# *Henderson* instructions: Do they enhance evidence evaluation?

Marlee Kind Dillon

*Psychology Department, John Jay College of Criminal Justice, New York, NY*

Angela M. Jones

*Department of Criminal Justice, Texas State University, San Marcos, TX*

Amanda N. Bergold

*Quattrone Center Fellow, University of Pennsylvania Law School, Philadelphia, PA*

Cora Y. T. Hui

*City University of Hong Kong*

Steven D. Penrod

*Psychology Department, John Jay College of Criminal Justice, New York, NY*

Forthcoming, *Journal of Forensic Psychology Research and Practice*

Corresponding author:

Dr. Marlee Kind Dillon

Address: 1370 Broadway New York, NY 10018.

Phone: (267) 446-6983

Email: mdillon@doar.com

Abstract

Mistaken eyewitness identifications are a leading cause of wrongful convictions. Even with procedural safeguards (e.g. attorney argument, cross examination of witnesses) in place, jurors still have difficulty evaluating the reliability of eyewitness identifications. The purpose of the current study was to test the New Jersey Supreme Court's assumptions that recently implemented research-based case-specific jury instructions sensitize jurors to unreliable eyewitness testimony. Four hundred sixty-eight jury-eligible adults watched a trial simulation in which estimator variables, system variables, and jury instruction were manipulated, and subsequently rendered a verdict. The *Henderson* instructions influenced mock jurors' perceptions of the eyewitness identification, but these perceptions did not translate to their verdict decisions. Rather than sensitizing jurors, the instructions induced an overall skepticism of eyewitness identification. Taken together, results indicate that the current *Henderson* instructions should be modified to improve juror sensitivity to various witnessing and identification conditions.

**Keywords**: eyewitness identification; judicial instruction; eyewitness variables; system variables; juror decision making

## *Henderson* instructions: Do they enhance evidence evaluation?

Although jurors tend to rely heavily on eyewitness evidence (Cutler & Penrod, 1999), eyewitness identifications are not always accurate. In fact, mistaken eyewitness identifications are a primary cause for wrongful convictions in the United States (Gross, Jacoby, Matheson, Montgomery, & Patil, 2005; Scheck & Neufeld, 2006; Scheck, Neufeld, & Dwyer, 2001; The National Registry of Exonerations, 2016). Courts have long been concerned about eyewitness error and have sought methods to reduce those errors, such as judicial instructions. Although research suggests that judicial instructions may not effectively sensitize jurors to eyewitness identification reliability, courts continue to rule in favor of using eyewitness jury instructions (see Massachusetts Supreme Court recommendation, 2015; *Commonwealth v. Gomes, 2009).*

In *New Jersey v. Henderson* (2011*),* the Court utilized a Special Master to review and report upon the extant empirical research on eyewitness identification. This research described the effect of both system variables (e.g., identification procedure, biased lineup instructions, confirmatory feedback) and estimator variables (e.g. viewing time, weapon focus, disguise, time lapse) on eyewitness identification accuracy (Wells, 1978; Wells, Memon, & Penrod, 2006). Due in part to a Special Master's report, the New Jersey Supreme Court determined that, if the defendant can show evidence of suggestibility, all system and estimator variables relevant to the identification should be investigated at a pretrial hearing (*New Jersey v. Henderson*, 2011). If the prosecution demonstrates that the identification is still reliable, the identification should be admitted. Upon admittance of the identification, however, courts must instruct the jury on the research concerning relevant system and estimator variables in order to help jurors evaluate the reliability of the identification (*New Jersey v. Henderson*, 2011). The *Henderson* Court proposed that "even with matters that may be considered intuitive, courts provide focused jury instructions" because "it is the Court's obligation to help jurors evaluate evidence critically and

objectively to ensure a fair trial" (p. 123-124). Specifically, the *Henderson* Court noted the added benefits of these enhanced jury instructions, such that 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" (p. 126).

Although New Jersey has adopted these new standards, researchers have yet to thoroughly investigate whether the *Henderson* instructions are actually an effective method for sensitizing jurors to eyewitness identification reliability. Sensitivity in this context refers to an improvement in jurors' ability to evaluate the quality of an identification, such that there is an increase in guilty verdicts when the identification was strong and a decrease in guilty verdicts when the identification was weak (Cutler, Dexter, & Penrod, 1989). Alternatively, the presence of judicial instructions may reveal evidence of a skepticism effect (Cutler, et al., 1989), wherein jurors, instead of discerning the reliability of an identification, become skeptical about all identifications. Papailiou, Yokum, and Robertson (2014) found only evidence of a skepticism effect, such that conviction rates decreased with *Henderson* instructions (compared to standard judicial instructions) regardless of identification quality. Papailiou et al., however, used an online sample of participants and presented the *Henderson*-type instruction at the end of the trial only. In another study, Laub, Kimbrough, and Bornstein (2014) did not find evidence of sensitivity or skepticism using a condensed version of the *Henderson*-type instruction. An abbreviated written version of the instructions, however, may not yield the same results as the full *Henderson* instruction read by a judge.

In related research, Pawlenko, Safer, Wise, and Holfield (2013) investigated whether jurors' ability to evaluate eyewitness identification accuracy could be improved through a teaching aid (Interview-identification-eyewitness; I-I-Eye; Wise, Fishman, & Safer, 2009). The purpose of this three-step method is to encourage jurors to evaluate whether the eyewitness

interview was conducted properly (e.g., could any of the police procedures have contaminated the eyewitness' memory or confidence in their identification?), to provide guidelines for assessing the eyewitness identification and subsequent interview(s), and to evaluate how variables at the crime scene may have affected the accuracy of the identification (Wise et al., 2009). Undergraduate mock jurors viewed three teaching aids via PowerPoint: the *Jury Duty aid* (basic instruction regarding the defendant's rights and the importance of being fair and impartial), the *Neil v. Biggers aid* (description of five eyewitness factors that jurors should consider in their evaluation of the identification), or the *I-I-Eye aid* (Pawlenko et al., 2013), and subsequently read a 27-page transcript and rendered a verdict. Results indicated that participants who received the I-I-Eye training rendered more guilty verdicts with the high quality identification compared to the poor quality identification (Pawlenko et al., 2013). Safer et al. (2016) sought to expand on the results of Pawlenko et al. (2013) by including circumstantial and forensic evidence in addition to eyewitness evidence. Results were consistent with Pawlenko et al. (2013) in that the I-I-Eye teaching aid increased sensitivity to eyewitness evidence. These results are promising because the I-I-Eye utilizes psychological research findings as a teaching aid and is thus similar to the *Henderson* instruction. In both studies, however, the researchers only manipulated system variables, so it is unknown whether jurors' sensitivity to estimator variables would be similarly enhanced. Moreover, the implementation of a teaching aid in the courtroom may not be readily embraced by the courts.

The *Henderson* Court dictates that enhanced instructions are "to be included in the Court's comprehensive jury charge at the close of evidence. In addition, instructions may be given during trial if warranted … Trial courts retain discretion to decide when to offer instructions" (p. 123). There is conflicting evidence regarding the efficacy of judicial instructions as a function of timing. In a study of pretrial publicity and judicial instructions, Prager (1991)

found that the timing of the instruction (either at the beginning of trial, at the end of trial, or both) had no effect on participants' verdicts. Previous research has shown that judicial instruction on the requirements of proof, presumption of innocence, burden of proof, and reasonable doubt prior to presentation of the evidence (compared to after presentation of the evidence and no instruction), can lower mock jurors' conviction rates (Kassin & Wrightsman, 1979). In another study, however, mock jurors who received a fact sheet outlining the research related to eyewitness identification prior to the eyewitness testimony were more likely to convict than those who received the fact sheet after hearing the eyewitness' testimony (Moore, 2010). As per the *Henderson* Court's recommendation, the current study examined whether presentation of the *Henderson* instructions prior to the eyewitness testimony would increase sensitivity to identification quality as compared to the presentation of the *Henderson* instructions at the end of the trial and compared to no judicial instruction on eyewitness testimony.

**Overview of experiment**

We sought to test the assumptions made by the *Henderson* Court. Given that past research demonstrates that eyewitness instructions do not generally work as intended (Dillon & Penrod, 2014), it is necessary to evaluate whether the *Henderson* instructions will be successful, especially as other states begin to implement similar instructions (see Massachusetts Supreme Court recommendation, 2015; *Commonwealth v. Gomes, 2009*). The current study examines the efficacy of the *Henderson* instructions in an ecologically-rich manner (see Diamond, 1997), using adult community member mock jurors and stimulus materials based on an actual wrongful-conviction case involving eyewitness testimony.

First, we hypothesized that there would be an interaction between instructions and witnessing/identification conditions. We further hypothesized that this sensitivity would occur only with the *Henderson* instructions. Specifically, jurors who heard the *Henderson* instructions

would perceive an identification as less accurate and convict less often when an eyewitness experienced poor witnessing and identification conditions and perceive the identification as more accurate and convict more often when an eyewitness experienced good witnessing and identification conditions (H1). We hypothesized that this effect would be strongest when jurors heard the *Henderson* instructions prior to the eyewitness testimony. We did not predict a sensitivity effect among jurors who did not hear *Henderson* instructions. The Special Master in *New Jersey v. Henderson* argued that "whether the science confirms commonsense views or dispels preconceived but not necessarily valid intuitions, it can properly and usefully be considered by both judges and jurors in making their assessments of eyewitness reliability" (p. 124). Based on the *Henderson* Court's implicit assumption that jurors should be able to evaluate eyewitness identification evidence more effectively after hearing the new instructions, we hypothesized that jurors' perceptions of the impact of the manipulated system and estimator variables on identification accuracy/eyewitness confidence would mediate the interaction of the *Henderson* instructions on verdict (H2). Thus, jurors who heard the *Henderson* instructions would rate the manipulated variables as having a greater impact (either positive or negative, depending on identification quality) on identification accuracy/eyewitness confidence (compared to jurors without the *Henderson* instructions), which would then influence jurors' final verdict decisions.

## Methods

### Participants

Four hundred sixty-eight jury-eligible community members (59% female; 41% male) were recruited via Craigslist.com. Participants came from diverse backgrounds (37.6% White, non-Hispanic; 15.4% Hispanic; 30.8% Black, non-Hispanic; 16.2% Other) and ranged in age

from 18-86 (*M* = 33.67, *SD* = 13.08). Participants received $30 in exchange for participating in the two-hour study.

**Design**

This study was a 3 (Instruction: No Instruction vs. *Henderson*-before eyewitness testimony vs. *Henderson*-after eyewitness testimony) X 2 (Quality of system variables: Good v. Poor) X 2 (Quality of estimator variables: Good v. Poor) factorial design.

**Trial Stimulus**

The trial video, *New York v. William Thomas Johnson* was loosely based on an eyewitness identification case that was later overturned (*Gregory v. City of Louisville*, 2006) and was filmed with professional actors. The video ran from 45-67 minutes long, depending on condition. The trial included opening and closing judicial instructions, opening and closing statements from both attorneys, as well as direct and cross-examination of a police officer, an eyewitness, an expert witness, the defendant, and friend of the defendant[1]. In the trial, the state charged the defendant with attempted rape in the first degree, alleging the defendant entered the victim's residence and attacked her in her bed. An expert forensic serologist testified that the hair found at the crime scene matched the hair taken from the defendant, but roughly 10% of males would also match. The defendant testified that he was at home sleeping at the time of the crime and his girlfriend served as an alibi.

[1] All manipulated system and estimator variable were mentioned three times: 1) in the opening statements, 2) during the eyewitness or police officer testimony, and 3) in the closing statements. The defense emphasized all variables that were good for the defense and the prosecution emphasized all variables that were good for the prosecution. In the *good system variable/poor estimator variable condition*, for instance, the prosecutor mentioned the good system variables in the opening and closing statements and also emphasized them during the police officer's testimony. The defense attorney mentioned the good estimator variables in his opening and closing statements and also emphasized them during the eyewitness' testimony. Thus participants were made aware of the positive or negative impact each variable may have had on the identification.

We conducted extensive pilot testing using 278 undergraduates, 30 community members, and 67 mTurkers ($N$ = 375)[2] to examine what we believed would lead to the highest and lowest conviction rates.[3] We observed an acceptable verdict split, with more guilty verdicts in the high conviction condition (50%) compared to the low conviction condition (25%, $n$ = 209).

**Manipulations**

**Instructions.** In all conditions, the judge provided standard instructions at the end of the trial, which included the definition of the charges brought against the defendant, as well as burden of proof and proof beyond a reasonable doubt. These were the only instructions jurors in the no instruction condition received. In the *Henderson*-before conditions, the judge provided research-based case-specific instructions prior to the eyewitness (victim) testimony (see Appendix A). In the *Henderson*-after conditions, the judge provided the same research-based case-specific instructions at the end of the trial (after defense closing statements). In all conditions, participants received a written version of the judicial instructions in addition to the verbal presentation.

**Estimator variables.** Each of the manipulated estimator variables (i.e., memory decay, weapon focus, and duration) were chosen based on the *Henderson* Court's decision that variations in these variables may have significant effects on eyewitness identification accuracy/eyewitness confidence. Specifically, the Court in *Henderson* notes that: "delays between the commission of a crime and the time an identification is made can affect reliability" (p. 79); "the presence of a weapon can impair a witness' ability to make a reliable identification and describe what the culprit looks like" (p. 73); and "the amount of time an eyewitness has to

[2] None of the participants in the pilot study participated in the full study.

[3] In the highest conviction condition, there were 6 *good* witnessing and identification variables present; in the lowest conviction condition, there were 6 *poor* witnessing and identification variables present.

observe an event may affect the reliability of an identification" (p. 74; *New Jersey v. Henderson, 2011*). The levels of each manipulated estimator variable were based on previous research showing significant differences in identification accuracy between levels. The manipulated estimator variables were the delay between the crime and subsequent identification (e.g., Shapiro & Penrod, 1986; Deffenbacher et al., 2006), weapon focus (e.g., Steblay, 1992; Fawcett, Russell, & Peace, 2013), and exposure duration (e.g., Shapiro & Penrod, 1986; Bornstein et al., 2012). In the good estimator conditions, the time delay was one day, there was no weapon present, and the exposure duration was 45 seconds. In the poor estimator conditions, the time delay was one month, there was a weapon present, and the exposure duration was 10-15 seconds.

**System variables.** Each of the manipulated system variables (i.e., identification procedure, lineup instructions, and confirmatory feedback) were chosen based on the *Henderson* Court's decision that variations in these variables may have significant effects on eyewitness identification accuracy and are the subject of new jury instructions. Specifically, the Court in *Henderson* notes that: showups, compared to lineups, "fail to provide a safeguard against witnesses with poor memories or those inclined to guess, because every mistaken identification in a showup will point to the suspect" (p. 69); informing the witness that the suspect may or may not be in the lineup is "regarded as one of the most useful techniques for enhancing the reliability of identifications" (p. 53); and confirmatory feedback can "reduce doubt and engender a false sense of confidence" and enhancement in a "witness' recollection of the quality" of the identification (p. 58; *New Jersey v. Henderson,* 2011). The levels of each manipulated system variable were based on previous research showing significant differences in identification accuracy/eyewitness confidence between levels. The manipulated system variables were the identification procedure (e.g., Steblay, Dysart, Fulero, & Lindsay, 2003), lineup instructions (e.g., Steblay, 1997; Quinlivan, 2012), and confirmatory feedback (e.g., Douglass & Steblay,

2006; Steblay, Wells, & Douglass, 2014). In the good system conditions, the police officer presented a lineup to the witness, informed the witness that the perpetrator may or may not be in the lineup, and did not tell the witness that she chose the suspect. In the poor system conditions, the police officer presented a showup to the witness, failed to mention that the perpetrator may or may not be present, and informed the witness that she chose the suspect.

## Measures

**Voir dire questionnaire.** Prior to watching the trial stimulus, all participants completed a questionnaire about their demographic characteristics. To ensure that all participants were jury eligible, participants indicated their age, and whether they were U.S. citizens, registered to vote, and had a driver's license.

**Post-trial measures.** Immediately following the trial, participants completed a post-trial questionnaire assessing verdict as well as evidence and witness perceptions.

**Verdict.** Participants indicated whether they believed the defendant was guilty or not guilty of attempted rape[4] (see Table 1).

**Perceptions of identification accuracy/eyewitness confidence (hypothesized mediators).** Participants indicated, on a scale of 1 to 9 (1 = Strongly Reduced Accuracy, 9 = Strongly Increased Accuracy), the impact of each of the manipulated variables (lineup instructions, duration, weapon, memory decay, identification procedure) on identification accuracy as well as the impact of confirmatory feedback on eyewitness confidence (see Table 2 for all descriptive statistics).

## Procedure

[4] Attempted rape cases bring with them unique juror perception issues and gender has been shown to play a significant role in rape trials (*see* Fischer, G. J., 1997; McNamara, Vattano, & Viney, 1993). We thus analyzed whether gender had an impact on verdict, which it did not, $p > .05$.

Participants came to the psychology lab to complete the study. After providing informed consent they completed a voir dire questionnaire. Participants were randomly assigned to condition and watched the simulated trial in groups of anywhere from five to fifteen, depending upon sign-up rate. Prior to the judge's instructions, whether before the eyewitness testimony or at the end of the trial, the video was paused and the experimenter handed out paper copies of the instructions and informed participants that they could follow along with the judge's instructions as the video resumed. After the trial, participants individually completed post-trial questionnaires and were informed that they could refer to their written copy of the instructions at their discretion. At the end of the study, all participants were debriefed, compensated, and dismissed.

**Data analysis plan**

Results were computed using a three-step hierarchical logistic regression with verdict as the dependent variable[5] (see Table 3).

**Step 1 variables.** All independent variables (system variables, estimator variables, and instruction type) were included in Step 1. Two variables were two-level independent variables (System variables, Estimator variables) and the other two were dummy variables (*Henderson*-before, *Henderson*-after) representing comparisons between instruction conditions with No Eyewitness Instruction as the baseline.

**Step 2 variables.** All two-way interactions involving *Henderson* were included in Step 2.

**Step 3 variables.** All hypothesized mediator variables were included in Step 3 (i.e., perceptions of identification accuracy/eyewitness confidence (see below)).

**Mediation analyses.** Six individual regressions were run on variables that we hypothesized would mediate the relationship between our independent variables and verdict.

[5] A power analysis (using G*Power) of our most complex hypothesis – an interaction between instruction type, system variables, and estimator variables – suggested a total sample size of 476 to have sufficient power $(1-\beta) = .80$ and $\alpha = .05$ to detect a small-medium sensitivity effect.

These variables concerned perceptions of identification accuracy/eyewitness confidence (lineup instructions, duration, weapon, memory decay, procedure, and confirmatory feedback; see Table 4).

## Results

### Manipulation checks

Significantly more people reported that the judge specifically mentioned all three of the manipulated estimator variables (i.e., memory decay, weapon focus, and duration) when they heard *Henderson* instructions (pre: 63.0%; post: 63.6%) compared to those that did not hear *Henderson* instructions (2.6%; $\chi^2(6, N = 227) = 169.31$, $p < .001$, $\varphi = .86$). More people reported that the judge specifically mentioned all three of the manipulated system variables (i.e., lineup instructions, confirmatory feedback, and showup) when they heard the *Henderson* instructions (pre: 44.6%; post: 50.7%) compared to those that did not hear the instructions (4.0%; $\chi^2(6, N = 222) = 134.07$, $p < .001$, $\varphi = .78$). The majority of participants correctly reported the quality of the manipulated system variables (95.5% got at least 2 of 3 questions correct) and estimator variables (97.5% got at least 2 of 3 questions correct). There were no differences in the pattern of results as a function of whether participants answered manipulation checks correctly.

### Verdict

The overall model was statistically significant ($\chi^2$ (6, 457) = 201.86, $p < .001$). Nagelkerke's $R^2$ of .54 indicated a moderately strong relationship between prediction and grouping. Prediction success overall was 83% (87.6% for not guilty and 76.2% for guilty). For simplicity, only results pertaining specifically to our hypotheses are presented here in the text; all results of the logistic regression are presented in Table 3, and individual mediator analyses are presented in Table 4. For all analyses, positive *Exp(β)* indicates an increase in the likelihood of convictions; negative *Exp(β)* indicates a decrease in the likelihood of convictions.

**Hypothesis one: Sensitivity to identification quality**

Overall, those who received either *Henderson*-before (32.2%) or *Henderson*-after (39.8%) instruction rendered fewer guilty verdicts compared to those who heard no eyewitness instruction (51.6%, $\chi^2$ (2, N = 448) = 12.22, $p$ = .002, φ = .17). There were no significant interactions between the *Henderson* instructions, and system and estimator variables conditions on verdict (all *p*s > .23). Thus, rather than increasing jurors' sensitivity to witnessing and identification conditions, the *Henderson* instructions induced skepticism by reducing convictions regardless of eyewitness quality.

**Ratings of Identification Accuracy and Eyewitness Confidence**

Results of the individual mediation analyses indicated that each of the independent variables significantly affected participants' ratings of identification accuracy/eyewitness confidence. When system variables were good, participants rated unbiased lineup instructions ($\beta$ = .51) and the use of a lineup ($\beta$ = .33) as increasing the likelihood of a correct identification, compared to when system variables were poor. Additionally, when system variables were poor, participants rated the confirmatory feedback as having increased the witness' confidence compared to when system variables were good ($\beta$ = -.56; all *p*s < .05; see Table 4 for all $\beta$ values). All of these effects were in the expected direction, such that perceptions of greater identification accuracy/eyewitness confidence were associated with good system variables.

When estimator variables were good, participants gave higher identification accuracy ratings for a long duration ($\beta$ = .15), absence of a weapon ($\beta$ = .16), and a short time delay ($\beta$ = .31), as compared to when estimator variables were poor. All of these effects were in the expected direction, such that perceptions of greater identification accuracy were associated with good estimator variables.

Participants who heard *Henderson*-before instructions, compared to no eyewitness instruction, rated lineup instructions ($\beta$ = -.10), duration ($\beta$ = -.13), weapon presence ($\beta$ = -.12), memory decay ($\beta$ = -.10), and identification procedure ($\beta$ = -.17) as more strongly reducing the accuracy of the identification, regardless of the quality of system and estimator variables present in the trial. These results indicate that the *Henderson*-before instructions produced a skepticism effect.

Participants who heard *Henderson*-after instructions, compared to no eyewitness instruction, rated lineup instructions ($\beta$ = -.11), memory decay ($\beta$ = -.10), and identification procedure ($\beta$ = -.13) as more strongly reducing the accuracy of the identification, regardless of the quality of system and estimator variables present in the trial (see Table 4 for all values), which further supports the presence of a skepticism effect.

There were no significant interactions between instruction and system/estimator variables on participants' ratings of identification accuracy and eyewitness confidence. [6]

**Hypothesis two: Mediation**

For the purposes of our study, a mediated relationship occurred when there was a total effect of an independent variable (ideally an interaction of instructions and system/estimator variables) on verdict (significance in Step 1), and when that independent variable had an effect on the mediator (significance in individual regression analyses) thus explaining away the total effect of the independent variable on verdict (non-significance in the Step 3); see Baron &

[6] While there were no significant interactions between instruction and system and estimator variables on verdict, results indicated there *was* a significant interaction between *Henderson*-before instructions and estimator variables on jurors' ratings of the impact of the weapon on the likelihood of an accurate identification. When estimators were poor, jurors who heard the *Henderson*-before instructions rated the weapon as having decreased identification accuracy significantly more so than those who heard no eyewitness instruction ($p < .01$). When estimators were good, there was no difference in jurors' ratings of the impact of the weapon as a function of instruction type (all *p*s >.65). However, given the number of tests included in the analysis, we suspect this interaction may be the result of Type I error.

Kenny, 1986). Since we did not find sensitivity effects, we sought an explanation for the skepticism effect through mediation analysis.

The effect of *Henderson*-before instructions on verdict was not mediated by participants' perceptions of the evidence (*p*s > .05). That is, *Henderson*-before instructions remained significant in Step 3, indicating that the impact variables did not account for the effect of instructions on verdict. The effect of *Henderson*-after instructions on verdict, however, was mediated by participants' perceptions of the evidence, reflected in the non-significance in Step 3 ($p = .36$). Specifically, the skepticism effect of *Henderson*-after instructions on verdict was mediated by participants' ratings of the identification procedure ($\beta = .15$, $p < .01$). Participants who heard the *Henderson* instructions at the end of the trial believed the identification procedure increased accuracy (regardless of whether the police used a showup or a lineup), which in turn increased convictions (see Table 3 for all stepwise $\beta$s, Table 4 for all mediation $\beta$s).

## Discussion

According to previous research on the efficacy of the *Henderson* instruction (e.g., Papailiou, et al., 2014; Laub et al., 2014) the *Henderson* instructions may be no more effective than previous versions of eyewitness identification instructions (i.e., *Telfaire* instructions[7]). The purpose of the *Henderson* instructions is to not only inform jurors about relevant system and estimator variables, but to directly instruct jurors about the potential impact of such variables on the accuracy of an eyewitness identification. Results of the current study indicate that jurors were aware of the effects of some system and estimator variables, but they failed to integrate this knowledge into their verdict decisions. Although the *Henderson* Court reasoned that these research-based instructions would help make jurors aware of the potential effects of system and

[7] Eyewitness identification instructions developed after decision in *United States v. Telfaire, 1972.*

estimator variables on eyewitness identification quality, the results of the current study suggest they are not quite as effective as anticipated.

**Sensitizing Effects**

If the *Henderson* instructions effectively sensitized jurors to identification quality (H1), there would have been interactions between instructions (before- and/or after-) and system and estimator variables on verdict. That is, participants who heard the *Henderson* instructions would, according to the *Henderson* Court's reasoning that these instructions would help jurors evaluate evidence "critically and objectively" (p. 124), render more guilty verdicts for the strong case (good system and estimator variables) and fewer guilty verdicts for the weak case (poor system and estimator variables) than those who heard no eyewitness instructions. There was no evidence of such an interaction on verdict. Rather than sensitizing participants, the *Henderson* instructions (both before- and after-) led to an overall skepticism effect among participants.

Future research should address whether this instruction, in its current and/or some alternative format, can influence perceptions of various system and estimator variables, including but not limited to the ones examined in this study. It is likely that even though jurors understood how the presence of a weapon, for example, affected the accuracy of the identification, this was simply not enough to impact their verdict. Had instructions interacted with multiple system and estimator variables, however, it is possible that we would have seen greater sensitivity in verdicts. Perhaps jurors weigh certain factors more heavily than others. That is, the presence of some poor witnessing conditions (e.g., weapon focus) may have a greater impact on perceptions of the identification accuracy compared to other witnessing conditions (e.g., duration). Moreover, jurors may interpret certain combinations of variables as having more of an impact on identification accuracy than other combinations of variables. Future research should examine

whether various combinations of system and estimator variables, in combination with the *Henderson* instructions, increase jurors' sensitivity.

We predicted there would be sensitivity effects in jurors' verdicts, but the desired sensitivity effect from judicial instructions may be quite difficult for jurors to achieve (Cutler et al., 1989). That is, although laypeople may understand some eyewitness issues (e.g., Desmarais & Read, 2010), if trained judges are not proficient integrators of knowledge (Slovic, 1969; Hoffman, Slovic, & Rorer, 1968) and are not particularly knowledgeable about eyewitness testimony (Wise & Safer, 2003), how can courts expect laypeople to be?

**Skepticism Effects**

Results of the current study were consistent with previous research (e.g., Cutler et al., 1989; Papailiou et al., 2014), in that the instruction induced skepticism of eyewitness evidence. Participants who heard either the *Henderson*-before or *Henderson*-after instruction rendered fewer guilty verdicts than those who heard no eyewitness instruction, regardless of the quality of the witnessing and identification conditions. The presence of this skepticism effect indicates that the instructions might have produced an overall more critical view of eyewitness evidence or confusion, and may thus be counter to the *Henderson* Court's hypothesis that the instructions would help to "avoid possible confusion to jurors created by dueling experts" (p. 126). Regardless of whether the conjecture regarding confusion in the face of dueling experts is correct (Levett & Kovera, 2008; Levett & Kovera, 2009), the *Henderson* instructions did not have their intended sensitizing effect. Thus, although our data does not support the *Henderson* Court's assumption that the instructions help jurors analyze and evaluate the credibility of eyewitness identification, it does indicate that the instructions lead jurors to become more critical of eyewitness evidence.

The skepticism effects and the absence of an interaction effect indicate that while jurors' impressions can be influenced by the quality of system and estimator variables, instructions improve neither the quality of evidence evaluations nor jurors' final judgments. However, while the *Henderson* Court did not explicitly hypothesize skepticism effects, such effects may result in a reduction in erroneous convictions (although at the cost of fewer convictions in cases that may warrant convictions), which *was* the intention of the *Henderson* Court.

**Mediation**

Our second hypothesis, that perceptions of the eyewitness variables would mediate the effect of instructions on verdict, was partially supported. The total effect of the *Henderson*-after instruction on verdict *was* mediated by perceived impact of the identification procedure on the likelihood of an accurate identification. Jurors believed the identification procedure increased accuracy and subsequently increased convictions when they heard *Henderson* instructions, regardless of whether the procedure was a showup or lineup. It appears that drawing attention to the identification procedure, via instructions, may increase jurors' credulity rather than sensitize them to differences in identification accuracy. That is, instead of becoming skeptical of the identification, jurors interpret all variations in the procedure as having increased identification accuracy (see similar findings of "overbelief" in Smalarz & Wells, 2014; Ramirez, Zemba, & Geiselman, 1996). Since we did not manipulate each system variable separately, we cannot discern whether instruction on a particular system variable is more confusing or misleading than another. We do know from previous research (e.g., Smalarz & Wells, 2014), however, that evaluators have a difficult time assessing the impact of confirmatory feedback on identification accuracy. Future research should focus on determining whether jurors are more capable of learning about certain system and estimator variables, compared to others, and incorporating this knowledge into their final verdict decisions.

**Limitations and Future Directions**

As is the case with most jury decision-making studies, the current study has its limitations. One limitation is the reliance on a single case in which the eyewitness was also the victim. While the research on eyewitness identification applies to bystanders as well as eyewitness-victims, some research suggests eyewitness-victims may make less accurate identifications compared to bystanders (Kassin, 1984). Thus, future research should utilize a case in which the eyewitness is a bystander, not a victim.

Another limitation is that with a large number of variables in the regression analyses, particularly in Step 3, there is an increased risk of committing a Type I error. Another limitation is the fact that many participants did not correctly report hearing the presented eyewitness instructions. This may have been a result of the wording of the manipulation check question, or simply the result of jurors having difficulty remembering an onslaught of new information contained in a trial. While this finding is somewhat disconcerting, we believe misremembering jury instructions is an adequate portrayal of real-world juries. A participant who recalls that the judge mentioned confirmatory feedback may not necessarily understand the instruction better than a participant who does not remember the instruction.

Despite these limitations, we believe we created the most ecologically valid test of the *Henderson* instructions to date. We recruited adult community members and created a video trial based on an actual case. The results of this study should thus be used as an impetus for replication using different stimuli and various system and estimator variables.

Overall, this study presents a significant contribution to the literature given that few have tested the *Henderson* instructions or the proposition that the New Jersey Supreme Court itself has suggested. Taken together, the results of this study suggest that the current *Henderson* instructions should be modified to improve juror sensitivity to various witnessing and

identification conditions. Jurors may either misinterpret (e.g., interpret feedback as an indication that the eyewitness made an accurate identification) or have difficulty grappling with the effects of these variables after hearing the *Henderson* instructions. If this is the case, future studies should focus on simplifying the instructions to make them more comprehensible for jurors (e.g., Greene 1988).

Future research should also focus on ways to make the instructions stand out more, to increase jurors' understanding of and memory for the instructions. According to research on the I-I-Eye teaching aid (e.g., Pawlenko et al., 2013; Safer et al., 2016), teaching jurors about the potential effects of system and estimator variables and providing guidelines for assessing these effects is a promising way to sensitize jurors to identification accuracy. Thus, modifications to the *Henderson* instructions to incorporate the I-I-Eye teaching style may be more effective than the current instructions. If researchers are able to determine how to sensitize jurors to each of these variables independently, we will be closer to developing the most effective eyewitness identification instructions yet.

## References

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 & Law, 18*, 473-490. doi:10.1080/1068316X.2010.508458

*Commonwealth v. Gomes*, 453 Mass. 506 (2009).

Cutler, B. L., Dexter, H. R., & Penrod, S. D. (1989). 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. & Penrod, S. D. (1999). *The case against traditional safeguards: Preventing mistaken convictions in eyewitness identification trials.* In R. Roesch & S. Hart (Eds.), *Psychology and Law: State of the Discipline*: New York.

Deffenbacher, K. A., Bornstein, B. H., & Penrod, S. D. (2006). Mugshot exposure effects: Retroactive interference, mugshot commitment, source confusion, and unconscious transference. *Law and Human Behavior, 30*, 287-307. doi:10.1007/s10979006-9008-1

Desmarais, S. L. & Read, J. D. (2010). 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

Dillon, M. K. & Penrod, S. D. (2014). Judicial instruction on eyewitness identification: A Meta-Analysis. *Unpublished Manuscript*.

Douglass, A. B. & Steblay, N. (2006). Memory distortion in eyewitnesses: A meta-analysis of the post-identification feedback effect. *Applied Cognitive Psychology, 20,* 859-869. doi:10.1002/acp.1237

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

Fischer, G. J. (1997). Gender effects on individual verdicts and on mock jury verdicts in a simulated acquaintance rape trial. *Sex Roles, 36,* 491-501.

Greene, E. (1988). Judge's instruction on eyewitness testimony: Evaluation and revision. *Journal of Applied Social Psychology, 18,* 252-276. doi:10.1111/j.15591816.1988.tb00016.x

*Gregory v. City of Louisville*, 444 F3d 725 (2006).

Gross, S. R., Jacoby, K., Matheson, D. J., Montgomery, N., & Patil, S. (2004). *Exonerations in the United States: 1989 through 2003.* Retrieved from: http://www.internationaljusticeproject.org/pdfs/ExonerationReport4.19.04.pdf

Hoffman, P. J., Slovic, P., & Rorer, L. G. (1968). An Analysis-of-Variance model for the assessment of configural cue utilization in clinical judgment. *Psychological Bulletin, 69,* 338-349. doi:10.1037/h0025665

Kassin, S. M. (1984). Eyewitness identification: Victims versus bystanders. *Journal of Applied Social Psychology, 14,* 519-529.

Kassin, S. M. & Wrightsman, L. S. (1979). On the requirements of proof: The timing of judicial instruction and mock juror verdicts. *Journal of Personality and Social Psychology, 37,* 1877-1887. doi:10.1037/0022-3514.37.10.1877

Laub, C. E., Kimbrough, C. D., & Bornstein, B. H. (2014). Mock juror perceptions of eyewitnesses vs. earwitnesses: Do safeguards help? *Manuscript under review.*

Levett, L. M. & Kovera, M. B. (2008). The effectiveness of opposing expert witnesses for educating jurors about unreliable expert evidence. *Law and Human Behavior, 32,* 363-374. doi:10.1007/s10979-007-9113-9

Levett, L. M. & Kovera, M. B. (2009). Psychological mediators of the effects of opposing expert testimony on juror decisions. *Psychology, Public Policy, and Law, 15,* 124-148. doi:10.1037/a0016309

McNamara, K., Vattano, F., & Viney, W. (1993). Verdict, sentencing, and certainty as a function of sex of juror and amount of evidence in a simulated rape trial. *Psychological Reports, 72,* 575-583.

Moore, D. R. (2010). *The effect of research-informed jury instructions on potential jurors' verdicts in eyewitness case vignettes* (Unpublished doctoral dissertation). Alliant International University, Fresno.

*Neil vs. Biggers,* 409 U.S. 188, 93 S. Ct. 375; 34 L. Ed. 2d 401 (1972).

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

Papailiou, A. P., Yokum, D. V. & Robertson, C. T. (2015). The novel New Jersey eyewitness instruction induces skepticism but not sensitivity. *PLoS ONE, 10: e0142695.* doi:10.1371/journal.pone.0142695

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

Quinlivan, D. S. (2012). Effects of pre-admonition suggestions on eyewitness' beliefs, expectations, choosing rates, and retrospective judgments. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 72,* 7673.

Ramirez, G., Zemba, D., & Geiselman, R. E. (1996). Judges' cautionary instructions on eyewitness testimony. *American Journal of Forensic Psychology, 14,* 31-66.

Safer, M. A., Murphy, R. P., Wise, R. A., Bussey, L., Millett, C., & Holfeld, B. (2016).

Educating jurors about eyewitness testimony in criminal cases with circumstantial and forensic evidence. *International Journal of Law and Psychiatry.* doi:10.1016/j.ijlp.2016.02.041

Scheck, B., & Neufeld, P. (2006). Toward the formation of innocence commissions in America. *Judicature, 86,* 104-105.

Scheck, B., Neufeld, P., & Dwyer, J. (2001). *Actual innocence: Five days to execution, and other dispatches from the wrongly convicted*. Oregon: Good Works on Earth.

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

Slovic, P. (1969). Analyzing the expert judge: A descriptive study of a stockbroker's decision processes. *Journal of Applied Psychology, 53,* 255-263. doi: 10.1037/h0027773

Smalarz, L. & Wells, G. (2014). Post-identification feedback to eyewitnesses impairs evaluators' abilities to discriminate between accurate and mistaken testimony. *Law and Human Behavior, 38,* 194-202.

Steblay, N. K., 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,* 523-450. doi: 10.1023/A:1025438223608

Steblay, N. K. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior, 16,* 413-424. doi:10.1007/BF02352267

Steblay, N. K. (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., Wells, G. L., & Douglass, A. B. (2014). The eyewitness post identification

feedback effect 15 years later: Theoretical and policy implications. *Psychology, Public Policy, and Law, 20*, 1-18. doi:10.1037/law0000001

Massachusetts Supreme Court. (2015). Statement of the Supreme Judicial Court: Model Jury Instructions on Eyewitness Identification.

*United States v. Telfaire*, 469 F.2d 552 (D.C. Cir. 1972).

Wells, G. L. (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

Wise, R. A., Fishman, C., & Safer, M. A. (2009). How to analyze the accuracy of eyewitness testimony in a criminal case. *Connecticut Law Review, 42*, 435-513.

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