Faculty Scholarship

2014

Hits, Misses, and False Alarms in Blind and Sequential Administration of Lineups

Roger C. Park

UC Hastings College of the Law, parkr@uchastings.edu

Hits, Misses, and False Alarms in Simultaneous and Sequential Lineups

ROGER C. PARK[1]

Scholars who study eyewitness identification have produced an impressive body of empirical research. On the basis of this research, they have also proposed lineup reforms and worked with policymakers to get the reforms implemented.[2] One of the most important reform proposals involves use of sequential instead of simultaneous lineups. This paper is an examination of costs and benefits of that reform.

Advocates of sequential lineups maintain that simultaneous lineups encourage witnesses to use a process of "relative judgment."[3] When lineup witnesses view a simultaneous array, they often pick the person who most closely resembles their memory of the culprit's appearance.[4] This relative judgment process results in false identifications when the culprit is absent from the array.[5] Risinger captures the

---

2. *See, e.g.*, Gary L. Wells et al., *Double-Blind Photo Lineups Using Actual Eyewitnesses: An Experimental Test of a Sequential Versus Simultaneous Lineup Procedure*, LAW & HUM. BEHAV., June 16, 2014, at 1 [hereinafter Wells, *Double-Blind*]. The authors note that many U.S. jurisdictions use sequential lineup procedures, including "the entire states of Connecticut, North Carolina, and New Jersey, as well as major cities such as Dallas and Boston." *Id.* A white paper adopted by the American Psychology/Law Society was influential in encouraging reforms. Gary L. Wells et al., *Eyewitness Identification Procedures: Recommendations for Lineups and Photospreads*, 22 LAW & HUM. BEHAV. 603, 603 (1998) [hereinafter Wells, *Eyewitness*].

3. *See, e.g.*, Nancy K. Steblay et al., *Seventy-Two Tests of the Sequential Lineup Superiority Effect: A Meta-Analysis and Policy Discussion*, 17 PSYCHOL. PUB. POL'Y & L. 99, 102 (2011). Gary L. Wells, Nancy K. Steblay, and Jennifer E. Dysart, *Double Blind Photo Lineups Using Actual Eyewitnesses: An Experimental Test of Sequential Versus Simultaneous Lineup Procedure*, 39 LAW & HUM. BEHAV 1, 2 (2015).

4. *See, e.g.*, Steblay et al. *supra* note 3. Wells, *Double-Blind*, *supra* note 2, at 2.

5. *See, e.g.*, Steblay et al., *supra* note 3.

essential point by quoting an expert who said that identification is supposed to be about recognition, not about figuring things out.[6] Skeptics question this model. For example, Clark disparages the psychological assumptions behind the relative judgment/recognition dichotomy and argues for a model of memory of culprit appearance as a continuous variable instead of creating a dichotomy between legitimate "recognition" hits and illegitimate "relative judgment" hits.[7]

The relative judgment process is one explanation for the loss of hits when sequential lineups are used. But there are other possible explanations. A witness might hold back from making an identification in a sequential lineup simply because the witness is in doubt about whether the person shown is the culprit, and has hopes that someone later in the sequence can be identified without doubt. The difference in performance in simultaneous and sequential lineups could simply be due to features of the sequential lineup that make the witness more cautious and less likely to attempt an identification.

If the sequential lineup avoided false alarms[8] without any loss of correct hits, then it would be easy to say that sequential lineups are better than simultaneous ones. Unfortunately, that is not the case. The experimental data indicates that the sequential lineup decreases the rate of false alarms at the cost of increasing the rate of lost hits. In other words, the reform reduces the rate of mistaken identification of non-culprits, at the cost reducing the rate of correct identification of culprits. A meta-analysis of studies of simultaneous and sequential lineups[9] yielded a hit rate of .54 for simultaneous lineups and .43 for

---

6. D. Michael Risinger, *At What Cost?: Blind Testing, Eyewitness Identification, and What Can and Cannot Be Counted as a Cost of Reducing Information Available for Decision*, 58 How. L.J. (forthcoming 2015).

7. *See* Steven E. Clark, *Eyewitness Identification Reform: Data, Theory, and Due Process*, 7 PERSP. ON PSYCHOL. SCI. 279, 281 (2012). For a hypothesis about how simultaneous lineups might enhance witness's abilities to distinguish faces they had previously seen and those they had not, *see* John T. Wixted & Laura Mickes, *A Signal-Detection-Based Diagnostic-Feature-Detection Model of Eyewitness Identification*, 121 PSYCHOL. REV. 262–76 (2014).

8. In the terminology of this article, a "hit" occurs when a witness makes a correct positive identification. A "miss" occurs when a witness fails to identify a culprit who is present in a lineup. A "false alarm" occurs when a witness identifies someone other than the culprit.

9. Steblay et al., *supra* note 3. The administrator of a simultaneous lineup shows the photographs or persons in the lineup to the witness simultaneously, so that all can be viewed at once. *See, e.g.*, Steblay et al., *supra* note 3. When the sequential procedure is used, they are shown one at a time, and the witness is asked to make an identification decision about each before going on to the next one. Steblay et al., *supra* note 3. Critics of the simultaneous lineup assert that it produces false identifications in culprit-absent lineups, because a witness using "relative judgment" may choose the person who looks most like the perpetrator. Steblay et al., *supra* note 3.

sequential lineups,[10] and a false alarm rate of .15 for simultaneous lineups and .09 for sequential lineups.[11]

Additional hits would not be a benefit if they were produced by a procedure that increased hits without adding anything of value. That is theoretically possible, if the hit rate were increased by use of arbitrary or random criteria.

By analogy, suppose that a polygraph operator created a lie detector test that caught all liars because it automatically labeled everyone who took it a liar. That test would have a sensitivity of 100%. It would have a higher proportion of true hits than real testing. But it would not be worth considering. It adds nothing to the other evidence, making no change in the probability that the subject was lying. The hit rate and the false alarm rate would both be 100%. In Bayesian terms, its likelihood ratio would be 1, and multiplying prior odds by 1 does not change posterior odds.[12]

The same principle would apply if a polygraph operator performed a genuine test on some suspects, and then increased the hit rate by randomly re-designating some of the results as positive. Some of the random re-designations would be true hits–the ground truth would be that the subject was lying–but the re-designated positives would have no probative value. There is no difference between them and the inconclusive results that were not re-classified. If the extra hits that are obtained when a simultaneous lineup is used were produced purely by random guessing, then they would have no value.

The data indicate that the extra hits are not produced purely by random guessing. If a purely random process were the reason why simultaneous lineups have a higher hit rate, then the random process

---

10. *See* Clark, *supra* note 7, at 242. Fifty-one studies were utilized for this analysis. Clark, *supra* note 7, at 242. Clark did not count identifications of foils (fillers) as a false identification. Clark, *supra* note 7, at 243. The reason is that foils are known innocents who would not be prosecuted if they were falsely identified. Clark, *supra* note 7, at 243. He criticizes Steblay et al. for counting foil identifications as false identifications in their comparison of simultaneous and sequential lineups. Clark, *supra* note 7, at 243. In culprit-absent lineup experiments, the distinction between suspect and fillers is that the "suspect" or "designated innocent" is the person who most closely resembles the actual culprit. Steblay et al., *supra* note 7, at 118.

11. *Id.*; *see also* Andrew E. Taslitz, *Eyewitness Identification, Democratic Deliberation, and the Politics of Science*, 4 CARDOZO PUB. L., POL'Y & ETHICS J. 271, 275, 306 (2006) (recognizing the trade-off between simultaneous and sequential lineups and discussing its relevance to the deliberative process related to consideration of lineup reforms).

12. The likelihood ratio is calculated by dividing the probability of finding the evidence if the condition is true by the probability of finding the evidence if the condition is false. Here the probability of getting a positive result if the subject is lying is 100%. The probability of getting a positive result if the subject is not lying is also 100%. 100% divided by 100% is 1.

that produced the higher hit rate would be expected to cause innocent suspects to be falsely identified at the same rate that guilty suspects are correctly identified.[13] If use of simultaneous instead of sequential lineups caused an increase in hits of .11 times the number of guilty suspects, then one would expect an increase in false alarms of .11 times the number of innocent suspects. In the Clark meta-analysis, hits increase by .11 when simultaneous lineups are used, but false alarms increase only by .06, a smaller amount than what be expected if the extra hits were produced by random guessing.[14] Nevertheless, a portion of the extra hits could be due to guessing. If so, the extra hits would be balanced by an increase in the false alarm rate that would be detected by the measures of probative value that are discussed later in this paper.

When weighing hits lost against false alarms avoided, commentators often refer to the principle that it is better for a guilty person to go free than for an innocent one to be convicted. Convicting the innocent causes two harms by punishing the innocent while also letting the guilty go free (the conviction of an innocent person will normally end, or at least impede, the search for the real perpetrator). Acquitting the guilty only causes one harm, freeing a guilty person.[15] So far so good. Unfortunately, the trip from the experimental data to inferences about the ratio of guilty freed to innocent convicted is fraught with hazards.

Suppose, for example, that a given reform would decrease the false alarm rate by 10% while at the same time decreasing the hit rate by 10%. Policymakers might be tempted to interpret this data as meaning that when the reform is put into place in the field, there will be a 10% decrease in false identification of the innocent, balanced by a 10% decrease in correct identifications of the guilty. Unfortunately, that is not the case.

The hit rates and false alarm rates yielded by experimental data do not tell us the ratio of hits lost to false alarms avoided when an

---

13. I am counting only identifications of suspects. I am not taking into account identifications of fillers because fillers are known innocents, and the false identification of a filler would not result in charges being filed. In lab experiments, the "suspect" in a culprit-present lineup is the culprit (the person who was observed by the witnesses during the scenario), and the "suspect" in culprit-absent lineups is a person designated by the investigator because of resemblance to the culprit.

14. *See* Clark, *supra* note 7, at 242; Gary L. Wells, Nancy K. Steblay & Jennifer E. Dysart, *Eyewitness Identification Reforms: Are Suggestiveness-Induced Hits and Guesses True Hits?*, 7 PERSP. ON PSYCHOL. SCI. 264, 268 (2012).

15. *See* Steblay et al., *supra* note 3, at 129.

identification is made in a lineup. In order to know the latter, we need to know how frequently suspects included in lineups are the culprit (the guilty base rate).[16]

For example, suppose that lineups have a guilty base rate of 90%. In other words, police investigation is so accurate that 90% of the suspects in lineups are guilty. In light of that assumption, consider the meta-analysis data indicating that under experimental conditions, the switch from simultaneous to sequential lineups results in a reduction in hits from .54 to .43 and a reduction in false alarms from .15 to .09. In 1000 lineups with a 90% guilty rate, there would be 900 guilty suspects and 100 innocent suspects. Using a simultaneous lineup, 486 of the guilty suspects will be correctly identified, and 15 of the innocent suspects will be falsely identified. Using sequential lineups, 387 of the guilty suspects will be correctly identified, and 9 of the innocent suspects will be falsely identified. Under the stated assumptions, the change from simultaneous to sequential would mean that avoiding 6 false identifications of the innocent has been purchased at the cost of missing 99 correct identifications of the guilty, yielding a lost guilty/saved innocent ratio of 16.5. In short, assessing the impact of a lineup reform that reduces false alarms while also reducing correct identifications is not a simple matter of comparing the false alarm rate to the hit rate, because impact in the field depends partly upon the guilty base rate, something we can only guess about.

Impact in the field also depends upon other features of the legal system. Suppose that a policymaker believes that the guilty base rate is 90% and that no more than ten guilty persons should be freed to exonerate one innocent person. For such a policymaker, should the sequential reform be rejected because the ratio is 16.5 to one instead of 10 to 1? The answer is no, not necessarily. The policymaker could still support the change from simultaneous to sequential lineups. The reason is that the 16.5 ratio of hits lost to false alarms avoided does do not tell us the ratio of true convictions lost to false convictions avoided, and there are reasons to think that that ratio might be lower.

One reason it does not tell us that ratio is that the failure to identify will often not affect the binary variable to conviction or acquittal. Instead, it will have an impact on the continuous variable of length of sentence. Most cases are not tried. When a witness fails to make an identification in the 90% guilty base rate condition, the prosecutor is

---

16. *See* Clark, *supra* note 7, at 246–48 (discussing the relevance of the guilty base rate).

may decline to drop the charges because there is substantial non-identification evidence against the defendant. However, the failure to identify will help the defendant in plea bargaining. The likely result is that the plea bargaining will result in a lower sentence, not that the defendant will be acquitted.

Even among the cases tried, the ratio of correct identifications to false identifications will not be the same as the ratio of justified convictions to false convictions. In the condition in which the guilty base rate is .90, a miss in lineup ID might well be detected. If the police have succeeded in presenting lineup participants with suspects who are guilty 90% of the time, that means that the police have a good deal of other evidence incriminating the suspects that they choose for lineups. If the witness fails to identify the suspect, the police can proceed with that other evidence, and seek to develop more of it. And although the failure to identify will hurt the prosecution, a failure to identify does not mean that the witness will testify at trial that the suspect is not the perpetrator. As the witness learns of other evidence and other witnesses, the witness may well come around to see things the way the police see them.

In contrast, a false identification in the 90% guilty condition seems less likely to be detected. The identification is corroborated by other evidence of guilt, giving the prosecution and the trier of fact confidence that a guilty verdict is the right result. Moreover, the witness's confidence that the identification is correct is likely to increase as the witness learns about other evidence against the defendant. Thus, a 16.5-1 ratio of hits lost to false alarms avoided might well translate into a less than 10-1 ratio of true convictions lost to false convictions avoided.

There is yet another complication that points in the other direction. In trials, the legal system already privileges one type of error over the other. The trier of fact is told to follow a decision rule that avoids false positives even if that means tolerating false negatives. That decision rule is expressed to juries in the form of instructions not to convict unless the prosecution has proven its case beyond a reasonable doubt. If false negatives are also preferred to false positives in decisions about providing evidence to the jury, the preference is multiplied. For example, assume that the best decision rule is one that lets ten guilty go free to save one innocent. If that concept is applied both

by evidence providers in screening evidence and by evidence consumers in evaluating it, then the ratio will be higher than 10 to 1.[17]

In short, it is difficult to know whether the avoidance of false convictions is being purchased at too high a price. Even if the policymaker specified exactly what the ratio of guilty freed to innocent exonerated should be, that would not answer the question. One cannot determine from the experimental data what ratio a lineup reform will lead to. First, the ratio of lost hits to false alarms avoided depends partly on the guilty base rate, something that requires guesswork and that will vary from place to place. Second, a false identification does not necessarily lead to a false conviction, nor does a lost hit necessarily lead to a false acquittal.

If the preference for false acquittals over false convictions does not give us the answer to which procedure is best, where else can we look for the answer? One possibility is to assess the probative value of a positive identification, and to choose the procedure whose identifications have the greatest probative value.

One method of assessing the probative value of a positive identification is by calculating a "diagnosticity ratio." The diagnosticity ratio is a Bayesian likelihood ratio calculated by dividing the probability of an identification given that the suspect is the culprit by the probability of an identification given that the suspect is not the culprit–in other words, by dividing the hit rate by the false alarm rate.[18] The higher the ratio, the more probative value the identification has. The Steblay et al meta-analysis yields a ratio of 7.72 for sequential lineups and 5.78 for simultaneous lineups.[19] A Bayesian analysis based on those figures would indicate when there is a positive identification, the prior odds that the suspect is the culprit should be multiplied by 7.72 if the identification was made in a sequential lineup, and by 5.78 if the identification was made in a simultaneous lineup.

Of course, the diagnosticity ratio would vary depending upon which studies were included in the meta-analysis. A more modest ad-

---

17. *Cf.* Erik Lillquist, *Improving Accuracy in Criminal Cases*, 41 U. RICH. L. REV. 897, 904 (2007) (assuming that informational or cognitive obstacles prevent evidence providers and consumers from understanding and properly weighing what their counterparts are doing).

18. *See, e.g.*, Richard O. Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021 (1977) (suggesting the use of likelihood ratios as a guide to probative value); Dale A. Nance, *Naturalized Epistemology and the Critique of Evidence Theory*, 87 VA. L. REV. 1551, 1610–11 (2001) (suggesting the use of likelihood ratios as a means of converting random match frequencies to source probabilities).

19. Steblay et al., *supra* note 3, at 114.

vantage appears if one casts a wider net in including studies[20] or if one uses the studies chosen by Clark for meta-analysis.[21]

The diagnosticity ratios noted above are an assessment of the probative value of a positive identification by a witness of a suspect who is in the lineup. Suppose that the eyewitness does not identify the defendant as the culprit when the defendant is in the lineup. The defendant offers the non-identification as evidence of innocence. Here the diagnosticity advantage of sequential lineups may disappear or run in the other direction.[22]

Recently, critics have questioned the value of the diagnosticity ratio as a measure of the probative value of lineup identifications. They have argued that a more appropriate tool for comparing lineup procedures is ROC analysis,[23] which has been used for many years in

---

20. *See id.* at 107 tbl.1.

21. For simultaneous lineups, Clark reports a correct identification rate of .54 and a false identification rate of .15. Clark, *supra* note 7, at 242 tbl.2. For sequential lineups, he reports a correct identification rate of .43 and a false identification rate of .09. *Id.* Calculating ratios in the fashion of Steblay et al. (2011), this yields a ratio of 3.6 for simultaneous lineups and 4.8 for sequential lineups.

22. The data in Steblay et al., *supra* note 3, at 113 tbl.3, is suggestive. In culprit-absent lineups, under the sequential procedure the witness refrained from identifying the designated innocent at a .85 rate. *Id.* In culprit-present lineups, the witness picked the culprit at a .44 rate. *Id.* Where E is the evidence that the suspect was not identified, the probability of E given the suspect is not the culprit is .85, and the probability of E given the suspect is the culprit is .56, yielding a likelihood ratio of 1.52. The equivalent calculation for the simultaneous lineup yields a likelihood ratio of 1.5. Using the more inclusive meta-analysis data reported in Steblay et al., *supra* note 3, at 107 tbl.1, the likelihood ratio is 1.4 for sequential lineups, compared to 1.56 for simultaneous lineups.

23. *See* Scott Gronlund, John Wixted, & Laura Mickes, *Evaluating Eyewitness Identification Procedures Using ROC Analyses*, 23 CURRENT DIRECTIONS IN PSYCHOL. SCI. 3, 5 (2014); Laura Mickes et al., *Receiver Operating Characteristic Analysis of Eyewitness Memory: Comparing The Diagnostic Accuracy of Simultaneous Versus Sequential Lineups*, 18 J. EXPERIMENTAL PSYCHOL.: APPLIED 361, 361 (2012) [hereinafter Mickes, *Operating Characteristic*]; Wixted & Mickes, *supra* note 7, at 265 fig.1; *cf.* Christian A. Meissner, Colin G. Tredoux, Janat F. Parker, & Otto H. MacLin, *Eyewitness Decisions in Simultaneous and Sequential Lineups: a Dual-Process Signal Detection Theory Analysis*, 33 MEMORY & COGNITION 783, 784 (2005). Gary L. Wells, a prominent proponent of the sequential advantage view, responded to the Gronlund et al. article by saying that in questioning the probative value advantage of sequential procedure, the authors "did so by pointing to a few selected contrasts in a few studies rather than relying upon broad meta-analyses. All literatures have some outliers." Gary L. Wells, *Eyewitness Identification: Probative Value, Criterion Shifts, and Policy Regarding the Sequential Lineup*, 23 CURRENT DIRECTIONS IN PSYCHOL. SCI. 11, 12 (2014). Clark notes six quantitative measures of probative value that could be applied to lineup research. *See* Clark, *supra* note 7, at 244–46. He indicates that the procedures recommended by lineup reformers, including the change from simultaneous to sequential lineups, "the probative value of a suspect identification was numerically higher" for the recommended procedures, though for the sequential lineup reform the effect size was small. Clark, *supra* note 7, at 246.

assessment of medical tests.[24]

To illustrate the concept of ROC analysis, suppose that an airport metal detector is being used as a test for detecting metal weapons such as guns and knives. When this hypothetical detector is set on "high," it is very sensitive. In other words, it has a good hit rate, and will rarely miss metal knives or guns. However, its specificity will be poor. It will issue many false alarms because it will be set off by non-weapon metal items, such as belt buckles, pens, and paper clips. When it is set on "low," its sensitivity will be less – it will miss some small metal weapons – but will issue fewer false alarms. There are a variety of intermediate settlings. The operators choose a setting based on factors such as the perceived threat level and the need to process fliers quickly.

One way to evaluate this test would be to ask the operator to choose a particular setting and derive a diagnosticity ratio (likelihood ratio) at that setting. The probative value of evidence that the detector issued an alarm at that setting could be assessed with a likelihood ratio derived by dividing the hit rate by the false alarm rate.[25] However, a fuller picture of the discriminatory power of the detector could be obtained by comparing the hit rate to the false alarm rate at various settings, not just the one chosen by the operator on that particular occasion. That is what ROC analysis aims to do.

As with the hypothetical metal detector, the results of a medical diagnostic test usually fall along a continuum. Mickes and her colleagues use the following hypothetical example.[26] Suppose that a blood test always yields a result of 0 to 100.[27] The doctor administering the test uses a score of 50 as the cutoff point for a disease.[28] At that point, 63% of the patients who have the disease will have positive results (scores of 50 or more), as will 16% of the patients who do not have the disease. The hit rate is 63% and the false alarm rate is 16%,

---

24. *See, e.g.,* Mark H. Zweig & Gregory Campbell, *Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine,* 39 CLINICAL CHEMISTRY 561, 561 (1993).

25. The likelihood ratio would be derived by dividing the probability that the detector would issue an alarm given that the passenger is carrying a metal weapon by the probability of an alarm given that the passenger is not carrying a metal weapon: P(A—W)/P(A—-W). Multiplying the prior odds that the passenger had a weapon by the likelihood ratio would yield the posterior odds. The degree to which evidence of an alarm by the detector increases the odds that a passenger has a weapon depends upon how large the likelihood ratio is. The larger the likelihood ratio, the greater the probative value of evidence that the detector issued an alarm.

26. *See* Mickes, *Operating Characteristics, supra* note 23, at 362–63.

27. Mickes, *Operating Characteristics, supra* note 23, at 362.

28. Mickes, *Operating Characteristics, supra* note 23, at 362.

yielding a likelihood ratio of 63/16 = 3.9.[29] An analyst using ROC analysis would not be satisfied with assessing the probative value of the test using only the likelihood ratio at that one operating point. She would also want to know how the hit rate compares to the false alarm rate at other cutoff points. She would plot an ROC curve showing how diagnostic the test was at various cutoff points, and evaluate its overall probative value by measuring how far the values on that curve deviated from those that would be obtained using a test that had no probative value, i.e. a test whose hit rate and false alarm rate were equal at every cutoff point.[30]

The blood test example involves a test that yields specific, objectively obtained blood count numbers. ROC analysis can also be applied when the test result has to be judged subjectively, relying upon the judgment of the person performing the test. Mickes et. al. (2012) gives the example of a radiologist seeking to use a mammogram to determine whether a malignant tumor is present.[31] In a study that whose objective was to compare the efficacy of judgments based on film and digital mammograms,[32] the investigators did not simply ask the radiologists to make a judgment whether a malignant tumor exists and then compare hits with false alarms.[33] They also asked the radiologists to supply confidence ratings on a 7 point scale, from 1 for "definitely not malignant" to 7 for "definitely malignant."[34] That allowed them to plot ROC curves showing the hit rate and the false alarm rate at different confidence levels.[35] Comparison of the ROC curves for digital and film mammography indicated that, for example, digital mammography was a superior technique when patients were women under 50 years old.[36]

Lineup procedures can be compared in a similar fashion. The witnesses observing a lineup are analogous to the radiologists evaluating a mammogram. The lineup can be viewed as a test in which the witnesses are asked to detect whether the culprit is present, just as radiologists are asked to detect whether a malignant tumor is present.

29. Mickes, *Operating Characteristics*, *supra* note 23, at 362–63.
30. *See* Appendix, Fig.1.
31. Mickes, *Operating Characteristics*, *supra* note 23, at 364.
32. *See* E.D. Pisano et al., *Diagnostic Performance of Digital versus Film Mammography for Breast-Cancer Screening*, 353 NEW ENG. J. OF MED. 1773, 1773–74 (2005).
33. Mickes, *Operating Characteristics*, *supra* note 23, at 364.
34. Mickes, *Operating Characteristics*, *supra* note 23, at 364.
35. Mickes, *Operating Characteristics*, *supra* note 23, at 364.
36. *See* Appendix, Fig.2.

Instead of simply asking participants in experiments comparing sequential and simultaneous lineups to decide whether they have seen the culprit in the lineup, investigators could also ask them to state their level of confidence. Then the hit rate and false alarm rates for the two procedures could be compared at different levels of confidence, and ROC curves could be constructed and compared.

The basic point of critics who urge use of ROC analysis is that the conventional diagnosticity ratio used in comparing the probative value of identifications made in simultaneous and sequential lineups is inadequate because it is based on the hit rates and false alarm rates obtained at only one point in the ROC curve. A more accurate picture could be obtained by asking experimental participants to state confidence levels and by comparing the ratio of hit rates to false alarm rates at various levels of confidence. This approach has been tried in a few experiments, and the results suggest that simultaneous lineups may be just as good as or better than sequential ones.[37]

If sequential lineups have no advantage over simultaneous ones, what accounts for the many studies indicating that the ratio of hits to false alarms – the diagnosticity ratio – is better for sequential lineups than for simultaneous ones? One possible explanation is that the sequential lineups cause witnesses to be more cautious in making identifications.[38] In other words, witnesses require a higher confidence level before making the identification when viewing a sequential lineup.

The probative value of a positive result on a test is influenced by both the discriminatory power of the test and the degree of caution used by the evaluator. The greater the caution – the greater the degree of confidence required before evaluator will say that a condition is present – the greater the probative value of a positive judgment. This proposition is intuitively plausible. Compare the probative value of a conviction as evidence that a defendant committed a forbidden act. If the jury is instructed not to convict unless it is convinced beyond a reasonable doubt, the observer is on firmer ground in believing that a conviction means that the defendant is guilty than if the jury is merely instructed to convict if it believes the defendant is probably

---

37. Mickes, *Operating Characteristics, supra* note 23, at 362; John T. Wixted & Laura Mickes, *A Signal-Detection-Based Diagnostic-Feature-Detection Model of Eyewitness Identification*, 121 PSYCHOL. REV. 262, 271 (2014).

38. *See* Christian A. Meissner et al., *Eyewitness Decisions in Simultaneous and Sequential Lineups: A Dual-Process Signal Detection Theory Analysis*, 33 MEMORY & COGNITION 783, 784 (2005).

guilty. Two factors affect the probative value of a conviction as evidence of guilt: the jury's astuteness in weighing the evidence, and the decision criterion employed by the jury.

Lineups are no exception to the principle that the more confident the judgment, the greater its probative value. The once-common notion that confidence has little or no relationship to accuracy has been debunked. The more confident a witness is that an identification is correct, the greater the probative value of the identification.[39]

Thus, the value of a lineup identification is affected by the witness's ability to correctly identify (which may be influenced by system variables such as whether the lineup is sequential or simultaneous) and the witness's decision criterion. The higher the degree of certainty that the witness requires before being willing to identify, the greater the probative value of the identification. Hence sequential lineups show higher hit rates and a lower rate of false alarms because the witnesses are employing more conservative decision criteria.

The diagnosticity ratio measures the probative value of a lineup at the decision point chosen by the participants. If the witnesses are more cautious in a sequential lineup than in a simultaneous one, then an identification made in a sequential lineup will have more probative value. Use of ROC analysis is a way of attempting to evaluate the probative value of lineup procedures independently of the decision criterion employed by the witnesses[40]. In other words, ROC analysis seeks to reach a judgment about how discriminating a test is apart from the effect of how cautious the witness is in applying it.[41]

As mentioned earlier, a few experiments have used ROC analysis to assess the probative value of lineup identifications, and the scholars conducting those experiments suggest that simultaneous lineups may have an advantage.[42] When there is control for the decision criteria used by lineup witnesses, identifications from simultaneous lineups may be more probative than identifications from sequential ones.[43]

What implications do these studies have for policy makers who are setting forth standards for conducting lineups? First, it is important to bear in mind that the ROC proponents have only conducted a

---

39. *See* Mickes, *Operating Characteristics, supra* note 23, at 366; N. Brewer et al., *The Confidence-Accuracy Relationship in Eyewitness Identification: the Effects of Reflection and Disconfirmation on Correlation and Calibration*, 8 J. EXPERIMENTAL PSYCHOL.: APPLIED 44, 45 (2002).
40. *See supra* notes 37–38.
41. *See supra* notes 37–38.
42. *See supra* notes 37–38.
43. *See supra* notes 35–36.

few experiments. Literally dozens of other experiments that have been conducted without ROC analysis[44] We do not have the data to perform an ROC analysis on those experiments, but the diagnosticity ratio approach to assessing probative value indicates an advantage for sequential lineups. Even supposing ROC analysis to be superior to likelihood ratio analysis, the latter cannot be dismissed as irrelevant to probative value. After all, it is based on a comparison hits to false alarms – the higher the proportion of hits compared to false alarms, the more probative the evidence. Even if this is not the best way to assess probative value, it's certainly a way that takes relevant factors into account and provides useful information[45]

Secondly, even if the advantage of sequential lineups is due to use of more conservative decision criteria by witnesses, one might want to use the sequential procedure precisely because it does cause use of more cautious decision criteria. Other ways of seeking more cautious criteria – instructions or exclusion of low-confidence identifications – might not be as feasible. The sequential procedure does at least provide one tested way of achieving the effect.[46]

The discussion above should give the reader a feel for the daunting problems involved in deciding whether sequential lineups are superior to simultaneous ones. (1) The preference for false acquittals over false convictions does not provide clear guidance. The comparison of correct hits to false alarms in the experimental data does not

---

44. *See, e.g.*, Nancy K. Steblay et al., *Sequential Lineup Laps and Eyewitness Accuracy*, 35 LAW. & HUM. BEHAV. 262, 262–74 (2011) (showing meta-analysis that includes 72 such studies).

45. Mickes argues that where the information needed to perform ROC analysis is not available, the answer is not to compute the diagnosticity ratio, but to use a difference score, based on the difference between the hit rate and the false alarm rate (d'). Laura Mickes et al., *Missing the Information Needed to Perform ROC Analysis? Then Compute d', Not the Diagnosticity Ratio*, 3 J. APPLIED RES. IN MEMORY & COGNITION 58, 60 (2014). If these authors are correct, the policy implications are unclear; Clark's review of fifty-one pre-ROC simultaneous versus sequential literature using d' as the outcome measure, showed that the simultaneous and sequential lineups yielded essentially identical average scores. *See id.* at 62; Clark, *supra* note 7, at 242 tbl.2. The seventy-two studies chosen by Steblay for their meta-analysis are more favorable to the sequential advantage than those chosen by Clark. *See* Steblay et al., *supra* note 3, at 99 tbl.1.

46. Wells argues that even if the increased probative value (diagnosticity) of sequential lineups is due to use of more conservative decision criteria instead of increased discrimination, identifications from sequential lineups can still be better trusted. Wells, *supra* note 23, at 12. "The simple way to think about this distinction is that eyewitnesses are less likely to make an identification with the sequential procedure than with the simultaneous procedure, but when they do make an identification with a sequential procedure, it is more trustworthy for the prosecutor." Wells, *supra* note 23, at 12. The fact that identifications are made with greater confidence increases their value as proof of guilt. But what about non-identifications that are offered as evidence of innocence? They are less valuable as evidence of innocence if witnesses would decline to make an identification unless they had a high degree of confidence.

even tell us what that ratio will be when the procedures are applied in the field, because that ratio depends upon the guilty base rate. Even if we knew the guilty base rate, the ratio of correct hits to false alarms does not tell us the ratio of correct convictions to false convictions. Based on current data about hits lost and false alarms avoided, it is difficult to say whether the protection of the innocent is purchased at a reasonable cost. (2) Estimates of the probative value of a correct identification are not conclusive, either, though they are certainly relevant. The proper measure of probative value is debatable, as illustrated by the controversy about ROC analysis and the diagnosticity ratio. Even if we knew how to measure probative value, the probative value of a positive identification is not the only relevant probative value. The probative value of a non-identification should also be considered. And probative value cannot be the only issue. However difficult they are to apply, the policy-maker also needs to take into account value judgments about the social cost of false alarms and lost hits.

SEQUENTIAL LINEUPS WITH A SECOND LAP

In jurisdictions that have adopted lineup reform that favors sequential lineups, it is common to allow the witness to see the sequence more than once at the request of the witness.[47] This "second lap" compromise adds a feature that is comparable to doing a sequential lineup and then following it with a simultaneous lineup if the witness makes no identification. The witness has already seen all the photos or individuals, and could make a relative judgment about which one looks most like the perpetrator. If proper records are kept, the trier of fact will learn that the witness did not make an identification in the first sequential lineup, even if it also learns that an identification was made in the later simultaneous lineup.

I have discussed in earlier paragraphs the difficulty of assessing the probative value of different lineup procedures. But for the purpose of argument, assume that the sequential lineup yields evidence with greater probative value than the simultaneous lineup. The second-lap compromise procedure could then be viewed as one that elicits evidence with high probative value (the sequential lineup) and then follows it with evidence of lesser probative value (the simultaneous

---

47. *See* Wells, *Double-Blind*, *supra* note 2, at 7; Nancy K. Steblay et al., *supra* note 44, at 262.

lineup). In the law of evidence, there is nothing unusual about allowing evidence with high probative value to be supplemented with evidence of lesser probative value. In general, evidence law has given up on the idea of restricting admissibility to evidence that is extremely accurate, in favor of a free proof approach that relies upon the trier to consider all of the evidence together and give it proper weight. For example, disinterested witnesses are better than interested witnesses, but under modern law both can testify. DNA evidence is more diagnostic than eyewitness identification, yet both are admissible. An identification by someone familiar with the perpetrator is more diagnostic in lab experiments than an identification by a stranger, but in trial situations, the fact that a witness familiar with the defendant says he was not the perpetrator (or gives him an alibi) does not preclude evidence of an identification by a stranger who testifies that he was the culprit.

Similarly, even if the sequential lineup is more probative than the simultaneous lineup, that consideration alone would not justify excluding the simultaneous lineup when offered as a supplement. To offer a second lap is to, in effect, follow a sequential lineup with a simultaneous one. The ultimate question is whether the second lap is prejudicial in the sense that the jury will not be able to properly evaluate a second lap hit, even when it is also informed that the witness did not identify the suspect during the first lap. This particular issue has not yet been studied systematically, and in the absence of evidence to the contrary, there seems to be no reason to depart from the usual practice of allowing the trier of fact to consider both the gold and the brass, when both have probative value.[48]

Even if sequential-only lineups are found to have an advantage in lab experiments over sequential-plus lineups,[49] that circumstance would not tell us how the evidence will impact determinations in the legal system. In cases that go to trial, the fact that the witness did not identify the defendant the first time could be effective impeachment material on cross-examination, and might fit in with other evidence of unreliability of the witness or of innocence of the defendant in a way

---

48. It is not clear whether it is practical to mandate a one-lap procedure. In cases in which the witness spontaneously requests a second lap, the administrator in the field might give one even if the procedure does not provide for it.

49. *See, e.g.*, Steblay et al., *supra* note 44, at 262–73 (2011) (demonstrating through two experiments the way in which sequential only lineups may be advantageous over sequential-plus lineups). The authors use the diagnosticity ratio (hit rate divided by false alarm rate) as the measure of probative value. Steblay et al., *supra* note 44.

that mitigates the effect of additional false alarms caused by the second lap.

In the trial context, the probative value of an item of evidence, such as an identification in a lineup, depends upon how it fits in with other evidence. For example, suppose that the defendant comes under suspicion because of a DNA match in a trawl through a database of persons arrested for felonies. The DNA evidence alone is incriminating, but it is not necessarily conclusive. The probability of a coincidental match because of an identical genetic profile on the loci tested is usually small, but there is always the danger of a match due to sample contamination or lab error. Nonetheless, it is undoubtedly true that a DNA identification, considered in isolation, has greater probative value than a lineup identification, considered in isolation. Nonetheless, despite the higher probative value of DNA identification evidence, lineup identification evidence is worth hearing when offered as a supplement to DNA evidence. If the victim also identifies the defendant out of a lineup, the two items of evidence combined – the DNA match and the identification – fit together in a way that makes them overwhelmingly incriminating.[50]

Now suppose that a victim is viewing a lineup that contains a suspect found through a database trawl. The lineup is a fair one: a non-witness would not know which participant is the suspect by reading the victim's description, there are enough fillers, and the victim has not been cued to the suspect because the administrator does not know who the suspect is. The value of an identification that fits with the DNA evidence is so great that it would seem wise to allow a victim a second lap if she requested it. The point here is not that this situation is the most common one, but that the value of lineup evidence cannot be assessed from experimental data alone, but only in view of other considerations, including whether it is corroborated by independently obtained evidence.

The second lap can be conceptualized in at least two ways. One is to view it as a variation of an identification procedure that makes the overall procedure less diagnostic. Another is to view it as a separate step that adds additional evidence to the sequential non-identification. If the trier of fact is able to give the two steps their proper weight no harm is done, and in some instances the extra evidence supplied by

---

50. *See* Williams v. Illinois, 132 S. Ct. 2221, 2244 (2012).

the second step will fit in with other evidence in a way that makes it highly probative.

One possible objection to the second lap is that it will lead to a higher identification rate that "spoils" witnesses. This point would be an extension of an argument by Steblay and her colleagues against simultaneous lineups.[51] If a witness identifies a filler in a culprit-absent simultaneous lineup, then the witness is "spoiled" and cannot be used to make a subsequent correct identification in a culprit-present lineup.[52] Because of the lower identification rate in sequential lineups, they argue, witnesses are less likely to be "spoiled" by them.[53]

But the spoiler effect has good and bad aspects. A witness who is spoiled for subsequent correct identification of the actual culprit is also spoiled for a subsequent false identification of an innocent person. Moreover, the fact that a witness is spoiled (or at least highly impeachable) for trial testimony does not mean that the witness is spoiled for investigatory purposes. The police can still show the witness other suspects and then use the identification as a reason to investigate the other subjects more thoroughly and develop other evidence of guilt.

Another reason why it might be acceptable to tolerate simultaneous lineups as a supplement to sequential lineups is that there are differences between experimental conditions and field conditions that affect the trade-offs involved in comparing simultaneous and sequential lineups. Professor Risinger has provided a thought-provoking example.[54] He suggests that lineup suspects are often chosen because of tips or non-identification evidence incriminating the suspect.[55] (A neighbor might report suspicions, or a traffic stop might yield a suspect because he had goods from a robbery in his car.[56]) In cases in which suspects are innocent, tip-based simultaneous lineups may be less error-prone than in experimental conditions. In an experiment, the investigator knows the identify of the culprit and can place someone who looks like the culprit in the lineup as the "designated innocent."[57] If the mock witnesses use relative judgment, they are likely to identify the "designated innocent" because he looks more like the cul-

---

51. Steblay et al., *supra* note 3, at 126.
52. Steblay et al., *supra* note 3, at 126.
53. Steblay et al., *supra* note 3, at 126.
54. Risinger, *supra* note 6.
55. Risinger, *supra* note 6.
56. Risinger, *supra* note 6.
57. Risinger, *supra* note 6.

prit than the fillers do.[58] In a tip-based field lineup, if best practices are followed, all of the fillers will fit the witness's description—or, if the suspect does not look like the description, they will resemble the suspect.[59] In the case of an innocent suspect in a tip-based lineup, the official choosing the fillers does not know what the real culprit looks like.[60] Everyone in the lineup, suspect and fillers, would have an equal chance of being the one who looks most like the culprit. If the witness makes an identification based on relative judgment, most of the time the witness will choose a filler instead of the suspect.[61]

* * * *

An academic lawyer cannot help but be awed by the many careful and thoughtful empirical studies of lineups. Nonetheless, despite decades of study, the question whether sequential lineups are superior to simultaneous lineups is still open to debate. A reasonable way to accommodate doubt is to use both types of lineups by following a sequential lineup with a second lap. That approach is consistent with the general presumption in evidence law that evidence should be admitted when it has probative value, trusting the trier of fact to sort the grain from the chaff.

---

58. Risinger, *supra* note 6.
59. Risinger, *supra* note 6.
60. Risinger, *supra* note 6.
61. If the tip-based lineups hypothesized by Professor Risinger were common, their existence would affect the inferences drawn from data from field studies describing the rate of identification of fillers. In field studies the investigators do not know which identifications are false alarms and which are hits. The proportion of filler identifications has, however, been used as a clue to the false alarm rate. If a procedure produces a higher rate of filler identifications than a comparison procedure, then one can infer that it has a higher rate of false alarms—that is, a higher rate of identifying not only fillers, but also of identifying innocent suspects. Viewed that way, a high rate of filler identification is a strike against a lineup procedure. *See* Wells, *Double-Blind, supra* note 2, at 13. But a high filler identification rate could also indicate that many of the lineups are tip-based lineups that include innocent suspects who are not more likely to resemble the culprit than the fillers.

APPENDIX 1
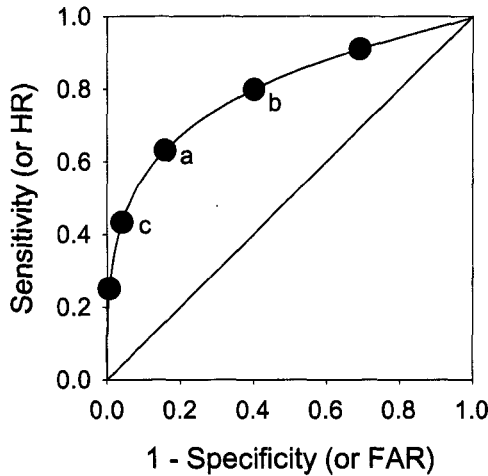
This figure is an hypothetical illustration of an ROC curve. The vertical axis shows the hit rate (HR) at various cutoffs. The horizontal axis shows the false alarm rate (FAR). The a, b, and c designations at the data points could, in the lineup context, represent the results obtained at different levels of confidence. Thus, data point "c" could refer to the results obtained when witnesses report a very high level of confidence on a confidence scale, while data point "b" could represent the results when witnesses report a much lower level of confidence. At point "c" the hit rate is just above .2, and the false alarm rate is close to zero. At point "b" the hit rate and the false alarm rate are both much higher. The solid diagonal line represents the no-probative-value line, where the hit rate and the false alarm rate are equal. The larger the area between the curved line and the no-probative-value line, the greater the probative value of the lineup procedure. The figure (but not this explanation) comes from figure 2 in Mickes et al. (2012).

APPENDIX 2

B Women Younger Than 50 Yr
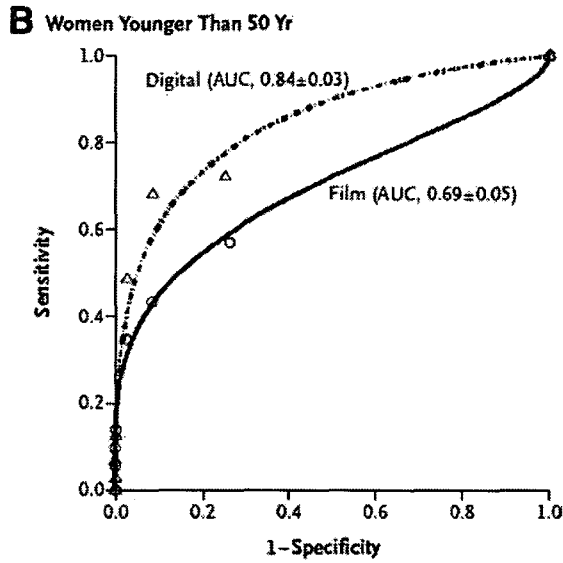
This figure shows the results of an actual experiment comparing the probative value of judgments made from digital mammogram displays compared to those made from film mammogram displays. The ROC curve indicates that judgments made from digital mammogram displays have greater probative value. The figure (but not this explanation) comes from Figure 3B in Mickes et al. (2012).