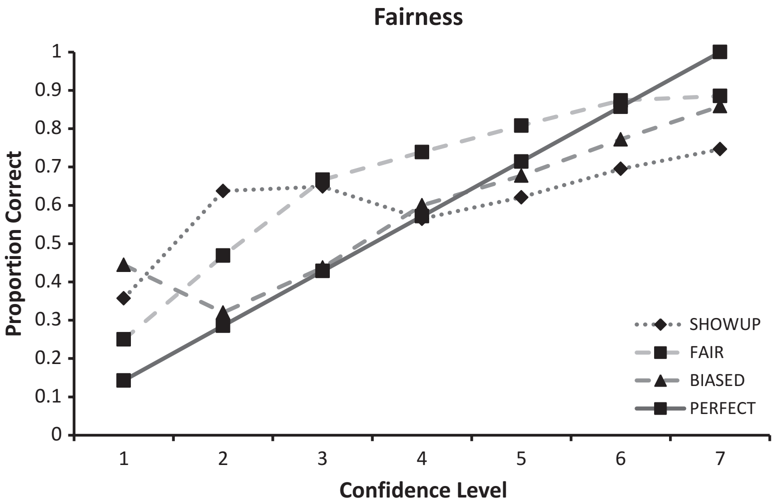

Fig. 7 Confidence calibration curves for suspect identifications comparing fair and biased lineups and showups

Conclusions

The goal of this chapter was to determine if there were situations in which a showup would outperform a lineup, thereby justifying the (occasional) use of showups. In every ROC comparison, lineups provided better discriminability and were the superior identification procedure. This was true regardless of lineup presentation method (simultaneous or sequential) and lineup fairness (fair or biased). Even in those situations for which showups might be thought to have a memorial advantage—at short retention intervals and when encoding and retrieval match—showups were worse. In sum, we have yet to find a situation where it would be more appropriate to conduct a showup if eyewitness accuracy is the primary goal. The current findings presented here provide a dismal portrayal of the most commonly used identification task.

The confidence data provide no solace for showups. When point-biserial correlations were calculated, the confidence–accuracy relationships were lower for showups than for lineups. More importantly, the forensically relevant suspect identification curves were better calibrated for lineups than showups. Witnesses viewing showups underestimated confidence at low levels and overestimated confidence at high levels. As mentioned earlier, confident eyewitnesses are most likely to testify in court: That means that for showups, the high confidence suspect identifications reflect eyewitness accuracy less well.

Wixted and Mickes (2014) provide a theoretical rationale for why showup performance is worse than lineup performance. They proposed a signal-detection-based diagnostic-feature model, arguing that simultaneous lineup performance is superior because witnesses can compare characteristics amongst the faces. This comparison process allows eyewitnesses to identify which characteristics are shared by all suspects (and thus not helpful for distinguishing the perpetrator), and those that are specific to the perpetrator. For example, a witness may look at a simultaneous lineup and determine that all lineup members are Caucasian with short brown hair; this redundancy makes race and hair color irrelevant for identifying the perpetrator. However, shifting attention away from these irrelevant (or redundant) characteristics allows an eyewitness to focus on more diagnostic characteristics (like face shape or the width of the eyes) that likely are unique to the perpetrator. Because only one face is viewed in a showup, this comparison process is not possible; eyewitnesses are unable to determine which characteristics are irrelevant and should be ignored, and which characteristics are diagnostic and worthy of attention. Poorer discriminability arises from utilizing the wrong cues. Consequently, according to Wixted and Mickes, performance from simultaneous lineups should exceed that of showups.

This is not to say that showups are the same as one-person sequential lineups; in fact, there are several sources of evidence that argue against this interpretation. Gonzalez et al. (1993) compared showup performance to a biased lineup with only one viable option and found that a showup was not equivalent. When we have compared showups to biased sequential lineups (Gronlund et al., 2009) and simultaneous lineups, biased lineups consistently outperformed showups. Witnesses exhibit better discrimination from a biased lineup with only one viable option than from a showup. According to the Wixted and Mickes (2014) hypothesis, this must mean that witnesses have some ability to determine what features to focus on even if viewing a lineup whose members do not resemble the perpetrator.

The Wixted and Mickes (2014) theory also predicts that an eyewitness should be better able to determine which characteristics are irrelevant, and which are diagnostic, from a fair than from a biased lineup. Because the fillers in a fair lineup, by definition, share more characteristics with the perpetrator, an eyewitness should be better able to discern irrelevant characteristics and focus on the key characteristics that point to the perpetrator. For example, in a fair lineup, all lineup members are about 30, Hispanic, with short dark hair, and skinny, allowing a witness to focus in on the narrow nose of the perpetrator. However, in a biased lineup, three lineup members are the right age, but three are not. A different three have short dark hair, two have shaved heads, and one has long hair. Two are heavyset, two are moderate in weight, and two are skinny. In this situation, it is more challenging to find the right combination of cues on which to focus. Although the present data show no discriminability difference as a function of lineup fairness, contrary to this theory, additional empirical work should be undertaken to verify this surprising result.

Another prediction that arises from the Wixted and Mickes (2014) theory is that an eyewitness in a sequential lineup can determine what characteristics are diagnostic after viewing several lineup members (see also Goodsell, Gronlund, & Carlson, 2010). For example, while viewing lineup member #1, a witness notes that the ears match the perpetrator but the nose is too big. After viewing a second lineup member, a witness gets a better sense of what the nose looked like, and perhaps remembers the perpetrator's heavy eyebrows, and so on. Consequently, discriminability improves as an eyewitness progresses through a sequential lineup. Indeed, this is exactly what Carlson et al. (2008) and Gronlund et al. (2009) found. Using the WITNESS computational memory model (Clark, 2003), Goodsell, Gronlund, and Buttaccio (2010) implemented several explanations for why a sequential lineup performance advantage seemed to arise when a suspect (guilty or innocent) was placed later in a sequential lineup. Specifically, Goodsell et al. modified WITNESS to fit data from two studies (Lindsay, Lea, & Fulford, 1991; Lindsay et al., 1991) that showed large sequential lineup advantages, both of which involved the placement of suspects into position 8 of 8. One of Goodsell et al.'s explanations in particular, the better cue model, was very similar to the ideas expressed in Wixted and Mickes.

There are many benefits to providing theoretical explanations for empirical findings (e.g., Bjork, 1973; Brewer, Weber, & Semmler, 2007; Clark, 2008). Including requiring researchers to operationalize and specify constructs of interest, viewing data in a more rigorous manner, and encouraging the generation of testable hypotheses (see Farrell & Lewandowsky, 2010; see also Clark & Gronlund, 2015). Only by evaluating data in the context of well-specified, testable theories can we make cumulative progress in understanding how eyewitnesses make decisions and what identification procedures might best support that decision-making.

One overlooked benefit of providing theoretical explanations is that a theoretical idea might point to a new method by which we can enhance performance. For example, if participants indeed can learn as a sequential lineup progresses, previewing a series of known innocent faces prior to a showup might engender the same comparative processes that allow an eyewitness to focus on diagnostic cues and ignore irrelevant cues. Goodsell, Gronlund and Buttaccio (2010) had participants view a mock crime and then evaluate known innocent faces just prior to a lineup identification. This pre-ID procedure showed beneficial effects for witnesses that evaluated faces with a high degree of match to the perpetrator prior to making an identification from a simultaneous lineup. It is important to determine if a method like this can enhance showup discriminability, as it seems likely that, despite the overwhelming evidence favoring lineups over showups, law enforcement will continue to use showups in the field.

In conclusion, the goal of this chapter was to compare showups and lineups using a variety of measures and lineup procedures (e.g., simultaneous, sequential, fair, biased). In all of our comparisons, we found that showups were at a disadvantage relative to lineups. More specifically, showups and lineups yielded similar correct identifications but showups consistently had more false identifications, as represented by significantly lower ROC curves. Even in situations in which showups should have, based on memory theory, a memorial advantage to lineups (e.g., clothing match and shorter retention interval), they still underperformed lineups. Furthermore, the calibration curves indicated that not only were showup witnesses less accurate, but at the highest levels of confidence, they were overconfident in the accuracy of their identification choices. It is likely that in forensic situations, those witnesses who have the highest confidence will be the ones that testify at trial. In our data, these were the most dangerous showup witnesses because although they were inaccurate they were highly confident. To the extent that the practice of showups is going to continue, and we believe it will, it is important to investigate ways to make this procedure yield more accurate identification choices.

Author Notes Jeffrey S. Neuschatz, Department of Psychology, The University of Alabama in Huntsville; Scott D. Gronlund and Stacy Wetmore, Department of Psychology, University of Oklahoma; Charles A. Goodsell, Department of Psychology, Canisius College. This work was supported by the National Science Foundation grant SES-1060921 to JSN, SES-1060902 to SDG, and SES-1060913 to CAG. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of the NSF.

References

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

Bjork, R. A. (1973). Why mathematical models? *American Psychologist, 28(5)*, 426–433. doi:10.1037/h0034623.

Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Behavior, 26*(3), 353–364. doi:10.1023/A:1015380522722.

Brewer, N., Weber, N., & Semmler, C. (2007). A role for theory in eyewitness identification research. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology* (Memory for people, Vol. II, pp. 201–218). Mahwah, NJ: Lawrence Erlbaum.

Carlson, C. A., & Carlson, M. A. (2014). An evaluation of lineup presentation, weapon presence, and a distinctive feature using ROC analysis. *Journal of Applied Research in Memory and Cognition, 3*, 45–53. doi:10.1016/j.jarmac.2014.03.004.

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

Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology, 17*, 629–654.

Clark, S. E. (2008). The importance (necessity) of computational modeling for eyewitness identification research. *Applied Cognitive Psychology, 22*, 803–813. doi:10.1002/acp.1484.

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspectives on Psychological Science, 7*(3), 238–259. doi:10.1177/174569161239584.

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

Clark, S. E., & Gronlund, S. D. (2015). Mathematical modeling shows that compelling stories do not make for accurate descriptions of data. In J. G. W. Raaijmakers, R. Goldstone, M. Steyvers, A. Criss, & R. M. Nosofsky (Eds.), *Cognitive modeling in perception and memory: A festschrift for Richard M. Shiffrin*. New York, NY: Psychology Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Dekle, D. J. (1997). Testing delays resulting in increased identification accuracy in line-ups and show-ups. *Journal of Offender Rehabilitation, 25*, 35–49. doi:10.1300/J076v25n03_03.

Dekle, D. J., Beal, C. R., Elliott, R., & Huneycutt, D. (1996). Children as witnesses: A comparison of lineup versus showup identification methods. *Applied Cognitive Psychology, 10*(1), 1–12. doi:10.1002/(SICI)1099-0720(199602)10:1<1::AID-ACP354>3.0.CO;2-Y.

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*(4), 345–357. doi:10.1037/a0034596.

Dysart, J. E., & Lindsay, R. C. L. (2007). Show-up identifications: Suggestive technique or reliable method? In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology* (Memory for people, Vol. II, pp. 137–153). Mahwah, NJ: Lawrence Erlbaum.

Dysart, J. E., Lindsay, R. C. L., & Dupuis, P. R. (2006). Show-ups: The critical issue of clothing bias. *Applied Cognitive Psychology, 20*, 1009–1023. doi:10.1002/acp.1241.

Farrell, S., & Lewandowsky, S. (2010). Computational models as aids to better reasoning in psychology. *Current Directions in Psychological Science, 19*(5), 329–335.

Gardner, T. J., & Anderson, T. M. (2004). *Criminal evidence: Principle and cases* (5th ed.). Belmont, CA: Thompson Wadsworth.

Garrett, B. (2011). *Convicting the innocent: Where criminal prosecutors go wrong*. Cambridge, MA: Harvard University Press.

Garrioch, L., & Brimacombe, E. C. A. (2001). Lineup administrators' expectations: Their impact on eyewitness confidence. *Law and Human Behavior, 25*, 299–315. doi:10.1023/A:1010750028643.

Gonzalez, R., Ellsworth, P. C., & Pembroke, M. (1993). Response biases in lineups and showups. *Journal of Personality and Social Psychology, 64*, 525–537. doi:10.1037/0022-3514.64.4.525.

Goodsell, C. A., Gronlund, S. D., & Buttaccio, D. (2010). *A test of the better memory probe hypothesis: Improving eyewitness identification accuracy*. Presented at the AP-LS meetings, Vancouver, March, 2010.

Goodsell, C. A., Gronlund, S. D., & Carlson, C. A. (2010). Exploring the sequential lineup advantage using WITNESS. *Law and Human Behavior, 34*, 445–459. doi:10.1007/s10979-009-9215-7.

Goodsell, C. A., Wetmore, S. A., Neuschatz, J. S., & Gronlund, S. D. (2013). Showups. In B. L. Cutler (Ed.), *Reform of eyewitness identification procedures* (pp. 45–63). Washington, DC: American Psychological Association.

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

Gronlund, S. D., Andersen, S. M., & Perry, C. (2013). Presentation methods. In B. Cutler (Ed.), *Reform of eyewitness identification procedures*. Washington, DC: APA Publications.

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

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

Gronlund, S. D., & Neuschatz, J. S. (2014). Eyewitness identification discriminability: ROC analysis versus logistic regression. *Journal of Applied Research in Memory and Cognition, 4*, 221–228.

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

Haw, R. M., & Fisher, R. P. (2004). Effects of administrator-witness contact on eyewitness identification accuracy. *Journal of Applied Psychology, 89*, 1106–1112. doi:10.1037/0021-9010.89.6.1106.

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. doi:10.1037/0278-7393.22.5.1304.

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research: A new survey of the experts. *American Psychologist, 56*, 405–416. doi:10.1037/0003-066X.56.5.405.

Key, K. N., Cash, D. K., Neuschatz, J. S., Price, J. L., Wetmore, S. A., & Gronlund, S. D. (2015). Age differences (or lack thereof) in discriminability for lineups and showups. *Psychology, Crime, & Law, 21*(9), 871–889. doi:10.1080/1068316X.2015.1054387.

Krug, K. (2007). The relationship between confidence and accuracy: Current thoughts of the literature and a new area of research. *Applied Psychology in Criminal Justice, 3*, 7–41.

Lawson, V. Z., & Dysart, J. E. (2014). The showup identification procedure: An exploration of systematic biases. *Legal and Criminalogical Psychology, 19,* 54–68. doi:10.1111/j.2044-8333.2012.02057.x.

Light, L. L. (1996). Memory and aging. In E. L. Bjork & R. A. Bjork (Eds.), *Memory: Handbook of perception and cognition* (2nd ed., pp. 444–477). San Diego, CA: Academic.

Lindsay, R. C. L., Lea, J. A., & Fulford, J. A. (1991). Sequential lineup presentation: Technique matters. *Journal of Applied Psychology, 76*, 741–745. doi:10.1037/0021-9010.76.5.741.

Lindsay, R. C. L., Pozzulo, J. D., Craig, W., Lee, K., & Corber, S. (1997). Simultaneous lineups, sequential lineups, and showups: Eyewitness identification decisions of adults and children. *Law and Human Behavior, 21*, 391–404. doi:10.1023/A:1024807202926.

Luus, E. C. A., & Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law and Human Behavior, 15*, 43–57. doi:10.1007/BF01044829.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Malpass, R. S., & Lindsay, R. C. L. (1999). Measuring lineup fairness. *Applied Cognitive Psychology, 13*, 1–8.

McQuiston, D., & Malpass, R. (2001, June). *Eyewitness identifications in criminal cases: An archival study*. Paper presented at the fourth biennial meeting of the Society for Applied Research in Memory and Cognition, Kingston, ON.

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. doi:10.3758/BF03193074.

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

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

Quinlivan, D. S., Neuschatz, J. S., Cutler, B. L., Wells, G. L., McClung, J., & Harker, D. L. (2011). Do pre-admonition suggestions moderate the effect of unbiased lineup instructions? *Legal and Criminological Psychology, 17*, 165–176. doi:10.1348/135532510X533554.

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), 77. http://dx.doi.org/10.1186/1471-2105-12-77.

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

Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York, NY: Appleton.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York, NY: Holt Reinhart and Winston.

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

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

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

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. doi:10.1037/a0021650.

Stephenson, G. M. (1984). Accuracy and confidence in testimony: A critical review and some fresh evidence. In D. J. Muller, D. E. Blackman, & A. J. Chapman (Eds.), *Psychology and law: Topics from an international conference* (pp. 229–248). Chichester, England: Wiley.

Stephenson, G. M., Clark, N. K., & Wade, G. S. (1986). Meetings make evidence? An experimental study of collaborative and individual recall of a simulated police interrogation. *Journal of Personality and Social Psychology, 50*, 1113–1122. doi:10.1037/0022-3514.50.6.1113.

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

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*(1), 1–26.

Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement*. Washington, DC: United States Department of Justice, Office of Justice Programs.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352–373. doi:10.1037/h0020071.

United States v. Wade, 388 U. S. 293 (1967)

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

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

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

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

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

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

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist, 70,* 515–526. doi:10.1037/a0039510.

Yarmey, A. D., Yarmey, A. L., & Yarmey, M. J. (1994). Face and voice identifications in showups and lineups. *Applied Cognitive Psychology, 8*, 453–464. doi:10.1002/acp.2350080504.

Yarmey, A. D., Yarmey, M. J., & Yarmey, A. L. (1996). Accuracy of eyewitness identification in showups and lineups. *Law and Human Behavior, 20*, 459–477. doi:10.1007/BF014989.

http://www.springer.com/978-3-319-29405-6

Advances in Psychology and Law
Volume 1
Miller, M.K.; Bornstein, B.H. (Eds.)
2016, XVII, 280 p. 9 illus., 1 illus. in color., Hardcover
ISBN: 978-3-319-29405-6