Comparing the Effectiveness of *Henderson* Instructions and Expert Testimony: Which Safeguard

Improves Jurors' Evaluations of Eyewitness Evidence?

by

Angela M. Jones[1]

Amanda N. Bergold[2]

Marlee Kind Dillon[2]

Steven D. Penrod[2]

[1]Texas State University

[2]John Jay College & the Graduate Center, CUNY

Address correspondence to Angela Jones, Department of Criminal Justice, Texas State University, 601 University Dr., San Marcos, TX 78666. E-mail: a_j194@txstate.edu

Angela M. Jones is an Assistant Professor in the School of Criminal Justice at Texas State University. Her research examines factors contributing to wrongful convictions, including eyewitness identification and false confession, and the effectiveness of safeguards designed to reduce wrongful convictions.

Amanda Nicholson Bergold is a Quattrone Center Fellow at the University of Pennsylvania Law School. Her research explores the role of biased decision-making in the legal system. Specifically, she researches jury decision-making and eyewitness decisions.

Marlee Kind Dillon is a Consultant at DOAR Litigation Consulting in NYC. She received a PhD from John Jay College and the Graduate Center, CUNY where she conducted research on jury decision making, jury instructions, and eyewitness identification. At DOAR, she specializes in pre-trial research and jury consulting.

Steven Penrod is Distinguished Professor of Psychology at John Jay College of Criminal Justice--CUNY. His research examines procedural, evidentiary, and extra-legal influences on jury decision-making, sources of eyewitness identification error, and plea bargaining.

Abstract

Objectives: The New Jersey Supreme Court recently determined that jurors may not be able to effectively evaluate eyewitness evidence on their own. As a result, the Court proposed the use of judicial instructions to assist jurors (called *Henderson* instructions) and suggested the implementation of these instructions would reduce the need for expert testimony. We tested the efficacy of these instructions compared to alternative instructions and expert testimony.

Methods: We utilized a mock trial paradigm, randomly assigning 452 participants to one of 20 trial conditions that varied the quality of eyewitness evidence (both witnessing and identification conditions) and the type of safeguard presented during the mock trial.

Results: Jurors were sensitive to the quality of police procedures on their own. Jurors were more likely to convict when police practices were good and less likely when police practices were poor. This relationship was mediated by eyewitness credibility ratings. Expert testimony resulted in skepticism by reducing the likelihood that jurors would convict regardless of the quality of witnessing and identification conditions. No variation of the instructions influenced verdicts.

Conclusions: While jurors were sensitive to the quality of identification conditions on their own, we observed no such effect for the quality of witnessing conditions, even with the aid of instructions and/or expert testimony. Both *Henderson* instructions and expert testimony may be insufficient for assisting jurors to effectively evaluate problematic witnessing conditions. Future research should examine the use of alternative safeguards.

Keywords. Expert testimony, judicial instructions, juror decision making, eyewitness identification, New Jersey Supreme Court

Comparing the Effectiveness of *Henderson* Instructions and Expert Testimony: Which Safeguard

Improves Jurors' Evaluations of Eyewitness Evidence?

Among documented exonerations in the United States, erroneous eyewitness

identifications have contributed to 30% of all wrongful convictions and 69% of wrongful

convictions for adult sexual assault cases (National Registry of Exonerations, 2016).[1] The New

Jersey Supreme Court recently conveyed their concern about eyewitness misidentifications

stating, "Recent studies-ranging from analyses of actual police lineups, to laboratory

experiments, to DNA exonerations—prove that the possibility of mistaken identification is real,

and the consequences severe (p. 881)." Troubled by these findings, the Court addressed issues

with the current standard for identification admissibility as well as in-court procedures dealing

with eyewitness evidence (*New Jersey v Henderson*, 2011). The *Henderson* Court emphasized

the important role of jurors in evaluating eyewitness identifications. To assist jurors, the Court

commissioned new eyewitness instructions. The focus of the current study is to test the

effectiveness of these instructions for drawing attention to the quality of witnessing and

identification conditions.

The *Henderson* Court stated, "We are convinced from the scientific evidence in the

record that memory is malleable, and that an array of variables can affect and dilute memory and

lead to misidentifications (p. 872)." In formulating their opinion, the *Henderson* Court relied on

the growing body of eyewitness identification research published in the past 40 years,

recognizing that factors both within (i.e., system variables) and beyond the control of the

criminal justice system (i.e., estimator variables) can influence the accuracy of eyewitness

identification (Wells, 1978; Wells & Olson, 2003; Wells, Memon, & Penrod, 2006). The current

study assesses jurors' ability to evaluate six of these estimator and system variables. For

example, the Court identified estimator variables such as short periods of exposure to a perpetrator and long delays between witnessing a crime and making an identification as reducing both the quality of an eyewitness' memory for an event and identification accuracy (see Deffenbacher, Bornstein, McGorty, & Penrod, 2008; Shapiro & Penrod, 1986). Similarly, the Court noted that weapon presence can also disrupt memory processes and decrease identification accuracy (see Fawcett, Russell, Peace, & Christie, 2013).

Aside from estimator variables, the New Jersey Supreme Court gave particular weight to system variables, as they are within the State's control (*New Jersey v. Henderson*, 2011). The Court pointed out that the showup is a suggestive type of identification procedure used by police where an eyewitness is aware that only one person will be shown to them (see Lindsay & Wells, 1985). The use of showups reduces identification accuracy relative to multi-person lineups (Steblay, Dysart, Fulero, & Lindsay, 2003). The Court also emphasized that failing to inform an eyewitness that the perpetrator may or may not be present in the lineup (i.e., biased lineup instructions) decreases identification accuracy compared to unbiased instructions (see Steblay, 1997; see also Rose, Bull, & Vrij, 2005). Finally, the Court indicated confirmatory feedback by police can falsely inflate an eyewitness' confidence, as well as self-reported view and attention paid during the crime (Steblay, Wells, & Douglass, 2014). All of the Court's observations were based on a report by a Special Master assigned to listen to the testimony of eyewitness identification experts (most of whom have been cited in this paper).

Eyewitness Identification Knowledge

While a large literature has established the impact of various system and estimator variables, juror knowledge of these factors is somewhat limited (see Leippe & Eisenstadt, 2009 for a review of studies). Indeed, the number of wrongful convictions due to misidentifications

suggests that the process of evaluating eyewitness evidence is not an easy one for jurors (Devenport, Cutler, & Penrod, 1997). A recent meta-analysis assessing lay knowledge concluded that 75% of 16 factors known to influence eyewitness identification accuracy are not common sense to jurors (Desmarais & Read, 2011; Benton, Ross, Bradshaw, Thomas, & Bradshaw, 2006; see also Houston, Hope, Memon, & Read, 2013 for performance of Scottish jurors, which depends on question format).

Beyond possessing knowledge of factors that influence identification accuracy, jurors must also be able to integrate such knowledge into their decision making. Several studies have demonstrated that jurors are capable of such a task. For example, mock jurors can distinguish between foil-biased and unbiased lineups (rating the former as more suggestive; Devenport, Stinson, Cutler, & Kravitz, 2002), biased and unbiased lineup instructions (reducing convictions with the former; Cutler, Penrod, & Stuve, 1988), and the number of perpetrator description errors an eyewitness makes (reducing eyewitness credibility as errors increase; Bruer & Pozzulo, 2014). These studies and others (e.g., Cowan, Read, & Lindsay, 2014; Neal, Christiansen, Bornstein, & Robicheaux, 2012) suggest jurors are somewhat sensitive to the quality of witnessing and identification conditions on their own. In contrast, others have found jurors unable to take into account system or estimator variables, such as disguise, weapon focus, or lineup size (Cutler et al., 1988; Cutler, Dexter, & Penrod, 1990). Eyewitness factors in these studies did not influence verdict decisions or perceptions that the identification was correct. Thus, jurors may need additional education to improve their ability to integrate knowledge concerning eyewitness factors into relevant legal decisions.

Procedural Safeguards

Procedural safeguards, such as judicial instructions and expert testimony, may provide the additional education jurors need to make informed decisions. Courts tend to prefer judicial instructions as they are less costly and time consuming compared to experts (Sheehan, 2011; see also Hoffheimer, 1989). To be effective, jurors must be willing and able to follow the instructions ([Devine, Clayton, Dunford, Seying, & Pryce, 2001](#)). Yet, much research attests to the difficulty jurors have in understanding and applying many types of judicial instructions, such as pattern instructions (Reifman, Gusick, & Ellsworth, 1992), those to disregard or limit evidence (Steblay, Hosch, Culhane, & McWethy, 2006), and instructions on eyewitness identification (Greene, 1988; see Bornstein & Hamm, 2012 for review). Overall comprehension of pattern instructions is rarely above chance levels (Reifman et al., 1992; see also Lieberman & Sales, 1997), even with the aid of deliberation (Ellsworth, 1989; see also Rose & Ogloff, 2001).

To improve jurors' comprehension of instructions, a few techniques have been employed. For example, while still low, comprehension rates do appear to increase when instructions are simplified, including death penalty instructions (Frank & Applegate, 1998), pattern instructions (Severeance & Loftus, 1982), civil jury instructions (Charrow & Charrow, 1979), and eyewitness instructions (Greene, 1988). Additionally, providing *written copies* of (simplified) death penalty instructions increases juror comprehension (Frank & Applegate, 1998). Others have not observed improved comprehension with written pattern instructions, though participants in a field study reported being more satisfied (compared to only oral instructions; Heuer & Penrod, 1989).

To be effective, safeguards should maximize juror sensitivity to factors that influence the accuracy of an identification (i.e., the ability to discriminate between good and bad witnessing and identification conditions) without inducing a general sense of skepticism regarding the ability of all eyewitnesses to make correct identifications (Cutler, Dexter, & Penrod, 1989a).

However, the timing of instructions may impact its effectiveness. In many states, such as New Jersey (*New Jersey v. Henderson,* 2011), judges have discretion on when to give instructions (although they are typically administered at the conclusion of the trial after all evidence has been presented; Ogloff & Rose, 2005). However, some evidence suggests that instructions are most effective prior to introducing evidence. For example, instructions on reasonable doubt and burden of proof, when given at the outset of the trial, increase jurors' threshold for conviction (Kassin & Wrightsman, 1979) and assist jurors in evaluating evidence in line with legal standards (Heuer & Penrod, 1989). Eyewitness instructions given prior to an eyewitness' testimony may similarly assist jurors to evaluate such evidence.

Aside from timing, the content of instructions can also influence its impact on legal decisions. One type of eyewitness instruction, *Telfaire*, is the most commonly used instruction cautioning jurors about the problems with eyewitness identification (*US v. Telfaire,* 1972). The *Telfaire* instruction has either failed to affect jurors' verdicts or simply resulted in skepticism (Cutler et al., 1990a; Greene, 1988), possibly due to several limitations. First, jurors are not provided with an explanation as to how memory works nor how the factors outlined in the instruction influence eyewitness accuracy. Second, the eyewitness factors outlined in the instruction are not tailored to the facts of the case and are not based on psychological research, but on previous case law (Cutler & Penrod, 1995). Moreover, suggestive system variables (e.g., confirmatory feedback) can influence three of the estimator variables identified in the instruction, making them unreliable indicators of identification accuracy (Wells & Quinlivan, 2009). However, a modification of the *Telfaire* instructions induced sensitivity to extreme witnessing and identification conditions (Ramirez, Zemba, & Geiselman, 1996). Researchers expanded the number of factors listed in the instruction from five to thirteen and then

subsequently asked mock jurors to evaluate an identification made under extremely good or extremely poor witnessing and identification conditions. By collapsing the quality of witnessing and identification conditions, it is unclear whether the instructions safeguard sensitized jurors to system variables, estimator variables, or both.

Recently, the New Jersey Supreme Court ruled that, in *all* cases involving eyewitness testimony, new instructions addressing case-specific factors must be presented to the jury (*New Jersey v. Henderson*, 2011). Only estimator and system variables that a judge deems problematic in a specific case would be included in the instruction to jurors.[2] The *Henderson* instructions address all of *Telfaire*'s shortcomings by outlining the three stages of memory (i.e., acquisition, retention, and retrieval) and explaining how case-specific factors based on sound psychological research influence eyewitness accuracy (see New Jersey Supreme Court, 2012a with link to full instructions).

New Jersey Chief Justice Stuart Rabner stated, "The [*Henderson*] instructions are designed to minimize the risk of wrongful convictions and help jurors reach informed, just decisions" (New Jersey Supreme Court, 2012b). An initial evaluation suggests *Henderson* instructions may minimize the risk of wrongful convictions, but may not provide the educational function that results in sensitivity to the quality of eyewitness factors. That is, mock jurors were less likely to convict when given the *Henderson* instructions, regardless of whether the identification in question was secured using good or poor police practices (e.g., lineup instructions; Papailiou, Yokum, & Robertson, 2014). While the Papailiou et al. (2014) study sheds some light on the potential impact of *Henderson*, there are a number of noteworthy limitations. First, estimator variables were not manipulated in this study. Thus, it is unclear whether the *Henderson* instructions sensitize jurors to the quality of witnessing conditions.

Second, participants were given *Henderson* instructions in their entirety, not one tailored to the

facts of the case as intended by the New Jersey Supreme Court. Perhaps not surprisingly, these

instructions had a similar effect as other general eyewitness instructions, such as *Telfaire* (see

Greene, 1988). Finally, the lack of a control group prevents evaluating the true effect of the

Henderson instructions relative to their absence. The current study addresses all of these

shortcomings by manipulating both system and estimator variables, tailoring instructions (and

expert testimony) to the facts of the case, and including a control group.

  In contrast to the skepticism effects observed with *Telfaire* (Greene, 1988) and (the

general version of) *Henderson* (Papailiou et al., 2014), the interview-identification-eyewitness (I-

I-Eye) teaching aid, may provide a promising avenue for sensitizing jurors to the quality of

eyewitness identifications (Pawlenko, Safer, Wise, & Holfeld, 2013). This aid relies on

established psychological research findings concerning police interviews with eyewitnesses as

well as the quality of estimator and system variables *not specific* to a case. Similar to the

Ramirez et al. (1996) study, Pawlenko et al. (2013) also created two extreme conditions under

which the identification occurred (i.e., a single block of 11 system variables was manipulated to

create one good and one poor identification condition). Compared to participants receiving

general information about the trial process or the *Biggers* criteria (contained in the *Telfaire*

instruction, see *Neil v. Biggers*, 1972), those who received the I-I-Eye teaching aid demonstrated

increased sensitivity to system variable manipulations (e.g., confirmatory feedback; Pawlenko et

al., 2013). Participants were more likely to convict when the eyewitness evidence was strong and

less likely to convict when the eyewitness evidence was weak *only* with the I-I-Eye. Given that

past research on eyewitness instructions, including the I-I-Eye, has only examined their impact

on jurors' ability to evaluate system variables alone (e.g., lineup procedures; Papailiou,

2014; Pawlenko et al., 2013) or a single combination of system and estimator variables (e.g., Ramirez et al., 1996), additional studies are needed to establish whether instructions can sensitize jurors independently to estimator and system variables.

Similar to jury instructions, research on the effectiveness of expert testimony is mixed. For example, some research suggests expert testimony sensitizes jurors to the problems of weapon focus and witness confidence (Cutler et al., 1989a), as well as biased lineup instructions (Devenport et al., 2002). However, others have found null effects of expert testimony concerning biased lineup instructions and foil bias (Devenport & Cutler, 2004) or skepticism effects for identifications that varied in a number of eyewitness factors (i.e., disguise, weapon, time delay, and lineup instructions; Cutler et al., 1990a).

The New Jersey Supreme Court opined that new eyewitness instructions will reduce the "need for expert testimony. Jury charges offer a number of advantages: they are focused and concise, authoritative (in that juries hear them from the trial judge, not a witness called by one side), and cost-free; they avoid possible confusion to jurors created by dueling experts; and they eliminate the risk of an expert invading the jury's role or opining on an eyewitness' credibility" (*New Jersey v. Henderson,* 2011, p. 998). While the opinion in *Henderson* suggested that the use of case-specific eyewitness instructions might diminish the need for expert testimony, only one study (that we are aware of) has directly compared the two. Cutler et al. (1990) found that (court-appointed) expert testimony produced skepticism towards identifications of varying quality, while *Telfaire* instructions had no impact on jurors' decisions (neither of which is ideal). Furthermore, no research has examined the combination of both instructions and expert testimony. In particular, it is unknown whether expert testimony is equivalent or superior to

Henderson instructions or if the combination of both could improve juror sensitivity (or enhance

skepticism) above and beyond that of either safeguard alone.

Sensitivity effects found with expert testimony may be due to the unique way (relative to

other safeguards) experts educate jurors on eyewitness issues (Desmarais & Read, 2011). An

expert may testify about the reliability of eyewitness identifications in general or speak about

specific factors that threaten the identification in question, though evidence is mixed on which is

more effective (see Cutler, Penrod, & Dexter, 1989b; Fox & Walters, 1986). Cutler and

colleagues suggested that jurors might assume the relationships between various eyewitness

factors and accuracy are larger than they actually are when evaluating this evidence on their own.

However, Leippe and Eisenstadt (2009) propose specific testimony might be more effective than

general testimony because such testimony provides a clear application to the case at hand. Yet, it

is unknown how this specific type of testimony, presented via judicial instructions, might impact

jurors' decisions. Some legal scholars have criticized the lack of effect sizes in *Henderson*

instructions, suggesting such information is necessary to properly educate the jury (Mnookin,

2015).

Considering the Court's embrace of psychological research-based instructions, which

read much like expert testimony, we surmised that the Court would welcome an effective means

of delivering this information to jurors, even if it varied slightly from the original format.

Furthermore, the success of the I-I-Eye (Pawlenko et al., 2013) indicates that instructions can be

an effective mechanism for conveying research findings to jurors without the need for expert

testimony, which is clearly the Court's preference. The *Henderson* Court did not, however, rule

out the use of expert testimony in conjunction with the current instructions, which suggests

advantages to providing jurors with additional research information. Thus, an enhanced version

of the *Henderson* instructions—one that includes effect sizes as advocated by some legal

scholars (Mnookin, 2015)—might accomplish the sensitivity effects observed with expert

testimony (but not found with the original, general version of *Henderson* instructions; see

Papailiou et al., 2014) while also appeasing the logistical concerns of the Court.

Present Study

The potential success of instructions and expert testimony on eyewitness identification

depends on whether these safeguards can assist jurors in developing a cohesive story. According

to the Story Model, story construction is central to jurors' decisions, essentially determining

one's verdict choice (Pennington & Hastie, 1992). If more than one story is constructed, jurors

will choose the story that has the best explanation, and is the most consistent, complete,

plausible, and unique (Levett, Danielsen, Kovera, & Cutler, 2005). The addition of either

instructions or expert testimony can guide jurors in determining which story has these

characteristics. It is less clear, however, whether hearing *both* instructions and expert testimony

would provide any additional benefit over either safeguard alone.

The *Henderson* Court anticipates that new instructions will sensitize jurors to the quality

of factors known to increase the risk of misidentification. However, initial findings suggest the

current *Henderson* instructions are not as effective as intended (see Papailiou et al., 2014). One

potential way to improve the effectiveness of the *Henderson* instructions would be to convey

information about the magnitude of the effect eyewitness variables have on performance. Such

research (typically meta-analyses) is often relied upon by experts and might explain sensitivity

effects observed in cases with expert testimony. Thus, we hypothesized the highest levels of

comprehension for eyewitness factors among those who hear either enhanced *Henderson*

instructions or expert testimony, followed by the original *Henderson* instructions (Hypothesis 1).

Accordingly, we hypothesized the greatest sensitivity to witnessing and identification conditions with the presentation of either enhanced *Henderson* instructions or expert testimony, followed by the original *Henderson* instructions (Hypothesis 2). An interaction between the quality of witnessing and identification conditions and safeguard type provides evidence of a sensitivity effect, such that convictions are lowest when witnessing and identification quality are poor and highest when witnessing and identification quality are good *only* when jurors hear instructions or expert testimony (relative to the absence of any safeguard).

We anticipated conditional indirect effects through witness ratings and belief in the eyewitness (Hypothesis 3), such that when jurors hear instructions or expert testimony and witnessing and identification conditions are good, they will increase credibility ratings for witnesses and their belief in the eyewitness, which in turn, will increase convictions. In contrast, when witnessing and identification conditions are poor, jurors will decrease credibility ratings for witnesses and their belief in the eyewitness, which in turn, will decrease convictions. We will explore the impact of the combination of expert testimony and instructions on jurors' decisions. Finally, as research is mixed concerning sample differences in legal comprehension and decisions (e.g., Berman & Cutler, 1996; Bornstein, 1999; Desmarais & Read, 2011; MacCoun & Kerr, 1988; Cutler, Penrod, & Dexter, 1990; Rose & Ogloff, 2001; see also Wiener, Krauss, & Lieberman, 2011 for overview of validity threats in mock jury research), we will compare the responses of a commonly used sample, students, to a more externally valid sample of community members.

Methods

Participants

We recruited a total of 452 jury-eligible participants (209 community members and 243 undergraduates) for this study. Community members were paid $20, while students received course credit for their participation. Overall, 61% of participants were female and averaged 27.38 years old (SD = 12. 47). This was a diverse sample (30% White, 29% Hispanic, 23% African American, 17% Other). Ethnicity did not vary by sample type. However, community members were significantly older (M= 35.71, SD = 13.85) than the student sample (M= 20.23, SD = 3.69), $t(233) = -15.68$, $p < .001$, and included significantly fewer women (52% vs. 69%), $t(428) = -3.83$, $p < .001$.

Design and Materials

This study utilized a 2 (Quality of witnessing conditions: good vs. poor) X 2 (Quality of identification conditions: good vs. poor) X 5 (Safeguard type: none vs. *Henderson* instructions only vs. enhanced instructions only vs. expert testimony only vs. *Henderson* instructions + expert testimony) factorial design.

Voir dire. Prior to viewing the trial film, participants completed a short voir dire questionnaire. These items included gender, age, citizenship, voter registration, ethnicity, political views, and jury duty history.

Trial stimulus. Mock jurors viewed a criminal trial simulation in which the prosecution alleged that the defendant committed an attempted rape. The mock trial for the control condition was 40 minutes long, while experimental conditions ranged from 55 - 75 minutes.[3] Expert testimony and instructions were of comparable lengths. This trial was based on the overturned case, *Kentucky v. Gregory* (1993). The Innocence Project worked to secure the defendant's exoneration and provided transcripts of the original trial to aide in developing the trial stimulation used for the current study.

Researchers filmed the trial stimulation in a mock courtroom using hired actors. All versions of the trial stimulations included New York standard judicial instructions, opening and closing statements from both attorneys, and direct and cross examination of the detective, eyewitness, serologist, defendant, and alibi witness. In addition, the inclusion of *Henderson* eyewitness instructions, enhanced *Henderson* eyewitness instructions, and/or expert testimony varied across trial stimulations. Any time instructions were included, the judge read them aloud prior to the eyewitness' testimony. In expert present conditions, the expert was the last witness to testify.

Witnessing and identification conditions were manipulated within the content of the detective and eyewitness' testimony and, when relevant, the expert's testimony. As cross race identification and stress were integral to the facts of the actual case, these factors remained present in all trial versions. Pilot testing of 375 participants indicated a sufficient verdict split between good and poor witnessing and identification conditions, even with cross race identification and stress as constants across trial stimulations (50% and 25% convictions, respectively). Furthermore, the manipulated system and estimator variables were chosen for three reasons: they were among the variables included in the *Henderson* instructions, much research attests to their impact on identification accuracy and confidence (see Bornstein, Deffenbacher, Penrod, & McGorty, 2012; Deffenbacher et al., 2008; Fawcett et al., 2013; Memon, Hope, & Bull, 2003; Steblay, 1997; Steblay et al., 2003; Steblay et al., 2014), and the combination of these variables were compatible given the basic facts of the case provided by the Innocence Project.

Independent Variables

Quality of witnessing conditions. A block of three estimator variables was manipulated: exposure duration (see Memon et al., 2003), weapon presence (see Fawcett et al. 2013), and time delay (see Deffenbacher et al., 2008). When witnessing conditions were good, the eyewitness saw the perpetrator's face for 45 seconds, no weapon was present, and there was only one day between when the crime occurred and when an identification was made. When witnessing conditions were poor, the eyewitness only saw the perpetrator's face for 12 seconds, a weapon was present, and there was a one month delay between when the crime occurred and when an identification was made.

Quality of identification conditions. A block of three system variables was manipulated: identification procedure type (see Steblay et al., 2003), lineup instructions (see Steblay, 1997), and confirmatory feedback (see Steblay et al., 2014). When identification conditions were good, the eyewitness made an identification from a six person lineup where the lineup administrator gave unbiased lineup instructions and did not give the eyewitness any confirmatory feedback after she made an identification. When identification conditions were poor, the eyewitness made an identification from a showup where the lineup administrator failed to explicitly instruct the eyewitness that this may or may not be the perpetrator and confirmed she identified the suspect before obtaining the eyewitness' confidence level.

Safeguard types. Four safeguards were manipulated in addition to a control group (no safeguards). Participants in the *Henderson* instructions condition heard the judge read aloud *Henderson* eyewitness instructions, beginning with an explanation on the process of memory (see New Jersey Supreme Court, 2012a with link to full instructions). Jurors were instructed on relevant estimator and system factors as the Court intended.

Participants in the enhanced instructions condition also heard *Henderson* eyewitness instructions on problematic estimator and system factors present in the trial they viewed (see Appendix A). Additionally, the judge elaborated on research findings specifically discussing how each factor can influence the likelihood of an accurate identification. The effect sizes in these instructions were taken directly from research presented to the Special Master in the *Henderson* case (see *New Jersey v. Henderson*, 2011). The content of expert testimony matched the enhanced instructions (i.e., a detailed explanation of the impact of different eyewitness factors). Thus, we did not create another condition that contained both as it would be too repetitive. Lastly, participants in the *Henderson* instructions + expert testimony condition heard both expert testimony and the original *Henderson* instructions on the same poor estimator and system variables. Except for those in the safeguard control conditions, all participants heard instructions and/or expert testimony on double blind administration, cross race identification, and stress as mandated in *Henderson*.

Mediators

Witness ratings. Participants rated their impressions of the detective and eyewitness on a series of seven-point bipolar adjective pairs. These items include trustworthy-untrustworthy, honest-dishonest, not believable-believable, convincing-unconvincing, and certain-uncertain. For each scale, we recoded items as necessary so that higher values represented higher credibility and averaged the items to create a single scale measure (detective, $\alpha = .87$; eyewitness, $\alpha = .84$).

Belief in eyewitness. Participants indicated the probability that the eyewitness correctly identified the defendant as the perpetrator (0% = not at all probable, 100% = completely probable).

Dependent Measures

Verdict. Mock jurors provided dichotomous judgments of guilt (0 = not guilty; 1 = guilty).

Comprehension ratings. Mock jurors indicated their perceptions of whether each of a total of 5 eyewitness factors *generally* influence the likelihood of an accurate identification using nine-point Likert scales (1 = Strongly Reduces Accuracy, 9 = Strongly Increases Accuracy). These factors included exposure duration, weapon presence, time delay, identification procedure type, and lineup instructions. For example, "As compared to an eyewitness who has 12 seconds to view a perpetrator's face, what impact would having 45 seconds to view a perpetrator's face have on the accuracy of an identification?" Jurors also indicated whether confirmatory feedback generally increases an eyewitness' confidence. Items were recoded as necessary so that higher values were indicative of increased comprehension.

Manipulation Checks

Participants completed 10 manipulation checks by indicating the type of identification procedure used by police, the time delay between witnessing the crime and making an identification, and the amount of time the victim had to see the perpetrator's face. Participants also indicated whether a weapon was present, and whether the police used biased lineup instructions or gave confirmatory feedback. Lastly, participants answered whether the judge gave instructions on system and estimator variables as well as whether an eyewitness expert gave testimony on system and estimator variables.

Procedure

Participants came to the psychology lab to complete the study. After providing informed consent, they completed the voir dire questionnaire. Community members and students completed the study separately. We grouped participants in randomly assigned sessions to view

the filmed trial. In the relevant conditions and prior to the eyewitness' testimony, the video was

paused and the experimenter handed out printed copies of the *Henderson* (or enhanced)

instructions and informed participants that they could follow along with the judge's instructions.

The video resumed with participants hearing the *Henderson* (or enhanced) instructions.

Participants then individually completed post-trial questionnaires, where they could refer to their

written copy of the instructions at their discretion. Participants were subsequently debriefed, paid

$20, and excused. Sessions were, on average, two hours.

Results

Are There Differences between Community Members and Students?

To ensure responses were consistent across participant type, we first ran a MANOVA to

test for the effect of participant type on witness ratings, comprehension ratings, and belief in the

eyewitness, and a logistic regression for verdict. Descriptives are listed in Table 1.

There was a significant multivariate effect of participant type, F (9, 399) = 4.39, $p < .001$,

Wilk's λ = .91, η_p^2 = .09. Compared to students, community members had higher comprehension

concerning exposure duration, F (1, 407) = 8.90, p = .003, η_p^2 = .02, d = .27, 95% CI [.08, .45],

time delay between witnessing a crime and making an identification, F (1, 407) = 5.84, p = .016,

η_p^2 = .01, d = .23, 95% CI [.04, .41], unbiased lineup instructions, F (1, 407) = 8.11, p = .005, η_p^2

= .02, d = .30, 95% CI [.11, .48], and a lineup procedure (versus a showup), F (1, 407) =4.50, p

= .035, η_p^2 = .01 d = .20, 95% CI [.01, .39] (see Table 1 for means). Community members also

rated eyewitness confidence higher when police gave confirmatory feedback, F (1, 407) = 7.98, p

= .005, η_p^2 = .02, d = .28, 95% CI [.09, .47].

Despite these more accurate responses to the *general* impact of witnessing and

identification conditions on eyewitness accuracy and confidence, community members rated the

eyewitness as more credible than students, $F(1, 407) = 3.90$, $p = .05$, $\eta_p^2 = .01$, $d = .19$, 95% CI

[.01, .38], and believed the eyewitness more, $F(1, 407) = 4.00$, $p = .05$, $\eta_p^2 = .01$, $d = .18$, 95%

CI [.00, .36], regardless of the actual quality of those conditions. Furthermore, any differences

observed in witness ratings or belief in the eyewitness did not translate into differences in

convictions, directly or in interaction with the quality of the witnessing and identification

conditions or safeguard types (all $ps > .05$). Therefore, participant type was not included in

subsequent analyses.

Manipulation Checks

We conducted chi-square analyses to evaluate performance on 10 manipulation checks.

First, participants were successful at noticing the manipulation of quality of identification

conditions. Compared to those who heard about poor identification conditions, participants who

heard about good identification conditions were more likely to respond that the police gave

unbiased lineup instructions (90% vs. 19%), $\chi^2(1, N = 448) = 227.56$, $p < .001$, $\varphi = .71$, that the

police conducted a lineup (77% vs. 8%), $\chi^2(2, N = 445) = 216.77$, $p < .001$, $\varphi = .70$, and that the

police did not give any confirmatory feedback, (89% vs. 10%), $\chi^2(3, N = 449) = 287.00$, $p <$

.001, $\varphi = .80$.

Second, participants were successful at noticing the manipulation of quality of witnessing

conditions. Compared to those who heard about poor witnessing conditions, participants who

heard about good witnessing conditions were more likely to respond that the victim had 45

seconds to view the perpetrator (88% vs. 6%), $\chi^2(3, N = 448) = 340.39$, $p < .001$, $\varphi = .87$, that the

perpetrator did not use a weapon (99% vs. 2%), $\chi^2(2, N = 448) = 425.64$, $p < .001$, $\varphi = .98$, and

that the victim made an identification one day after the crime occurred (89% vs. 6%), $\chi^2(3, N =$

$446) = 339.21$, $p < .001$, $\varphi = .87$.

Third, participants were successful at noticing the manipulation of safeguard type. Compared to those who did not hear expert testimony, those who did were more likely to respond that an eyewitness expert testified about identification conditions, (77% vs. 34%), χ^2(1, N = 450) = 81.86, $p < .001$, φ = .43, and witnessing conditions, (89% vs. 44%), χ^2(1, N = 448) = 95.87, $p < .001$, φ = .46. Additionally, compared to those who did not hear *Henderson* instructions, those who did were more likely to respond that the judge instructed them on identification conditions (81% vs. 16%), χ^2(1, N = 448) = 183.94, $p < .001$, φ = .64, and witnessing conditions (89% vs. 25%), χ^2(1, N = 449) = 192.27, $p < .001$, φ = .65.

The majority of participants passed nine of the 10 manipulation checks (55.1%; $M = 8.45$, $SD = 1.44$) and we observed no differences between community members and students concerning the number of manipulation checks passed, t(436) = -1.08, $p = .28$, $d = .10$, 95% CI [-.09, .28]. Thus, we retained all participants for subsequent analyses.

Do Safeguards Increase Eyewitness Knowledge? (Hypothesis 1)

As mentioned in the methods section, we recoded the comprehension items so that higher values were equal to higher comprehension. With that in mind, jurors, on their own, appeared to be knowledgeable concerning the six eyewitness factors we assessed with all means above the mid-point (see overall means, Table 1). To determine if any of our safeguards served as an educational tool concerning various eyewitness factors, we ran a MANOVA to examine the main effect of safeguard on comprehension ratings. We observed a significant multivariate effect of safeguard on comprehension, F (24, 1532.70) = 1.74, $p = .014$, Wilk's λ = .91, η_p^2 = .02. There was a main effect of safeguard type on lineup instruction comprehension, F (4, 444) = 4.48, $p = .001$, η_p^2 = .04. The Tukey test for mean differences indicated that *Henderson* instructions, d = .47, 95% CI [.17, .77], expert testimony, d = .54, 95% CI [.24, .83], and enhanced instructions, d

= .53, 95% CI [.23, .82] all had significantly higher means than the no safeguard condition. That

is, participants in these conditions correctly noted that the likelihood of identification accuracy is

generally increased with the use of unbiased lineup instructions. None of the other means

differed from one another. Furthermore, none of the safeguards influenced comprehension of

exposure duration, weapon presence, time delay, identification procedure type, or feedback on

identification accuracy or confidence, all Fs $(4, 444) < 2.07$, ps $> .05$, η_p^2s $< .02$. Thus, with the

exception of lineup instructions, we did not find that any of the safeguards improved jurors'

eyewitness comprehension as hypothesized (Hypothesis 1).

Do Safeguards Sensitize Jurors? (Hypothesis 2)

We ran a MANOVA to test for the main and interactive effects of our manipulated

variables on witness ratings and belief in the eyewitness. Correlations among all variables are

presented in Table 2. Conditional means are presented in Table 3. There was a significant

multivariate effect for quality of identification conditions, F $(3, 428) = 8.95$, $p < .001$, Wilk's $\lambda =$

.94, $\eta_p^2 = .06$. Participants viewed the detective as more credible when identification conditions

were good (vs. poor), F $(1, 430) = 25.72$, $p < .001$, $\eta_p^2 = .06$, $d = .46$, 95% CI [.27, .64] (see

Table 1). Additionally, participants viewed the eyewitness as more credible when identification

conditions were good (vs. poor), F $(1, 430) = 6.80$, $p = .009$, $\eta_p^2 = .02$, $d = .23$, 95% CI [.04,

.41]. There were no significant effects on belief in the eyewitness, Fs $(1, 430) < 1.81$, $p > .05$,

η_p^2s $< .02$.

We also regressed the dichotomous verdict on quality of witnessing and identification

conditions, safeguard type, and all two and three way interactions involving safeguard type in a

logistic regression analysis. The omnibus test with all partial effects was significant, $\chi^2(6, N =$

$448) = 19.52$, $p = .003$, Nagelkerke $R^2 = .06$. Quality of identification conditions and expert

testimony (vs. no safeguard) each predicted verdict. The odds of a guilty verdict were 1.56 times

greater when identification conditions were good (vs. poor), Wald χ^2(1, N = 448) = 5.22, p = .02,

Exp(B) = 1.56, 95% CI [1.07, 2.28]. When jurors heard expert testimony (vs. no safeguard), the

odds of a guilty verdict were .39 times lower, Wald χ^2(1, N = 448) = 9.12, p = .003, *Exp(B) =*

.39, 95% CI [.21, .72]. Furthermore, the odds of conviction were lower with expert testimony

compared to *Henderson* instructions, Wald χ^2(1, N = 448) = 8.85, p = .003, *Exp(B)* = .39, 95%

CI [.21, .73], enhanced instructions, Wald χ^2(1, N = 448) = 4.67, p = .03, *Exp(B)* = .51, 95% CI

[.28, .94], and expert + *Henderson*, Wald χ^2(1, N = 448) = 7.30, p = .01, *Exp(B)* = .43, 95% CI

[.23, .79]. With the exception of expert testimony, none of the safeguards differed from one

another or the no safeguard condition. All interactions were non-significant, Wald χ^2(1, N = 448)

< 2.24, ps > .05, *Exp(B)* < 3.16. Due to these non-significant interactions, as well as the non-

significant interactions for witness ratings and belief in the eyewitness, we failed to find support

for Hypothesis 2 (safeguard sensitivity).

Are Verdict Effects Mediated by Witness Credibility? (Hypothesis 3)

We did not observe any interactions in the previous analyses. We did, however, find that

the quality of identification conditions was predictive of both detective and eyewitness

credibility, as well as verdict. Therefore, we conducted a mediational analysis (PROCESS Model

4; Hayes, 2013) using bias-corrected 10,000 sample bootstrap confidence intervals. Following

Hayes' (2013) recommendation, we report unstandardized regression coefficients. This model

examined the potential mediating roles of detective and eyewitness credibility on the relationship

between quality of identification conditions and verdict, with safeguard (dummy coded) and

quality of witnessing conditions included as covariates (see Table 4 for total, direct, and indirect

effects).

Recall that quality of identification conditions had a positive effect on both detective credibility, $b = .55$, 95% CI [.32, .78], and eyewitness credibility, $b = .29$, 95% CI [.05, .53]. Additionally, eyewitness credibility was a positive predictor of verdict, $b = 1.09$, 95% CI [.84, 1.34]. However, detective credibility did not predict verdict after accounting for eyewitness credibility, $b = .05$, 95% CI [-.17, .28]. There was an indirect effect of the quality of identification conditions on verdict through eyewitness credibility, $b = .32$, 95% CI [.04, .61], such that good identification conditions increased perceptions of the eyewitness' credibility which, in turn, increased the odds of a conviction (see Figure 1). Thus, due to the lack of interactions, we only observed partial support for Hypothesis 3.

Discussion

The current study sought to test three hypotheses concerning the effectiveness of instructions and expert testimony safeguards. First, we expected that safeguards, in particular enhanced *Henderson* instructions and expert testimony, would improve jurors' comprehension for eyewitness factors relative to no safeguard at all. The presentation of *Henderson* instructions, expert testimony, or enhanced *Henderson* instructions (compared to no safeguard) equally improved jurors' comprehension of the impact of lineup instructions on identification accuracy. None of the safeguards improved comprehension for the remaining factors (weapon presence, time delay, exposure duration, showup, and confirmatory feedback). Across safeguard conditions, however, eyewitness knowledge was above the scales' mid-point for all six system and estimator variables, suggesting jurors were not oblivious to factors influencing identification accuracy.

Second, following the Court's assumptions, we hypothesized that safeguards, in particular enhanced *Henderson* instructions or expert testimony, would sensitize jurors to

variations in witnessing and identification quality. However, we did not observe any interactions

between the quality of witnessing and identification conditions and safeguard type for witness

ratings, belief in the eyewitness, or verdict. Thus, all forms of our safeguards (*Henderson*

instructions, enhanced instructions, expert testimony, and expert testimony + *Henderson*

instructions) were ineffective in sensitizing jurors to identification quality.

The *Henderson* instructions (both the original and enhanced versions) not only failed to

induce sensitivity to identification quality, they failed to have any effect at all. These null

findings support previous research examining the *Telfaire* instructions (Cutler et al., 1990a; see

also Greene, 1988 Experiment 1), but stands in contrast to the skepticism effects found with the

(general version of) *Henderson* instructions in Papailiou et al. (2014). In contrast to the null

effects for both versions of the *Henderson* instructions, expert testimony appeared to induce

skepticism concerning the accuracy of the eyewitness identification. Relative to all of our other

safeguards (*Henderson*, enhanced instructions, expert + *Henderson* instructions, and no

safeguard), participants who heard expert testimony were less likely to convict regardless of the

quality of witnessing and identification conditions. This skepticism effect replicates previous

findings (Cutler et al., 1990a).

Though not what we (and the Court) had hypothesized, the odds of conviction also

depended on the quality of identification conditions. Our mock jurors, on their own, were able to

align their verdict decisions with the quality of police practices by convicting more often when

identification conditions were good and less often when conditions were poor. Eyewitness

credibility mediated this relationship. Specifically, the credibility of the eyewitness hinged on

how police performed; when police gave unbiased lineup instructions, conducted a lineup

procedure (instead of a showup), and withheld confirmatory feedback from the eyewitness after

she identified the defendant, jurors increased their ratings of the eyewitness' credibility and, in turn, were more likely to convict. While not what we had hypothesized (Hypothesis 3), these findings shed light on the underlying mechanism driving changes in jurors' verdict decisions. Furthermore, a few studies have observed juror sensitivity to system variables, such as lineup instructions (Cutler et al., 1988) and foil-biased lineups (Devenport et al., 2002) without any safeguards. We believe this is the first study to demonstrate juror sensitivity to identification procedure (lineup vs. showup) and confirmatory feedback, though the confounding of these factors indicates additional studies need to be conducted to tease apart findings.

While jurors were sensitive to the quality of identification conditions on their own, we observed no such effect for the quality of witnessing conditions. Comprehension of the impact of exposure duration, weapon focus, and time delay were above the mid-point (and quite similar to comprehension levels for the system variables we assessed). Yet, jurors' verdicts did not differ between good and poor witnessing conditions. Thus, while jurors in our study had high levels of estimator knowledge, they were unable to apply that knowledge to evaluate specific trial facts (see Desmarais & Read, 2011) and none of our safeguards were able to assist jurors in this regard. We are unaware of any eyewitness instruction studies that have observed sensitivity effects for estimator variables (cf., Ramirez et al, 1996, though even they cannot say with certainty that jurors were sensitized to estimator variables with eyewitness instructions because they confounded system and estimator variables in one manipulation). The lack of safeguard sensitivity in the current study could be due to a variety of reasons, including, but not limited to, the specific facts of the case, the eyewitness factors we manipulated, and/or the length of the experiment (which could have induced fatigue—although this is unlikely given the high number of manipulation checks correctly answered).

Beyond testing our hypotheses, we had an opportunity to examine potential differences between community members and students. Both appeared to pay equal attention to the trial stimulus, answering the majority of manipulation checks correctly. Community members were more knowledgeable than students on five out of six system and estimator variables (see also Desmarais & Read, 2011). The only factor community members did not understand better than students was weapon focus. However, community members (compared to students) believed the eyewitness was more credible and accurate regardless of the actual witnessing and identification conditions. Additionally, we did not observe any differences in convictions as a function of participant type. Thus, any superior knowledge that community members may have had over students was not integrated into their decision making. The courts' hesitation to rely on studies conducted solely with college students may be overly cautious as verdict decisions, at least in this study, did not differ between these samples (see also Bornstein, 1999; Cutler et al., 1990b).

Limitations and Future Directions

The current study suffered from a few limitations. First, we manipulated a block of three system and three estimator variables. While we did not observe any effect of estimator variables, we found that the quality of identification conditions did predict verdict. However, we are unable to disentangle whether any one or the combination of all three system variables (i.e., identification procedure type, lineup instructions, and confirmatory feedback) were driving the effect on verdict. Future research should examine the individual impact of system and estimator variables discussed in the *Henderson* instructions.

Second, considering the New Jersey Supreme Court emphasized that instructions should be "tailored to the facts of the case," we worked on the assumption that judges would only include specific factors considered *problematic* in a case (*New Jersey v. Henderson*, 2011, p.

1002). Therefore, the instructions given under varying witnessing and identification conditions were not equivalent. We suspect determining which factors are problematic will be a difficult task for judges to determine in actual cases given the subjective nature of many eyewitness factors. For example, how far should the distance between the witness and perpetrator be before judges give the distance instruction? What about the length of time an eyewitness viewed the perpetrator? Should judges only give instructions when factors are poor to draw attention to the possibility of a misidentification or should they also draw attention to factors that increase the likelihood of an accurate identification? For example, if a perpetrator does not use a weapon, should the jury be informed that the absence of a weapon increases the likelihood of an accurate identification? These are all issues that we struggled with and suspect court judges will as well. For simplicity purposes, we decided to include instructions only on factors where research has clearly established that reliability is reduced, following the format typically used by experts who testify for the defense. Of course, it is possible that delivering instructions in good witnessing conditions could produce a sensitizing effect reflected in increased convictions. Future research should examine general *Henderson* instructions on all factors regardless of whether they are problematic in a given case to determine its impact on jurors' decisions.

Relatedly, the *Henderson* Court granted judges discretion in when to administer the instructions. Thus, we chose to have the judge give the *Henderson* instructions prior to the eyewitness' testimony because we suspected this would be most effective (see Kassin & Wrightsman, 1979). Consequently, the timing of the instructions and the testimony of the expert (who testified for the defense, and thus later than when the instructions were administered) did not match. A recency effect may explain why expert testimony affected verdicts, but not instructions. This is unlikely, however, because jurors had access to a written copy of the

instructions for the duration of the study. We sought to provide the best balance of internal and ecological validity with this novel test, though future researchers may wish to align the timing of various safeguards, even at the cost of ecological validity. Furthermore, while there is some research on the timing of instructions (e.g., Kassin & Wrightsman, 1979), none have explicitly tested the impact of the timing of *Henderson* instructions.

Finally, we attempted to assess comprehension of various estimator and system variables by having participants respond on a nine-point Likert scale. Some of these questions may have been leading (e.g., "As compared to an eyewitness who has 10 seconds to view a perpetrator's face, what impact would having 45 seconds to view a perpetrator's face have on the accuracy of an identification?"), inflating comprehension rates. Future research should measure eyewitness knowledge in alternative ways, including open ended responses.

Conclusions

As a leader in criminal law, the New Jersey Supreme Court's decision to implement case-specific eyewitness instructions has already encouraged other states to follow suit (e.g., Massachusetts; *Commonwealth v. Gomes*, 2015). The widespread adoption of instructions similar to *Henderson* should depend on its effectiveness. While we applaud the efforts of the New Jersey Supreme Court and their desire to ensure justice, we suggest that states who are considering adopting *Henderson* style eyewitness instructions give pause as the current results suggest these instructions have little impact on jurors' decisions. Overall, it appears the *Henderson* Court overestimated the ability of these case-specific eyewitness instructions to assist jurors to better evaluate eyewitness evidence or to lessen the need for expert testimony.

The finding that jurors adjusted their verdicts based on the quality of some system variables suggests they may not need any safeguards to help them better evaluate the quality of

identification conditions as it concerns lineup instructions, identification procedure, and/or

confirmatory feedback. However, we did not detect this same effect on verdict for estimator

variables, suggesting jurors are not sensitive on their own to the quality of witnessing

conditions—despite demonstrating high comprehension for such factors. Therefore, safeguards

could provide some assistance to jurors to evaluate this type of eyewitness evidence.

Unfortunately, none of the safeguards tested in this study was able to induce sensitivity to

estimator variables.

Since none of the variations of *Henderson* instructions sensitized jurors (and courts prefer

instructions over expert testimony), future studies should seek additional ways to modify and

improve the effectiveness of *Henderson* instructions (or seek out other cost effective alternatives

such as premade modular expert testimony; see Mnookin, 2015). The sensitivity effects found

with the I-I-Eye teaching aid (see Pawlenko et al., 2013) may be due to its brevity, general (vs.

case-specific) nature, or the rhetorical questions concerning eyewitness evidence quality that

resulted in jurors evaluating the evidence more critically. Future researchers should examine

whether incorporating these features into the *Henderson* instructions will improve its

effectiveness. In addition, researchers should continue examining ways to improve the

effectiveness of safeguards assumed by courts to assist jurors in evaluating eyewitness evidence.

As Chief Justice Rabner indicated, "Only with a fully informed and properly instructed jury can

justice be served" (New Jersey Supreme Court, 2012b).

Appendix A
Enhanced *Henderson* Instructions

The Court instructs you as follows:

It is your function to determine whether the witness's identification of the defendant is reliable and believable or whether it is based on a mistake or for any reason is not worthy of belief. You must decide whether it is sufficiently reliable evidence that this defendant is the person who committed the offenses charged.

Eyewitness identifications must be scrutinized carefully. Human beings have the ability to recognize other people from past experiences and to identify them at a later time. But research has shown that there are risks of making a mistake in identifications. That research has focused on the nature of memory and the factors that affect the reliability of witness's identifications.

Human memory is not foolproof. Research has revealed that human memory is not like a video recording that a witness need only replay to remember what happened.

Memory is far more complex. The process of remembering consists of three stages:

- **Acquisition -**The perception of the original event.

- **Retention -** The period of time that passes between the event and the eventual recollection of a piece of information. And

- **Retrieval** – The stage during which a person recalls stored information.

At each of these stages memory can be affected by a variety of factors.
Relying on some of the research that has been done, I will instruct you on specific factors you should consider in this case determining whether the eyewitness identification evidence is reliable.

In evaluating misidentification, you should consider the observations and perceptions on which the identification was based, the witness's ability to make those observations and perceive those events and the circumstances under which the identification was made.
Although nothing may appear more convincing than a witness's categorical identification of a perpetrator, you must critically analyze such testimony. Such identifications, even when made in good faith, may be mistaken. Therefore, when analyzing such testimony be advised that a witness's level of confidence, standing alone, may not be indication of reliability of the identification.

If you determine that the out of Court identification is not reliable you may still consider the witness's in Court identification of the defendant. If you find that it resulted from the witness's observations or perceptions of the perpetrator during the commission of the offense and that the identification is reliable.

If you find that the in Court identification is the product of an impression gained at the out of Court identification procedure, it should be afforded no weight. The ultimate question of the reliability of both the in Court and out of Court identifications is for you to decide.
To decide whether the identification testimony is sufficiently reliable evidence to conclude that the defendant is the person who committed the offenses charged, you should evaluate the testimony of the witness in light of the factors for considering credibility that I have already explained to you.

In addition, you should consider the following factors related to the witness, the alleged perpetrator, and the criminal incident itself.

In particular you should consider:

- **Stress** – Even under the best viewing conditions, high levels of stress can reduce the eyewitness's ability to recall and make an accurate identification. Therefore, you should consider a witness's level of stress and whether that stress, if any, distracted the witness or made it harder for him or her to identify the perpetrator. Additionally a recent meta-analysis demonstrates that when the perpetrator is present in the lineup, the identification accuracy rate under high stress was only 39% compared to 59% under low stress.

- **Duration** – The amount of time the eyewitness has to observe an event may affect the reliability of an identification. Although there is no minimum time required to make an accurate identification, or brief or fleeting contact is less likely to produce an accurate identification than are more prolonged exposure to the perpetrator. In addition, time estimates given by witnesses may not always be accurate because witnesses tend to think events last longer than they actually did. Research demonstrates reliable differences between identifications made when the witness sees the perpetrator for 12 seconds and 45 seconds. Those viewing a perpetrator for only 12 seconds made correct identifications in 30% of perpetrator present lineups and false identifications in 85% of perpetrator absent lineups. On the other hand, those viewing a perpetrator for 45 seconds had 90% correct identifications in perpetrator present lineups and 45% false identifications in perpetrator absent lineups.

- **Weapon focus** – You should consider whether the witness saw a weapon during the incident and the duration of the crime. The presence of a weapon can distract the witness and take the witness's attention away from the perpetrator's face. As a result, the presence of a visible weapon may reduce the reliability of subsequent identification if the crime is of short duration. In considering this factor you should take into account the duration of the crime because the longer the event, the more time the witness may have to adapt the presence of the weapon and focus on other details. One study found that the presence of a weapon increased false identifications from 33% to 64%.

- **Confidence and accuracy** – You heard testimony that Marie Sands made a statement at the time she identified the defendant asserting her level of certainty that the person she selected is in fact the person that committed the crime. As I explained earlier, a witness's level of confidence standing alone may not be an identification of the reliability of the identification. Although some research has found that highly confident witnesses are more likely to make an accurate identification, eyewitness confidence is generally an unreliable indicator of accuracy. It is not unusual for witnesses to make confidence judgments that are 15-20% higher than actual accuracy rates.

- **Time Elapsed:** Memories fade with time. As a result, delays between the commission of a crime and the time an identification is made can affect the reliability of the identification. In other words, the more time that passes, the greater the possibility that a witness's memory of a perpetrator will weaken. A review of the literature found that longer delays can lead to 10% fewer correct identifications and 8% more false identifications. Specifically, research has demonstrated that eyewitness accuracy declines 20-25% in the first 2 hours. In the next 10 hours, another 5% of memory loss is experienced with more gradual losses as time extends beyond that.

- **Cross-racial effects** – Research has shown that people may have greater difficulty accurately identifying members of a different race. You should consider whether the fact that the witness and the defendant are not of the same race may have influenced the accuracy of the witness's identification. In particular, research indicates that overall accuracy rates are 60% for same-race identifications, compared to 40% for cross-race identifications.

In evaluating the reliability of the witness's identification you should also consider the circumstances under which any out of court identification was made and whether it was the result of a suggestive procedure. In that regard, you may consider everything that was done or said by law enforcement to the witness during the identification process. You should consider the following factors:

- **Showups** – In this case, the witness identified the defendant during a showup, that is, the defendant was the only person showed to the witness at that time. Even though such a procedure is suggestive in nature it is sometimes necessary for the police to conduct a showup or one-on-one identification procedure. Although the benefits of a fresh memory may balance the risk of undue suggestion, showups conducted more than 2 hours after an event present a heightened risk of misidentification. Also, police officers must instruct witnesses that the person they are about to view may or may not be the person who committed the crime and they should not be compelled to make an identification. In determining whether the identification is reliable or the result of unduly suggestive procedure you should consider how much time elapsed since witness saw the perpetrator,

whether appropriate instructions were given and all circumstances surrounding the showup. A recent review of the literature found that false identifications are 23% in single person showups but only 10% in six person lineups.

In determining the reliability of the identification you should also consider whether the identification procedure was properly conducted.

- **Double blind** – A lineup administrator who knows which person or photo in the lineup is the suspect may intentionally or unintentionally convey that knowledge to the witness. That increases the chance that the witness will identify the suspect even if the suspect is innocent. For that reason, whenever feasible, live lineups and photo arrays should be conducted by an officer who does not know the identity of the suspect. Research has shown that witnesses behaved in a fashion consistent with the lineup administrator's beliefs. One study found that 30% of witnesses who saw and interacted with the administrator selected the target substitute in a target-absent lineup compared with 3% of the witnesses who did not see the administrator.

You may consider this factor when you consider the circumstances under which the identification was made and when you evaluate the overall reliability of the identification.

Instructions-You should consider what was or was not said to the witness prior to viewing a lineup or showup. Identification procedures should begin with instructions to the witness that the perpetrator may or may not be present and the witness should not feel compelled to make an identification. Failure to give this instruction can increase the risk of misidentification. If you find that the police did or did not give this instruction to the witness you may take this factor into account when evaluating the identification evidence. A recent review of the literature found that when the witness was told that the perpetrator may or may not be present in lineups, error rates dropped from 60 to 35%.

Feedback: Feedback occurs when police officers, or witnesses to an event who are not law enforcement officials, signal to eyewitnesses that they correctly identified the suspect. That confirmation may reduce doubt and engender or produce a false sense of confidence in a witness. Feedback may also falsely enhance a witness's recollection of the quality of his or her view of an event. It is for you to determine whether or not a witness's recollection in this case was affected by feedback or whether the recollection instead reflects the witness's accurate perception of the event. A recent review demonstrates that confirmatory feedback about the accuracy of an identification can substantially increase confidence and inflate a witness' sense of the quality of their memory.

You may consider whether the witness was exposed to opinions, descriptions, or identifications given by other witnesses, to photographs, newspaper accounts, or to any other information or influence that may have affected the independence of her identification. Such information can affect the independent nature and reliability of a witness's identification and inflate the witness's

confidence in the identification. You are also free to consider any other factor based on the evidence or lack of evidence in the case that you consider relevant to your determination of whether the identifications were reliable.

Keep in mind that the presence of any single factor or combination of factors however, is not an indication that a particular witness is incorrect. Instead, you may consider the factors that I have discussed as you assess all of the circumstances of the case including all of the testimony and documentary evidence in determining whether a particular identification was made by a witness as accurate and thus worthy of your consideration as you decide whether the State has met its burden to prove identification beyond a reasonable doubt.

If you determine that the in Court or out of Court identifications resulted from the witness's observations or perceptions of the perpetrator during the commission of the offense, you may consider that evidence and decide how much weight to give it. If you instead decide that the identification is, or identifications are, a product of an impression gained at the in court or out of court identification procedures the identifications should be afforded no weight. The ultimate issue of the trustworthiness of an identification is for you to decide.

If after consideration of all the evidence you determine that the state has not proven beyond a reasonable doubt that Mr. William Johnson was the person who committed these offenses then you must find him not guilty. If, on the other hand, after consideration of all the evidence you are convinced beyond a reasonable doubt that Mr. William Johnson was correctly identified you will then consider whether the State has proven each and every element of the offense's charge beyond a reasonable doubt.

Acknowledgments

Endnotes

1. The National Registry of Exonerations is a comprehensive database of exonerations that identifies contributing factors of wrongful convictions by type of crime and race, distinguishes between DNA and non-DNA exonerations, and tracks year of conviction and exoneration for all known exonerees.

2. The current instructions mandate that double blind and lineup instructions always be given to jurors when a lineup procedure was conducted.

3. The longest trial contained both expert testimony and instructions. We examined whether attention varied by condition. Participants answered slightly more manipulation checks correctly in the expert only condition than the control condition, $F(4, 418) = 3.19$, $p = .01$, $\eta_p^2 = .03$, $d = .52$, 95% CI [.22, .82], which suggests fatigue was not an issue. In addition, there was an interaction between system and estimator variable manipulations, $F(1, 418) = 5.47$, $p = .02$, $\eta_p^2 = .01$. When the quality of witnessing conditions was poor, there was no difference between good and poor identification conditions concerning the number of correct manipulation checks. In contrast, when the quality of witnessing conditions was good, the number of correct manipulation checks were higher with poor identification conditions than good, $F(1, 418) = 5.06$, $p = .03$, $\eta_p^2 = .01$, $d = .35$, 95% CI [.08, .62]. As trial length was similar across these conditions, fatigue cannot account for the differences in manipulation check responses.

References

Benton, T.R., Ross, D.F., Bradshaw, E., Thomas, W.N., & Bradshaw, G.S. (2006). Eyewitness memory is still not common sense: Comparing jurors, judges, and law enforcement to eyewitness experts. *Applied Cognitive Psychology, 20,* 115-129. doi: 10.1002/acp.1171

Berman, G.L., & Cutler, B.L. (1996). Effects of inconsistencies in eyewitness testimony on mock-juror decision making. *Journal of Applied Psychology, 81*, 170-177. doi: 10.1037/0021-9010.81.2.170

Bornstein, B.H. (1999). The ecological validity of jury simulations: Is the jury still out? *Law and Human Behavior* 23: 75-91. doi: 10.1023/A:1022326807441

Bornstein, B.H., Deffenbacher, K.A., Penrod, S.D., & McGorty, E.K. (2012). Effects of exposure time and cognitive operations on facial identification accuracy: A meta-analysis of two variables associated with initial memory strength. *Psychology, Crime and Law, 18,* 473-490.

Bornstein, B.H., & Hamm, J.A. (2012). Jury instructions on witness identification. *Court Review, 48*, 48-53.

Bruer, K., & Pozzulo, J.D. (2014). Influence of eyewitness age and recall error on mock juror decision-making. *Legal and Criminological Psychology, 19*, 332-348. doi: 10.1111/lcrp.12001

charrow, R. P., & Charrow, V. R. (1979). Making legal language understandable - Psycholinguistic study of jury instructions. *Columbia Law Review, 79*, 1306-1374. doi: 10.2307/1121842

Commonwealth v. Gomes, 470 Mass. 352 (2015).

Cowan, S., Read, J.D., & Lindsay, D.S. (2014). Predicting and postdicting eyewitness accuracy and confidence. *Journal of Applied Research in Memory and Cognition, 3*, 21-30. doi: 10.1016/j/jarmac.2014.01.002

Cutler, B.L., & Penrod, S. (1995). *Mistaken identification: The eyewitness, psychology, and the law*. New York: Cambridge University Press.

Cutler, B.L., Dexter, H., & Penrod, S. (1989a). Expert testimony and jury decision making: An empirical analysis. *Behavioral Sciences and the Law, 7*, 215-225. doi: 10.1002/bsl.2370070206

Cutler, B.L., Dexter, H., & Penrod, S. (1990a). Nonadversarial methods for sensitizing jurors to eyewitness evidence. *Journal of Applied Social Psychology, 20*, 1197-1207. doi: 10.1111/j.1559-1816.1990.tb00400.x

Cutler, B.L., Penrod, S.D., & Dexter, H.R. (1990b). Juror sensitivity to eyewitness identification evidence. *Law and Human Behavior, 14*, 185-191. doi: 10.1007/BF01062972

Cutler, B.L., Penrod, S.D., & Dexter, H. (1989b). The eyewitness, the expert psychologist, and the jury. *Law and Human Behavior, 13*, 311-332. doi: 10.1007/BF01067032

Cutler, B.L., Penrod, S.D., & Stuve, T.E. (1988). Juror decision making in eyewitness identification cases. *Law and Human Behavior, 12*, 41-55. doi: 10.1007/BF01064273

Deffenbacher, K.A., Bornstein, B., McGorty, E.K., & Penrod, S. (2008). Forgetting the once-seen face: Estimating the strength of an eyewitness's memory representation. *Journal of Experimental Psychology: Applied, 14*, 139-150. doi: 10.1037/1076-898X.14.2.139

Desmarais, S., & Read, D. (2011). After 30 years, what do we know about what jurors know? A meta-analytic review of lay knowledge regarding eyewitness factors. *Law and Human Behavior, 35*, 200-210. doi: 10.1007/s10979-010-9232-6

Devenport, J.L., & Cutler, B.L. (2004). Impact of defense-only and opposing eyewitness experts

on juror judgments. *Law and Human Behavior, 28,* 569-576. doi:

10.1023/B:LAHU.0000046434.39181.07

Devenport, J.L., Cutler, B.L., & Penrod, S.D. (1997). Eyewitness identification evidence:

Evaluating commonsense evaluations. *Psychology, Public Policy, and Law, 3,* 338-361.

doi:10.1037/1076-8971.3.2-3.338

Devenport, J.L., Stinson, V., Cutler, B.L., & Kravitz, D.A. (2002). How effective are the cross-

examination and expert testimony safeguards? Jurors' perceptions of the suggestiveness

and fairness of biased lineup procedures. *Journal of Applied Psychology, 87*, 1042-1054.

doi: 10.1037/0021-9010.87.6.1042

Devine, D.J., Clayton, L.D., Dunford, B.B., Seying, R., & Pryce, J. (2001). Jury decision

making: 45 years of empirical research on deliberation groups. *Psychology, Public

Policy, and Law, 7,* 622-727. doi: 10.1037/1076-8971.7.3.622

Ellsworth, P.C. (1989). Are twelve heads better than one? *Law and Contemporary Problems, 52*,

205-224. doi: 10.2307/1191911

Fawcett, J.M., Russell, E.J., Peace, K.A., & Christie, J. (2013). Of guns and geese: A meta-

analytic review of the 'weapon focus' literature. *Psychology, Crime, & Law, 19*, 35-66.

doi: 10.1080/1068316X.2011.599325

Fox, S.G., & Walters, H.A. (1986). The impact of general versus specific expert testimony and

eyewitness confidence upon mock juror judgment. *Law and Human Behavior, 10*, 215-

228. doi: 10.1007/BF01046211

Frank, J., & Applegate, B.K. (1998). Assessing juror understanding of capital-sentencing

instructions. *Crime & Delinquency, 44*, 412-433. doi: 10.1177/0011128798044003005

Greene, E. (1988). Judge's instructions on eyewitness testimony: Evaluation and revision.

    Journal of *Applied Social Psychology, 18,*151-276. doi: 10.1111/j.1559-

    1816.1988.tb00016.x

Hayes, A. F. (2013*). Introduction to mediation, moderation, and conditional process analysis: A*

    regression-based approach. New York: The Guilford Press.

Hoffheimer, M.H. (1989). Requiring jury instructions on eyewitness identification evidence at

    federal criminal trials. *Journal of Criminal Law and Criminology, 80*, 585-672.

Houston, K.A., Hope, L., Memon, A., & Read, J.D. (2013). Expert testimony on eyewitness

    evidence: In search of common sense. *Behavioral Sciences and the Law, 31*, 637-651.

    doi: 10.1002/bsl.2080

Heuer, L., & Penrod, S.D. (1989). Instructing jurors: A field experiment with written and

    preliminary instructions. *Law and Human Behavior, 13*, 409-430. doi:

    10.1007/BF01056412

Kassin, S.M., & Wrightsman, L. (1979). On the requirements of proof: The timing of judicial

    instruction and mock juror verdicts. *Journal of Personality & Social Psychology, 37*,

    1877-1887. doi: 10.1037/0022-3514.37.10.1877

Leippe, M.R., & Eisenstadt, D. (2009). The influence of eyewitness expert testimony on jurors'

    beliefs and judgments. In B.L. Cutler (Ed.), *Expert testimony on the psychology of*

    eyewitness identification (pp. 169-199). Mahwah, NJ: Erlbaum.

Levett, L.M., Danielsen, E.M., Kovera, M.B., & Cutler, B.L. (2005). The psychology of jury and

    juror decision making. In N. Brewer & K.D. Williams (Eds.), *Psychology and law: An*

    empirical perspective (pp. 365-406). NY: Guilford Press.

Lieberman, J.D., & Sales, B.D. (1997). What social science teaches us about the jury instruction process? *Psychology, Public Policy, and Law, 3*, 589-644. doi: 10.1037/1076-8971.3.4.589

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](10.1037/0021-9010.70.3.556)

MacCoun, R.J., & Kerr, N.L. (1988). Asymmetric influence in mock jury deliberation: Jurors' bias for leniency. *Journal of Personality and Social Psychology, 54,* 21-33. doi: 10.1037/0022-3514.54.1.21

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology, 94*, 339-354. doi: 10.1348/000712603767876262

Mnookin, J.L. (2015). Constructing evidence and educating juries: The case for modular, made-in-advance expert evidence about eyewitness identification and false confessions. *Texas Law Review, 93*, 1811-1848.

National Registry of Exonerations. (2016). Retrieved October 28, 2016, from http://www.law.umich.edu/special/exoneration/Pages/learnmore.aspx

Neal, T.M.S., Christiansen, A., Bornstein, B.H., & Robicheaux, T.R. (2012). The effects of mock jurors' beliefs about eyewitness performance on trial judgments. *Psychology, Crime, & Law, 18*, 49-64. doi: 10.1080/1068316X.2011.587815

Neil v. Biggers, 409 U.S. 188 (1972).

New Jersey v. Henderson, 27 A.3d 872 (2011).

New Jersey Supreme Court. (2012a). *Expanded jury instructions*. Retrieved August 11, 2014, from http://www.judiciary.state.nj.us/pressrel/2012/jury_instruction.pdf

New Jersey Supreme Court. (2012b). **Supreme Court releases eyewitness identification criteria for criminal cases [Press release]. Retrieved from** http://www.judiciary.state.nj.us/pressrel/2012/pr120719a.htm

Ogloff, J.R.P., & Rose, V.G. (2005). The comprehension of judicial instructions. In N. Brewer & K.D. Williams (Eds.), *Psychology and law: An empirical perspective* (407-444). New York: Guilford Press.

Papailiou, A.P., Yokum, D.V., & Robertson, C.T. (2014). The novel New Jersey eyewitness instruction induces skepticism but not sensitivity. *Arizona Legal Studies Discussion Paper No. 14-17*. doi: 10.2139/ssrn.2475217

Pawlenko, N.B., Safer, M.A., Wise, R.A., & Holfeld, B. (2013). A teaching aid for improving jurors' assessment of eyewitness accuracy. *Applied Cognitive Psychology, 27*, 190-197. doi: 10.1002/acp.2895

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology, 62*, 189-206. doi: 10.1037/0022-3514.62.2.189

Ramirez, G., Zemba, D., & Geiselman, R. (1996). Judges' cautionary instructions on eyewitness testimony. *American Journal of Forensic Psychology, 14,* 31-63.

Reifman, A., Gusick, S.M., & Ellsworth, P.C. (1992). Real jurors' understanding of the law in real cases. *Law and Human Behavior, 16*, 539-554. doi: 10.1007/BF01044622

Rose, R.A., Bull, R., & Vrij, A. (2005). Non-biased lineup instructions do matter-a problem for older witnesses. *Psychology, Crime, and Law, 11*, 147-159. doi: 10.1080/10683160512331316307

Rose, V. G., & Ogloff, J. R. P. (2001). Evaluating the comprehensibility of jury instructions: A method and an example. *Law and Human Behavior, 25*, 409-431. doi: 10.1023/A:1010659703309

Severance, L.J., & Loftus, E.F. (1982). Improving the ability of jurors to comprehend and apply criminal jury instructions. *Law & Society Review, 17,* 153-197. doi: 10.2307/3053535

Shapiro, P.N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100,* 139-156. doi: 10.1037/0033-2909.100.2.139

Sheehan, C. (2011). Making the jurors the "experts": The case for eyewitness identification jury instructions. *Boston College Law Review, 52*, 651-693.

Steblay, N. (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., Dysart, J., Fulero, S., & Lindsay, R.C.L. (2003). Eyewitness accuracy rates in police showup and lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 27,* 523-540. doi: 10.1023/A:1025438223608

Steblay, N., Hosch, H., Culhane, S., & McWethy, A. (2006). The impact on juror verdicts of judicial instruction to disregard inadmissible evidence: A meta-analysis. *Law and Human Behavior, 30,* 469-492. doi: 10.1007/s10979-006-9039-7

Steblay, N.K., Wells, G.L., & Douglass, A.B. (2014). The eyewitness post identification feedback 15 years later: Theoretical and policy implications. *Psychology, Public Policy, and Law, 20*, 1-18. doi: 10.1037/law0000001

United States v. Telfaire, 469 F.2d 552 (D.C. Cir. 1979).

Wells, G. (1978). Applied eyewitness testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology*, *36*, 1546–1557. doi: 10.1037/0022-3514.36.12.1546

Wells, G., 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., & Olson, E. (2003). Eyewitness testimony. *Annual Review of Psychology, 54,* 277-295. doi: 10.1146/annurev.psych.54.101601.145028

Wells, G., & Quinlivan, D. (2009). Suggestive eyewitness identification procedures and the Supreme Court's reliability test in light of eyewitness science: 30 years later. *Law and Human Behavior, 33,* 1-24. doi: 10.1007/s10979-008-9130-3

Wiener, R.L., Krauss, D.A., & Lieberman, J.D. (2011). Mock jury research: Where do we go from here? *Behavioral Sciences and the Law, 29*, 457-479. doi: 10.1002/bsl.989