Costs and Benefits of Eyewitness Identification Reform : Psychological Science and Public Policy

Steven E. Clark

Perspectives on Psychological Science 2012 7: 238
DOI: 10.1177/1745691612439584

Perspectives on Psychological Science
7(3) 238–259

DOI: 10.1177/1745691612439584
http://pps.sagepub.com

Costs and Benefits of Eyewitness Identification Reform: Psychological Science and Public Policy

Steven E. Clark
University of California, Riverside

Abstract

Psychological science has come to play an increasingly important role in the legal system by informing the court through expert testimony and by shaping public policy. In recent years, psychological research has driven a movement to reform the procedures that police use to obtain eyewitness identification evidence. This reform movement has been based in part on an argument suggesting that recommended procedures reduce the risk of false identifications with little or no reduction in the rate of correct identifications. A review of the empirical literature, however, challenges this *no-cost view*. With only one exception, changes in eyewitness identification procedures that reduce the risk of false identification of the innocent also reduce the likelihood of correct identification of the guilty. The implication that criminals may escape prosecution as a result of procedures implemented to protect the innocent makes policy decisions far more complicated than they would otherwise be under the no-cost view. These costs (correct identifications lost) and benefits (false identifications avoided) are discussed in terms of probative value and expected utility.

Keywords

eyewitness identification, law enforcement, public policy

The story is well-known. An eyewitness to a crime looks carefully at a photographic lineup and points to the suspect and says, "That's him." The suspect is prosecuted and convicted, and then years later—sometimes decades later—new evidence is discovered that shows, often quite convincingly, that the person is innocent. The link between mistaken identification and wrongful convictions in the United States is empirically well-established. There is no doubt: Eyewitnesses make mistakes that send innocent people to prison (Gross, Jacoby, Matheson, Montgomery, & Patil, 2005).

Four decades of scientific psychological research have shown that the risk of such false identifications could be significantly reduced if the criminal justice system were to change the procedures it uses to obtain eyewitness identification evidence. This process of reform is already under way. The states of New Jersey, Wisconsin, North Carolina, Ohio, and West Virginia have passed legislation, or developed guidelines, to change the procedures used by law enforcement to obtain eyewitness identification evidence. In addition, local jurisdictions including Dallas, TX, Denver, CO, Suffolk County (including Boston, MA), Northampton, MA, and Santa Clara, CA, have also enacted eyewitness identification reforms.

The reforms are directed at fundamental aspects of the identification process: How lineups are constructed, what witnesses are told and how they are instructed prior to the lineup, the way that the lineup is presented, and what police officers should and should not say and do during the identification procedure. Five specific recommendations are addressed in this article: (a) the witness should be instructed prior to the lineup that the perpetrator of the crime might not be present in the lineup; (b) the individuals in a lineup should be presented to the witness sequentially, rather than simultaneously; (c) the police officer or detective who conducts the lineup should not say or do anything that could influence the witness's decision; (d) the lineup should be composed in such a way that the suspect does not stand out; (e) the presentation of a single suspect, alone, in a procedure called a *one-person showup*, should be avoided in favor of full lineups that typically present the suspect along with a minimum of five fillers. These recommendations, which will be discussed in detail, are based on empirical data from laboratory experiments and are closely tied to fundamental issues regarding human memory, decision-making, and social influence.

The reform movement has also been driven by an assertion that the recommended procedures can reduce the false

Corresponding Author:
Steven E. Clark, Department of Psychology, University of California, Riverside, CA 92521
E-mail: clark@ucr.edu

identification rate with little or no reduction in the correct identification rate (e.g., Wells, Memon, & Penrod, 2006; Wells & Seelau, 1995; Wells et al., 1998).[1] This view has not only been expressed by eyewitness research psychologists, but also by legal scholars (Findley, 2008; Garrett, 2008),[2] and it is referred to here as the *no-cost view*.[3] As will be shown, the no-cost view is convincingly contradicted by data. To the contrary, the data show (with one exception) a general trade-off pattern. Correct identifications of the guilty are lost as false identifications of the innocent are avoided.

This trade-off makes policy decisions regarding the implementation of the recommended procedures more complicated. To be clear, it is not the purpose of this article to argue for or against any of the recommended procedures, but rather to light the path between data and public policy. For psychological science to inform public policy, it is essential to understand the no-cost view and its implications, properly assess the relevant data, and plot a new course for assessing the benefits and costs of eyewitness identification procedural reform. It is also important to emphasize at the outset that the evaluation of identification procedures presented here is closely tied to empirical data, rather than existing or evolving law or legal theory. However, as will also become clear, data alone do not tell the full story.

This article is organized into four sections. The first section describes the no-cost view and its implications. The second section describes eyewitness identification procedures and the experimental research paradigm, focusing on the relevant experimental data. The third section assesses the recommendations, and their policy implications, in light of the data presented in section two. The fourth and final section of the paper discusses the analyses in the broader context of criminal justice policy and the role of psychological science in shaping policy.

The No-Cost View and Its Implications

The no-cost argument can be illustrated with what may be the most well-known and well-documented case of false identification in recent history. In 1984, Jennifer Thompson identified Ronald Cotton, not once, but three times, as the man who broke into her apartment and raped her. Cotton was convicted and sentenced to life plus 50 years in prison. But DNA evidence analyzed only after Cotton's conviction showed that Cotton did not rape Jennifer Thompson—Bobby Poole did (Thompson-Cannino, Cotton, & Torneo, 2009).

The Cotton case provides a useful model to illustrate and define the basic elements of the no-cost view. Consider if the investigation had led police to Poole, rather than Cotton, and Poole had been placed in the lineup shown to Jennifer Thompson. Using the standard terminology of the research literature, a lineup with Poole in it would be called a *perpetrator-present* lineup, and an identification of Poole would be referred to as a *correct identification*. Of course, the investigation led police to Cotton, and again, using the standard terminology of the research literature, a lineup with Cotton, instead of Poole, is called a *perpetrator-absent* lineup, and the identification of Cotton is referred to as a *false identification*.[4]

The no-cost argument for eyewitness identification reform asserts that recommended identification procedures reduce the risk of a false identification of an innocent suspect such as Ronald Cotton, but have little or no effect on the likelihood of a correct identification of actual criminals such as Bobby Poole. Put another way, the recommended procedures would protect Ronald Cotton, without letting Bobby Poole get away.

This provides a compelling argument for reform, as it claims that recommended procedures provide substantial benefits with no costs. In the purest version of the no-cost view, eyewitness identification procedural reform offers a policy choice between dominating and dominated alternatives. Alternative A dominates Alternative B if A is superior to B on at least one dimension, and inferior to B on no dimensions. (See von Neumann & Morgenstern, 1944, p. 264; Keeney & Raiffa, 1976, p. 69, for more formal treatments of dominance among choice alternatives.) According to the no-cost view, recommended procedures dominate standard or nonrecommended procedures because they are superior on the dimension of false identification rate and no worse on the dimension of correct identification rate. Thus, the recommended procedure is obviously better, the nonrecommended procedure is a "noncontender," and the choice of the recommended procedure over the nonrecommended procedure is a demonstrably correct decision. Failure to implement recommended procedures would be irrational.

This version of the no-cost view has been expressed many times over the last 30 years. More recently, for some reforms, the no-cost view has evolved into a weaker version. The weaker version acknowledges some losses of correct identifications associated with recommended procedures; however, those losses are described as uncertain (i.e., there *may* be a loss of correct identifications) and so small as to be functionally irrelevant.

The no-cost view has profound implications. Specifically, by denying or minimizing the trade-off between correct identifications lost and false identifications avoided, the no-cost view implies that policymakers need not confront the possibility that citizens within a society may have conflicting, but equally legitimate, values (Berlin, 1958, 1969). Thus, those who are primarily concerned with protecting the Constitutional due-process rights of criminal defendants should embrace the recommended procedures because they lower the risk of false identification, and those who are primarily concerned with crime control and the conviction of the guilty should be satisfied in knowing that recommended procedures will not increase the risk that the truly guilty will escape justice. The no-cost view implies that there is no tension between due process and crime control models of criminal justice (Findley, 2008; see Packer, 1964 for a broader discussion of due process and crime control models). Unfortunately, this view is contradicted by data presented in the next section.

Eyewitness Identification Procedures and Experimental Research

This section of the article describes the research paradigm for eyewitness identification research, with examples to illustrate how the relevant data are defined and calculated. With that background, the empirical basis of the no-cost claim is evaluated.

Research on eyewitness identification follows one of three paradigms: field studies, retrospective archival analyses, and experiments. Field studies (Behrman & Davey, 2001; Wright & McDaid, 1996; Wright & Skagerberg, 2007) examine eyewitness identification procedures in actual cases, but they are generally unable to distinguish between correct identifications of the guilty versus false identifications of the innocent. For most crimes, with the exception of those that may have definitive DNA evidence, one cannot be certain whether the suspect is guilty or innocent. This problem is at the core of the Cotton case. Two juries were convinced beyond a reasonable doubt that Jennifer Thompson's identification of Ronald Cotton was a correct identification. It was not. Retrospective archival analyses typically start with a well-established or proven wrongful conviction, like that of Ronald Cotton, and then work backwards to identify factors associated with that wrongful conviction. These retrospective analyses provide the empirical basis for the link between eyewitness identification and wrongful convictions (Garrett, 2008; Gross et al., 2005). However, such analyses are based on a small and specific subset of cases–those for which eyewitness errors are known. Also, because these analyses have not systematically compared cases with known false identifications to cases with known correct identifications, they cannot establish the extent to which a given factor is associated with both correct and false identifications.

Experimental research, by contrast, allows one to know the identity of the perpetrator who is often an actor paid by the experimenter to play the role. In a typical experiment, participants become witnesses to a staged crime, either live or on video, and are later shown a suspect who is either guilty (perpetrator present) or innocent (perpetrator absent).

The research has focused on two kinds of identification procedures: showups and lineups.[5] Showups present the witness with a single person who is suspected of having committed the crime. Showups have two response options: The witness can identify that person as the perpetrator or not identify that person as the perpetrator. Lineups present the witness with a suspect, plus some number of foils or fillers who are known to be innocent. The response options are the same as those for a showup, with the additional response option of identifying a foil. There are a number of ways that a witness can make an identification error, and it is important to distinguish one kind of error—the false identification of an innocent suspect—from all other errors (i.e., foil identifications or incorrect nonidentifications). False identifications can lead to prosecution and wrongful convictions of the innocent; foil identifications and incorrect nonidentifications do not. If Jennifer Thompson had positively identified the person standing next to Ronald Cotton (i.e., a lineup filler) instead of Ronald Cotton, that person would not have been prosecuted. Thus, the focus here is on suspect identifications: correct identifications of guilty suspects and false identifications of innocent suspects.

Table 1 illustrates how response rates are calculated from data, for showup and lineup procedures. Assume that 100 participant/witnesses are shown a perpetrator-present showup, 100 are shown a perpetrator-absent showup, 100 are shown a perpetrator-present lineup, and 100 are shown a perpetrator-absent lineup. In a typical experiment, each participant will see only one showup or lineup and thus provides only a single response. The response rates are given simply by the number of participant-witnesses who make each identification response divided by the total number of participant-witnesses in that condition. Thus, in the example the correct and false identification rates are .40 (40/100) and .20 (20/100) for the showup and .50 (50/100) and .10 (10/100) for the lineup.

Based on the data from such experiments, eyewitness identification researchers have made several recommendations regarding best practices as well as practices to avoid. A landmark in the research was the publication of a paper, commissioned by the American Psychology and Law Society, as an

Table 1. Example Calculation of Response Proportions for Showups and Lineups for Eyewitness Identification Experiments

Variable	Perpetrator present (suspect is guilty)	Perpetrator absent (suspect is innocent)
Showup		
Suspect identification	40/100 = .40 (correct identification)	20/100 = .20 (false identification)
Nonidentification	60/100 = .60	80/100 = .80
Lineup		
Suspect identification	50/100 = .50 (correct identification)	10/100 = .10 (false identification)
Foil identification	20/100 = .20	40/100 = .40
Nonidentification	30/100 = .30	50/100 = .50

official "white paper" on recommended procedures (Wells et al., 1998). Shortly after its publication the U.S. Department of Justice, with input from eyewitness identification researchers, law enforcement, and defense and prosecuting attorneys, published *Eyewitness Evidence: A Guide for Law Enforcement* (U.S. Department of Justice, 1999). Shortly thereafter, the American Bar Association (2004) published a statement of best practices for eyewitness identification procedures. All three papers made formal recommendations regarding the use of unbiased instructions, the selection of foils, and the control of administrator influence. Each paper also raised concerns about the inherent suggestiveness and risk of false identifications due to the use of showup procedures. None of the papers formalized a preference for sequential lineup presentation, but the ABA report described the sequential procedure as "promising" (p. 7), the Department of Justice report stated that sequential procedures "produce more reliable evidence" (p. 9), and the Wells et al. (1998) white paper stated that, "The evidence in support of the sequential procedure . . . is rather impressive," noting further that, "sensitivity to the presence versus absence of the culprit in the lineup is far greater with the sequential procedure than it is with the simultaneous procedure," (p. 617). These three documents have provided solid foundation for efforts to reform eyewitness identification procedures in the United States.

We now turn to the relevant data. Each review presented here was conducted as a random effects analysis (Hedges & Vevea, 1998) in which each comparison constitutes a unit of analysis and the summary statistics simply average across the comparisons. The analyses take a "many small slices" approach. In other words, the analyses did not aggregate data across conditions within a study prior to aggregating data across studies. This is important for two reasons: (a) Aggregating over conditions within a study averages away the variability in the data, and (b) many of the analyses utilize non-linear dependent measures, and averaging prior to the calculation of those dependent measures[6] can distort the results. (For example, a ratio of averages is not equal to an average of ratios.) The citations for all of the studies and the citations and data for all of the analyses may be found online as supplemental material at http://pps.sagepub.com/supplemental.

Unbiased versus biased lineup instructions

Prior to a lineup or showup, the administrator (typically a police officer or detective) will instruct the witness regarding what is about to happen. Researchers have recommended that the instructions be unbiased with respect to the presence or absence of the perpetrator and unbiased with respect to the possible response options. Simply put, the recommendation is that the lineup administrator should warn the witness that the perpetrator may not be in the lineup and acknowledge that "the person I saw isn't there" is an acceptable response. Biased instructions explicitly state or imply that the perpetrator is in the lineup and that it is the witness's "job" to pick him out.

According to the no-cost view, unbiased instructions decrease the risk of false identification with no change in the correct identification rate. This has been stated many times in the eyewitness identification literature—for example, by Steblay (1997), "biased instructions produced a moderate effect on accuracy in target-absent lineups. . ., but minimal effect in target-present lineups," (p. 289), and by Wells et al.'s (1998) summary of the Steblay meta-analysis:

> A recent meta-analysis of instruction effects shows that the "might or might not be present" instruction has the effect of reducing identifications when the perpetrator is absent from the lineup while having *no effect* on identifying the perpetrator when the perpetrator is in the lineup (p. 615, emphasis added).

In 2005, Clark reexamined the Steblay (1997) meta-analysis and concluded that correct and false identifications both decreased with unbiased instructions. Additional data, published after 2005, have also shown decreases in correct and false identification rates, and the no-cost view has evolved accordingly but hesitatingly, noting that correct identification rates "*might* be slightly harmed" by unbiased instructions (Wells et al., 2006, p. 62) and that biased instructions, "*sometimes* result in a higher proportion of culprit selections," (Brewer & Palmer, 2010). In a case recently decided by the New Jersey Supreme Court (*State of New Jersey v. Larry R. Henderson*, 2010), the claim was made that, "the loss in accurate identifications [due to unbiased instructions] pales in comparison to the drop in mistaken identifications," (Scheck, Edwards, & McNamara, 2010). In its opinion, the Court noted the effects of biased instructions on false identifications, but made no mention of the effects for correct identifications.[7] Some recent reviews continue to claim that correct identifications are not affected by the "not there" warning (Fulero, 2009), and others focus only on the reduction of false identifications, without mention of the correct identification rates (Hope, 2010).

The average correct and false identification rates, comparing biased and unbiased instructions, are shown in the two leftmost columns of Table 2. False identification rates are lower with unbiased instructions (.09) than with biased instructions (.15); correct identification rates are also lower for unbiased instructions (.50) than with biased instructions (.59). Thus, the empirical results do not support the no-cost view. They also do not support the assertion that the loss of correct identifications pales in comparison to the reduction in false identifications.

Presentation format: Sequential versus simultaneous lineups

Wells (1984) hypothesized that false identifications are often a product of witnesses making relative judgments, specifically choosing, "the lineup member who most resembles the witness's memory (of the perpetrator) relative to other lineup members" (p. 92). To minimize witnesses' reliance on relative

Table 2. Analysis of Different Lineup Conditions Using Correct and False Identification Rates, Correct/False (C/F) Ratios, log (C/F), Innocence Risk (IR), log of the Risk Ratio, Pearson's *r*, *d′*, and log(β)

Condition	Correct	False	C/F	log(C/F)	IR	log(RR)[a]	*r*[a]	*d′*	log(β)
Lineup instructions (*n* = 23)									
Biased	.59	.15	6.66	1.61	.19	.11	.03	1.40	.66
Unbiased	.50	.09	6.85	1.73	.17			1.38	.86
t	2.56	2.51	.11	.72	.68	.74	.81	.19	1.70
p	.02	.02	.92	.48	.50	.47	.43	.86	.10
Presentation format (*n* = 51)									
Simultaneous	.54	.15	7.27	1.54	.21	.14	.03	1.27	.70
Sequential	.43	.09	7.74	1.71	.18			1.24	.97
t	4.48	4.09	0.44	1.72	1.65	1.76	1.71	.38	3.76
p	<.001	<.001	.67	.09	.10	.09	.09	.71	<.001
Lineup foil similarity (*n* = 18)									
Lower	.67	.31	4.00	1.05	.29	.37	.09	1.09	.21
Higher	.59	.16	7.07	1.54	.20			1.36	.61
t	3.14	4.39	1.33	2.81	3.79	2.45	3.51	2.20	2.84
p	.006	<.001	.20	.01	.001	.03	.003	.04	.01
Administrator influence (*n* = 11)									
More	.58	.21	7.61	1.27	.26	.40	.08	1.14	.46
Less	.45	.11	8.92	1.76	.19			1.25	1.03
t	3.06	2.70	.31	1.45	1.79	1.37	1.61	.57	2.10
p	.01	.02	.77	.18	.10	.20	.14	.58	.06
Showups vs. lineups (*n* = 15)									
Showup	.41	.18	3.67	1.08	.27	.32	.08	.81	.60
Lineup	.43	.11	5.86	1.50	.21			1.13	.83
t	.45	1.88	2.28	2.95	2.86	2.80	1.93	2.61	3.09
p	.66	.08	.04	.01	.01	.01	.01	.02	.008

Note. Correct = correct identification rate, False = false identification rate, C/F = ratio of correct to false identifications, log (C/F) = the log of the C/F ratio, IR = false identification rate / (correct identification rate + false identification rate), log(RR) = the log of the risk ratio of two IR calculations, *r* = Pearson's product-moment correlation; *d′* and log(β) are the discriminability and response bias parameters from signal detection theory.
[a]log(RR) and *r* have only one entry representing the effect size for each comparison.

judgments, Lindsay and Wells (1985) proposed that lineup members be presented sequentially, such that witnesses must decide yes ("that's him") or no ("that's not him") for each lineup member as he or she is presented. Because the lineup members are not all presented together, the tendency to make comparisons between lineup members and to make identifications based on relative judgments should be minimized.

The first experiment that compared simultaneous and sequential lineups showed a substantial decrease in the false identification rate (from .43 to .17) and a much smaller decrease in the correct identification rate (from .58 to .50). Lindsay and Wells (1985) emphasized this asymmetry in the results, noting a "reduction of inaccurate identifications without loss of accurate identifications" (p. 562). This view has been echoed by others (e.g., American Bar Association, 2004; Devenport, Penrod, & Cutler, 1997; Lindsay et al., 1991).[8]

A meta-analytic review by Steblay, Dysart, Fulero, and Lindsay (2001) reported a .15 decrease in correct identifications in sequential lineups in comparison with simultaneous lineups, contrary to the no-cost view. However, the loss of correct identifications continues to generate controversy. The .15 decrease in the correct identification rate was dismissed by the researchers who reported it. "Under the most realistic simulations of crimes and police procedures . . . the differences between the correct identification rates for simultaneous and sequential lineups are likely to be small or non-existent" (p. 471). It is not clear, however, which studies were deemed to be most realistic.

More recent assessments have acknowledged the loss of correct identifications, and the controversy regarding the loss of correct identifications due to sequential lineup presentation appeared settled (Lindsay, Mansour, Beaudry, Leach, &

Bertrand, 2009; Wells, 2006). However, Steblay, Dysart, and Wells (2011) have recently published an updated meta-analysis which reports only an 8% loss in correct identifications and suggests that a "better estimate" of the loss is 5% (p. 127). Although the cost is not zero,[9] it appears quite small in comparison to the 22% reduction in errors in perpetrator-absent lineups reported by Steblay, Dysart, and Wells (2011). Their conclusion is not that there is zero cost in terms of lost correct identifications, but a 5%–8% loss in correct identifications certainly does appear pale in comparison to a 22% reduction in errors arising from perpetrator-absent lineups. However, these results require careful examination.

Most important, the 22% reduction in errors for perpetrator-absent lineups, reported by Steblay et al. in their abstract, refers to all identification errors: false identifications of the innocent suspect plus identifications of lineup foils. As noted earlier, the incorrect identification of a lineup foil is a very different kind of error than the false identification of an innocent suspect. The false identification of an innocent suspect can result in prosecution and false conviction. The identification of a foil, because it is a known error, almost never leads to prosecution of the identified person. Thus, the 22% decrease in "errors" provides the wrong comparison and greatly overestimates the reduction in false identifications.

In addition, the present analysis includes data from five studies that were not included in the Steblay et al. analysis and excludes, due to a confound, one study that was included in the Steblay et al. analysis. Although the conclusions of the present analysis do not hinge on these inclusion–exclusion differences, they are described briefly in Appendix A.

The results of the present analysis show lower false identification rates for sequential lineups (.09) than for simultaneous lineups (.15) but also lower correct identification rates for sequential lineups (.43) than for simultaneous lineups (.54). Thus, the empirical results, showing roughly equivalent decreases in both correct and false identification rates, contradict the no-cost view.

Lineup composition: More versus less similar foils

Another source of false identifications arises from the composition of the lineup. Specifically, if the foils are chosen in such a way that the suspect stands out, then the suspect may be easily identified. The problem, of course, is that if the suspect is innocent, this "easy" identification will be a false identification. To avoid such false identifications, lineup foils should be selected in such a way that the suspect does not stand out (Farmer, 2001; U.S. Department of Justice, 1999; Wells et al., 1998). Fundamentally, the issue is about the similarity of the foils.

The no-cost view with regard to foil similarity arose from the results of the first study conducted by Lindsay and Wells (1980). They compared lineups with foils that mismatched the perpetrator on salient dimensions (ethnicity, hair color, and facial hair) and lineups with foils that matched the perpetrator on those salient dimensions. Lineups with better matched foils showed a steep reduction in the false identification rate (from .70 to .31) and a smaller, statistically nonsignificant reduction in the correct identification rate (from .71 to .58). In light of these results, Lindsay and Wells (1980) wrote, "Our findings provide the criminal justice system with a reasonable means of improving the reliability of eyewitness testimony at apparently little cost" (p. 310). Nineteen years later, Lindsay and Pozzulo (1999) echoed the point.

> The results (of Lindsay & Wells, 1980) establish a clear pattern that is repeated in *all* (emphasis added) lineup bias studies. Manipulation of foil quality did not significantly influence the rate of correct identification of the criminal. . . . The impact of foil quality on false identifications was significant and dramatic (p. 352).

In the eyewitness identification literature, foil similarity has been defined in various ways—for example, in terms of the match to a description of the perpetrator (Lindsay & Wells, 1980; Wells et al., 1993), in terms of clothing worn by the perpetrator (Lindsay, Wallbridge, & Drennen, 1987), and in terms of similarity ratings (e.g., Brewer & Wells, 2006; Leippe, Eisenstadt, & Rauch, 2009). The analyses presented here include all of these different ways of varying foil similarity. The results in Table 2 show that false identification rates are lower with more similar (.16) than with less similar (.31) foils, and correct identification rates are also lower with more-similar (.59) than less-similar (.67) foils. These results clearly are inconsistent with the no-cost view.[10]

Administrator influence: More versus less influence

Eyewitness identification researchers have also recommended that lineups be presented by an administrator who is blind as to the position of the suspect in the lineup. This is analogous to blind testing procedures that have been used in medical and psychological research for over 100 years (Dean, 2006; Rivers & Webber, 1907). The rationale for blind testing is that lineup administrators who know the position of the suspect and are highly motivated to obtain an identification of that suspect may deliberately or unwittingly direct witnesses toward making that identification.

Only one published study directly compares blind and nonblind procedures (Greathouse & Kovera, 2009).[11] However, several studies have made similar comparisons between conditions in which lineup administrators were free, or even encouraged to influence witnesses, versus conditions that made it difficult or impossible to influence witnesses. The results averaging over 11 such comparisons are presented in Table 2, which shows a decrease in the average false identification rate, from .21 to .11, and a decrease in the average correct identification rate from .58 to .45. Procedures that minimize administrator influence have both benefits and costs. Fewer innocent suspects are misidentified, and fewer guilty suspects are correctly identified.

Lineups versus showups

As described earlier, a showup procedure presents a single person to a witness for a yes or no response. There are no foils. Showup procedures have been the target of concern and criticism for decades, not only from eyewitness identification researchers (Wagenaar & Veefkind, 1992, Yarmey, Yarmey, & Yarmey, 1996), but also in court rulings (*Stovall v. Denno*, 1967; *Wisconsin v. Dubose*, 2005) and police procedure manuals (Calandra & Carey, 2005).

The showup–lineup comparison is presented last because it is the one case in which the results are consistent with the no-cost view. Averaging over 15 comparisons, lineups show lower false identification rates (.11) and slightly higher correct identification rates (.43) than showups (.18 and .41, respectively). The choice between lineup and showup procedures does indeed seem like an easy choice between a dominating and a dominated alternative. There are, however, practical matters that potentially complicate the comparison between lineups and showups. Showups are typically conducted in circumstances in which police identify a suspect quickly after the crime occurred. For example, within minutes after an armed robbery, police may spot a person who matches the description of the robber. The argument made by law enforcement (Calandra & Carey, 2005; Penrod, 2007) and prosecutors (Cooley, 2007) is that a showup can be conducted within minutes, whereas the presentation of a lineup requires additional time to select foils and compose the lineup. During this additional time, memories fade, identifications may be less accurate, and perpetrators may escape.[12]

Summary

The results in Table 2 show a consistent pattern. With the exception of showup–lineup comparisons, identification procedures that reduce the risk of false identifications also reduce the probability of correct identifications. The decreases in correct identification rates are not trivial, ranging from .09 to .13, and in many cases they are numerically as large or larger than the decrease in false identification rates. Clearly, policy decisions about eyewitness identification procedures are not easy choices between dominating and dominated alternatives. Rather, they involve choices between alternatives with cost–benefit trade-offs. This raises the question as to how much cost in terms of lost correct identifications should policymakers accept in exchange for the benefit of a reduction in false identifications. This is taken up next.

Assessment of Identification Procedures

The assessment of eyewitness identification procedures requires the development of a metric to weigh the costs (correct identifications lost) against the benefits (false identifications avoided) of recommended procedures. Two sets of analyses are presented here. The assessment starts out close to the data, comparing identification procedures based on the expected *probative value* of a suspect identification. I will define probative value more precisely (and in more than one way) later in this article, but in short, probative value measures the usefulness of the evidence in discriminating between the guilty and the innocent. There is more than one way to calculate probative value, and the conclusions that one draws will depend on precisely how probative value is calculated.

The measure of probative value is but one way to compare and evaluate identification procedures, and other information may need to be considered in making policy decisions. Thus, I present additional analyses that extend the probative value analyses and build on economic decision models of evidence (Cullison, 1969; DeKay, 1996; Kaplan, 1968; Laudan, 2006; Laudan & Saunders, 2009; Lillquist, 2002; Tribe, 1971) to compare identification procedures by the index of their long-run expected utility.

Probative value of eyewitness identification evidence

Evidence, including eyewitness evidence, is only useful to the extent that it is diagnostic, and that it allows the trier of fact (i.e., the judge or jury) to evaluate the critical question before the court. In a criminal case, that critical question is typically about whether the defendant is guilty of the crime. The identification of the suspect (who may later become the defendant) has probative value to the extent that it is likely to occur when the suspect is guilty (correct identification rate is high) and unlikely to occur when the suspect is innocent (false identification rate is low).

Probative value, as it is discussed here, is the expected probative value of a suspect identification, based on expected correct and false identification rates as they are estimated from laboratory experiments. Thus, the experimental paradigm allows the calculation of what is uncalculable outside of the laboratory—a numerical index of the probable accuracy of an eyewitness's identification of the suspect.

Six measures of probative value are considered here: (a) likelihood ratios; (b) logarithmic transformations of likelihood ratios; (c) risk differences; (d) logarithmic transformations of risk ratios (RRs); (e) Pearson's r, as a general effect size measure of proportional changes in correct and false identification rates; and (f) a signal detection theory analysis assessing both the discriminability between guilty and innocent suspects, as well as shifts in response tendencies across identification procedures. These measures are commonly used in psychological and medical research (see Rosenthal & Rosnow, 2008; Wickens, 2002).

Why measure the same thing six different ways? There is no single measure of probative value that is agreed upon by scholars of legal evidence. Moreover, the various measures have different statistical properties and different advantages and disadvantages. Consequently, different measures may not lead to the same pattern of results. There is much to be gained by noting and understanding this variation across measures, rather than choosing one measure and arguing for why it is "the right method." This lack of loyalty to a particular measure of probative value is also critical for policy analysis. Recommendations and policy decisions, to the extent that they are based on assessments of probative value, should not be tied to a single measure of probative value but should generalize across different measures of probative value. A lack of generalization across measures should not be argued away or ignored, but rather should be brought out into the daylight for policy analysis.

For each analysis, the relevant measures of effect size were calculated and the averages are shown in Table 2. To assess the reliability of the effects across studies, *t* tests were conducted on the relevant effect size measures, and these are also summarized in Table 2. Random effects analyses tend to have very low statistical power as each comparison counts as one observation, irrespective of whether that comparison is based on data from 10 or 10,000 participants. Because of that lack of power, and so that we do not overlook a good procedure that increases the probative value of evidence, we used an alpha level of .10 as our measure of statistical significance. We must be cautious, of course, in the consideration of these statistical analyses so that we do not fall into the trap of interpreting significance tests as dichotomous tests of whether effects exist or do not exist. To avoid a repetitive recitation of statistical analyses, the analyses across all measures will be summarized together, after all of the measures have been described.

Likelihood ratios. A common index of probative value is the likelihood ratio (Friedman, 1986; Kaye, 1986; Kaye & Koehler, 2003; Koehler, 1996; Lempert, 1977; Lyon & Koehler, 1996; Wells & Lindsay, 1980; Wells & Olson, 2002). Here the probative value of a suspect identification is given by the correct identification rate divided by the false identification rate, denoted as *C/F*. The interpretation of the C/F ratio is straightforward. If the correct identification rate is .4 and the false identification rate is .1, the C/F ratio is 4.0, implying that a guilty suspect is four times more likely to be identified than an innocent suspect. The drawback of a ratio measure is that it is extremely sensitive to small values in the denominator such that very small decreases in the false identification rate can produce very large increases in the C/F ratio, making it unstable (Clark, Howell, & Davey, 2008). Also, ratio distributions are asymmetric with one side of the distribution stretching from 1 to infinity, whereas the other side is compressed between 0 and 1. A standard solution to these problems is to use the log of the C/F ratio rather than the C/F ratio itself (Kaye, 1986). For a correct identification rate of .4 and a false identification rate of .1, the log of the C/F ratio is 1.39. C/F ratios and log(C/F) ratios were calculated for each study, and the means are given in the third and fourth columns of Table 2.

Risk differences and RRs. Another measure of eyewitness identification responses is based on the risk that the suspect who has been identified is innocent. Clark and Godfrey referred to this as innocence risk (IR), calculated as the proportion of suspect identifications that are false identifications of an innocent person—specifically, IR is the false identification rate divided by the sum of correct and false identification rates: F / (F + C). To illustrate, using the same .4 and .1 correct and false identification rates, IR = .1 / (.1 + .4) = .20. There are problems that arise when comparing differences in very small risk values (see Kaye & Koehler, 2003; Rosenthal & Rosnow, 2008, pp. 324–327). For example, the difference between two risk values of .01 and .001 is very small, although the risk is 10 times higher in one case than the other. The arcsine transformation reduces this problem somewhat but not entirely. Another means of comparing low risk values is to take their ratio rather than their difference. Of course, the RR calculation suffers from the same problems as C/F ratios, and the statistical analysis depends on the arbitrary decision of which value of IR is placed in the numerator. Thus, it is essential that the RRs be analyzed by a logarithmic transformation: log (RR).

To illustrate the calculation of the RR, assume two cases, one where the correct and false identification rates are .4 and .1, respectively, and one where the correct and false identification rates are .6 and .2, respectively. IR for the two cases is .20 and .25.

The log of the RR is calculated as log(.25/.20) = .22. Thus, decreases in risk are denoted by positive values (reducing risk is a good or positive outcome).

Proportional changes in correct and false identification rates. Changes in correct and false identification rates were also assessed with an all-purpose measure of effect size Pearson's *r*.[13] If correct and false identification rates change proportionally, *r* will be zero. If the changes are disproportional, such that false identification rates decrease proportionally more than correct identification rates, *r* will be positive. If false identifications decrease proportionally less than correct identifications, *r* will be negative. Positive values indicate an increase in the probative value of the evidence, whereas negative values indicate a decrease in the probative value of the evidence. The advantage to Pearson's *r*, in addition to other measures of probative value, is that *r* is very general and flexible, allowing it to be evaluated and compared across a wide range of circumstances (Rosenthal & Rosnow, 2008; Rosnow & Rosenthal, 2003).

Detection theory analysis: d′ and log(β). Eyewitness identification can also be viewed as a signal detection task in which the witness must distinguish between guilty and innocent suspects. The detection theory analysis provides two measures of performance: d' as a measure of the discriminability between guilty and innocent suspects, and log(β) as a measure of the degree to which changes in response probabilities reflect a change in response bias. In the present analyses, log(β) measures the tendency to identify guilty and innocent suspects (rather than distinguishing between them), and it decreases as the guilty and innocent suspect identification rates increase.

Detailed descriptions of the theory and its many applications are given by Green and Swets (1966), Macmillan and Creelman (2005), and Wickens (2002), in addition to many others.[14]

Results of the probative value analyses

For every comparison, and with every measure of probative value, the probative value of a suspect identification was numerically higher for the recommended procedure relative to the nonrecommended comparison. In general, the effect sizes were larger for the lineup–showup, lineup foils, and administrator influence analyses than for the lineup instructions and simultaneous-sequential analyses. The analyses for administrator influence, however, were generally not statistically reliable, due to the much smaller number of comparisons.

There was considerable variation across the various measures of probative value. The analyses are described in terms of the proportional changes in correct and false identification rates (Pearson's r), IR and RRs, and C/F ratios and d'.

The analyses of all five procedures showed that the decrease in false identifications was proportionally larger than the decrease in correct identifications, although these effect sizes varied from small to very small. Pearson's r was .09 comparing more-similar and less-similar foils, .08 comparing lineups and showups, and .08 comparing less versus more administrator influence on lineups. Pearson's r for unbiased instructions and sequential lineups were both .03. These are small, although certainly not trivial effects.

All of the recommended procedures reduced IR as measured by the risk difference (IR) as well as the log of the RR. The RR showed the same pattern as the risk difference measure, so only the latter is discussed. The risk differences were larger for the analyses of lineups and showups (.06), lineup foils (.09), and administrator influence (.07) and smaller for the analyses of lineup instructions (.02) and sequential and simultaneous lineups (.03). The reductions were not statistically reliable for the lineup instructions and administrator influence analyses, for different reasons. The reduction was simply very small for the lineup instructions analysis. The reduction was relatively large for the administrator influence analysis, but it was based on few comparisons. The mismatch between the size of the risk reduction and the statistical significance is a sure sign that more research is needed to assess the effects of administrator influence on correct and false identification rates.

There were also notable inconsistencies across measures. Several analyses that showed significant reductions in IR did not show significant increases in the C/F ratios. In addition, the d' statistics in the comparison of simultaneous and sequential lineups and biased versus unbiased instructions were nearly identical. Additional analyses suggest that the nonsignificant differences for C/F and d' arose for completely different reasons. As expected, the distributions of C/F ratios were very positively skewed because the ratio goes to infinity as the false identification rate goes to zero, making C/F unstable with very large variances. These statistical properties are less pronounced for the calculation of IR and the log of the RR. The near-zero differences in d' suggest that the effects of unbiased instructions and sequential lineups are best described as a criterion shift rather than a change in discriminability. Analyses of log(β) are consistent with this criterion shift account. All recommended procedures showed conservative shifts in the criterion for making an identification.

Policy implications of probative value analyses

The policy recommendations based on these analyses depend not only on the data, but also on the decision rule applied to those data. If one is concerned only with false identifications, with no concern for the loss of correct identifications, all of the procedures would be recommended. However, one need not adopt such a one-sided decision rule to endorse most or all of the recommended procedures. Another reasonable decision rule gives preference to procedures that reduce the false identification rate, provided that there is no loss in the probative value of the evidence. This decision rule allows for decreases in the correct identification rate, but only if they do not undermine the value of the evidence. All of the recommended procedures pass the test on this decision rule as well.

Other decision rules would support the endorsement of some identification procedures, but not others. In particular, the analyses presented here suggest that unbiased instructions and sequential lineups are not associated with an increase in discriminability between suspects who are guilty versus suspects who are innocent. If the decision rule is to recommend procedures that exhibit greater discriminability between guilt and innocence, there is no clear recommendation to make.

Limitations of probative value analyses. The analyses discussed so far are limited in at least three ways: (a) They do not take into consideration the *guilty base rate* (i.e., the relative proportion of identification procedures that involve a guilty suspect rather than an innocent suspect), (b) they do not consider the utilities attached to the various responses, and (c) it may be possible to increase probative value through procedures that clearly violate even the most basic standards of justice and due process. The first two limitations correspond to important parameters of expected utility theory (Savage, 1954; von Neumann & Morgenstern, 1944), which is discussed in the next section. The third problem reflects a fundamental limitation of expected utility.

The guilty base rate in eyewitness identification

A cost–benefit analysis of eyewitness identification requires one to consider: How often are innocent people presented to witnesses for the purpose of identification? The question concerns the base rates of guilty versus innocent suspects, denoted here as $p(G)$ and $p(I)$. Since they must sum to 1.0, we will refer only to the guilty base rate, $p(G)$.

If the guilty base rate is very high, such that innocent suspects are rarely presented to witnesses for the purposes of identification, then procedures that are designed to reduce the false identification rate protect against an error that has very little opportunity to occur. Conversely, if the guilty base rate is low, such that innocent suspects are frequently subjected to such procedures, then the recommended procedures provide increased protection against an error that has greater opportunity to occur.

This is illustrated in Figure 1 by comparing two real-world scenarios: one in which the guilty base rate is fairly low, $p(G) = .6$, and one in which the guilty base rate is much higher, $p(G) = .9$. Assume in both cases that there are 1,000 identification procedures, and assume two different kinds of identification procedures, A and B, in which both the correct and false identification rates are higher for A (correct = .5 and false = .2) than for B (correct = .4 and false = .12).

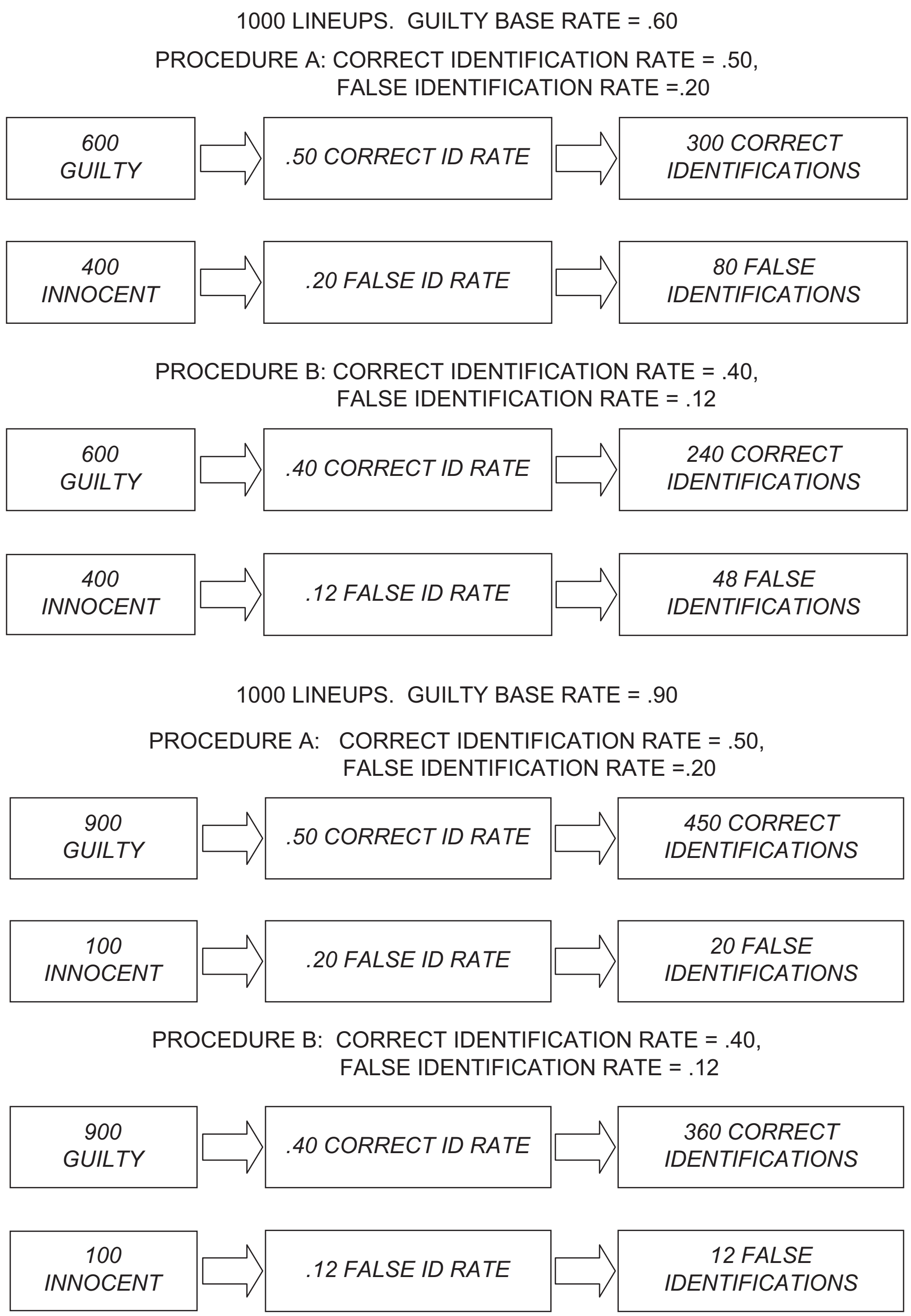

Fig. 1. Example calculation of the number of correct and false identifications, for Procedure A and B, in which the correct and false identification rates are higher for A than for B, and the guilty base rate is assumed to be .6 or .9.

For $p(G) = .6$, there will be 600 guilty-suspect lineups and 400 innocent-suspect lineups. Applying the correct (.5) and false identification (.2) rates for Procedure A, there will be 300 correct identifications and 80 false identifications. If Procedure B were implemented instead of Procedure A, there would be 240 correct identifications and 48 false identifications. Thus, by implementing Procedure B, the number of correct identifications would be reduced by 60 and the number of false identifications would be reduced by 32. Thus, the "exchange rate" of correct identifications lost in exchange for false identifications avoided is 60 to 32, or 1.88 to 1.

Now, consider the case where $p(G) = .9$, and there are 900 guilty-suspect lineups and 100 innocent-suspect lineups. For Procedure A, there would be 450 correct identifications and 20 false identifications, whereas for Procedure B, there would be 360 correct identifications and 12 false identifications. By implementing Procedure B, the number of correct identifications would be reduced by 90 and the number of false identifications would be reduced by 8. The exchange rate here is 90 to 8, or 11:25 to 1.

Clearly the cost of protecting the innocent is substantially higher if 1 out of 10 lineups includes an innocent suspect than if 4 out of 10 lineups include an innocent suspect. This is true even though the probative value of a suspect identification arising from Procedure A is higher than the probative value of a suspect identification arising from Procedure B. The example suggests that an evaluation of the costs and benefits of different lineup procedures must consider not only the probative value of identification evidence, but also the guilty base rate.[15]

Utilities and disutilities of expected outcomes

The question remains: What should the exchange rate be for correct identifications lost versus false identifications avoided? To answer the normative "should" question, one must consider the costs and benefits associated with the possible outcomes of an identification procedure. This is easier said than done, of course, as the questions can be daunting. What are the costs associated with a false identification that may lead to a false conviction and years of false imprisonment? What are the costs associated with a false nonidentification that may allow a criminal to escape justice and commit additional crimes?

Such questions have occupied philosophers and legal scholars for centuries (Blackstone, 1769; Volokh, 1997). A less philosophical analysis considered here is based on expected utility theory. As noted by DeKay (1996), expected utility theory may be viewed as a normative theory of choice and decision making for two reasons: (a) The decision-maker necessarily maximizes utility if he or she follows a small number of "intuitively appealing" axioms, and (b) the application of these axioms will maximize utility in the long run (DeKay, 1996, p. 110; see also von Neumann & Morgenstern, 1944, pp. 24–29, 617–628). The expected utility associated with a specific identification procedure may be calculated simply as the sum of the utilities for each of the possible outcomes, weighted by the probability of each of the possible outcomes. At its simplest level, in choosing between two eyewitness identification procedures, one can simply calculate the expected utility of each procedure and choose the procedure with the highest long-run expected utility.

Ceci and Friedman (2000) derived a preference rule from the axioms of expected utility theory that allows one to compare two lineup procedures, A and B, in which the correct and false identification rates are lower for B than for A. The derivation of the preference rule is somewhat lengthy, and a more formal description is given in Appendix B. For present purposes, it is sufficient to note that the preference rule reduces to a simple inequality: Procedure B should be preferred over Procedure A if and only if:

$$\frac{\text{Correct identifications lost}}{\text{False identifications avoided}} < \frac{\text{Cost of a false identification (innocent suspect)}}{\text{Cost of a false nonidentification (guilty suspect)}}$$

The left-hand side of the inequality represents the ratio of the number of correct identifications lost versus the number of false identifications avoided, if one implements Procedure B, rather than Procedure A. The right-hand side of the inequality is a ratio that represents the cost of a false identification relative to the cost of a false nonidentification. (Note that for these analyses, a non-identification refers to any response that is not an identification of the suspect, including foil identifications as well as "not there" responses.) The inequality shows that Procedure B should be preferred over Procedure A if the ratio of correct identifications lost to false identifications avoided is lower than the cost of a false identification relative to the cost of a false non-identification.

The left-hand term can be estimated with empirical data from eyewitness identification experiments, following the example in Figure 1. Thus, if $p(G) = .6$, the ratio of correct identifications lost to false identifications avoided by implementing Procedure B rather than Procedure A (the left-hand term in the inequality) is 1.88. However, if $p(G) = .9$, then that ratio is much higher (11.25).

The right-hand side of the inequality represents a normative ratio based on the relative costs of the two kinds of identification errors. These relative costs were perhaps most famously considered by the English legal scholar William Blackstone who wrote, ". . . better that ten guilty persons escape than that one innocent suffer," (Blackstone, 1769, p. 352). This 10-to-1 *Blackstone ratio* provides one possible estimate for the right-hand side of the Ceci–Friedman inequality. Using the calculations above, if $p(G) = .6$, one should prefer Procedure B over Procedure A because the ratio of correct identifications lost (60) to false identifications avoided (32) is 1.88—considerably less than 10. However, if $p(G) = .9$, one should prefer Procedure A over Procedure B, because the ratio of correct

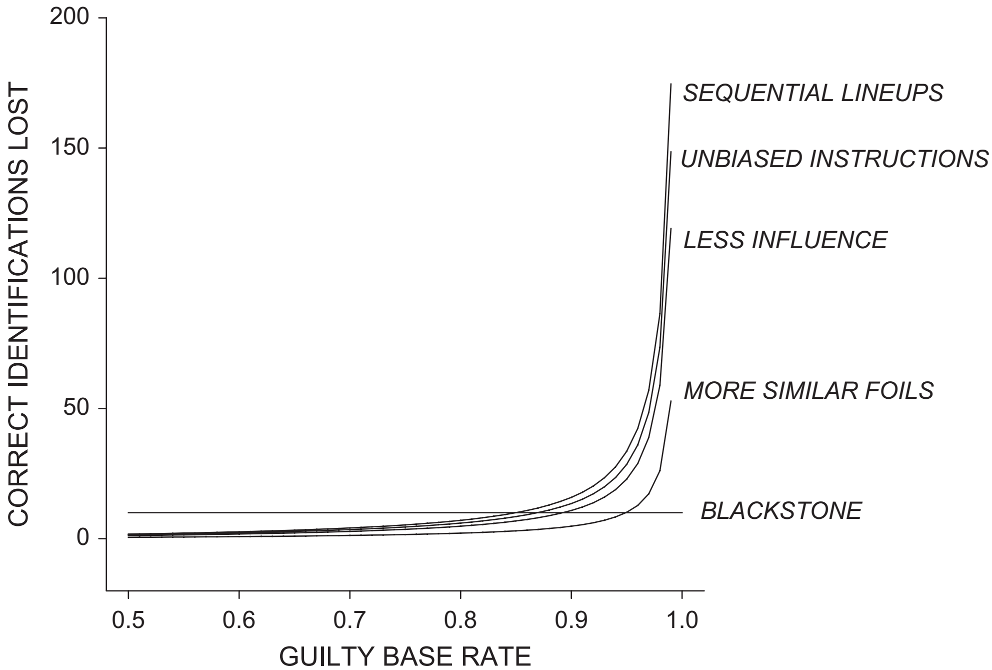

Fig. 2. Ceci–Friedman utility functions plotting reduction of correct and false identifications as a function of guilty base rate, *p(G)*, for foil similarity, lineup instructions, simultaneous and sequential lineups, and degrees of administrator influence. Functions cross the Blackstone ratio line at *p(G)* = .95 (more similar foils), .87 (unbiased lineup instructions), .85 (sequential lineups), and .89 (less administrator influence).

identifications lost (90) to false identifications avoided (8) is 11.25—just above 10. The Blackstone ratio is not the only standard that has been offered regarding the relative utilities of trial outcomes, but it is a useful benchmark to begin the discussion of lineup procedures.

The Ceci–Friedman inequality is illustrated in Figure 2 for four of the eyewitness identification procedures considered here. (The lineup–showup analysis is excluded because there is no trade-off between correct identifications lost and false identifications avoided.) Each plot shows the number of correct identifications lost in exchange for each false identification avoided, as a function of the guilty base rate $p(G)$, using the correct and false identification rates shown in Table 2. The Blackstone ratio is shown as the flat line (at 10:1). From the figure, one can determine the value of $p(G)$ at which each trade-off function crosses the Blackstone ratio: .85 when comparing simultaneous lineups with sequential lineups, .87 when comparing biased instructions with unbiased instructions, .95 when comparing more-similar lineup foils with less-similar lineup foils, and .89 when comparing more administrator influence with less administrator influence. These crossing points allow specific recommendations based on explicitly stated underlying assumptions. For example, using experimental data to estimate the correct and false identification rates obtained with simultaneous and sequential lineups, one should prefer sequential lineups over simultaneous lineups if the following conditions hold: (a) the normative cost ratio comparing false identifications with missed identifications is 10 to 1, and (b) at least 15% of identification procedures include a suspect who is innocent.

The analyses suggest that with certain assumptions, one can express utility-based preferences for one identification procedure over another. The calculations may also be used to bring policy decisions into the daylight of their underlying assumptions. For example, if a jurisdiction mandates the use of sequential lineups based on the data garnered by experimental psychologists over the last 25 years and further assumes a 10:1 cost ratio, then the implicit assumption is that at least 15% of identification procedures involve a suspect who is innocent.

This, of course, is not the final word in the consideration of costs and benefits in the eyewitness identification reform. The next two sections discuss the guilty base rate and the estimation of utilities in more detail.

Considerations and cautions regarding the guilty base rate. The guilty base rate estimation is critical in that it constrains the opportunity for false identifications. It follows that if the guilty base rate is very high (above 90%) identification procedures that have higher suspect identification rates (both guilty and innocent) should have higher long-run utility than identification procedures that have lower suspect identification rates. However, this analysis is not as simple as it may first appear, for several reasons.

First, it is misleading to describe $p(G)$ as a single base rate, because the base rate of guilt will vary across jurisdictions, investigators, and cases and depends on the diagnosticity of the evidence prior to the identification. Thus, $p(G)$ for cases with DNA matches to the suspect should be extremely high, whereas $p(G)$ for cases in which suspects are identified through anonymous tips is certainly much lower. A utility analysis suggests that one should use a more liberal identification procedure for DNA cases, because one has more (correct identifications) to lose, and a more conservative procedure for anonymous tip cases. However, such a policy would obviously violate the independent assessment of evidence and effectively institutionalize an "echo chamber" approach to evidence (see Risinger, Saks, Thompson, & Rosenthal, 2002).

Also, to the extent that one is evaluating a procedure, the consideration of prior probabilities confuses the utility of a particular procedure in a way not unlike the confusion between ability and inheritance. If the suspect in the lineup is almost always guilty, then the procedure that obtains more suspect identifications (both guilty and innocent) will result in more correct identifications, not because of the probative value achieved by the procedure, but rather because of the probative value of the evidence inherited by the procedure.

The consideration of guilty base rates significantly complicates policy decisions regarding eyewitness identification reform, and it is not clear how base rates should be considered. However, given that the costs and benefits of reform depend on assumptions regarding base rates, policy makers must consider base rate information in their policy decisions.

Estimating utilities. It is clear from Figure 2 that the preference for one identification procedure over another depends on the relative cost of a false identification versus a missed identification. The Blackstone ratio is one possible benchmark; however, there is nothing demonstrably correct about the 10:1 ratio. In an expansive and somewhat tongue-in-cheek review, Volokh (1997) noted that this ratio has been declared to be as low as 1:1 and as high as 5000:1. Some of the ratios are little more than sound bites, with little or no justification. However, more carefully reasoned analyses of the utilities have been given by Tribe (1971), Milanich (1981), Risinger (2007), and Lillquist (2002) for trial outcomes, by Biedermann, Bozza, and Teroni (2008) for forensic analysis outcomes, and by Malpass (2006) for eyewitness identification outcomes. Although the utilities vary considerably across the various treatments, the ratios of utility differences vary much less. Indeed, they all lie between 1.25 and 12.5.[16,17]

It is also important to note that the utilities considered by Tribe (1971), Lillquist (2002), Risinger (2007), and Milanich (2006) are the utilities associated with trial outcomes, not eyewitness identification outcomes. Although it is reasonable to assume that the utilities associated with eyewitness identification outcomes are correlated with the utilities associated with trial outcomes, one should not assume that they are equivalent. To the extent that correct identifications do not always lead to correct guilty verdicts or guilty pleas, their utility is reduced. Likewise, to the extent that false identifications do not always lead to incorrect guilty verdicts or guilty pleas, their disutility is also reduced.[18] These distinctions raise important questions regarding the downstream consequences of a suspect identification. How does an eyewitness's identification of a suspect alter the trajectory of a criminal case as it proceeds through plea bargaining and trial? How does the criminal justice system correct eyewitness errors? There is little in the way of data that address these questions (see however, Bushway & Redlich, 2010; Elder, 1989; Gross, 1987); however, the questions highlight the point that one cannot simply attach the utility estimates for the outcomes of trials to the outcomes of eyewitness identification procedures. Notably, some legal scholars have argued that innocence-emphasizing rules such as the Blackstone ratio should not be applied to the investigative stage of the criminal investigation. Specifically, Laudan (2006) and Lillquist (2007) have both suggested that the ratio of exculpatory to inculpatory errors should be closer to 1:1 rather than 10:1.

Beyond outcomes: Other social psychological costs and benefits. There is, of course, a fundamental limitation of utility analyses in that they may overlook costs and benefits that arise independently of specific outcomes. The problem is illustrated in Tribe's (1971) rhetorical question.

> "How much would you regret the erroneous conviction of [a] defendant for armed robbery?" The answer must surely be, "It depends." It depends in part upon the *character* of the error itself. . . . And it depends even more significantly upon the process that led to the error; one cannot equate the lynching of an innocent man with his mistaken conviction after a fair trial. Indeed, it is at least arguable that there is nothing good or bad about *any* trial outcome as such; that the process, and not the result *in any particular case*, is all important (Tribe, 1971, p. 1381, emphasis in the original).

Tribe's query illustrates that undesirable outcomes may arise from more or less desirable procedures. The opposite problem may be even more troublesome: Desirable outcomes may arise from decidedly undesirable procedures. One can imagine a correct conviction based in part on an identification procedure in which a police officer, convinced of the suspect's guilt, simply tells the witness to circle and initial the suspect's photograph in a photo lineup "or else." If the police officer's hunches are right most of the time, such that the guilty base rate is very high, $p(G) = .95$, such a procedure should lead to a large number of correct identifications. Of course, such a procedure would also produce some false identifications and even some false convictions, but on the basis of a probative value or utility analysis, one can imagine that, overall, the correct outcomes outnumber the incorrect outcomes at a sufficiently high rate that the expected utility of the procedure justifies its use. In other words, the ends justify the means.

This distinction between outcomes and procedures is at the core of psychological theories of procedural justice as they have been developed by Thibaut and Walker (1975, 1978; see also Thibaut, Walker, LaTour, & Houlden, 1974), and by Tyler and colleagues (Tyler, 1990; Tyler & Huo, 2002). A fundamental premise in Tyler's theory of procedural justice is that people trust legal authorities (Tyler & Huo, 2002), obey the law (Tyler, 1990), and cooperate with the police (Sunshine & Tyler, 2003; Tyler & Fagan, 2008) to the extent that they perceive legal institutions and authorities to be trustworthy and legitimate.

A detailed analysis of eyewitness identification procedures, within a procedural justice framework, has not been conducted, and a complete and thoughtful analysis is beyond the scope of the present article. In lieu of that detailed analysis, it seems clear that some police procedures—procedures that are manipulative or coercive, or lineups that are clearly biased due to the poor selection of fillers—would land out of bounds. To the extent that the use of such procedures undermines the legitimacy of the criminal justice system, they may allow police to win the battle by clearing a particular case but lose the war by losing public trust. Specifically, eyewitnesses may be less willing to come forward to participate in the criminal justice process, and juries may become sufficiently skeptical that they become unwilling to convict defendants who are guilty.

Synthesis and Policy Implications

The no-cost view implies that policy decisions regarding eyewitness identification evidence are simple. With the (incorrect) assumption that correct identification rates are invariant across eyewitness identification procedures, policy decisions can be entirely data driven by a simple rule: Implement the procedure that has the lower false identification rate.

This article shows that such a simple decision rule is untenable. At a minimum, the assessment of eyewitness identification procedures involves the following considerations: (a) false identification rates, (b) correct identification rates, (c) the utilities and disutilities associated with identification outcomes, (d) the guilty base rates, and (e) social costs and benefits beyond specific eyewitness identification outcomes.

Reformed identification procedures are more attractive in terms of a cost–benefit analysis to the extent that innocent suspects are likely to be presented in identification procedures (i.e., the guilty base rate is low), and to the extent that the cost of a false identification is higher than the cost of a lost correct identification. In this analysis, I considered the Blackstone ratio as one measure of the relative costs. However, the Blackstone ratio is not a natural law, and the fact that it is well-known among legal scholars does not establish it to be true. It is essential that the estimation of these costs rises above slogans, anecdotes, and emotional appeals.

Policy analysis based on these multiple constraints is consistent with the aims of evidence-based policing (Sherman, 1984, 1998) in that policy decisions are based on empirical data in the form of expected outcomes estimated from eyewitness identification experiments. The additional considerations of social costs and values echoes a point made by Thatcher (2001):

> Police will clearly benefit from instrumental knowledge (data) such as that produced through experiments. But they will also benefit from better forms of practical reasoning, including better interpretations of ambiguous values and better ideas about how trade-offs among values should be made . . . (Thatcher, 2001, p. 388).

It is critical in such a multiply determined policy analysis that the components do not become confused or conflated. In particular, the interpretation of empirical data must not be influenced by social values. More to the point, the decreases in correct identification rates that come with recommended eyewitness identification procedures must not be ignored or interpreted away in pursuit of the worthy goal of reducing the risk of false identifications that can lead to wrongful convictions of the innocent.

It is also critical that policy decisions not be based on shifting criteria. In the worst case, such shifting criteria can lead to a situation in which some recommendations are based on empirical data when the data are helpful, whereas other recommendations are based on other considerations, such as fairness or logical arguments, when the data are missing or unhelpful.

Policy decisions, of course, must be made by policy makers. However, research psychologists play an important role. Policy makers cannot weigh the costs of lost correct identifications against the benefits of fewer false identifications unless they are provided with accurate relevant data. Moreover, psychological science can improve the decisions that policymakers make. As Swets, Dawes, and Monahan (2000) note, "apart from improving accuracy, there is a need to produce decisions that are in tune . . . with the benefits and costs . . . of correct and incorrect decisions" (p. 1).

Broader implications

The analyses and principles discussed here have implications that extend far beyond eyewitness identification, to other aspects of witness statements and to other forms of forensic identification. In the area of eyewitness testimony, experiments show that witnesses will report more correct information and more false information when such information is obtained from an interviewer (Loftus, Miller, & Burns, 1978) or from another witness (Rush & Clark, 2010). Likewise, a meta-analysis comparing the cognitive interview with standard interview procedures also shows an increase in both correct and incorrect information (Memon & Higham, 1999). Experimental data have also shown results suggesting that police will obtain more correct and more false confessions by employing procedures that are more manipulative and coercive (Russano, Meissner, Narchet, & Kassin, 2005). These same principles apply to some extent to other forms of forensic identification, including fingerprints, bite marks, handwriting, DNA analysis (see Biedermann et al., 2008; Cole, 2001; Dror & Rosenthal, 2008; Miller, 1984; Risinger et al., 2002), and other forms of forensic analysis, some of which are now in dispute (see Saks & Faigman, 2008). All of these cases—from eyewitness statements to eyewitness identification to confessions to forensic identification—are connected in that they involve trade-offs and various degrees of uncertainty (Biedermann et al., 2008).

Future directions

The data in Table 2 represent a work in progress and are not the last word on any of the reforms or recommendations. Surely, new experiments will be conducted, and our knowledge will evolve with future experiments. The data in Table 2 should be viewed as a snapshot of our current knowledge, not as established truths.

Also, this article evaluated lineup procedures at the level of the forest rather than the trees (i.e., based on general patterns rather than the specifics of particular experiments). However, the devil often lurks in the details. For example, comparisons between simultaneous and sequential lineups may depend on the position of the suspect in sequential lineups (Carlson, Gronlund, & Clark, 2008; Clark & Davey, 2005; McQuiston-Surrett, Malpass, & Tredoux, 2006). Other theoretical work suggests that probative value differences due to instructions or presentation format may depend on how lineups are constructed (Clark, Erickson, & Breneman, 2011).

More research is needed on the antecedents and consequences of eyewitness identification. The antecedents are concerned with law enforcement decisions that precede the lineup, including decisions about whether to conduct an identification procedure. For example, at what point and with what evidence will police determine that a lineup should be conducted? The decision to conduct a lineup presumably reflects law enforcement's assessment of the likely guilt of the suspect in addition to other factors, such as public pressure to solve crimes. These antecedents are relevant to estimates of the guilty base rate. It is also important to examine the downstream consequences of eyewitness identification outcomes. How do identification outcomes carry forward through the various stages of the criminal justice system, including plea negotiations, admissibility hearings, and trial outcomes? How do the participants in the criminal justice process—police, attorneys, judges, and jurors—distinguish between correct and false identifications to correct errors, convict the guilty, and free the innocent?

Finally, policy decisions regarding eyewitness identification would benefit from a procedural justice analysis. As Tyler and Huo (2002) have stated, when police interact with the public (including eyewitnesses), they "are also engaging in political socialization, that is, they are educating those in their community about the legal system" (p. 130). Thus, these police–witness interactions may have consequences far beyond the identification outcomes in a particular case.

Conclusion

The no-cost view of eyewitness identification reform is contradicted by data. In general (perhaps with one exception for lineup–showup comparisons), procedural changes that reduce the false identification rate also reduce the correct identification rate. Thus, policy decisions regarding the implementation of recommended eyewitness identification procedures are not specified by empirical data and are not easy choices between dominating and dominated alternatives. To the contrary, these policy decisions require the careful consideration of expected outcomes and expected utilities, as well as procedural justice considerations that extend beyond expected outcomes and utilities. These multiple constraints on policy analysis must be considered in a coherent and transparent decision-making framework.

Appendix A

Simultaneous and Sequential Lineups: Inclusions and Exclusions

The present analysis of simultaneous and sequential lineups differs from that of Steblay et al. (2011) regarding the data that were included or excluded. Steblay et al. included a study by Lindsay et al. (1991) that confounded the simultaneous–sequential variable with lineup instructions and lineup composition. Simultaneous lineups were presented with biased instructions and were composed of foils that were "not the best available," whereas sequential lineups were presented with unbiased instructions and were composed of "the best available foils." Because of this confounding, the Lindsay et al. study was not included in the present analysis.

Steblay et al. (2011) also excluded the data from five studies. Two of those studies, by Haw and Fisher (2004) and by Smith, Lindsay, Pryke, and Dysart (2001), appear to have simply been overlooked. Steblay et al. excluded data from three other studies, by Douglass and McQuiston-Surrett (2006), Gronlund, Carlson, Dailey, and Goodsell (2009), and Steblay, Dietrich, Ryan, Raczynski, and James (2011), arguing that the level of performance in each case was at chance or near chance.

Their assessment of chance performance was based on four difference scores: the correct identification rate minus the false identification rate for simultaneous lineups, $(CI - FI)_{SIM}$; the correct identification rate minus the false identification rate for sequential lineups, $(CI - FI)_{SEQ}$; the correct nonidentification rate minus the false nonidentification rate for simultaneous lineups $(CN - FN)_{SIM}$; and the correct nonidentification rate minus the false nonidentification rate for sequential lineups, $(CN - FN)_{SEQ}$. They excluded a study if all four of these differences were less than .10.

However, for two of these studies (Douglass & McQuiston-Surrett, 2006; Steblay, Dietrich et al., 2011) the critical differences actually did exceed the .10 criterion. The critical differences for Gronlund et al. were all smaller than .10, but that does not mean that performance was at or near chance. To make that determination, the relevant chi-square analyses were calculated. The critical differences and chi-square statistics are as follows:

$$(\mathrm{CI}-\mathrm{FI})_{\mathrm{SIM}} = .48 - .40 = .08, \quad \chi^2\,(1, N = 1276) = 7.26, p = .008,$$
$$(\mathrm{CI}-\mathrm{FI})_{\mathrm{SEQ}} = .36 - .30 = .06, \quad \chi^2\,(1, N = 1250) = 5.05, p = .025,$$
$$(\mathrm{CN}-\mathrm{FN})_{\mathrm{SIM}} = .36 - .27 = .09, \quad \chi^2\,(1, N = 1276) = 11.30, p = .0008,$$
$$(\mathrm{CN}-\mathrm{FN})_{\mathrm{SEQ}} = .49 - .41 = .08, \quad \chi^2\,(1, N = 1250) = 9.77, p = .002.$$

The claim that the Gronlund et al. (2009) data were not reliably above chance is clearly contradicted by these analyses. However, although the Gronlund et al. results, collapsed over all conditions, were clearly above chance, many of the conditions did show extremely low performance, such that the perpetrator was identified from the perpetrator-present lineup at a rate less than chance and the correct identification rate was lower than the false identification rate. In the original article, these are referred to as the "guilty–weak" condition and the "innocent–strong" condition. These conditions are the source of the low performance noted by Steblay et al. (2011). Rather than speculate as to why performance was so low in those conditions, my analysis includes data from conditions for which performance was quite good, for which no speculation is required. The data from the guilty–weak and innocent–strong conditions are excluded from the analyses presented in the main text, such that correct identification rates are based on the guilty–strong condition and false identification rates are based on the innocent–weak condition. To assess the effect of the exclusion of the guilty–weak and innocent–strong conditions, an analysis was conducted that included all 12 of the simultaneous–sequential comparisons from the Gronlund et al. study. With the low performance conditions from Gronlund et al. included, the correct identification rates are .50 for simultaneous lineups and .40 for sequential lineups, $t(56) = 4.54$, $p < .001$, and the false identification rates are .20 for simultaneous lineups and .13 for sequential lineups, $t(56) = 4.54$, $p < .001$.

Appendix B

Expected Utility Theory Applied to Eyewitness Identification

In this appendix, I present a more quantitative description of expected utility theory as it is applied to the trade-off between correct identifications lost and false identifications avoided. In theory, one can calculate the long-run expected utility of any two identification procedures A and B. The procedure with the higher long-run expected utility should be preferred. Calculation of expected utility is given as follows.

There are four outcomes of an eyewitness identification procedure, defined by whether the suspect is guilty or innocent and whether that suspect is identified or not identified. Each outcome has a utility. Thus, there are four parameters, $u(\mathrm{CI})$, $u(\mathrm{CN})$, $u(\mathrm{FI})$, and $u(\mathrm{FN})$, that denote the utilities of a correct identification (CI) and a correct nonidentification of the suspect (CN) and the negative utilities, referred to here as the disutilities, of a false identification (FI) and a false nonidentification of the suspect (FN). For these analyses, nonidentifications include both "he's not in the lineup" responses and foil identifications. The expected utility of a suspect identification is given by the sum of the (Probability × Utility) products for each outcome. Thus, for two eyewitness identification procedures, A and B, the expected utilities are given as:

$$E(U)_{\mathrm{A}} = [p(\mathrm{CI_A})\,p(G)\,u(\mathrm{CI})] + [p(\mathrm{FN_A})\,p(G)\,u(\mathrm{FN})] + [p(\mathrm{CN_A})\,p(\mathrm{I})\,u(\mathrm{CN})] + [p(\mathrm{FI_A})p(\mathrm{I})\,u(\mathrm{FI})]$$

and

$$E(U)_{\mathrm{B}} = [p(\mathrm{CI_B})\,p(G)\,u(\mathrm{CI})] + [p(\mathrm{FN_B})\,p(G)\,u(\mathrm{FN})] + [p(\mathrm{CN_B})\,p(\mathrm{I})\,u(\mathrm{CN})] + [p(\mathrm{FI_B})p(\mathrm{I})\,u(\mathrm{FI})]$$

where $p(\mathrm{I}) = 1 - p(G)$, $p(\mathrm{CI})$ and $p(\mathrm{CN})$ refer to the probabilities of correct identifications and correct nonidentifications, and $p(\mathrm{FI})$ and $p(\mathrm{FN})$ denote the probabilities for false identifications and false nonidentifications. These probabilities can be estimated from the correct and false identification rates from eyewitness identification experiments, given in Table 2. This leaves five parameters that must be estimated to calculate the expected utility associated with each procedure. Of course, such a five-parameter model is unwieldy. However, the number of parameters can be reduced from five to two. The utilities associated with each of the four outcomes may be reduced to a single parameter that is the ratio of the two utility differences, which represents the cost of a false identification of an innocent suspect relative to the cost of a missed identification of a guilty suspect. These costs are determined by subtracting the disutility from the utility. For example, the cost of a false identification is $u(\mathrm{CN}) - u(\mathrm{FI})$—that is, the utility of a correct nonidentification of the innocent minus the disutility of a false identification of the innocent. Similarly, the cost of a false nonidentification is given by the difference between the utility of a correct identification minus the disutility of a false nonidentification of the guilty: $u(\mathrm{CI}) - u(\mathrm{FN})$.

Ceci and Friedman (2000) derived a preference rule from the axioms of utility theory that allows one to compare two lineup procedures, A and B, where the correct and false identification rates are lower for B than for A. Ceci and Friedman (2000) showed that procedure B should be preferred over procedure A if and only if the following inequality holds.

$$\frac{[p(\mathrm{FN_B})p(G)] - [p(\mathrm{FN_A})p(G)]}{[p(\mathrm{FI_A})p(I)] - [p(\mathrm{FI_B})p(I)]} < \frac{u(\mathrm{CN}) - u(\mathrm{FI})}{u(\mathrm{CI}) - u(\mathrm{FN})}$$

In the lefthand term, the response proportions $p(\mathrm{FN_B})$ and $p(\mathrm{FN_A})$ denote the false nonidentification rates, and $p(\mathrm{FI_A})$ and

$p(FI_B)$ denote the false identification rates for procedures A and B. These response proportions are multiplied by the appropriate base rates $p(G)$ or $p(I)$.

Although the formal notation may appear somewhat complicated, the meaning is not. The lefthand term represents the number of correct identifications that are lost (numerator) relative to the number of false identifications avoided (denominator). The righthand term represents the relative cost of a false identification compared with a false nonidentification. These cost–benefit functions are shown in Figure 2.

Acknowledgments

The research was supported by Grant 064797 from the National Science Foundation. The author has many people to thank for their comments, suggestions, and readings of earlier drafts: Lauren Bingham, Scott Gronlund, Jonathan Koehler, Richard Leo, Tom Lyon, Jodi Quas, Michael Risinger, Michael Saks, Dan Simon, and Bobbie Spellman. Thanks also to my graduate students Ryan Rush, Rakel Larson, Molly Moreland, and Marie Hicks.

Declaration of Conflicting Interests

The author declared that he had no conflicts of interest with respect to his authorship or the publication of this article.

Notes

1. Wells, Memon, and Penrod (2006) wrote that, "'the might or might not be present' instructions have little effect on accurate identifications. . . . Similarly, the use of a filler-biased lineup has little effect on accurate identifications. . . . Also suggestive influences from lineup administrators appear to have little effect when eyewitnesses view a [lineup that includes the perpetrator]" (p. 63). Wells and Seelau (1995) wrote, "When the culprit is present in the lineup, the rate of accurate identifications is affected little, if any, by the instructions variable, . . . the distractors used, . . . or sequential or simultaneous presentation. . . . Manipulations of instructions, distractors, and presentation procedures have strong effects, however, in cases where the culprit is not in the lineup" (p. 773). Wells et al. (1998) wrote, "We have taken great care to recommend procedures that do not serve to reduce the chances that the guilty party will be identified" (p. 637).

2. Findley (2008) wrote, "Research shows that instructing eyewitnesses that the perpetrator may or may not be in the lineup lowers rates of [false] identifications. . . but has little effect on reducing identifications when the offender is present in the lineup." (p. 18). Garrett (2008) wrote, "research increasingly suggests that procedures such as . . . conducting double blind and sequential eyewitness identifications . . . could have prevented many such costly miscarriages, without reducing correct conviction rates." (p. 122).

3. Some authors have also discussed the financial costs of implementing recommended procedures (Thompson, 2008; Wells, 2006). To avoid confusion, the costs discussed in the present article concern only the loss of correct identifications.

4. The terms *target-present* and *target-absent*, and *culprit-present* and *culprit-absent*, are also sometimes used in the literature.

5. Police also conduct mugshot searches where witnesses may look through a large number of photographs. Mugshot searches differ from showups and lineups in that they may be used when the police have not identified anyone as a suspect. There is far less research on mugshot search procedures than on showups and lineups, and consequently, they are not considered here.

6. Some dependent measures cannot be calculated if the false alarm rate is zero. For these cases non-zero false alarm rates were estimated using the procedure by Murdock and Ogilvie (1968). Also, some studies did not report a false identification rate, but rather reported the total identification rate for perpetrator-absent lineups. For these studies the false identification rate was estimated by dividing the total identification rate by the number of lineup members.

7. The New Jersey Supreme Court may also have had an inflated view of the effects of lineup instructions, noting "In one experiment [Malpass & Devine, 1981], 45% more people chose innocent fillers in target-absent lineups when administrators failed to warn that the suspect may not be there." (p. 54). The 45% increase, distributed across the five members of the lineup, would suggest a 9% increase in the risk of false identification of an innocent suspect. The Court also incorrectly described the comparison as having been between an unbiased instruction and a failure to give an unbiased instruction, rather than between an unbiased instruction and a very biased instruction.

8. The American Bar Association (2004) report noted that, "The vast majority of researchers also conclude that sequential methods result in "little loss" of accuracy when the perpetrator is present." (p. 11). The report cites Saks, Constantine, Dolezal, and Garcia (2001, p. 686), which in turn cites Wells et al. (1998, pp. 639–640). The ABA report continues, "However, what constitutes 'little loss' is debatable." Lindsay et al. (1991) noted that their own experiments "replicated Lindsay and Wells' (1985) finding that sequential lineup presentation decreased false identification of innocent suspects. . ., but did not influence correct identification rates" (p. 800). Devenport, Penrod, & Cutler, (1997) noted that "empirical research has demonstrated that sequentially presented lineups reduce the rate of false identifications without reducing the rate of correct identifications. . ." (p. 345).

9. However, Wells, Steblay, and Dysart (2011), refer to "decades of research showing that the sequential procedure reduces mistaken identifications with little or no reduction in accurate identifications" (p. x).

10. The recommendation regarding foil similarity is often expressed in terms of the match to the description of the perpetrator, rather than the similarity to the suspect (Luus & Wells, 1991; Wells & Seelau, 1995). Discussion of this distinction is beyond the scope of this article, although it should be noted that description-matched foil selection appears to produce an increase in both correct and false identifications, relative to suspect-matched foil selection (see Clark & Godfrey, 2009). Nonetheless, it is useful to consider separately only those studies that varied foil similarity in terms of a description of the perpetrator and set aside those that varied foil similarity based on similarity ratings. Lineups with foils selected to match the description of the perpetrator produced a lower false identification rate—.16 versus .39, $t(10) = 5.58$, $p < .001$—and a lower correct

identification rate—.61 versus .73, $t(10) = 3.06$, $p = .01$—relative to lineups with foils that mismatched the description of the perpetrator.

11. See Russano, Dickinson, Greathouse, & Kovera (2006) for a review of several unpublished studies that showed no differences between blind and nonblind lineup administration. Because there are so few published studies, one unpublished study, by Brower, Godfrey, Clark, and Rosenthal (2009) is included in this analysis.

12. The argument implies an accuracy trade-off between conducting a showup sooner versus conducting a lineup later. This trade-off may apply less often as technology is developed that would allow police officers to compose and present photographic lineups in the field, perhaps within minutes (MacLin, Meissner, & Zimmerman, 2005).

13. In some cases, false identification rates were estimated from the total identification rate, resulting in proportions that do not correspond to whole numbers. However, r can still be calculated by multiplying each probability by 1000 and submitting those counts to a 2×2 χ^2 statistic. The χ^2 statistic is meaningless, of course. However, r as a measure of effect size can be accurately calculated from the χ^2 statistic.

14. The application of detection theory, and the calculation o*f d* and log(β), involves additional assumptions about the eyewitness identification task. In its most constrained form, it is assumed that discrimination—in this case, between guilty and innocent—is based on a single underlying variable. For the present case, a reasonable underlying variable would be the match between lineup members and the witness's memory of the perpetrator (see, e.g., Clark, 2003). For mathematical convenience, it is also often assumed that the underlying distributions are normal with equal variance. The extent to which these assumptions hold for the eyewitness identification task may depend on a number of factors, including the type of identification procedure and the decision rule that witnesses use to make their decisions (see Clark, Erickson, & Breneman, 2011). One may question the calculation of d' and log(β), based on the extent to which these assumptions may not hold, and yet, the insights gained from detection theory application may outweigh the imperfections of its application.

15. Steblay, Dysart, & Wells (2011) correctly note that the *ordering* of probative values is invariant with respect to $p(G)$. Thus, for Procedures A and B, if the probative values for B are greater than those for A for *any* value of $p(G)$, then the probative values of B are greater than those for A for *all* values of $p(G)$. However, this does not mean that $p(G)$ is irrelevant. The costs and benefits, measured in terms of the exchange between correct identifications lost in exchange for false identifications avoided, depend on the value of $p(G)$.

16. Of course, there are higher estimates as well as other very low estimates of the cost ratio. An example at the high end is the 99:1 ratio proposed by Starkie (1876), which is featured prominently in the review by Ceci and Friedman (2000). The context of the proposal suggests that the specific figure of 99 may be a proxy for absolute certainty. "To hold that any finite degree of probability shall constitute proof adequate to the conviction of an offender, would in reality be to assert, that out of some finite number of persons accused, an innocent man should be sacrificed for the sake of punishing the rest; a proposition which is as inconsistent with the humane spirit of our law as it is with the suggestions of reason and justice. The maxim of the law is, that it is better that ninety-nine (i.e., an indefinite number of) offenders should escape, than one innocent man should be condemned" (p. 859).

17. In another line of research, estimates of the cost ratio have been provided by college students and lay people (i.e., the jury pool) rather than legal scholars. These studies consistently show very low cost ratios that are close to 1:1 (see Hastie, 1993, for an excellent review). The comparison between lay estimates and those of legal scholars raises a question as to who should provide the answer to the "should" question regarding the exchange between correct identifications lost versus false identifications avoided.

18. It is often noted that a false acquittal counts as one error, whereas a false conviction counts as two; not only is an innocent person convicted, the actual perpetrator is not convicted. This does not apply as clearly for the pretrial investigation. Not all people who are identified are arrested (Behrman & Davey, 2001). Also, to the extent that the criminal justice system can correct its errors, a false identification does not necessarily imply a false nonidentification. It is safe to say, however, that a false identification reduces the likelihood of a subsequent correct identification, to the extent that the initial error is not detected.

References

American Bar Association. (2004, August). *American Bar Association statement of best practices for promoting the accuracy of eyewitness identification procedures* [Report #111C]. Retrieved from http://meetings.abanet.org/webupload/commupload/CR209700/relatedresources/ABAEyewitnessIDrecommendations.pdf. Washington, DC: Author.

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

Berlin, I. (1958). *Two concepts of liberty: An inaugural lecture delivered before the University of Oxford on 31 October, 1958*. Oxford, UK: Clarendon Press.

Berlin, I. (1969). *Four essays on liberty*. Oxford, England: Oxford University Press.

Biedermann, A., Bozza, S., & Taroni, F. (2008). Decision theoretic properties of forensic identification: Underlying logic and argumentative implications. *Forensic Science International, 177*, 120–132.

Blackstone, W. (1769). *Commentaries on the Laws of England, Vol. II, Book IV*. New York, NY: Duyckinck, Long, Collins & Hannay, and Collins & Co.

Brewer, N., & Palmer, M. A. (2010). Eyewitness identification tests. *Legal and Criminological Psychology, 15*, 77–96. doi:10.1348/135532509X414765

Brewer, N., & 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. doi: 10.1037/1076-898X.12.1.11

Bushway, S. D., & Redlich, A. (2010, November). *The role of evidence in the value of plea bargains.* Conference on Empirical Legal Studies, Yale University, New Haven, CT.

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

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

Ceci, S. J., & Friedman, R. D. (2000). The suggestibility of children: Scientific research and legal implications. *Cornell Law Review, 86*, 33–108.

Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology, 17*, 629–654. doi:10.1002/acp.891

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior, 29*, 395–424. doi:10.1007/s10979-005-5690-7

Clark, S. E., & Davey, S. L. (2005). The target-to-foils shift in simultaneous and sequential lineups. *Law and Human Behavior, 29*, 151–172. doi:10.1007/s10979-005-2418-7

Clark, S. E., Erickson, M. A., & Breneman, J. (2011). Probative value of absolute and relative judgments in eyewitness identification. *Law and Human Behavior, 35*, 364–380. doi:10.1007/s10979-010-9245-1

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

Clark, S. E., Howell, R. T., & Davey, S. L. (2008). Regularities in eyewitness identification. *Law and Human Behavior, 32*, 187–218. doi:10.1007/s10979-006-9082-4

Cole, S. A. (2001). *Suspect identities: A history of fingerprinting and criminal investigation*. Cambridge, MA: Harvard University Press.

Cooley, S. (2007, July). *Letter to Senator Mark Ridley-Thomas, opposing California SB 756*. Retrieved from ccfaj.org/legislation-2007-SB756.html

Cullison, A. D. (1969). Probability analysis of judicial fact-finding: A preliminary outline of the subjective approach. *Toledo Law Review, 1*, 538–598.

Dean, M. E. (2006). An innocent deception: Placebo controls in the St. Petersburg homeopathy trial, 1829-1830. *Journal of the Royal Society of Medicine, 99*, 375–376.

DeKay, M. L. (1996). The difference between Blackstone-like error ratios and probabilistic standards of proof. *Law & Social Inquiry, 21*, 95–132. doi:10.1111/j.1747-4469.1996.tb00013.x

Devenport, J. L., Penrod, S. D., & Cutler, B. L. (1997). Eyewitness identification evidence: Evaluating commonsense evaluations. *Psychology, Public Policy, and Law, 3*, 338–361. doi:10.1037/1076-8971.3.2-3.338

Douglass, A. B., & McQuiston-Surrett, D. (2006). Post-identification feedback: Exploring the effects of sequential photospreads and eyewitnesses' awareness on the identification task. *Applied Cognitive Psychology, 20*, 991–1007. doi:10.1002/acp.1253

Dror, I., & Rosenthal, R. (2008). Meta-analytically quantifying the reliability and biasability of forensic experts. *Journal of Forensic Sciences, 53*, 900–903. doi:10.1111/j.1556-4029.2008.00762.x

Elder, H. W. (1989). Trials and settlements in the criminal courts: An empirical analysis of dispositions and sentencing. *Journal of Legal Studies, 18*, 191–208.

Farmer, J.J., Jr. (2001). *Attorney General guidelines for preparing and conducting photo and live lineup identification procedures*. Retrieved from http://www.state.nj.us/lps/dcj/agguide/photoid.pdf

Findley, K. A. (2008). Toward a new paradigm of criminal justice: How the innocence movement merges crime control and due process. *Texas Tech Law Review, 41*, 1–41.

Friedman, R. D. (1986). A close look at probative value. *Boston University Law Review, 66*, 733–759.

Fulero, S. M. (2009). System and estimator variables in eyewitness identification: A review. In D. A. Krauss & J. D. Lieberman (Eds.), *Psychological expertise in court* (pp. 57–78). Burlington, VT: Ashgate.

Garrett, B. (2008). Judging innocence. *Columbia Law Review, 108*, 55–142.

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. doi:10.1007/s10979-008-9136-x

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

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

Gross, S. R. (1987). Loss of innocence: Eyewitness identification and proof of guilt. *Journal of Legal Studies, 16*, 395–453.

Gross, S. R., Jacoby, K., Matheson, D. J., Montgomery, N., & Patil, S. (2005). Exonerations in the United States 1989 through 2003. *Journal of Criminal Law and Criminology, 95*, 523–560.

Hastie, R. (1993). Algebraic models of juror decision processes. In R. Hastie (Ed.), *Inside the juror: The psychology of juror decision making* (pp. 84–115). New York, NY: Cambridge University Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*, 486–504. doi:10.1037/1082-989X.3.4.486

Hope, L. (2010). Eyewitness testimony. In G. J. Towl & D. A. Crighton (Eds.), *Forensic psychology* (pp. 161–177). Chichester, UK: Wiley & Sons.

Kaplan, J. (1968). Decision theory and the fact-finding process. *Stanford Law Review, 20*, 1065–1092.

Kaye, D. H. (1986). Quantifying probative value. *Boston University Law Journal, 66*, 761–766.

Kaye, D. H., & Koehler, J. J. (2003). The misquantification of probative value. *Law and Human Behavior, 27*, 645–659. doi:10.1023/B:LAHU.0000004892.94380.88

Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*. New York, NY: Wiley.

Koehler, J. J. (1996). On conveying the probative value of DNA evidence; Frequencies, likelihood ratios and error rates. *University of Colorado Law Review, 67*, 859–886.

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

Laudan, L., & Saunders, H. D. (2009). Re-thinking the criminal standard of proof: Seeking consensus about the utilities of trial outcomes. *International Commentary on Evidence, 7*, 1–34.

Leippe, M. R., Eisenstadt, D., & Rauch, S. M. (2009). Cueing confidence in eyewitness identifications: Influence of biased lineup instructions and pre-identification memory feedback under varying lineup conditions. *Law and Human Behavior, 33*, 194–212. doi:10.1007/s10979-008-9135-y.

Lempert, R. O. (1977). Modeling relevance. *Michigan Law Review, 75*, 1021–1057.

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

Lillquist, E. (2007). Improving accuracy in criminal cases. *University of Richmond Law Review, 41*, 897–933.

Lindsay, R. C. L., Lea, J. A., Nosworthy, G. J., Fulford, J. A., Hector, J., LeVan, V., & Seabrook, C. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology, 75*, 791–802.

Lindsay, R. C. L., Mansour, J. K., Beaudry, J. L., Leach, A., & Bertrand, M. I. (2009). Sequential lineup presentation: Patterns and policy. *Legal and Criminological Psychology, 14*, 13–24. doi:10.1348/135532508X382708

Lindsay, R. C. L., & Pozzulo, J. D. (1999). Sources of eyewitness identification error. *International Journal of Law and Psychiatry, 22*, 347–360. doi:10.1016/S0160-2527(99)00014-X

Lindsay, R. C. L, & Wells, G. L. (1980). What price justice? Exploring the relationship of lineup fairness to identification accuracy. *Law and Human Behavior, 4*, 303–313. doi:10.1007/BF01040622

Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70*, 556–564. doi:10.1037/0021-9010.70.3.556

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information in a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 19–31. doi:10.1037/0278-7393.4.1.19

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

Lyon, T. D., & Koehler, J. J. (1996). The relevance ratio: Evaluating the probative value of expert testimony in child sexual abuse cases. *Cornell Law Review, 82*, 43–78.

MacLin, O. H., Meissner, C. A., & Zimmerman, L. A. (2005). PC eyewitness: A computerized framework for the administration and practical application of research in eyewitness psychology. *Behavior Research Methods, 37*, 324–334. doi:10.3758/BF03192700

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

Malpass, R. S. (2006). A policy evaluation of simultaneous and sequential lineups. *Psychology, Public Policy, and Law, 12*, 394–418. doi:10.1037/1076-8971.12.4.394

McQuiston-Surrett, D., Malpass, R. S., & Tredoux, C. G. (2006). Sequential vs. simultaneous lineups: A review of methods, data, and theory. *Psychology, Public Policy, and Law, 12*, 137–169. doi:10.1037/1076-8971.12.2.137

Memon, A., & Higham, P. A. (1999). A review of the cognitive interview. *Psychology, Crime & Law, 5*, 177–196. doi:10.1080/10683169908415000

Milanich, P. G. (1981). Case/comment: Decision theory and standards of proof. *Law and Human Behavior, 5*, 87–96.

Miller, L. S. (1984). Bias among forensic document examiners: A need for procedural changes. *Journal of Police Science and Administration, 12*, 407–411.

Murdock, B. B., & Ogilvie, J. C. (1968). Binomial variability in short-term memory. *Psychological Bulletin, 70*, 256–260. doi:10.1037/h0026259

Packer, H. L. (1964). Two models of the criminal process. *University of Pennsylvania Law Review, 113*, 1–68.

Penrod., G. (2007). *Letter to California Senator Mark Ridley-Thomas, by the San Bernardino County Sheriff, opposing California Senate Bill 756*. Retrieved from ccfaj.org/legislation-2007-SB756.html.

Risinger, D. M. (2007). Innocents convicted: An empirically justified factual wrongful conviction rate. *Journal of Criminal Law and Criminology, 97*, 761–806.

Risinger, D. M., Saks, M. J., Thompson, W. C., & Rosenthal, R. (2002). The Daubert/Kumho implications of observer effects in forensic science: Hidden problems of expectation and suggestion. *California Law Review, 90*, 1–56.

Rivers, W. H. R., & Webber, H. N. (1907). The action of caffeine on the capacity for muscular work. *Journal of Physiology, 36*, 33–47.

Rosenthal, R., & Rosnow, R. L. (2008). *Essentials of behavioral research: Methods and data analysis* (3rd ed.). New York, NY: McGraw-Hill.

Rosnow, R. L., & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology, 57*, 221–237. doi:10.1037/h0087427

Rush, R. A., & Clark, S. E. (2010). *Social contagion of correct and incorrect information in memory*. Unpublished manuscript.

Russano, M. B., Dickinson, J. J., Greathouse, S. M., & Kovera, M. B (2006). Why don't you take another look at number three? Investigator knowledge and its effects on eyewitness confidence and identification decisions. *Cardozo Public Law, Policy, and Ethics Journal, 4*, 355-379.

Russano, M. B., Meissner, C. A., Narchet, F. M., & Kassin, S. M (2005). Investigating true and false confessions within a novel experimental paradigm. *Psychological Science, 16*, 481–486.

Saks, M. J., Constantine, L., Dolezal, M., & Garcia, J. (2001). Model prevention and remedy of erroneous convictions act. *Arizona State Law Journal, 33,* 669–718.

Saks, M. J., & Faigman, D. L. (2008). Failed forensics: How forensic science lost its way and how it might yet find it. *Annual Review of Law and Social Science, 4*, 149–171.

Savage, L. J. (1954). *The foundations of statistics*. New York, NY: Wiley.

Scheck, B. C., Edwards, E. R., & McNamara, J. M. (2010). *Eyewitness identification science post-Manson: An explosion of robust and rigorous research*. State of New Jersey v. Larry R. Henderson, Remand Hearing.

Sherman, L. (1984). Experiments in police discretion: Scientific boon or dangerous knowledge? *Law and Contemporary Problems, 47*(4), 61–81.

Sherman, L. (1998). *Evidence-based policing: Ideas in American policing*. Washington, DC: Police Foundation.

Starkie, T. (1876). *A practical treatise on the law of evidence*. Philadelphia, PA: Johnson.

State of New Jersey v. Larry R. Henderson, A-8–08 (2010).

Steblay, N. M. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior, 21*, 283–297. doi:10.1023/A:1024890732059

Steblay, N. K., Dietrich, H. L., Ryan, S. L., Raczynski, J. L., & James, K. A. (2011). Sequential lineup laps and eyewitness accuracy. *Law and Human Behavior, 35*, 252–274. doi:10.1007/s10979-010-9236-2

Steblay, N. K., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 25*, 459–473. doi:10.1023/A:1012888715007

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law, 17*, 99–139. doi:10.1037/a0021650

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

Sunshine, J., & Tyler, T. R. (2003). The role of procedural justice and legitimacy in shaping public support for policing. *Law and Society Review, 37*, 555–589.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*, 1–25. doi:10.1111/1529-1006.001

Thatcher, D. (2001). Policing is not a treatment: Alternatives to the medical model of police research. *Journal of Research in Crime & Delinquency, 38*, 387–415.

Thibaut, J. W., & Walker, L. (1975). *Procedural justice: A psychological analysis*. Hillsdale, NJ: Erlbaum.

Thibaut, J., & Walker, L. (1978). A theory of procedure. *California Law Review, 66*, 541–566.

Thibaut, J., Walker, L., LaTour, S., & Houlden, P. (1974). Procedural justice as fairness. *Stanford Law Review, 26*, 271–1289.

Thompson, S. G. (2008). What price justice? The importance of costs to eyewitness identification reform. *Texas Tech Law Review, 41*, 33–63.

Thompson-Cannino, Cotton, R., & Torneo, E. (2009). *Picking cotton: Our memoir of injustice and redemption*. New York, NY: St. Martin's Press.

Tribe, L. H. (1971). Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review, 84*, 1329–1393.

Tyler, T. R. (1990). *Why people obey the law*. New Haven, CT: Yale University Press.

Tyler, T. R., & Fagan, J. (2008). Legitimacy and cooperation: Why do people help the police fight crime in their communities? *Ohio State Journal of Criminal Law, 6*, 231–275.

Tyler, T. R., & Huo, Y. J. (2002). *Trust in the law: Encouraging public cooperation with the police and courts*. New York, NY: Russell-Sage Foundation.

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

U.S. House of Representatives. (2008). *Federal rules of evidence*. Washington, DC: Author.

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

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Wagenaar, W. A., & Veefkind, N. (1992). Comparison of one-person and many-person lineups: A warning against unsafe practices. In F. Losel, D. Bendre, & T. Bliesner (Eds.), *Psychology and law: International perspectives* (pp. 275–285). Berlin, Germany: Walter De Gruyter.

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14*, 89–103. doi:10.1111/j.1559-1816.1984.tb02223.x

Wells, G. L. (2006). Eyewitness identification: Systemic reforms. *Wisconsin Law Review, 2*, 615–643.

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin, 88*, 776–784. doi:10.1037/0033-2909.88.3.776

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest, 7*, 45–75. doi:10.1111/j.1529-1006.2006.00027.x

Wells, G. L., & Olson, E. A. (2002). Eyewitness identification: Information gain from incriminating and exonerating behaviors. *Journal of Experimental Psychology: Applied, 8*, 155–167. doi:10.1037/1076-898X.8.3.155

Wells, G. L., & Seelau, E. P. (1995). Eyewitness identification: Psychological research and legal policy on lineups. *Psychology, Public Policy, and Law, 1,* 765–791. doi:10.1037/1076-8971.1.4.765

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. doi:10.1023/A:1025750605807

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2011). *A test of the simultaneous vs. sequential lineup methods: An initial report of the AJS national eyewitness identification field studies.* Retrieved from http://www.psychology.iastate.edu/~glwells/Wells_articles_pdf/EWID_PrintFriendly.pdf

Wickens, T. D. (2002). *Elementary signal detection theory*. New York, NY: Oxford.

Wisconsin v. Dubose. (2005). 272 Wis. 2d 856, 679 NW 2d 927.

Wright, D. B., & McDaid, A. T. (1996). Comparing system and estimator variables using data from real line-ups. *Applied Cognitive Psychology, 10*, 75–84. doi:10.1348/000712600161781

Wright, D. B., & Skagerberg, E. M. (2007). Postidentification feedback affects real eyewitnesses. *Psychological Science, 18*, 172–178. doi:10.1111/j.1467-9280.2007.01868.x

Yarmey, A. D., Yarmey, M. J., & Yarmey, A. L. (1996). Accuracy of eyewitness identification in showups and lineups. *Law and Human Behavior, 20*, 459–477. doi:10.1007/BF01498981