Journal of Applied Research in Memory and Cognition 5 (2016) 277–283

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Eyewitness Identification: Research, Reform, and Reversal☆

Molly B. Moreland* and Steven E. Clark

University of California, Riverside, United States

Eyewitness identification research and reform are being reconsidered in light of research suggesting that reforms that were once thought to increase identification accuracy may have little effect on accuracy or may actually decrease accuracy. This article addresses three questions: How should eyewitness identification procedures be evaluated? How can the research-policy collaboration prevent policy revisions and reversals? And how can the research-policy collaboration prepare for revisions and reversals?

Keywords: Eyewitness identification, Public policy, Policy reversals, Scientific reversals, Eyewitness identification reform and best practices

Mistaken eyewitness identification errors have long been implicated as a major contributing factor in false convictions of the innocent (Gross & Shaffer, 2012; Munsterberg, 1908). The relationship between mistaken eyewitness identifications and false convictions has been established most dramatically in post-conviction appeals in which DNA evidence established the error of the identification and the innocence of the defendant (Garrett, 2012; Gross & Shaffer, 2012).

This linkage between false identifications and false convictions has given energy to a movement to reform the police procedures used to obtain eyewitness identification evidence as well as the legal procedures that regulate how that evidence is evaluated by judges and juries. Many of these reforms have already been adopted through legislation or as law enforcement best practices. But now the road to reform seems less clear, as recent scientific evidence suggests that some recommended procedures may not increase the accuracy of eyewitness evidence or criminal justice outcomes—and may even lead to decreases in accuracy. The new evidence implies possible policy revisions and reversals—specifically that new procedures that were once recommended will no longer be recommended.

This article provides a brief review of some of the key reforms and raises three questions: How should proposals for reform be evaluated? How can policy reversals be avoided? And—on the view that some policy reversals and revisions are inevitable—how can policy makers prepare for them?

Research and Reform

There are two general procedures used by police to collect eyewitness identification evidence. In a one-person showup, the police present a single suspect to the witness and ask if that suspect is the person who committed the crime. In a typical lineup, police present the witness with a single suspect along with some number of fillers—other lineups members who are known to be innocent.

Eyewitness identification experiments simulate these conditions by presenting participants with a staged crime, live or on video, followed by a lineup or showup. Participants make an identification response and provide an indication of their confidence in that response. Researchers typically compare two conditions, one in which the suspect is guilty and one in which the suspect is innocent; thus the true identity of the perpetrator is known.

Eyewitness researchers have made dozens of recommendations designed to increase the reliability of eyewitness evidence. In recent years, the field has zeroed in on a few key recommendations that have been identified as best practices and in some

Authors Note

Molly B. Moreland, University of California, Riverside, United States; Steven E. Clark, University of California, Riverside, United States.

☆ Please note that this paper was handled by the current editorial team of JARMAC.

* Correspondence concerning this article should be addressed to Molly B. Moreland, University of California, Riverside, United States. Contact: mmore010@ucr.edu

cases incorporated into legislation. These recommendations include:

1. Avoid the use of one-person showup identification procedures.
2. Construct lineups so that the suspect does not stand out.
3. Provide instructions to the witness, stating that the perpetrator of the crime may not be present, and that they are not required to identify anyone.
4. Present lineups sequentially, one person at a time, rather than simultaneously with all lineup members at the same time.
5. Lineups should be administered by a person who is blind or blinded as to the position of the suspect in the lineup.
6. Ask the witness to provide a statement of confidence in his or her identification response.

These recommendations have been incorporated into police procedures through various mechanisms. In some cases, states have signed certain police procedures into law. For example, New Jersey, Connecticut, and North Carolina require that police departments present lineups sequentially, rather than simultaneously. Other states, including California, New York, and Rhode Island, have not implemented reform through legislation, but have developed guidelines and model policies at the state or county level. For many U.S. states the policies are non-existent or unclear. This variation across states raises the question: What standard should be applied in deciding which procedures to use? This question is important, because, depending on the standard that is applied, one could accept or reject all of the current reforms. We explore this question regarding the assessment of eyewitness identification procedures next.

Assessment of Eyewitness Identification Reforms

Increased Accuracy With No Cost–Benefit Trade-Off

One very high standard for the adoption of a new procedure is that it should produce benefits, with little or no cost. In the realm of eyewitness identification reform the claim has been made that the recommended reforms increase accuracy either by reducing the risk of false identifications with little or no reduction in correct identifications, or by increasing the correct identification rate with no increase in false identifications.

If these no-cost claims were true, policy decisions to implement the reforms would be relatively simple. The reform procedures would formally dominate the old procedures such that a decision to not implement them would appear irrational. However, the no-cost claim is unambiguously contradicted by data for most of the proposed reforms (Clark, 2012). Thus, the policy decisions almost always involve cost–benefit trades-offs, making the policy decisions much more complicated.

Increased Accuracy With Cost–Benefit Trade-Off

In the 20/20 hindsight of the data, the no-cost standard seems unrealistic. However the goal of increased accuracy can still be achieved even with trade-offs in errors—to the extent that recommended procedures reduce false identifications more than they reduce correct identifications, or increase correct identifications more than they increase false identifications. This raises questions about how one measures overall accuracy and what trade-offs are acceptable.

Measurement of Overall Accuracy

Simply put, the overall accuracy of identification outcomes will be high to the extent that the correct identification rate is high and the false identification rate is low. Despite that simple principle, the discussion of the various measures of overall accuracy (also referred to as diagnostic accuracy) may become rather technical. Prior to 2012, overall accuracy was typically measured as the ratio of correct identifications to false identification, C/F (Wells & Lindsay, 1980). However, the C/F ratio measure conflates accuracy and response bias. Specifically, when witnesses become more conservative, lowering both the correct and false identification rates, the C/F ratio increases. More recently, researchers have begun reporting measures of overall accuracy derived from signal detection theory, including d' and the area under the receiver operating characteristic (ROC) curve. The measure d' generally approximates the ROC curve, and can be used in lieu of ROC analysis in the absence of confidence or bias partitions (Mickes, Moreland, Clark, & Wixted, 2014). Clark (2012) and Palmer and Brewer (2012) used different calculations of d' for meta-analyses where ROCs could not be calculated. Rather than expand on the technical details, we simply note that these signal detection measures, ROC analysis and d', are broadly used in the medical and social sciences as measures of diagnostic accuracy and risk assessment. However, their specific application to eyewitness identification has not been embraced by all researchers (see Wells, Smalarz, & Smith, 2015 and Wixted & Mickes, 2015a, 2015b for a debate on this point).

Assessment of Acceptable Trade-Offs

When considering cost–benefit trade-offs one needs to take into account the relative costs of the two kinds of errors, false identifications of the innocent versus missed identifications of the guilty, as well as the opportunities for those two kinds of errors to occur (Clark, 2012). There is widespread agreement that the costs associated with a false conviction are greater than those associated with a false acquittal (see Volokh, 1997). To the extent that this cost inequality applies to identification evidence the costs associated with a false identification of an innocent suspect will be greater than the costs associated with a false non-identification of a guilty suspect. Optimal identification procedures will reflect this cost inequality by making the more costly error less likely than the less costly error. In assessing these error rates one must also consider the guilty and innocent base rate: the proportions of identification procedures for which the suspect is guilty or innocent. Optimal identification procedures should also reflect these error opportunities defined by the guilty and innocent base rates.

Importantly, these considerations are not about the diagnostic accuracy or the discriminability between suspects who are guilty versus innocent. Rather, they are about the criterion

placement and the likelihood of making positive versus negative responses—identifying someone versus identifying no one. A high criterion makes positive responses less likely and therefore reduces the likelihood of a false identification error whereas a low criterion makes negative responses less likely and therefore reduces the likelihood of a false non-identification error. Although the costs and base rates may be difficult to estimate, the calculations for the optimal decision criterion are well-known (Clark, 2012; Peterson, Birdsall, & Fox, 1954).

Decisions of criterion placement, which are at their core decisions about the trade-off in errors, should reflect important legal and social values, but within a statistical decision theory framework, they are secondary to considerations about discriminability. Specifically, in deciding which procedure to implement, policy makers should follow two steps: First, choose the procedure that yields the highest ROC, and then identify the criterion—that is, the position on that ROC curve—to determine the appropriate error trade-off for that procedure. We will have more to say about criterion placement later.

We should be clear that we are assessing identification procedures in terms of the discriminability between suspects who are guilty and suspects who are innocent. In making this point explicit, we acknowledge that there may be circumstances in which social, legal, or moral values might trump discriminability. For example, our Fifth Amendment protections from self-incrimination may reduce the diagnostic accuracy (discriminability) of criminal trials (see Laudan, 2006); yet this is a trade-off between core social values and accuracy that we are willing to make. A complete theory that integrates social values and discriminability is beyond the scope of the present paper. For now we apply the decision framework based on discriminability and error trade-offs to eyewitness identification reform.

Application to Recommended Eyewitness Identification Reform

Showups and Lineups

Legal scholars have criticized showups as highly suggestive (Gross, 1911; Wall, 1965). The U.S. Department of Justice provided best-practices guidelines on eyewitness identification (U.S. Department of Justice, Office of Justice Programs, National Institute of Justice, 1999) that assumed an "inherent suggestiveness" of showup procedures and, although the guidelines did not endorse a complete ban on showup procedures, they prescribed specific procedures to minimize bias (e.g., use of unbiased instructions, documentation of witness descriptions and confidence). Meta-analytic literature reviews have shown that, relative to showups, lineups reduce the false identification rate with no loss of correct identifications (Clark, 2012). This is the one recommendation that appears to show the no-cost pattern. To that extent, a recommendation to avoid the use of showup identification procedures seems to be on solid ground.

However, beyond this general comparison things get more complicated, largely as a result of the differences between experiments and actual criminal investigations. It is, of course, a basic rule of experimental methodology that when comparing Condition A to Condition B, all other variables should be held constant. Thus, experimental comparisons typically compare one-person photographic showups to multi-person photographic lineups, with the interval between the staged crime and the identification held constant. However, one of the potential benefits of showups is that they can be conducted more easily and more quickly than a lineup, suggesting that the key comparison may be between showups conducted sooner versus lineups conducted later. The few studies that have compared early showups to later lineups have produced mixed results: higher accuracy for showups in some cases (Yarmey, Yarmey, & Yarmey, 1996), higher accuracy for lineups in other cases (Wetmore et al., 2015). The sooner-later dimension is only one way in which showups and lineups typically differ. Showups involve live suspects, viewed from head-to-toe, whereas lineups are often conducted through head-and-shoulders photographs, a factor that may favor showup procedures. However, witnesses who participate in showup procedures may experience more stress during the identification, and may view the suspect in the suggestive context of police custody, in poor lighting, from a distance, or from the back of a police car. These factors, which operate in real investigations, may decrease the accuracy of showup procedures.

Although the extant experimental literature would seem to support a policy to avoid the use of showups, the apples-to-oranges differences inherent in showup and lineup procedures make it difficult to draw conclusions about their relative diagnostic accuracy as they are conducted in real-world criminal investigations. Expert testimony should follow this point as well. Given the totality of the circumstances it may not necessarily be the case that an identification from a showup is less reliable than an identification from a lineup.

Lineup Composition

The U.S. Justice Department guidelines on the composition of lineups instruct that the suspect should not "unduly stand out" and that fillers should "generally fit the witness' description." The rule that the suspect should not stand out is supported by meta-analytic reviews of the filler-similarity literature (Clark, 2012; Fitzgerald, Price, Oriet, & Charman, 2013), which suggest that higher-similarity lineups produce higher accuracy compared to lower-similarity lineups. However, there is some evidence to suggest that the relationship between similarity and accuracy is non-monotonic, such that accuracy increases with similarity, but then decreases if similarity is "too high" (Fitzgerald et al., 2013). Further research to determine an optimal level of filler-to-suspect similarity is warranted. The general guideline that the suspect should not stand out may be a useful rule for police practice to the extent that the guideline implies that fillers should be similar in appearance to the suspect.

However, the more specific recommendation that fillers should fit the description provided by the witness, rather than the appearance of the suspect, is not clearly supported by data, but may be based largely on the outlier results of one study (Wells, Rydell, & Seelau, 1993), which showed a strong accuracy advantage of description-matched fillers (chosen solely based on features described by the witness) compared to suspect-matched fillers (chosen based on a visual comparison of the

fillers to the suspect). Their results showed a no-cost pattern with much higher correct identification rates for lineups with description-matched fillers, with no difference at all in the false identification rates. However, this pattern of results is not replicated in the broader literature (Clark, Moreland, & Gronlund, 2014). Analyses that include all the relevant studies show the ubiquitous trade-off, but in the opposite direction: description-matched lineups produce higher correct identification rates, but also higher false identification rates. Importantly, this error trade-off produces a reduction in overall accuracy: d' is lower for description-matched lineups than for suspect-matched lineups.

Instructions

It is common practice—whether it is legislated or not—for police departments to instruct witnesses that the perpetrator may not be present (Police Executive Research Forum, 2013). In the foundational research, the comparison condition states that the perpetrator *is* present. Thus, the comparison is not between the "not present" instruction versus no instruction, but rather between the "not present" instruction and an affirmative "is present" instruction. The "not present" instruction appears to decrease both correct and false identification rates relative to an affirmative "is present" instruction, with no change in overall accuracy as measured by d' (Clark, 2012). This pattern of results is consistent with a criterion shift—the "not present" condition induces people to raise their standard for making a positive identification, resulting in a tradeoff of fewer correct identifications, but also fewer false identifications compared to the "is present" condition. These results speak to the issue of acceptable trade-offs rather than discriminability or diagnostic accuracy. Importantly, these experiments suggest a simple means of shifting a witness's decision criterion and hence the kinds of errors that are obtained through simple instructions.

Sequential Presentation

Sequential lineups present lineup members one-at-a-time, and require a yes (that's him) or no (that's not him) response as each lineup member is presented. This contrasts with simultaneous lineups, which present all lineup members at the same time, allowing witnesses to see all lineup members prior to making any response. The sequential procedure was designed to prevent comparisons between lineup members, which were thought to increase false identification errors. Beginning in 1985, early studies showed a large accuracy advantage for the sequential procedure, but later studies up until 2012 showed that the sequential advantage dwindled and disappeared at about the same time that the state of New Jersey began using the sequential procedure (Clark et al., 2014). The simultaneous – sequential debate is far from settled; however, most studies that have been conducted since 2012 have shown that sequential lineups may be less accurate than simultaneous lineups (Carlson & Carlson, 2014; Dobolyi & Dodson, 2013; Gronlund et al., 2013; Mickes, Flowe, & Wixted, 2012).

Blind Lineup Administration

Researchers have recommended that lineups be administered by a person who is blind to the position of the suspect in the lineup. The small number of studies comparing blind and non-blind lineup administration suggests that blind lineup administration does not consistently increase accuracy, and may actually decrease the overall accuracy of suspect identifications. Clark, Brower, Rosenthal, Hicks, and Moreland (2013) have suggested that, to the extent that non-blind administrators may influence witness's decisions, it may be easier to steer a witness toward a correct identification of the guilty than a false identification of the innocent.

This creates an awkward policy decision. If Clark et al.'s (2013) view is correct, and if the overriding goal is to increase the diagnostic accuracy of a suspect identification, this goal would be better achieved with non-blind lineup administrators. To the extent that blind lineup administration also reduces both correct and false identifications, as suggested by Clark's (2012) meta-analysis on lineup administrator influence, a societal preference for the more conservative procedure may suggest that lineup administrators use blind lineup procedures. However, the current state of the literature does not clearly show that blind procedures are more accurate than non-blind procedures.

The problem with non-blind lineup administration may not lie in the accuracy of the evidence, but rather in the inaccuracy of the assumptions of the jury about the accuracy of the evidence. An example illustrates the problem: a person becomes a suspect in a murder investigation because the victim's property is found in the suspect's car. Based on this evidence, the non-blind police officer administering the lineup is so convinced of the suspect's guilt that he intervenes when a witness hesitates to make an identification. Simple, innocuous sounding comments that do not mention any specific lineup member can increase suspect identification rates. However, the jury may either not know about these interventions or not appreciate their role in producing the identification. As a result the jury may believe that the two pieces of evidence—the victim's property and the witness's identification—are independent, when in fact they are not independent. The identification evidence was not obtained *independently* of the property evidence, but rather—to a certain extent—*because* of the property evidence.

Assessment of Confidence

The confidence expressed by a witness is considered by the trial court in determining whether the identification is sufficiently reliable that the evidence should be admitted at trial (*Manson v. Brathwaite*, 1977) and by juries in determining the weight to be given to the evidence (from *U.S. v.Telfaire*, 1972). However, eyewitness researchers have long argued that the confidence of the witness is such a weak predictor of accuracy that it should be given little consideration by the trial judge in deciding the admissibility of the evidence and little consideration by the jury (Penrod & Cutler, 1995; Wells & Quinlivan, 2009). This view has led some state courts such as Utah and Georgia to remove confidence from jury instructions. New Jersey has gone

a step further by explicitly instructing jurors that confidence is an unreliable predictor of accuracy.

This broadly negative assessment of confidence is incorrect. The association between confidence and accuracy, measured by the point-biserial correlation or by calibration curves, is actually quite strong. The confidence of the witness, expressed at the time of the identification, is a better predictor of accuracy than any information about the similarity of the fillers, the instructions given to the witness, or whether the lineup was sequential or simultaneous, blind or non-blind. As effect sizes go, the average $r = .41$ (Sporer, Penrod, Read, & Cutler, 1995) is strong. Other measures of the confidence-accuracy relationship—specifically calibration curves—show even clearer evidence that eyewitness confidence is a very strong predictor of eyewitness accuracy. This has led some researchers to call for a policy reversal—reversing *back* to the U.S. Supreme Court's view that the confidence expressed at the time of the identification is a useful indicator of accuracy. To be clear, it is the initial expression of confidence that is strongly associated with accuracy, not the confidence expressed months or years later at trial (Wixted, Mickes, Clark, Gronlund, & Roediger, 2015; National Research Council, 2014). Such expressions of confidence at trial may be distorted by feedback and post-event information.

This strong relationship between confidence and accuracy has two important implications for the present discussion. First, in order to benefit from this relationship, police officers must obtain and document statements of confidence at the time the identification is made. Asking the witness two weeks later in a follow-up interview is likely to be of little diagnostic value. Second, these confidence judgments provide another mechanism for adjusting the criterion on the ROC curve. Just as witnesses may set a criterion for deciding whether to make an identification, the trial court can set a criterion based on witness confidence for deciding whether to admit the evidence at trial, and jurors can use initial expressions of confidence to determine how much weight to give to the identification evidence.

This review suggests an ill-defined path between data and policy. Depending on how one assesses the reforms, one could argue to implement them all, implement some, or implement none:

- Research comparing showups and lineups suggests an advantage for lineup procedures, but the critical conditions, as they are likely to occur in actual criminal investigations have not been fully explored.
- There is no clear guiding principle for finding the sweet spot in foil similarity that will maximize accuracy.
- Selection of fillers based on their match to a verbal description, rather than their similarity to the suspect, may increase false identification rates and reduce overall accuracy.
- Cautionary instructions do not increase the diagnostic accuracy of a suspect identification, although they may shift the witness's decision criterion and hence the trade-off of errors.
- The results comparing simultaneous and sequential lineups are inconsistent, and the most recent research and meta-analytic reviews suggest that sequential lineups produce less reliable evidence than simultaneous lineups.
- The results of experiments comparing blind and non-blind lineup administration are also inconsistent, as some results suggest that non-blind lineup administration may improve accuracy, which seems at odds with social values.
- Confidence at the time of the initial identification, long argued to be such a poor predictor of accuracy as to be of little use, now appears to be a very strong predictor of eyewitness identification accuracy, although with some likely exceptions.

Despite this somewhat murky assessment of the eyewitness identification research literature, there are clear recommendations expressed as if–then contingencies: if the goal is to reduce the risk of false identification errors, without regard to the effects on the correct identification rate, most of the reforms seem appropriate—with the exception of the preference for description-matched filler selection over suspect-matched filler selection. However, if the goal of eyewitness identification reform is to increase the accuracy of the evidence and the accuracy of the decisions based on that evidence, some of the reforms should be reconsidered, and policies should be revised or reversed.

We should note at the outset that revisions in best practices are not uncommon, and to the extent that eyewitness researchers "got it wrong," the error is not unique to our discipline. Recent meta-analyses and literature reviews, across a wide range of disciplines, have shown changes in empirical results over time. In the literature on standards of medical care, Prasad, Cifu, and Ioannidis (2012) have noted many reversals in which established medical procedures were challenged by data. In the remainder of this article we will discuss the issue of policy reversals and how eyewitness researchers can both prevent and prepare for them.

Preventing Policy Reversals

To minimize the chances of a policy reversal, the research needs to get it right the first time. It is an empty truism to say that we should only recommend new procedures when the evidence is strong. It begs the question: How strong is strong enough? Certainly, one or two studies is not enough, so should the number of studies be three? Eight? Twenty? There is no metric that will assure error-free policy-making. Certainly, the likelihood of a policy error will decrease as the evidentiary standard for implementing policy is raised. But we must not fall into the trap of cautious inaction that forever awaits "one more study." There is a cost of doing nothing.

Yet, it is clear that advocacy often got ahead of the data. Researchers began "selling" the sequential lineup as early as the late 1980s when there were only a few published studies that had compared simultaneous and sequential lineups (Lindsay, 1999). Warnings about the inherent bias of showup procedures, traced back to the turn of the 19th century, were widely accepted decades before empirical studies were conducted. Recommendations for blind lineup administration are routine, and although such recommendations are based on hundreds of social psychological experiments, the recommendation is informed by very few eyewitness identification experiments. Early research led to

conclusions that confidence was so poorly connected to accuracy as to be of little value to trial courts or jurors. We know now that this view is incorrect.

In many cases, strong conclusions preceded strong data. These strong conclusions were so well-accepted that researchers may have seen an illusory accumulation of consistent results that did not in fact exist. The research community may have contributed to that illusion in several ways. As we (Clark et al., 2014) have previously noted, the ostensibly well-established conclusions may have been both a cause and an effect of misinterpreted data, the choice of the C/F ratio as a measure of accuracy, and a selection bias in publication. Strong conclusions, made before their time, can distort the research enterprise and be resistant to change. To prevent policy reversals researchers must resist the "pull of the policy audience" (Sarat & Silbey, 1988, p. 97) that operates within an adversarial legal system.

Preparing for Policy Reversals

Policy reversals will occur. A procedure that is widely viewed as a best practice today may not be viewed as a best practice five years from now. A healthy science invites debate, controversy, and revision. Woody Allen famously noted that romantic relationships are like sharks. They have to keep swimming forward, or they die. The same is true for eyewitness science. Criminal justice policy must be nimble enough to keep pace with a fast, forward-moving science. Toward that goal, we raise a question (and encourage discussion) about whether eyewitness identification policies should be implemented informally through voluntary compliance with recommended best practices, or by the hammer of legislative mandate. Lindsay (1999) and Clark (2015) have both argued for a best practices approach rather than legislation, on the view that legislation may be slow to change and hard to modify. However, it is not clear that a policy mandated by legislation is any more (or any less) resistant to modification than a policy adopted voluntarily by a police department. To the best of our knowledge, none of the jurisdictions – at the state, county, or city level – has reversed or modified its decision to implement sequential lineup presentation. Indeed, there may be both advantages and disadvantages to the legislative approach, depending on how the legislation is written.

One of the advantages of legislation is that it requires public debate and a transparent public record. The potential for revision or reversal can be addressed very simply through two important mechanisms: first, all reform legislation must include provisions (and funding) for evaluation. Otherwise, how will the legislature, legal community, and citizens of the state know the impact of the legislation? Second, all reform legislation must include sunset provisions that require policy-makers to evaluate the reforms and provide or withhold reauthorization in light of that evaluation.

One objection to this approach is that sunset clauses are unnecessary because a governing body can always repeal or modify any law. A complementary objection is that it *should* be difficult to undo hard-won reforms, lest they be too easily undone when the political balance changes. The response to these objections lies in the requirement for independent, scientific evaluations.

Conclusion

The eyewitness identification research literature is in the midst of an important revision. This provides an opportunity to improve eyewitness identification research practices. This revision also provides an opportunity for academics, policy makers, and law enforcement to discuss and reconsider the mechanism by which eyewitness identification research becomes policy. An open dialog between researchers and participants in the criminal justice system, coupled with stronger research may advance the aims of better recommendations for reforms, fewer scientific reversals, and improved eyewitness accuracy.

Author Contributions

Both authors contributed equally to all aspects of manuscript preparation.

References

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. http://dx.doi.org/10.1016/j.jarmac.2014.03.004

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

Clark, S. E. (2015). Eyewitness identification: California reform redux. *Policy Matters*, *7*, 1–12.

Clark, S. E., Brower, G. L., Rosenthal, R., Hicks, J. M., & Moreland, M. B. (2013). Lineup administrator influences on eyewitness identification and eyewitness confidence. *Journal of Applied Research in Memory and Cognition*, *2*, 158–165.

Clark, S. E., Moreland, M. B., & Gronlund, S. D. (2014). Evolution of the empirical and theoretical foundations of eyewitness identification reform. *Psychonomic Bulletin and Review*, *21*, 251–267.

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

Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. D. (2013). The effect of suspect-filler similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law*, *19*, 151–164.

Garrett, B. (2012). Eyewitnesses and exclusion. *Vanderbilt Law Review*, *65*, 451–506.

Gronlund, S. D., Carlson, C. A., Neuschatz, J. S., Goodsell, C. A., Wetmore, S. A., Wooten, A., & Graham, M. (2013). Showups versus lineups: An evaluation using ROC analysis. *Journal of Applied Research in Memory and Cognition*, *1*, 221–228. http://dx.doi.org/10.1016/j.jarmac.2012.09.003

Gross, H. (1911). *Criminal psychology: A manual for judges, practitioners, and students*. Boston, MA: Little, Brown.

Gross, S. R., & Shaffer, M. (2012). Exonerations in the United States, 1989–2012. *University of Michigan Law School*.

Laudan, L. (2006). *Truth error, and criminal law*. New York: Cambridge University Press.

Lindsay, R. C. L. (1999). Applying applied research: Selling the sequential line-up. *Applied Cognitive Psychology*, *13*, 219–225.

Manson v. Brathwaite (1977). 32 U.S. 98.

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

Mickes, L., Moreland, M. B., Clark, S. E., & Wixted, J. T. (2014). Missing the information needed to perform ROC analysis? Then compute d', not the diagnosticity ratio. *Journal of Applied Research in Memory and Cognition*, *3*, 58–62. http://dx.doi.org/10.1016/j.jarmac.2014.04.007

Munsterberg, H. (1908). *On the witness stand*. New York, NY: McClure.

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

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.

Penrod, S., & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law*, *1*, 817–845.

Peterson, W., Birdsall, T., & Fox, W. (1954). The theory of signal detectability. *Transactions of the IRE Professional Group on Information Theory*, *4*, 171–212.

Police Executive Research Forum (2013). *A national survey of eyewitness identification procedures in law enforcement agencies*.

Prasad, V., Cifu, A., & Ioannidis, J. P. A. (2012). Reversals of established medical practices: Evidence to abandon ship. *Journal of the American Medical Association*, *307*, 37–38.

Sarat, A., & Silbey, S. (1988). The pull of the policy audience. *Law & Policy*, *10*, 97–166.

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. *Psychonomic Bulletin*, *118*, 315–327. http://dx.doi.org/10.1037/0033-2909.118.3.315

U.S. v. Telfaire (1972). 469 F.2d 522.

Volokh, A. (1997). *n* guilty men. *University of Pennsylvania Law Review*, *146*, 173–211.

U.S. Department of Justice. (1999). *Eyewitness evidence: A guide for law enforcement*. Washington, DC: National Institute of Justice.

Wall, P. M. (1965). *Eyewitness identification in criminal cases*. Springfield, IL: Thomas.

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin*, *88*, 776–784.

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.

Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). On the selection of distractors for eyewitness lineups. *Journal of Applied Psychology*, *78*, 835–844.

Wells, G. L., Smalarz, L., & Smith, A. M. (2015). ROC analysis of lineups does not measure underlying discriminability and has limited value. *Journal of Applied Research in Memory and Cognition*, *4*, 313–317.

Wetmore, S. A., Neuschatz, J. S., Gronlund, S. D., Wooten, A., Goodsell, C. A., & Carlson, C. A. (2015). Effect of retention interval on showup and lineup performance. *Journal of Applied Research in Memory and Cognition*, *4*, 1–7.

Wixted, J. T., & Mickes, L. (2015a). Evaluating eyewitness identification procedures: ROC analysis and its misconceptions. *Journal of Applied Research in Memory and Cognition*, *4*, 318–323.

Wixted, J. T., & Mickes, L. (2015b). ROC analysis measures objective discriminability for any eyewitness identification procedure. *Journal of Applied Research in Memory and Cognition*, *4*, 329–334.

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.

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.

Received 9 May 2016;
received in revised form 7 July 2016;
accepted 7 July 2016