Eyewitness Identification and the Accuracy of the Criminal Justice System

Policy Insights from the Behavioral and Brain Sciences
2015, Vol. 2(1) 175–186
© The Author(s) 2015
DOI: 10.1177/2372732215602267
bbs.sagepub.com
SAGE

Steven E. Clark[1], Aaron S. Benjamin[2], John T. Wixted[3], Laura Mickes[4], and Scott D. Gronlund[5]

Abstract

This article addresses the problem of eyewitness identification errors that can lead to false convictions of the innocent and false acquittals of the guilty. At the heart of our analysis based on signal detection theory is the separation of diagnostic accuracy—the ability to discriminate between those who are guilty versus those who are innocent—from the consideration of the relative costs associated with different kinds of errors. Application of this theory suggests that current recommendations for reforms have conflated diagnostic accuracy with the evaluation of costs in such a way as to reduce the accuracy of identification evidence and the accuracy of adjudicative outcomes. Our framework points to a revision in recommended procedures and a framework for policy analysis.

Keywords

eyewitness identification, eyewitness memory, criminal justice, public policy, theories of memory, legal decision making

Tweet

Psychological science can help to reduce eyewitness and criminal justice errors by distinguishing between accuracy and response biases.

Key Points

- Eyewitnesses sometimes make identification errors that can have profound consequences. False identifications of the innocent can lead to false convictions, and false non-identifications of the guilty can allow criminals to commit additional crimes.
- Research has led to several proposals for reforms, but many of the proposed reforms result in a trade-off of errors as correct and false identification rates covary.
- A theoretical framework that separates accuracy from response bias is critical to understanding the error trade-off and producing true overall reductions in error.
- Eyewitness confidence can be helpful in reducing errors as eyewitness evidence flows downstream in the criminal justice process.

Introduction

Eyewitnesses make mistakes. They sometimes fail to identify the guilty, and they sometimes falsely identify the innocent. Social scientists have studied the problem of eyewitness error for more than 100 years (Arnold, 1906; Munsterberg, 1908). In recent years, this research effort has coalesced into a number of concrete recommendations for reform of the criminal justice system, from the procedures that police officers use to obtain identification evidence to the laws that regulate how identification evidence is considered and evaluated within the legal system. These reforms have arisen from social scientists (Wells, Memon, & Penrod, 2006; Wells et al., 1998), legal scholars (Wall, 1965), and scientific and legal organizations, including the American Bar Association (2004), the Innocence Project, the U.S. Department of Justice (1999), and most recently the National Academy of Sciences (National Research Council, 2014).

This article addresses a fundamental question about how these reforms—and future reforms—should be evaluated. Criminal justice policy is shaped by the careful consideration of social science, due process, procedural justice, and social values. Importantly, reasonable people can draw different conclusions and disagree about social values. However, there are two values about which there is little disagreement: (a)

[1]University of California, Riverside, USA
[2]University of Illinois, Champaign, USA
[3]University of California, San Diego, La Jolla, USA
[4]Royal Holloway, University of London, UK
[5]University of Oklahoma, Norman, USA

Corresponding Author:
Steven E. Clark, Psychology Department, University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA.
Email: clark@ucr.edu

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/49353

the justice system must strive for accuracy and the reduction of error and (b) some errors are worse than others.

The fundamental importance of accuracy was succinctly noted by the U.S. Supreme Court in *Tehan v. U.S.* (1966), in stating that "the basic purpose of a trial is the determination of the truth." As Laudan (2006) has persuasively argued, the legitimacy of the criminal justice system requires that the government gets it right—by minimizing error, which means convicting the guilty and acquitting the innocent. There is also little disagreement that a false conviction of the innocent is a more costly and more troubling error than a false acquittal of the guilty. Although some aspects of this cost asymmetry are difficult to specify, one point is clear: A false conviction implies two errors—the false conviction itself, of course, and also a hidden and unadjudicated false acquittal. The conviction of an innocent person often allows the actual perpetrator to remain free.

Consistent with this cost asymmetry, eyewitness identification research and reform has focused almost exclusively on one kind of error—the false identification of the innocent—while giving much less consideration to the other error—the false non-identification of the guilty. This asymmetric focus has led to confusion about the goals of eyewitness identification reform and, we argue, has misled researchers and policy makers to recommend and adopt new procedures that are no more accurate, or even less accurate, than the procedures they replaced. As we will show, police and legal procedures can and should be evaluated with respect to accuracy independently of how one chooses to weigh the relative costs of correct and false identification errors. These relative costs are determined in part by how eyewitness identification evidence carries forward in the criminal justice process. On this point, it is important to note that eyewitness identification outcomes are not legal outcomes, rather they provide evidence that contributes to legal outcomes. Therefore, this article will consider not only the accuracy of eyewitness identification evidence but also the accuracy of adjudicative outcomes that rely on eyewitness evidence.

This article is organized as follows: First, we describe eyewitness identification procedures as they are conducted in criminal cases and simulated in mock-crime experiments. Second, we outline a theoretical framework that describes the memory and decision processes that witnesses utilize in making eyewitness identification decisions. Third, we examine the accuracy of recommended and non-recommended eyewitness identification procedures. Fourth, we consider the accuracy of legal outcomes based on eyewitness evidence. We end with recommendations about how to go forward.

Eyewitness Identification and Experimental Simulation

We focus on two kinds of procedures used by the police to obtain identification evidence: showups and lineups. A showup is typically conducted when the police have identified a suspect shortly after the crime. For example, the police may observe a person who matches the victim's description of the perpetrator of an armed robbery in close proximity to the crime, only minutes after the crime occurred. This single suspect is typically shown to the witness who is asked to make a judgment as to whether the suspect is or is not the perpetrator of the crime. A typical lineup also presents the witness with a single suspect, plus some number of *fillers*, that is, people who are known to be innocent. A lineup is similar to a multiple-choice exam question, except that the correct answer is only *suspected* to be correct but is not known to be correct. Thus, a positive identification of the suspect may be a *correct identification* of a person who is guilty but might instead be a *false identification* of a person who is innocent. Because of this ground-truth uncertainty in real criminal cases, researchers have relied primarily on experimental simulations of crimes and criminal investigations. In these simulations, the crime is staged and presented live or on video. The "perpetrator" of the crime is an actor and confederate of the experimenter, and thus the accuracy of any suspect identification is known. In experimental simulations, as in real cases, the witness may identify the suspect, identify a filler, or make no identification. Our focus here is on suspect identifications because these are the responses that carry forward in actual criminal cases as direct evidence of the suspect's guilt.

Theoretical Framework

Our theoretical framework borrows heavily from signal detection theory, which was developed in the 1940s and 1950s and applied to research in psychophysics (Green & Swets, 1966) and recognition memory (Egan, 1958; Wixted, 2007), as well as many other decision tasks across a wide array of disciplines, including diagnostic medicine (Lusted, 1971), violence risk analysis (Mossman, 1994; Rice & Harris, 1995), crime investigation (Bennell, Jones, & Melnyk, 2009), and law (Bell, 1987; DeKay, 1996; Laudan, 2006; Lillquist, 2002).

One of the useful properties of signal detection theory is that it clearly separates the diagnostic accuracy of decision making from the consideration of the relative costs of different decision errors. We will illustrate this distinction for the identification decisions of eyewitnesses and also more broadly for the decisions made by the criminal justice system.

Eyewitness Identification

The application of signal detection theory to eyewitness identification (see Clark, 2003; Wixted & Mickes, 2014) is a straightforward extension of its application to human memory (Clark & Gronlund, 1996; Egan, 1958; Wixted, 2007). It is particularly straightforward for showups. The witness compares the suspect to his or her memory of the perpetrator to decide whether to identify that suspect as the perpetrator.

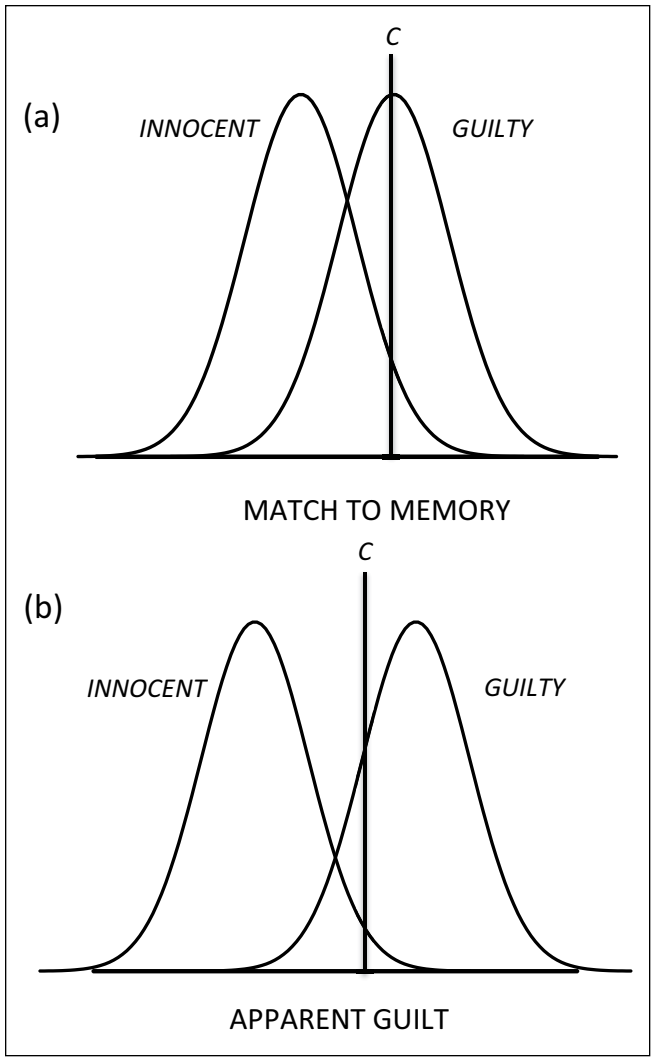


Figure 1. (a) Distributions of match to memory for eyewitness identification of guilty and innocent suspects and (b) distribution of apparent guilt for adjudication of innocent and guilty defendants.
Note. The decision criterion for identification and guilty verdicts is represented by the vertical line labeled *c*. Note that the overlap of the distributions is less in (b) than (a) reflecting greater diagnosticity accuracy for adjudicative outcomes than for eyewitness identification outcomes.

It is important to note that both guilty and innocent suspects will vary in terms of how well they match the witness's memory of the perpetrator. This variation is illustrated in Figure 1a.[1] Some guilty suspects will match memory quite well, but others will match less well, due in part to failures to store or retrieve the relevant information. On the whole, we should expect that innocent suspects will match the witness's memory less well than guilty suspects do, but some innocent suspects will be relatively good matches. The theory proposes that witnesses apply a simple rule to make an identification decision: If the match between the suspect and the witness's memory of the perpetrator is relatively high—above some criterion *c*—the witness will identify that suspect as the perpetrator, and if the match is below that criterion the witness will not identify that suspect as the perpetrator. False identification errors will occur for innocent suspects whose match is above that decision criterion, and false non-identification errors will occur for guilty suspects whose match is below that decision criterion.

Diagnostic accuracy—the ability to discriminate between suspects who are guilty versus suspects who are innocent—is determined solely by the overlap of the guilty-suspect and innocent-suspect distributions. The only way to increase accuracy is to develop procedures that decrease the overlap. The relative cost of the two kinds of errors, false identifications and false non-identifications, is addressed by the placement of the decision criterion *c* (Peterson, Birdsall, & Fox, 1954). It is important to emphasize that shifting the decision criterion has no effect on diagnostic accuracy but only introduces a trade-off between the kinds of errors that are made. Raising the decision criterion will not only reduce the false identification rate but will also reduce the correct identification rate (thus, increasing the false non-identification rate). Lowering the decision criterion will introduce a similar trade-off by increasing the correct identification rate while also increasing the false identification rate. Peterson et al. (1954) had shown that the optimal placement of the decision criterion is determined by evaluating the relative costs associated with this trade-off.

These same principles hold for a lineup. Each lineup member is compared with memory, and an identification is made of the person who is the best match to memory if that best match is above a criterion. Diagnostic accuracy, the ability to discriminate between guilty and innocent suspects, still depends on the overlap of the distributions; shifting the decision criterion only affects the trade-off between false identifications avoided and correct identifications lost.

As it is described so far, the theory provides a framework for understanding the outcomes of eyewitness identification procedures. Again, it is important to note that eyewitness identification outcomes do not represent or specify legal outcomes. Rather, they provide evidence that carries forward in the criminal justice process and contributes to legal outcomes. Eyewitness evidence can provide law enforcement with grounds for arrest, or with a basis for seeking search warrants. Eyewitness evidence contributes to the plea negotiations between prosecution and defense, and the jury verdicts for cases that go to trial. Thus, it is important to not only consider eyewitness identification outcomes but also consider how those outcomes flow downstream to contribute to the accuracy of the criminal justice outcomes which we discuss next.

Criminal Justice Outcomes

Legal scholars have also adapted signal detection theory as a framework for understanding criminal justice outcomes (Bell, 1987; DeKay, 1996; Laudan, 2006; Lillquist, 2002). Guilty

and innocent defendants will vary in terms of the evidence against them or their "apparent guilt" (Laudan, 2006), as shown in Figure 1b. The defendant will be found guilty if the jury finds that the evidence is above a criterion *c*, or standard of proof (beyond a reasonable doubt). Some innocent defendants will incorrectly be found guilty if there is compelling evidence linking them to the crime, and some guilty defendants will incorrectly be found not guilty if there is little evidence that links them to the crime. Again, diagnostic accuracy—in this case the criminal justice system's ability to discriminate between defendants who are guilty and defendants who are innocent—is based on the overlap of the guilty defendant and innocent defendant distributions. Adjusting the decision criterion, or, in legal terms, adjusting the standard of proof, does not affect the ability to discriminate between guilt and innocence but again only affects the trade-off of the two error types. Raising the standard of proof will not only reduce the number of false convictions but will also reduce the number of true convictions.

One of the key questions concerns the extent to which the criminal justice system can detect and recover from eyewitness identification errors. Such error correction allows the guilty to be convicted even if an eyewitness fails to identify that person and allows the innocent to be released from suspicion even if an eyewitness falsely identifies that person. To the extent that the criminal justice system can correct its mistakes, there should be less overlap in the evidence distributions for guilty and innocent defendants (and less error) as the case proceeds through the criminal justice system than there is in the memory match distributions that underlie eyewitness identification.

Accuracy of Eyewitness Identification Procedures

It is almost cliché to aspire to the goal of increasing the accuracy of eyewitness identification. Who could disagree with such a goal? However, the controversy arises in how one defines accuracy, and policy decisions based on accuracy will depend on how accuracy is defined.

Much of the research, and the main thrust of the reform movement, has focused specifically on false identifications of the innocent. This focus is clearly seen in the published literature in which many papers illustrate the problem of eyewitness error through case studies of false identifications that led to false convictions but rarely mention cases involving false non-identifications.

If the reduction, or even the complete elimination of false identifications, was the only goal, there would be little need for research. One could eliminate *all* false identifications by simply banning all eyewitness identification procedures. Of course, this would eliminate all correct identifications as well. Instead, the approach taken by eyewitness identification researchers has been to develop procedures that reduce the false identification rate as much as possible while affecting the correct identification rate as little as possible.

For decades, there was widespread agreement that this strategy had in fact succeeded and had led to a number of reforms that had been shown to reduce the false identification rate with little or no loss of correct identifications (Wells et al., 2006; Wells et al., 1998). If this *no-cost* claim were true, then there would be little doubt that the proposed reforms increase overall accuracy. Policy decisions would be trivially easy because the recommended procedures would produce clear benefits (reduction of false identifications) with no costs (no loss of correct identifications).

However, the data from eyewitness memory experiments unambiguously refute this "no-cost" claim. With few exceptions, correct and false identification rates covary, as predicted by signal detection theory; procedures that reduce the false identification rate also reduce the correct identification rate, and procedures that increase the correct identification rate also increase the false identification rate (Clark, 2005, 2012; Clark, Moreland, & Gronlund, 2014; Palmer & Brewer, 2012).

When correct identifications of suspects who are guilty and false identifications of suspects who are innocent both decrease, it is no longer obvious whether accuracy has changed (and, if it has, whether it has increased or decreased). Clearly, a measure of eyewitness identification accuracy must consider both outcomes. The question is *how* these two responses should be considered together in an overall measure of accuracy.

On this point, there has been much recent disagreement. For decades, eyewitness researchers have measured accuracy as a *diagnosticity ratio* of correct and false identification rates (Wells & Lindsay, 1980). The calculation and its interpretation are straightforward. For example, if the correct identification rate is .6 and the false identification rate is .1, the ratio is 6, meaning that a correct identification is 6 times more likely than a false identification (assuming equal numbers of guilty-suspect and innocent-suspect lineups). Despite its longevity as the accuracy measure of choice, researchers and legal scholars have long been aware of its statistical and interpretive limitations (Kaye, 1986). These problems have been most clearly articulated in the recent work of Wixted and Mickes (2012); Gronlund, Wixted, and Mickes (2014); and Clark (2012). The problem is that as the correct and false identification rates approach zero as responding becomes more conservative, the ratio approaches infinity. Thus, the ratio confuses and conflates diagnostic accuracy and criterion placement. This can be seen in the fact that the ratio will increase with nothing more than an increase in the bias to make no identification. Adjusting the decision criterion does not affect the overlap of the underlying distributions and no change in true diagnostic accuracy—only an apparent one based on an ill-chosen measure.

The field's longstanding—and in some cases continuing—reliance on the diagnosticity ratio illustrates a fundamental

mixing of social science and social justice. Again, most people—laypeople, researchers, and legal scholars—would agree that a false conviction of the innocent is a worse error than a false acquittal of the guilty. Consistent with that social and legal value, the diagnosticity ratio will tend to give preference to procedures that lower false identification rates—*even if those procedures decrease accuracy*. Thus, the preference for a particular identification procedure may appear to be based on overall accuracy, when in fact it is based on a response bias that conforms to a particular social value.

More recently, researchers have started to apply measures of accuracy based on signal detection theory, including the area under the receiver operating characteristic (ROC) curve and d'. Our discussion of these measures is at a conceptual rather than mathematical level to highlight the usefulness of the approach rather than the statistical details. Most importantly, and in contrast to the diagnosticity ratio, the area under the ROC curve and d' both separate accuracy from the placement of the decision criterion.

ROC analysis provides a measure of accuracy across the full range of possible decision criteria. In that sense, it addresses questions such as the following: What if witnesses responded very conservatively (with a high criterion) such that they only made an identification when the match of the suspect to memory was very close? What if witnesses responded less conservatively (with a lower criterion), to make identifications of suspects who matched memory less closely? ROC analysis allows one to consider all of the possible decision criteria in between as well, producing a *family* of correct and false identification rates, rather than just one correct and one false identification rate per procedure. To the extent that the criminal justice system's primary concern is the determination of the truth, policy makers should implement procedures that produce the highest ROC with the greatest area under the curve.

ROC analysis is the preferred method for measuring diagnostic accuracy, but it requires data that are not available in eyewitness identification research reports published before 2012. The d' statistic, on the contrary, can be calculated for any study that provides correct and false identification rates. It provides a measure of the extent to which the distributions for guilty and innocent suspects do not overlap and has been shown to be a good proxy for the area under the ROC curve (Mickes, Moreland, Clark, & Wixted, 2014). Thus, we discuss ROC analyses when they are available and summarize results in terms of d' when ROC analyses are not available.

Application of Accuracy Standard to Eyewitness Identification Procedures

Police make a number of important decisions relevant to eyewitness identification: Whether to conduct a one-person showup, or a lineup? If conducting a lineup, how should it be put together? What should the witness be told about the procedure? How should the procedure be conducted? We apply the accuracy standard in addressing these questions below.

Lineup or Showup

Showups have been criticized by legal scholars for more than 100 years (Gross, 1911; see also Borchard, 1932), as "the most grossly suggestive identification procedure now or ever used by the police." (Wall, 1965, p. 28). This view was so well accepted that it remained untested until Gonzalez, Ellsworth, and Pembroke published the first study in 1993. The results of a meta-analysis of showup–lineup comparisons (Clark, 2012) actually showed a no-cost pattern. False identification rates for lineups are lower than for showups, and the correct identification rates are about equal, resulting in greater guilty-innocent discriminability for lineups than for showups. Results of more recent studies are also generally consistent with that pattern (Gronlund et al., 2012; Mickes, 2015; Wetmore et al., 2015).

Neither the long history of legal criticism nor the more recent history of empirical evidence has led any state or local jurisdiction to ban showup identification procedures. Showup procedures continue to be used by law enforcement based on arguments of necessity, efficiency, and timeliness. For example, the *California Peace Officers Legal Sourcebook* states that "a showup held shortly after the offense benefits everyone: The witness has the culprit's image fresh in mind, so an innocent suspect gets cut loose immediately . . ." (Calandra & Carey, 2005). However, the empirical results suggest that showups may not benefit *everyone*; false identification rates for innocent suspects are higher for showups than for lineups.

However, the key comparison based on the timeliness argument is between a showup conducted immediately versus a lineup conducted later. Three studies that allow this comparison show mixed results. Yarmey, Yarmey, and Yarmey (1996) showed higher accuracy for the immediate showup ($d' = 1.446$) than for the delayed lineup ($d' = 0.624$); Dekle, Beal, Elliott, and Huneycutt (1996) showed equivalent performance comparing immediate showups ($d' = 1.190$) and delayed lineups ($d' = 1.006$); and Wetmore et al. (2015) had shown higher accuracy for delayed lineups, irrespective of whether the lineup was fair ($d' = 1.705$) or biased ($d' = 1.263$), compared with an immediate showup ($d' = 0.581$).

There is more research to be done on the showup–lineup question. In addition to the timing issue, the experimental research needs to capture important differences between showups and lineups as they are conducted in actual criminal investigations. In many of the experimental comparisons, the lineups and showups are both conducted with photographs. In actual criminal investigations, the showup procedure is conducted live, which should provide more information than is available in a head-and-shoulders photograph (which should increase accuracy), but showups are also sometimes

conducted from a distance or in poor lighting (which should decrease accuracy).

Lineup Composition

A general principle for constructing a lineup is to select fillers in such a way that the suspect does not stand out and is not readily identifiable by non-witnesses. (If non-witnesses can pick out the suspect, the lineup may be biased.) Two procedures have been considered for achieving this goal, both of which are based on the similarity of the fillers to the perpetrator. Because the identity of the perpetrator is unknown (or at least uncertain), the similarity of the fillers to the perpetrator cannot be assessed directly but must be assessed indirectly either through their similarity to the suspect (who may or may not be the perpetrator) or through their match to the witness's description of the perpetrator (which may vary in its accuracy). Researchers have expressed a strong preference for selecting fillers based on their match to the witness's description of the perpetrator, rather than their similarity to the suspect (Luus & Wells, 1991; Wells, Rydell, & Seelau, 1993). The data, however, do not support this preference. Although the first published study (Wells et al., 1993) showed an enormous d' advantage for description-matched lineups, the entire corpus of studies shows a d' advantage for suspect-matched lineups (Clark et al., 2014).

In actual criminal cases, lineups are typically presented with photographs rather than with live human beings. This preference for photo lineups over live lineups is due in part to practical and legal considerations. A suspect in a live lineup has a right to have an attorney present, whereas a suspect in a photo lineup does not. As a practical matter, fillers for photo lineups can be selected from enormous databases that are searchable with sophisticated image processing software, and photo lineups can be carried to a witness anywhere and anytime, whereas fillers for live lineups are constrained by the current jail population, and the procedure must be conducted at the jail.

These legal and practical considerations, of course, do not directly address the issue of accuracy. For many years, researchers have asserted that live lineups are no more accurate than photo lineups (Cutler & Fisher, 1990), but a recent review by Clark, Moreland, and Rush (2015) suggested that diagnostic accuracy may be higher for live lineups than for photo lineups. This makes sense, as live lineups provide witnesses with more information than photo lineups, which would increase diagnostic accuracy by moving the underlying memory distributions apart.

Lineup Instructions

It is generally recommended that witnesses be instructed that the person who committed the crime "may or may not" be present and that witnesses are not required to make an identification. In the research, these "unbiased" instructions are typically compared with "biased" instructions that imply that the perpetrator *is* present and that the witness's task *is* to pick him. The recommended "unbiased" instructions reduce both correct and false identification rates, with no change in accuracy as measured by d' (Clark, 2012). These results suggest that the biased and unbiased instructions conditions lie on the same ROC curve, and that the difference between them is best described in terms of a criterion change rather than a change in accuracy.

Lineup Presentation

Researchers have recommended two reforms to the standard lineup procedure: (a) that the lineup be presented sequentially rather than simultaneously and (b) that the lineup be administered by a person who is uninformed or *blind* to important aspects of the investigation. In the reforms, these two procedures have often been packaged together as the blind-sequential lineup; however, we will describe them separately below.

Sequential Lineup

A sequential lineup requires the witness to respond "yes" (that is him) or "no" (that is not him) as each lineup member is presented, one at a time, precluding comparisons between lineup members. Studies published prior to 2012 showed equivalent levels of accuracy for simultaneous and sequential lineups, using d' as the measure of accuracy (Clark, 2012; Palmer & Brewer, 2012). Studies utilizing ROC analysis, published after 2012, show either no difference or a sequential lineup *disadvantage* (see Gronlund, Mickes, Wixted, & Clark, 2015, for a review). Why would diagnostic accuracy be higher for simultaneous lineups than for sequential lineups? One explanation is that the opportunity to make comparisons among lineup members helps witnesses determine the most diagnostic features, that is, those features that best distinguish between the perpetrator and similar fillers. By focusing on those diagnostic features, the memory match for innocent people will decrease (because the diagnostic features will mismatch memory) and the memory match for guilty people will increase (because the diagnostic features will match memory), thus separating the memory distributions for guilty and innocent suspects (Wixted & Mickes, 2014).

Blind Lineup Administration

Blind lineup administration requires, at a minimum, that the lineup administrator not know which position the suspect is in the lineup, making it impossible to deliberately or inadvertently steer the witness toward the suspect. In some cases, the blind administrator may be completely uninvolved in the case and thus also unaware of any other evidence of the

suspect's guilt. There are very few published studies that have examined accuracy in blind and non-blind lineups, and those studies have made the critical comparison (blind vs. non-blind) across a number of other variables, complicating the interpretation of the results. Here, we present the simple comparison in a standard identification procedure with unbiased instructions. An important caveat is that by focusing on a subset of conditions, the sample sizes become very small.

The first published study comparing blind versus non-blind lineup administration (Phillips, McAuliff, Kovera, & Cutler, 1999) considered simultaneous and sequential lineups for innocent-suspect lineups only. Because they did not include a guilty-suspect lineup condition, we cannot assess diagnostic accuracy via d' or ROC analysis. Nonetheless, the results are important to consider, as they do not show a consistent reduction in false identifications. For sequential lineups, blind administration did reduce the false identification rate relative to non-blind lineup administration, but for simultaneous lineups, blind administration produced a small increase in the false identification rate.

Greathouse and Kovera (2009) showed a similar pattern. For simultaneous lineups, blind administration produced a small increase in the false identification rate, a small decrease in the correct identification rate, and a small decrease in d'. For sequential lineups, blind administration produced decreases in both the correct and false identification rate and again a small decrease in d'. Again, we note that the sample sizes were quite small. The d' differences consistently favor the non-blind procedure, although they are not statistically significant.

Clark, Brower, Rosenthal, Hicks, and Moreland (2013) approached this issue in another way. The lineup administrators were trained to steer witnesses in subtle ways to identify the suspect in a lineup. One example of how lineup administrators would steer witnesses toward the suspect was that they would ask for clarification if the witness mentioned a lineup filler but accept any mention of the suspect without question. This steering condition was compared with a condition in which lineup administrators could not influence witnesses' decisions. Unsurprisingly, when lineup administrators steered witnesses to the suspect, suspect identification rates increased for both guilty and innocent suspects. What was surprising is that the correct and false identification rates increased in such a way as to *increase* the diagnostic accuracy of the eyewitness identification evidence. Witness manipulation did not simply increase suspect identification rates; it increased accuracy. An unpublished study by Alberts (2007) showed the same pattern of results: Witness steering improved eyewitness accuracy. The explanation for the increased accuracy is that it is easier to steer witnesses toward a correct answer (the guilty suspect) than it is to steer them toward an incorrect answer (the innocent suspect). Clark et al. showed that efforts to steer witnesses to the guilty suspect were more likely to be successful, whereas efforts to steer witnesses to the innocent suspect would often result in witnesses identifying a filler instead.

These results create an uneasy conflict between the goal of increased accuracy and perceptions of justice and fairness. The accuracy standard would suggest that police officers should conduct lineups with full knowledge of the position of the suspect so that they can better steer witnesses toward that suspect. This uncomfortable conclusion led Clark (2012) and Wells, Steblay, and Dysart (2012) to pose the same question (in slightly different ways): If the goal is simply to increase accuracy, what is to keep police from simply telling the witness whom to identify ("identify #4 or else")? Why not place a red arrow above the suspect in the lineup? One response to such questions is to appeal to procedural justice rather than accuracy. However, there may be an accuracy-driven solution to the "or else" and "red arrow" lineup procedures, discussed in the next section.

Accuracy in Criminal Justice Outcomes

Eyewitness identification responses are only relevant to the extent that they contribute to the outcome of legal proceedings against the defendant. On this point, it is important to note that an eyewitness identification does not mean that the suspect will be arrested, does not mean that the suspect will be prosecuted, and does not mean that the suspect will be convicted. Gould, Carrano, Leo, and Hail-Jares (2013) compared two sets of real-world cases—those that resulted in false convictions versus those that resulted in "near misses," that is, cases in which the charges were eventually dropped, or in which juries acquitted the defendant. Many of the false convictions—but also many of the near misses—involved false eyewitness identification evidence. Importantly, false identifications did not inevitably lead to false convictions. The impact of eyewitness identification evidence will depend on how it is weighed and evaluated by the police, by prosecuting and defense attorneys who resolve the vast majority of criminal cases through plea agreements, by the judge who may be asked to make a ruling on whether the evidence is admissible at trial, and by the members of the jury who must render a verdict in those rare cases that go to trial.

In this section of our article, we consider how the accuracy of criminal justice outcomes can be increased through the evaluation of eyewitness evidence. We focus specifically on the confidence expressed by the witness, the utility of video recording identification procedures, and the education of the jury through jury instructions and expert testimony.

Confidence

One way in which accuracy may be increased downstream in the criminal justice system is by considering witnesses' own assessments of their accuracy through their expressions of certainty or confidence. The underlying intuition is that a

confident witness is more likely to be correct than a less confident witness.

At the outset, we note that, contrary to this intuition, researchers have for more than 30 years held the opposite view—that trial courts and juries should *not* consider a witness's expressions of confidence when assessing the accuracy of that witness's identification because the relationship between confidence and accuracy is too weak to be helpful (Deffenbacher, 1980; Wells & Murray, 1983; Wells & Quinlivan, 2009). More recent analyses show that this view is incorrect and that confidence is a very strong predictor of accuracy (Juslin, Olsson, & Winman, 1996; Sporer, Penrod, Read, & Cutler, 1995; Wixted, Mickes, Clark, Gronlund, & Roediger, in press). In both laboratory experiments and field studies of actual criminal cases, eyewitnesses who are more confident in their identifications are much more likely to be correct than witnesses who are less confident in their identifications. The recommendation to ignore witness confidence effectively deprives the criminal justice system of one of the most diagnostic cues to eyewitness accuracy. Trial courts and juries *should* consider the confidence of the witness when assessing the accuracy of the witness's identification. With that, we explore two questions: How will assessments of confidence affect the accuracy of adjudicative outcomes, and what are the limitations of confidence in assessing accuracy?

The downstream effects of confidence assessments will depend on who is making the assessment and for what purpose. Specifically, trial judges are often required to make decisions about whether the identification evidence should be admitted at trial, whereas juries are asked to make decisions about the defendant's guilt. These are very different decisions, and an assessment of the witness's confidence can usefully contribute to both, although in different ways.

The admissibility of eyewitness identification evidence follows a procedure outlined by the U.S. Supreme Court in *Manson v. Brathwaite* (1977). Under *Manson*, if the trial court determines that the identification procedure was not suggestive then the evidence will be admissible at trial. If the procedure is determined to have been suggestive, the trial court may determine that the evidence is nonetheless reliable based on an evaluation of five factors, one of which is the witness's confidence. Such rulings are to be based on the reliability of the identification evidence independent of any other corroborative evidence (however, see Koch, 2003, for examples of cases in which courts have erred on this point). In other words, an identification is not reliable because the suspect was found to be in possession of the victim's wallet, for example. Thus, for an admissibility decision the judge's consideration of the witness's confidence may be viewed as being based on the (hopefully) optimal criterion on the ROC curve based on the relative costs of false identification and false non-identification errors.

The jury, however, is asked to consider all the evidence in rendering a verdict. Consequently, the consideration of confidence should have a different effect. In a simple weighting-of-evidence model of juror decision making (Hastie, 1993), a juror would weigh each piece of evidence based on its relevance and reliability. In such a model, a juror should give less weight to a low-confidence identification than a high-confidence identification when making a decision about the suspect's guilt in the same way that an eyewitness should give less weight to a non-diagnostic feature when making a decision about whether the suspect is the perpetrator. Importantly, by giving less weight to the less reliable low-confidence identification, the juror can give more weight to other more reliable evidence. This reduces the overlap in the distributions illustrated in Figure 1b and increases the diagnostic accuracy in discriminating between defendants who are guilty versus defendants who are innocent. This analysis should also hold for plea negotiations—to the extent that such negotiations are conducted in the "shadow of the jury," that is, they are based on an expectation of what the jury would do if the case went to trial (Bushway, Redlich, & Norris, 2014). The accuracy of plea agreements should increase through the assessment of witness confidence by prosecuting and defense attorneys.

We note two important limitations in the use of confidence to assess accuracy. First, the confidence–accuracy relationship is not as strong for negative responses (i.e., identifying no one) as it is for positive responses (i.e., identifying someone; Brewer & Wells, 2006; Sporer et al., 1995). The implication of the differing relationships for positive and negative responses is that lawyers, judges, and juries may be able to reliably distinguish between correct and false identifications based on witness confidence but may be much less able to distinguish between correct and false non-identifications based on witness confidence. Thus, confidence may provide a useful means for reducing the adjudicative accuracy costs associated with false identifications of the innocent but not for reducing the adjudicative accuracy costs associated with false non-identifications of the guilty.

Second, the strong relationship between confidence and accuracy holds for confidence judgments made at the time of the initial identification but may not hold for confidence judgments made long after the identification. There is a large body of research showing that memory—and judgments about memory—can be distorted by events that transpire *after* the remembered event (Loftus, 2005). For example, witnesses may become more confident about their identifications if they later receive feedback indicating that they were correct (Wells & Bradfield, 1998). Thus, a witness who initially is not very confident about his or her identification may become very confident by the time of the trial. For this reason, Wixted et al. (in press) have recommended that jurors consider the confidence of the witness *at the initial identification* but not the confidence expressed at trial.

These more recent assessments of the strong confidence–accuracy relationship suggest that the U.S. Supreme Court, in *Manson v. Brathwaite* (1977), was exactly right when it included witness certainty at the time of the identification as

a factor to be considered in evaluating an identification's reliability and that state courts in Georgia (*Brodes v. State*, 2005) and Utah (*State v. Long*, 1986) have erred by removing witnesses' expressions of confidence as a factor to consider when assessing the reliability of eyewitness evidence. The most serious error in this regard is found in the recently revised New Jersey jury instructions that did not simply remove confidence as a factor to be considered but went further to explicitly—and incorrectly—instruct jurors that "eyewitness confidence is generally an unreliable indicator of accuracy" (New Jersey Model Criminal Jury Charges, 2012).

Videotaping

The goal of increased accuracy in adjudicative outcomes requires an accurate record of the identification procedures. Disagreements between witnesses, police, and attorneys regarding the nature of the procedures or the statements made by witnesses can only add noise to the adjudicative process (increasing the overlap of the distributions shown in Figure 1b).

The National Academy of Sciences recently recommended that video recording of eyewitness identification procedures become a standard practice (National Research Council, 2014). This recommendation was made years before by Kassin (1998) on the argument that the video recording of identifications would reduce both deliberate and inadvertent misrepresentations by police. Simon (2012) has likewise recommended more complete and transparent records of criminal investigations.

Will video recordings allow jurors to more accurately discriminate correct from incorrect identifications? Two studies bear directly on this question. In both studies, one group of participants witnessed a staged crime and made identifications from a lineup and another group of participants, that is, mock jurors, judged the accuracy of those identifications. Reardon and Fisher (2011) found that mock jurors were better able to discriminate between correct and incorrect identifications when they were presented with a video recording of the identification along with the witness's testimony than when they were presented with the witness's testimony alone. In contrast, Beaudry et al. (2013) concluded that the opportunity to view a video recording of the identification did not improve mock juror's ability to distinguish between correct and incorrect identifications. However, there are several problems with this study that make the results difficult to interpret. The one condition free of interpretive problems—using simultaneous lineups, with a blind administrator, and without on-camera feedback to the witness—provided a mixed pattern of results (albeit with a very small number of participants' data that precluded meaningful statistical analyses). Mock jurors were best able to distinguish between correct and incorrect identifications when they were shown the video of the identification *only*, second best when they were shown the witness's testimony *only*, and worst when they were shown both the video recording of the identification and the testimony. These results suggest that the video recordings provided the most useful cues to witnesses' identification accuracy. One tentative explanation for the poor performance of mock jurors who were presented with both the video recording and the witness testimony is that the two sources of information may have provided conflicting cues.

We return now to the problem of lineup administrator influence. There is no evidence that the lineup administrators in these studies engaged in any behaviors that might be viewed as suggestive or coercive. Presumably, none of the experimenters told the witnesses whom to identify, "or else," and none used a red arrow to mark the position of the suspect. What if such suggestive and coercive procedures had been used? Would observers have discounted or dismissed those identifications based on the video evidence? The broader question is about the extent to which a complete and accurate record of the identification procedure is an antidote to the problems associated with such procedures.

On this point, the problem with police officers steering witnesses to identify the suspect may not be about the accuracy of the evidence (which may actually increase) but rather about how juries interpret the evidence in the context of other evidence. For example, a victim's property may be found in the trunk of the suspect's car, prompting the police officer to give the witness an extra nudge to identify that suspect from a lineup. The jury may erroneously believe that these two pieces of evidence are independent when in fact they are not. To the contrary, the identification of the suspect came about in part *because* of the property found in the suspect's car. One promise of video-recorded identification procedures is that they may help jurors consider the non-independence of the evidence.

Educating the Jury

Researchers have argued that laypeople—the people who make up the jury pool—lack sufficient understanding of eyewitness identification to be able to distinguish between correct and false identifications and have recommended that juries be "educated" through jury instructions or through expert testimony. There are two important issues to raise here: First, jurors may not be as incompetent as assumed (Desmarais & Read, 2011). Among other things, jurors have been criticized for putting too much stock in witness confidence as a cue to witness accuracy when in fact their underlying intuition about the confidence–accuracy relationship is largely correct. Jurors should consider witness confidence in assessing accuracy, although it should be the confidence expressed at the time of the identification, rather than the confidence expressed at trial.

Importantly, jury instructions and expert testimony must accurately reflect the advances in the scientific literature. The New Jersey jury instructions are incorrect in declaring that blind lineup administration leads to more reliable identification evidence and that confidence is not a useful cue to identification

accuracy. In addition, it is inaccurate for experts to testify that unbiased instructions, sequential lineups, or blind lineup administration increase the accuracy of eyewitness identification evidence; none of these claims has a solid empirical foundation. It is inaccurate for experts to testify that a witness's initial statement of confidence is not a useful predictor of that witness's accuracy; this claim is clearly contradicted by a solid empirical foundation. Adjudicative accuracy cannot increase by misinforming jurors about the social science.

Recommendations

If the purpose of the criminal justice process is adjudicative accuracy and the determination of the truth, then based on the current research, police should continue to use simultaneous lineups and should consider ways to increase the information available in lineups beyond the static, head-and-shoulders photo lineups that are commonly used. Lineups should be created with fillers that are similar in appearance to the suspect. As eyewitness evidence carries forward, the criminal justice system should carefully consider the useful information in witnesses' initial statements of confidence. Police should obtain a clear statement of confidence from the witness at the time the identification is made and should be cautious about identifications that are made with low confidence. Courts should continue to consider initial expressions of confidence in deciding the admissibility of the evidence but should give little or no weight to expressions of confidence made by witnesses at admissibility hearings. Likewise, pattern jury instructions should instruct juries that the witness's initial statement of confidence, but not subsequent statements of confidence, provide useful information for evaluating the likely accuracy of a witness's identification. States that have removed confidence as a factor to be considered by the jury instructions should consider adding it. The New Jersey instructions that explicitly instruct jurors that confidence is an unreliable indicator of accuracy should be corrected.

The accuracy standard provides no basis upon which to make a recommendation about the instructions provided to witnesses. Although it seems reasonable to tell witnesses that the perpetrator may not be in the lineup and that they are not required to make an identification, because in fact both of these statements are true, there is no clear evidence to suggest that including this instruction will increase accuracy.

Other considerations require additional research. For example, it is explicitly assumed that a showup sooner will provide more accurate identification evidence than a lineup later. As reasonable as this assumption may seem (and at some point it must be true—if "later" means months or years later), there are very few studies that allow the relevant comparison, and the results are not consistent.

Policy

Some states have implemented specific police procedures through legislation. In New Jersey and North Carolina, for example, police officers are required to present lineups sequentially, but the most recent research suggests that sequential lineup presentation decreases the diagnostic accuracy of suspect identification evidence. Thus, the laws in these states mandate that police use a procedure that has been shown to reduce the accuracy of identification evidence. We should expect that best practice recommendations will change as the research literature develops and evolves. Legislation that mandates the use of particular procedures may force police to use procedures that were considered best practices 10 years ago rather than procedures that are considered to be best practices now.

The practice of mandating particular identification procedures as a matter of law should stop. It is perfectly reasonable for legislation to mandate that current best practices should be used. However, it is neither productive nor forward thinking to mandate as a matter of law what those best practices should be. Changing laws can be exceedingly slow and fraught with political maneuvering that has little to do with the merit of the proposed law. Eyewitness research is entering a new phase in which past conclusions are being challenged and new views are evolving. The criminal justice system must be nimble enough that it can incorporate best practices as our understanding of eyewitness evidence and its role in the criminal justice system evolve and advance.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Presley Center for Crime and Justice Studies at the University of California, Riverside and by National Science Foundation grant 1061183 to Steven E. Clark.

Note

1. Although our framework borrows from signal detection theory, it need not be tied specifically to it, and similar theoretical frameworks may also be usefully applied. We note also that for simplicity, Figure 1 shows distributions that are normal with equal variances for guilty and innocent. These properties are not critical, however, and particularly for the application to memory, there is evidence suggesting that the variances are unequal (see Malmberg, 2008; Ratcliff, Sheu, & Gronlund, 1992; Wixted, 2007).

References

Alberts, W. (2007). *Steering in the eyewitness identification procedure* (Unpublished master's thesis). University of Leiden, The Netherlands.

American Bar Association. (2004, August). *American Bar Association statement of best practices for promoting the accuracy of eyewitness identification procedures* (Report No. 111C).

Arnold, G. F. (1906). *Psychology applied to legal evidence and other constructions of law*. Calcutta, India: Thacker, Spink.

Beaudry, J. L., Lindsay, R. C. L., Leach, A., Mansour, J. K., Bertrand, M. I., & Kalmet, N. (2015). The effect of evidence type, identification accuracy, line-up presentation, and line-up administration on observers' perceptions of eyewitnesses. *Legal and Criminological Psychology*, *20*, 343-364.

Bell, R. S. (1987). Decision theory and due process: A critique of the Supreme Court's lawmaking for burdens of proof. *Journal of Criminal Law and Criminology*, *78*, 557-585.

Bennell, C., Jones, N. J., & Melnyk, T. (2009). Addressing problems with traditional crime linking methods using receiver operating characteristic analysis. *Legal and Criminological Psychology*, *14*, 293-310.

Borchard, E. M. (1932). *Convicting the innocent*. New Haven, CT: Yale University Press.

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

Brodes v. State (2005). 279 GA 435.

Bushway, S. D., Redlich, A. D., & Norris, R. J. (2014). An explicit test of plea bargaining in the "shadow of the trial." *Criminology*, *52*, 723-754.

Calandra, D., & Carey, J. E. (2005). *Field guide for the California peace officers legal sourcebook*. Sacramento, CA: California District Attorneys Association.

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

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior*, *29*, 395-424.

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

Clark, S. E., Brower, G. L., Rosenthal, R., Micks, 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., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*, 37-60.

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

Clark, S. E., Moreland, M. B., & Rush, R. A. (2015). Lineup composition and fairness. In T. Valentine & J. P. Davis (Eds.), *Forensic facial identification: Theory and practice of identification from eyewitnesses, composites and CCTV* (pp. 129-158). New York, NY: Wiley.

Cutler, B. L., & Fisher, R. P. (1990). Live lineups, videotaped lineups, and photo arrays. *Forensic Reports*, *3*, 439-448.

Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship? *Law and Human Behavior*, *4*, 243-260.

DeKay, M. L. (1996). The difference between Blackstone-like error ratios and probabilistic standards of proof. *Law and Social Inquiry*, *21*, 95-132.

Dekle, D. J., Beal, C., Elliott, R., & Huneycutt, D. (1996). Children as witnesses: A comparison of lineup versus showup identification methods. *Applied Cognitive Psychology*, *10*, 1-12.

Desmarais, S. L., & Read, J. D. (2011). After 30 years, what do we know about what jurors know? A meta-analytic review of lay knowledge regarding eyewitness factors. *Law and Human Behavior*, *35*, 200-210.

Egan, J. P. (1958). *Recognition memory and the operating characteristic* (Technical Note AFCRC-TN-58-51). Bloomington: Indiana University Hearing and Communication Laboratory.

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

Gould, J. B., Carrano, J., Leo, R. A., & Hail-Jares, K. (2013). Predicting erroneous convictions. *Iowa Law Review*, *99*, 471-522.

Greathouse, S. M., & Kovera, M. B. (2009). Instruction bias and lineup presentation moderate the effects of administrator knowledge on eyewitness identification. *Law and Human Behavior*, *33*, 70-82.

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

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

Gronlund, S. D., Mickes, L., Wixted, J. T., & Clark, S. E. (2015). Conducting and eyewitness lineup: How the research got it wrong. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 63, pp. 1-43). New York: Academic Press.

Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using receiver operating characteristic analysis. *Current Directions in Psychological Science*, *23*, 3-10.

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

Hastie, R. (1993). *Inside the juror: The psychology of juror decision making*. New York: Cambridge University Press.

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *5*, 1304-1316.

Kassin, S. M. (1998). Eyewitness identification procedures: The fifth rule. *Law and Human Behavior*, *22*, 649-653.

Kaye, D. H. (1986). Quantifying probative value. *Boston University Law Review*, *66*, 71-81.

Koch, R. (2003). Process v. outcome: The proper role of corroborative evidence in due process analysis of eyewitness identification testimony. *Cornell Law Review*, *88*, 1097-1141.

Laudan, L. (2006). *Truth, error, and criminal law: An essay in legal epistemology*. New York, NY: Cambridge.

Lillquist, E. (2002). Recasting reasonable doubt: Detection theory and the virtues of variability. *University of California Davis Law Review*, *36*, 85-197.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*, 361-366.

Lusted, L. B. (1971). Signal detectability and medical decision-making. *Science*, *171*, 1217-1219.

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

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*, 335-384.

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

Mickes, L. (2015). Receiver operating characteristic analysis and confidence-accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition*, *4*, 93-102.

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.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, *62*, 783-792.

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.

New Jersey Model Criminal Jury Charges. (2012). Retrieved from http://www.judiciary.state.nj.us/ressrel/2012/jury_instructions.pdf

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.

Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. In *Information theory, transactions of the IRE professional group* (Vol. *4*, pp. 171-212).

Phillips, M. R., McAuliff, B. D., Kovera, M. B., & Cutler, B. L. (1999). Double-blind photoarray administration as a safeguard against investigator bias. *Journal of Applied Psychology*, *84*, 940-951.

Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, *99*, 518-535.

Reardon, M. C., & Fisher, R. P. (2011). Effect of viewing the interview and identification process on juror perceptions of eyewitness accuracy. *Applied Cognitive Psychology*, *25*, 68-77.

Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, *63*, 737-748.

Simon, D. (2012). *In doubt: The psychology of the criminal justice system*. Cambridge, MA: Harvard University Press.

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

State v. Long (1986). 721 P.2d 483.

Tehan v. U.S. (1966) 382 U.S. 406.

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

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

Wells, G. L., & Bradfield, A. L. (1998). "Good, you identified the suspect": Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology*, *83*, 360-376.

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

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, *7*, 45-75.

Wells, G. L., & Murray, D. M. (1983). What can psychology say about the Neil v. Biggers criteria for judging eyewitness accuracy? *Journal of Applied Psychology*, *68*, 347-362.

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). The selection of distractors for eyewitness lineups. *Journal of Applied Psychology*, *5*, 835-844.

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

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2012). Eyewitness identification reforms: Are suggestiveness-induced hits and guesses true hits? *Perspectives on Psychological Science*, *7*, 264-271.

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*, 8-14.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*, 152-176.

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

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

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L. (in press). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist*.

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.