Evaluating Eyewitness Identification Procedures Using ROC Analysis

Journal:	*Current Directions in Psychological Science*
Manuscript ID:	CDPS-12-0155.R2
Manuscript Type:	Invited Manuscript
Date Submitted by the Author:	n/a
Complete List of Authors:	Gronlund, Scott; Oklahoma University, Psychology Wixted, John; UCSD, Psychology; Mickes, Laura; Royal Holloway, University of London, Psychology
Keywords:	eyewitness identification, simultaneous and sequential lineups, Receiver Operating Characteristic (ROC) analysis, diagnosticity ratio, probative value
Abstract:	Eyewitness identification is a pivotal issue in applied research because, in practice, a correct identification can help to remove a dangerous criminal from society, but a false identification can lead to the erroneous conviction of an innocent suspect. Consequently, psychologists have tried to ascertain the best procedures for collecting identification evidence, evaluating them using measures based on the ratio of correct to false identification rates. Unfortunately, ratio-based measures are ambiguous because they change systematically as a function of a witness's willingness to choose (e.g., Wixted & Mickes, 2012). In other words, a measure thought to index discriminability is instead fully confounded with response bias. A better method involves constructing Receiver Operating Characteristic (ROC) curves. ROC curves trace out discriminability across levels of response bias for each procedure. We illustrate the shortcomings of ratio-based measures and demonstrate why ROC analysis is required. Recent research comparing simultaneous and sequential lineup procedures using ROC analyses provides no evidence for the sequential superiority effect and instead indicates that the simultaneous procedure may be diagnostically superior. It is not yet clear which lineup procedure will prove to be generally superior, but it is clear that ROC analysis is the only way to make that determination.

COVER PAGE

Scott D. Gronlund
Department of Psychology
University of Oklahoma
455 W. Lindsey St.
Norman, OK 73019
sgronlund@ou.edu
Phone: 405 325-4553 FAX: 405 325-4737

Word Count: 3180

No permissions required for table or figures.

Evaluating Eyewitness Identification Procedures Using ROC Analysis

Scott D. Gronlund[1], University of Oklahoma

John T. Wixted, University of California, San Diego

Laura Mickes, Royal Holloway, University of London

Abstract

Eyewitness identification is a pivotal issue in applied research because, in practice, a correct identification can help to remove a dangerous criminal from society, but a false identification can lead to the erroneous conviction of an innocent suspect. Consequently, psychologists have tried to ascertain the best procedures for collecting identification evidence, evaluating them using measures based on the ratio of correct to false identification rates. Unfortunately, ratio-based measures are ambiguous because they change systematically as a function of a witness's willingness to choose (e.g., Wixted & Mickes, 2012). In other words, a measure thought to index discriminability is instead fully confounded with response bias. A better method involves constructing Receiver Operating Characteristic (ROC) curves. ROC curves trace out discriminability across levels of response bias for each procedure. We illustrate the shortcomings of ratio-based measures and demonstrate why ROC analysis is required. Recent research comparing simultaneous and sequential lineup procedures using ROC analyses provides no evidence for the sequential superiority effect and instead indicates that the simultaneous procedure may be diagnostically superior. It is not yet clear which lineup procedure will prove to be generally superior, but it is clear that ROC analysis is the only way to make that determination.

Keywords: eyewitness identification, simultaneous and sequential lineups, Receiver Operating Characteristic (ROC) analysis, diagnosticity ratio, probative value

Evaluating Eyewitness Identification Procedures Using ROC Analysis

To date, the Innocence Project has used DNA evidence to exonerate over 300 innocent people. Faulty eyewitness evidence played a significant role in nearly 75% of these wrongful convictions. As a result, eyewitness memory researchers have developed new lineup procedures that have the positive effect of reducing false identifications, but one difficulty is that they often have the negative effect of also reducing correct identifications (Clark, 2012). What is the proper way to determine the best procedure in those circumstances?

Before answering this question, it is important to draw a distinction between discriminability and response bias. Discriminability refers to the degree to which eyewitnesses who are tested using a lineup can tell the difference between innocent and guilty suspects. The highest level of discriminability occurs when guilty suspects always are identified and innocent suspects never are identified. The lowest level of discriminability occurs when innocent suspects are identified as often as guilty suspects. In practice, discriminability usually falls between these two extremes. Response bias, on the other hand, refers to the inclination of eyewitnesses to identify someone from the lineup, and it can vary over a wide range while holding discriminability constant. When responding is conservative, both correct and false identifications tend to be rare. When responding is liberal, both correct and false identifications are more frequent.

The best identification procedure is the one that maximizes the ability of eyewitnesses to discriminate between innocent and guilty suspects. Science can establish which procedure yields higher discriminability in laboratory studies that use forensically-relevant experimental designs, and scientists should recommend that procedure to policymakers to the extent such studies are judged to apply to the real world. But once the superior lineup procedure is identified, a separate question concerns how liberal or how conservative responding should be using that procedure. Liberal responding can be induced by instructing witnesses to make an identification even if they have to guess or, equivalently, by counting identifications made with any level of confidence (including low confidence). Conservative responding can be induced by instructing witnesses not to make an identification unless they are certain to be correct or, equivalently, by only counting identifications made with high confidence. (See Roediger, Wixted, & DeSoto, 2012 for an extensive discussion of how confidence data are analyzed.) Encouraging conservative responding would mean fewer innocents being accused, but at the expense of fewer guilty suspects being implicated, whereas encouraging liberal responding would have the opposite effect. The optimal balance between conservative and liberal extremes is not something that can be settled by science because it is largely a function of subjective values and unknown factors (the base rate of innocent suspects being placed into lineups). But the procedure that yields greater discriminability is always preferred, a point that is well understood in the field of diagnostic medicine (e.g., Swets, 1979). We illustrate these issues by considering the debate over simultaneous versus sequential lineup procedures (for a review see Gronlund, Andersen, & Perry, 2013).

Simultaneous versus Sequential Lineups

In the US, eyewitness memory typically is tested by simultaneously presenting one suspect together with several fillers (i.e., known innocents), usually in a photo spread. One decision is required, either a decision to choose an individual from the lineup, or a decision to reject the lineup because the witness cannot identify the perpetrator. Presenting lineup members in a sequential manner has been proposed as a partial solution to the unreliability of eyewitness identifications (Wells et al., 1998). In the sequential procedure, lineup members are presented one at a time, and a decision is required for lineup member i before lineup member i+1 is presented. An identification decision is recorded the first time the witness chooses someone, and a ‘reject’ decision is recorded when the lineup ends with no one being identified.

In the laboratory, eyewitness identification procedures are studied in the following manner. Participants first observe a mock crime. After a delay, they are presented with either a simultaneous or a sequential lineup and informed that the perpetrator may or may not be present. The lineup contains (typically) six individuals, one of whom is the (innocent or guilty) suspect, plus five fillers. Because the selection of a filler is not a dangerous error, laboratory studies focus on the selection of guilty suspects (correct identifications from target-present lineups) and innocent suspects (false identifications from target-absent lineups). The benefit of studying eyewitness identification in the laboratory is that the researcher knows whether the suspect is guilty or innocent, and therefore knows if a participant has made a correct or false identification, whereas police investigators only know whether an eyewitness has selected their suspect.

Lindsay and Wells (1985) reported that the sequential lineup reduced false identifications (IDs) (*M* simultaneous = .43; *M* sequential = .17, $p < .01$) far more than it reduced correct IDs (*M* simultaneous = .58; *M* sequential = .50, *ns*).[2] In fact, because the difference between the correct ID rates for the two lineup procedures was not significant, it was initially thought that it was nonexistent. However, later research has clearly established that both the false ID rate and the correct ID rate are lower for sequential lineups compared to simultaneous lineups (Steblay, Dysart, & Wells, 2011). This pattern is widely attributed to the fact that sequential lineups induce more conservative responding than simultaneous lineups. But which procedure yields higher discriminability? With rare exceptions (e.g., Meissner, Tredoux, Parker, & MacLin, 2005; Palmer & Brewer, 2012), the field has not focused on this question. Instead, it has focused on a statistic called the diagnosticity ratio, the ratio of correct IDs to false IDs.[3] As noted by Lindsay and Wells (1985), the diagnosticity ratio favored the sequential lineup 2.94 to 1.35. That meant that a suspect identified from a sequential lineup was nearly 3 times as likely to be guilty than innocent, but a suspect identified from a simultaneous lineup was only slightly more likely to be

---

[2] One could simply compute d' from Lindsay and Wells' (1985) data as a proxy for a full ROC analysis; it shows a sequential advantage (1.58 vs. .61). Had it been performed, ROC analysis may have shown a sequential advantage as well. However, the only way to be certain about that is to actually perform ROC analysis, which is superior to computing d' because it is an assumption-free method. By contrast, d' is based on a specific model that applies to a standard single-item old/new recognition procedure, not to a lineup recognition procedure. d' also makes assumptions regarding normally-distributed memory strength values. Although d' is almost certainly more informative than the more typically used diagnosticity ratio, it is less informative than ROC analysis.

[3] There are several measures of probative value of a suspect identification that utilize the ratio of correct and false IDs in various configurations. All of these measures are subject to the criticism we level below.

guilty than innocent. Because the odds of guilt are higher (and the likelihood of misidentification is lower) if the suspect is identified using the sequential procedure, the scientific case in favor of replacing the simultaneous procedure with the sequential procedure is intuitively compelling.

In light of these findings, it is not surprising that the view that the sequential lineup is superior to the simultaneous lineup has gained considerable traction. In the US, a number of states and municipalities have switched to using sequential lineups (Jonsson, 2007). In addition, the Innocence Project has endorsed sequential lineups (http://www.innocenceproject.org/Content/Sequential_Presentation_of_Lineups.php; retrieved May 14, 2013), and the sequential advantage has been advanced in textbooks (e.g., Goldstein, 2008), and popular culture (e.g., *Law and Order: SVU,* McCreary, Wolf, & Forney, 2009).

Challenging the Sequential Advantage

Although there is a general consensus that sequential lineups yield lower correct and false ID rates than simultaneous lineups, there is no consensus about the effect that sequential lineups have on the diagnosticity ratio. For example, Carlson, Gronlund, and Clark (2008) replicated the experiment by Lindsay and Wells (1985) and found no evidence that the sequential procedure yields a higher diagnosticity ratio. In a second experiment, Carlson et al. varied the degree to which the fillers resembled the suspect (i.e., lineup fairness) and again found no sequential advantage as measured by the diagnosticity ratio. In the most comprehensive study to date, Gronlund, Carlson, Dailey, and Goodsell (2009) conducted a very large study (N = 2,529) that varied a number of different factors including lineup fairness, suspect position, and quality of the suspect photos. In agreement with past work (and with the idea that the sequential procedure induces conservative responding), both the correct and false ID rates were generally lower for the sequential procedure. However, across 24 possible comparisons of sequential and simultaneous lineups, there were only two significant sequential advantages and three significant simultaneous advantages. These findings suggest that the two lineup procedures generally yield similar diagnosticity ratios.

What do the diagnosticity ratio data suggest about the ability of eyewitnesses to distinguish between innocent and guilty suspects using simultaneous or sequential lineups? That is, what do they suggest about *discriminability*? As we discuss next, a higher diagnosticity ratio is a natural consequence of more conservative responding and is not, by itself, an indication of higher discriminability. Thus, the empirical pattern reported by Lindsay and Wells (1985) – a pattern that Steblay et al. (2011) described as being representative of the current literature – is consistent with the idea that the sequential procedure induces conservative responding without increasing discriminability (see Palmer & Brewer, 2012). However, if the sequential procedure does not yield a higher diagnosticity ratio despite inducing more conservative responding (e.g., Gronlund et al., 2009), it would imply that the sequential procedure actually yields *lower* discriminability than the simultaneous procedure. Although this would be the implication, it would not be a conclusive result. The only way to conclusively determine whether one lineup procedure is diagnostically superior to the other in terms of discriminability is to perform Receiver Operating Characteristic (ROC) analysis.

ROC analysis

ROC analysis is widely used to measure the accuracy of diagnostic systems in fields as varied as medical imaging, weather forecasting, and materials testing (for reviews, see Swets, 1988, and Swets, Dawes & Monahan, 2000). Although ROC analysis and signal detection theory are mainstays of basic recognition memory research, the approach is new to the eyewitness memory literature. Here, we illustrate ROC analysis by showing how it is performed in the eyewitness memory domain, using the results reported by Lindsay and Wells (1985) to guide our illustration. This study was chosen to illustrate ROC analysis because, intuitively, it appears to show that the sequential procedure is diagnostically superior to the simultaneous procedure. ROC analysis shows why the reported data are instead ambiguous. In the study reported by Lindsay and Wells (1985), after the participants made an identification decision from a lineup, they also made a confidence rating using a 1-to-7 scale, where 1 reflects very low confidence and 7 reflects very high confidence. Although these confidence data were not reported in enough detail to perform ROC analysis, we can use hypothetical confidence rating data to illustrate how it works (see Table 1; see also the Supplementary Materials). We chose hypothetical confidence ratings for the sequential condition such that when they are aggregated together, the obtained overall correct and false ID rates correspond to the values reported by Lindsay and Wells (1985). For simplicity, we assume that 100 participants viewed target-present lineups and 100 participants viewed target-absent lineups. Thus, in Panel 1 of Table 1, the 50 correct IDs summed across varying levels of confidence correspond to an overall correct ID rate of 50/100 = .50, and the 17 false IDs summed across varying levels of confidence correspond to an overall false ID rate of 17/100 = .17.

As shown in the first line of Panel 1 (Table 1), 4 correct IDs of the guilty suspect and 3 false IDs of the innocent suspect were, hypothetically, made with the lowest level of confidence (i.e., with a rating of 1, essentially a guess). Under ordinary circumstances, these low-confidence identifications would not play a significant role in the courtroom (i.e., they would be excluded from consideration before reaching that point). The first step of ROC analysis is to treat these low-confidence identifications the same way the legal system does by only counting identifications made with a higher level of confidence (which is tantamount to inducing slightly more conservative responding on the part of witnesses before they make an ID). As shown in Panel 2 of Table 1, when these guesses are removed from the analysis by treating them as effective non-identifications, the correct and false ID rates both decrease, creating a second ROC point. In addition, the diagnosticity ratio goes up. This exercise immediately shows that more than one pair of correct and false ID rates and more than one diagnosticity ratio characterizes the performance of a lineup procedure in a single study. In fact, more than two pairs of correct and false IDs (and more than two diagnosticity ratios) characterize its performance because, if even more certainty is desired, it makes sense to also remove the near-guess ratings of 2 from the analysis, thereby creating a third point on the ROC (Panel 3 of Table 1). Repeating this process to its logical conclusion yields a whole family of correct and false ID rates, which, when plotted on a graph, yields the ROC in Figure 1. A curve traced out by the family of correct and false ID rates reflects a *single* level of discriminability associated with a lineup procedure for the particular memory conditions tested in an experiment (e.g., for a 20-s viewing of a mock crime video). The different points along the ROC curve represent different levels of bias, ranging from the point associated with the most liberal decision rule on the upper right (which includes all IDs, regardless of the level of confidence) to the point associated with the most conservative decision rule on the lower left (which includes only high-confidence IDs). Critically, the diagnosticity ratio steadily increases as you move down the curve, a phenomenon that is invariably true of real data as well.

Table 1. *Hypothetical Data that Corresponds to the Correct and False ID Rates from Lindsay and Wells (1985, Sequential Procedure) Summed across all 7 Levels of Confidence*

Confidence	Correct ID	False ID		PANEL 1
1	4	3		
2	3	2		
3	6	3	Correct ID rate:	.50
4	13	4	False ID rate:	.17
5	3	1	Diagnosticity ratio:	2.94
6	15	3		
7	6	1		
Sum	50	17		

Confidence	Correct ID	False ID		PANEL 2
2	3	2		
3	6	3		
4	13	4	Correct ID rate:	.46
5	3	1	False ID rate:	.14
6	15	3	Diagnosticity ratio:	3.29
7	6	1		
Sum	46	14		

Confidence	Correct ID	False ID		PANEL 3
3	6	3		
4	13	4	Correct ID rate:	.43
5	3	1	False ID rate:	.12
6	15	3	Diagnosticity ratio:	3.58
7	6	1		
Sum	43	12		

Confidence	Correct ID	False ID		PANEL 4
4	13	4		
5	3	1	Correct ID rate:	.37
6	15	3	False ID rate:	.09
7	6	1	Diagnosticity ratio:	4.11
Sum	37	9		

Confidence	Correct ID	False ID		PANEL 5
5	3	1	Correct ID rate:	.24
6	15	3	False ID rate:	.05
7	6	1	Diagnosticity ratio:	4.80
Sum	24	5		

Confidence	Correct ID	False ID		PANEL 6
6	15	3	Correct ID rate:	.21
7	6	1	False ID rate:	.04
Sum	21	4	Diagnosticity ratio	5.25

Confidence	Correct ID	False ID		PANEL 7
7	6	1	Correct ID rate:	.06
			False ID rate:	.01
Sum	6	1	Diagnosticity ratio:	6.00

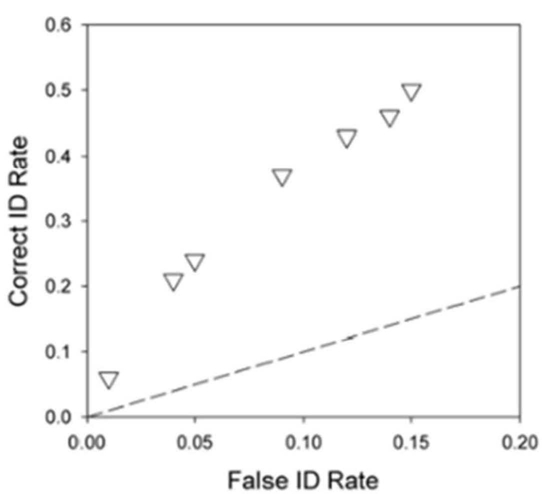

Figure 1. Hypothetical ROC data for the sequential procedure. The data represent the seven correct ID versus false ID rates presented in Table 1. The rightmost point of the ROC represents the correct ID rate and false ID rate obtained by collapsing across all levels of confidence (Panel 1 of Table 1). This point matches the ID rates reported by Lindsay and Wells (1985) for the sequential procedure (correct ID rate = .50, false ID rate = .17). The remaining points are hypothetical. The dashed line indicates chance performance (correct ID rate = false ID rate).

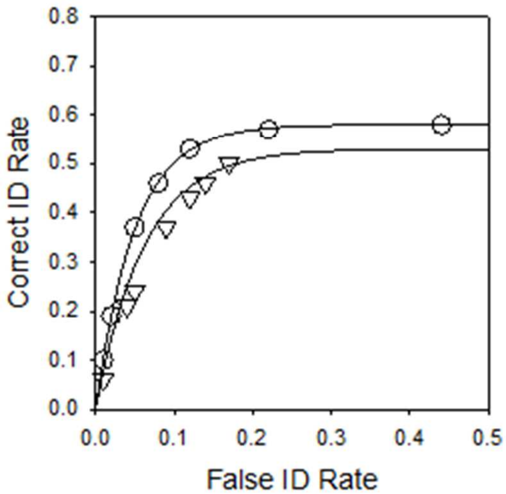

Figure 2. Two hypothetical ROCs, one for the simultaneous procedure (circles) and one for the sequential procedure (triangles). The sequential ROC data are the same as those shown in Figure 1. The rightmost point of both ROCs represents the correct ID rate and false ID rate obtained by collapsing across all levels of confidence, as is typically done when comparing lineup procedures. For this example, the rightmost point of each ROC was chosen to match the ID rates reported by Lindsay and Wells (1985). That is, for the simultaneous procedure, the rightmost point corresponds to a correct ID rate of .58 and a false ID rate of .43. For the sequential procedure, the rightmost point corresponds to a correct ID rate of .50 and a false ID rate of .17. Thus, those two points represent what has been taken to reflect the "sequential superiority effect." Nevertheless, if the rest of the ROC followed the paths traced out by the hypothetical data, the results would be consistent with a *simultaneous* superiority effect. For any false ID rate in this hypothetical example (e.g., .10), a higher correct ID rate can be achieved using the simultaneous procedure.

The fact that the diagnosticity ratio increases from right to left on the ROC indicates that it is not a measure of discriminability but is instead more like a measure of bias, with higher values reflecting a more conservative decision rule. The diagnosticity ratio increases as an increasingly conservative decision rule is used because the lower-confidence IDs that are excluded to create each new point on the ROC tend to be less accurate than the remaining IDs made with higher confidence. It is not possible to say whether the increase in accuracy associated with higher confidence IDs occurs because witnesses who are generally more confident also tend to be generally more accurate or because witnesses who realize they got a good look at the perpetrator (and therefore formed a good memory representation) expressed higher confidence than eyewitnesses who realized they did not get a good look at the perpetrator (and therefore formed a poor memory representation). Either way, when lower-confidence IDs are excluded to compute each new point on the ROC, less accurate decisions are excluded. In this regard, ROC analysis mimics what the legal system typically does. That is, as a criminal case moves from the investigation stage to a court of law, the emphasis is increasingly placed on identifications made with higher levels of confidence, which is to say that the emphasis shifts from points that fall towards the upper right of the ROC to points that fall towards the lower left of the ROC curve. Note, we are not talking about the inflation of confidence that arises due to confirming feedback.

Although the diagnosticity ratio does not provide useful information about the key issue of discriminability, ROC analysis does. Higher discriminability is indicated by an ROC curve that bows farther up and away from the diagonal line of chance performance. As illustrated in Figure 2, if a different lineup procedure yields a higher ROC, the procedure that yields the higher ROC is (objectively) the diagnostically superior procedure because, for any given false ID rate, that procedure can be used to achieve a higher correct ID rate. Equivalently, for any given correct ID rate, the diagnostically superior procedure can be used to achieve a lower false ID rate.

As illustrated in Figure 3, the correct and false ID rate data reported by Lindsay and Wells (1985), which were collapsed over confidence ratings, are compatible with a sequential superiority effect (top panel), a simultaneous superiority effect (middle panel), or a simple criterion shift resulting in more conservative responding (bottom panel), depending on how the ROC data for each procedure actually trace out. With these considerations in mind, we now turn to a key question that the field of eyewitness memory must confront: when put to an empirical test, does the simultaneous or the sequential procedure yield the higher ROC? In other words, does one lineup procedure facilitate the discrimination between innocent versus guilty suspects more than the other?

Thus far, only two studies have used ROC analysis to compare simultaneous and sequential lineups (Figure 4). Mickes, Flowe, and Wixted (2012) conducted three experiments comparing simultaneous and sequential lineups using ROC analysis. They found that the sequential lineup procedure was never better than and was sometimes significantly inferior to the simultaneous lineup procedure. Gronlund et al. (2012) conducted ROC analyses on the simultaneous and sequential data from Gronlund et al. (2009) and also found that the sequential lineup was never better than the simultaneous lineup. Both of these studies also illustrate the dependency of the diagnosticity ratio on response biases (see Tables 1 and 3 in Mickes et al., 2012, and Table 4 in Gronlund et al., 2012). As a more conservative decision rule is used and the choosing rates decrease, the diagnosticity ratios increase (cf., Brewer & Wells, 2006).

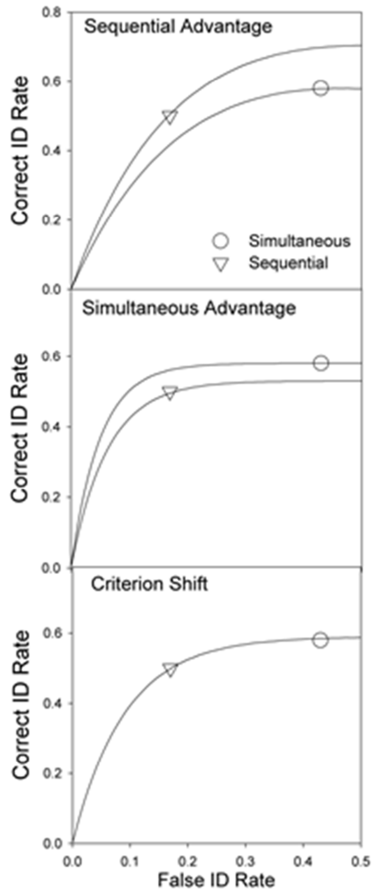

Figure 3. An illustration of three different possibilities consistent with the Lindsay and Wells (1985) data. The two data points in each panel represent the correct ID rate and false ID rate for simultaneous (circle) and sequential (triangle) lineup procedures from Lindsay and Wells. The curves drawn through the data represent the full range of correct ID-false ID rate pairs that might be associated with each procedure. The pair of correct ID and false ID rates might fall on different ROCs, with the sequential procedure yielding the higher ROC (top panel). This would indicate a sequential superiority effect. Alternatively, the same two points might fall on different ROCs, with the simultaneous procedure yielding the higher ROC (middle panel, corresponding to the hypothetical example shown in Figure 2), which would indicate a simultaneous superiority effect. Finally, the same two points might fall on the same ROC (bottom panel), a result that would support a conservative criterion shift interpretation and no discriminability difference.

Implications for Practice and Policy

More studies are needed that perform ROC analyses comparing sequential and simultaneous lineups, as a function of lineup fairness, quality of view, suspect position, etc. But the first two such studies do not support the notion of a sequential superiority effect and instead raise the possibility that the simultaneous procedure is diagnostically superior. Until such time as ROC analysis establishes sequential lineups as being superior to simultaneous lineups, police departments should refrain from switching to the sequential procedure. The available evidence simply does not support the claim for a sequential superiority effect (even if, collapsed across confidence ratings, the sequential procedure often yields a higher diagnosticity ratio). Finally, it is time to look beyond the simultaneous-sequential debate and begin to examine promising alternative procedures (e.g., Brewer, Weber, Wootton, & Lindsay, 2012; Weber & Perfect, 2012). But when evaluating these alternative procedures, ROC analysis (not the diagnosticity ratio) will reveal whether the new procedures are diagnostically superior to the procedures they would replace.

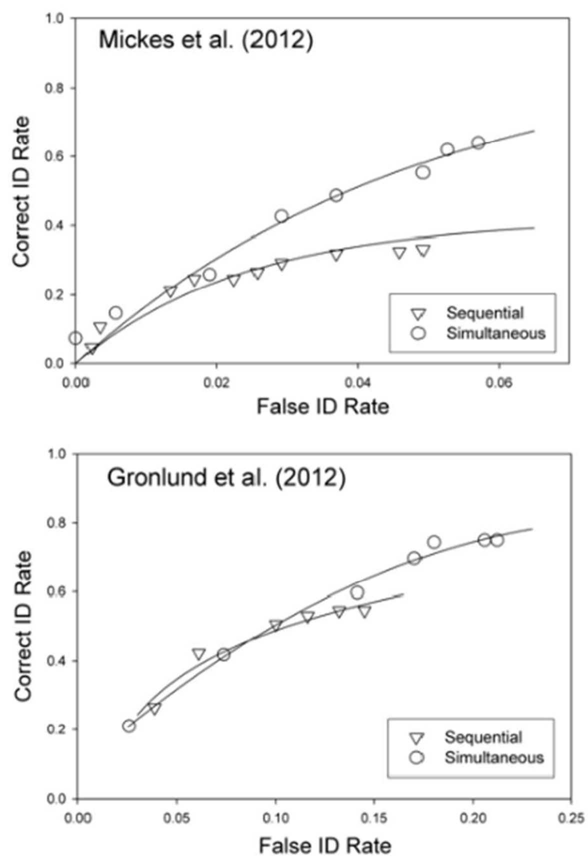

Figure 4. ROC data from Experiment 1a of Mickes et al. (2012) (top panel) and Gronlund et al. (2012) (bottom panel). Note the difference in the range of the *x*-axis across the two graphs. Mickes et al. (2012) reported two other studies that also yielded an advantage for the simultaneous procedure, but the effect was not significant in either case.

Author Note

[1]Correspondence concerning this article can be addressed to: Scott D. Gronlund (sgronlund@ou.edu), Department of Psychology, University of Oklahoma, 455 W. Lindsey St., Norman, OK 73019; or John T. Wixted (jwixted@ucsd.edu), Department of Psychology, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093.

Funding

This article was supported by National Science Foundation (NSF) grant SES-1060902 to Scott Gronlund and SES-1155248 to John Wixted and Laura Mickes. 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

Brewer, N., Weber, N., Wootton, D., & Lindsay, D. S. (2012). Identifying the bad guy in a lineup using deadlined confidence judgments. *Psychological Science*, *23*, 1208-1214.

Brewer, N. & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness accuracy: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied, 12*, 11–30.

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.

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

Clark, S. E., Erickson, M. A., & Breneman, J. (2011). Probative value of absolute and relative judgments in eyewitness identification. *Law and Human Behavior*, *35*, 364–380.

Goldstein, E. B. (2008). *Cognitive psychology: Connecting mind, research, and everyday experience* (2nd ed.). Belmont, CA: Thomson Wadsworth.

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

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.

Jonsson, P. (2007, February 6). The police lineup is becoming suspect practice. *Christian Science Monitor.* Retrieved Nov. 20, 2012, from http://www.csmonitor.com/2007/0206/p01s02-usju.html

Lindsay, R. C., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, *70*, 556–564.

McCreary, J., Wolf, D., & Forney, A. W. (2009). Unstable. In Wolf Films, *Law & Order: Special Victims Unit*. Universal City, California: NBC.

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.

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

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

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*. New York: Oxford University Press.

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

Swets, J. A. (1979). ROC analysis applied to the evaluation of medical imaging techniques. *Investigative Radiology, 14*, 109-121.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science, 240*, 1285-1293.

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

Weber, N., & Perfect, T. J. (2012). Improving eyewitness identification accuracy by screening out those who say they *Don't Know*. *Law and Human Behavior, 36,* 28-36.

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*, 603–647.

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.

Recommended Reading

Gronlund, S. D. et al. (2013). (See References). Review of the debate surrounding simultaneous and sequential lineups.

Gronlund, S. D. et al. (2012). (See References). ROC analysis evaluated performance from simultaneous and sequential lineups and showups.

Mickes, L., Flowe, H. D. & Wixted, J. T. (2012). (See References). ROC analysis showed no evidence for sequential advantage.

Wixted, J. T. & Mickes, L. (2012). (See References). Illustrates the limitations of ratio-based probative value measures and the advantages of ROC-based analysis of simultaneous and sequential lineup performance.