Effects of Evidence Evaluation Placement on Juror Sensitization to Non Case-Specific New Jersey *Henderson* Instructions

Jayson R. Chojar

CUNY John Jay College of Criminal Justice

Author Note

Jayson R. Chojar, Under Marlee K. Berman, Department of Psychology, John Jay College

National Science Foundation PD 98-1372 awarded to Distinguished Professor Steven D. Penrod supported this research.

Correspondence concerning this article should be addressed to Jayson R. Chojar

Abstract

Eyewitness testimony is crucial to the proceedings of a criminal trial. However, eyewitnesses are often subject to a number of factors that diminish their ability to recall events accurately. The New Jersey State Supreme Court recently adopted the *Henderson* instructions to sensitize jurors to the possibility of eyewitness inaccuracies. Few previous studies have demonstrated the success of *Henderson* instructions. As such, this study hypothesized that the current, case-specific New Jersey *Henderson* instructions should be amended in order to better sensitize participants to the system and estimator variables that influence eyewitness identification accuracy. This study also proposed that these new, non-case-specific instructions should be evaluated alongside trial evidence prior to a mock juror's rendering of a verdict. The results showed a statistically significant two-way interaction between instruction type and variable quality when verdict occurred after the evidence evaluation that highlighted the non-case-specific *Henderson* instructions. The amended instructions effectively alerted jurors to stimuli that commonly alter eyewitness perceptions. The *Henderson* instructions were shown to be most effective when administered as non-case-specific and when jurors were prompted to consider the factors affecting eyewitness accuracy prior to rendering a verdict. Implementing these practices may help jurors evaluate eyewitness identification accuracy and decrease the prevalence of false convictions if other state or federal courts adopt them.

Introduction

Under the United States Judicial System, all accused persons are entitled to a trial by a jury of their peers. The Seventh Amendment to the Constitution of the United States dictates that "In suits at common law, where the value in controversy shall exceed twenty dollars, the right of trial by jury shall be preserved, and no fact tried by a jury, shall be otherwise re-examined in any Court of the United States, than according to the rules of the common law" (U.S. Const. amend. VII). This amendment guarantees all citizens of the United States the right to stand trial and have their fate decided by laypersons rather than by a judge. The law was originally intended to reassure the people that the newly formed and increasingly powerful federal government would not exploit them. Today, the Seventh Amendment has endured as one of our nation's legal system's most notable characteristics. There are two major types of trials in the United States Justice System: civil and criminal. In civil cases, each juror is instructed to render his verdict based solely on a preponderance of the evidence; criminal cases require additional proof.

In criminal trials, the burden of proof is relegated to the prosecution, whose evidence must substantiate guilt beyond a reasonable doubt. Legal professionals and jurors alike disagree on what constitutes this term—"reasonable doubt. Furthermore, jurors in criminal cases are often uncertain of the meaning or application of reasonable doubt (Shapiro, 1991). A juror who submits a guilty verdict does not need to be certain that the defendant committed the crime (Wright 2007). Guilt beyond a reasonable doubt merely dictates that guilt can be safely assumed despite the possibility of a situation in which the defendant may in fact be innocent. An accepted level of certainty, which is considered "beyond reasonable doubt," is a probability of 0.9 (Magnussen, S., Eilertsen, D., Teigen, K., & Wessel, E., 2014). The figure 0.9 is determined by a juror's percent chance of guilt. A 90% chance of guilt as deemed by the juror corresponds to a probability of 0.9. It is important for jurors to be accurately informed of what constitutes reasonable doubt and what level of doubt permits a guilty verdict. However, when a verbal "reasonable doubt" instruction is administered to the jury, jurors tend to reconsider the possibility of the defendant's innocence or reassure themselves that their belief in guilt is enough to convict (Wright, 2007). In addition, the defense's tactic of shifting blame onto another suspect via a potential cause for motive has been shown to decrease juror confidence due to its diminishing effect on assurance. Reasonable doubt further comes into play when considering the validity of evidence presented to the jury.

Eyewitness testimony is of paramount importance to a criminal trial, for it may be the single most compelling piece of evidence in the eyes of the jury (Loftus, 1979). While eyewitness testimony is central to a clear and developed reconstruction of a crime, witnesses may present an incomplete or incorrect recollection of events, which can invalidate an otherwise fair trial. Jurors are influenced greatly by eyewitness testimony in terms of their decisions to convict or acquit (Pozzulo & Dempsey, 2009). The testimony they observe is perhaps the greatest factor leading to their eventual verdict. As the U.S. Supreme Court has noted, "There is almost nothing more convincing than a live human being who takes the stand, points a finger at the defendant and says, 'That's the one'" (Loftus 1979, p. 19). Eyewitness accounts, however credible they may seem, have a propensity to convey a poorly represented or inaccurate version of events, and this severely discounts their reliability (McGonigle & Emily, 2008; *United States v. Wade*, 388 U.S. 218, 228, 1967). Furthermore, many jurors fail to recognize the often-unreliable nature of eyewitness testimony and they often favor it despite the presence of more reliable evidence (Devenport, Penrod, & Cutler, 1997). Psychological effects on human perception are triggered by certain stimuli. These mental cues create fallacies in eyewitness testimony.

Among the many variables affecting eyewitness testimony are estimator variables, which are variables out of the hands of the criminal justice system (Wells, 1978). Estimator variables affecting the accuracy of eyewitness testimony are caused by psychological influences on recall and awareness rather than post-crime intervention by police (Fradella, 2007). Estimator variables elicit predictable responses from eyewitnesses; jury members must be sensitized to these reactions. Past research has found that eyewitnesses have frequently been shown to affect the accuracy of an eyewitness identification when faced with the following estimator variables: lighting, crime duration, weapon focus, disguise, identification delay, stress, estimation of time passage, and distance. While these variables may seem intuitive, jurors are often unaware of the impact of their presence and fail to adequately consider each variable when rendering a verdict. Furthermore, estimator variables are important to understanding the mistakes made by witnesses (Wells, Memon, & Penrod, 2006).

An eyewitness's ability to commit a certain perpetrator to memory is influenced by several physical and psychological cues. For instance, in a crime scenario in which lighting conditions are poor, eyewitnesses are less accurate (Wells, Memon & Penrod, 2006). When sight lines are compromised, eyewitnesses are more likely to make mistakes in identification. Additionally, at long distances, eyewitness ability to decipher facial features and remember the identities of suspects decrease. Another factor that tends to decrease eyewitness reliability is the presence of a disguise. This inhibits the eyewitness's ability to encode memory in regards to facial features (Mansour, et al., 2012). Time influences considerations in determining the accuracy of identification as well, whether during or after the crime. When witnesses have only a short duration of time (e.g., 10-15 seconds or less) to view the perpetrator of a crime, they are much more likely to make an inaccurate identification (Kassin, Ellsworth, & Smith, 1986). Note that exposure duration played a significant role in this study. Also, if there is a long period of time between the events of the crime and when the witness is asked to recall those events, accuracy declines; this is a result of decreased memory retention over time (Shapiro & Penrod, 1986). Eyewitnesses consistently overestimate short durations of time when asked to recall the events of a crime (Wells, Memon, & Penrod), and especially in short-duration crimes, the presence of a weapon is detrimental to a witness's credibility. If a weapon is present, eyewitnesses have a tendency to focus on it rather than on the appearance and face of the culprit, thereby decreasing their identification accuracy (Wells, Memon & Penrod, 2006). In cases of larger exposure duration, accuracy is greater even though the weapon is present. Certain research finds that high stress scenarios inhibit accurate recall and identification (Deffenbacher, et al., 2004). The effect of stress, however, is widely contested, with no definite consensus. Estimator variables directly affect witness perceptions, and they also serve as mechanisms by which the validity of eyewitness testimony can be judged.

Additionally, system variables, or those variables in the control of the administrators who question an eyewitness after a crime, have been shown to affect identification accuracy (Wilford & Wells, 2013). The lineup process has enormous potential for undue influence on the eyewitness if not handled properly. The Supreme Court cites the Fifth and Sixth Amendments as rationale for protections against unfairly suggestive police lineup procedures" (*Stovall v. Denno*, 1967; *US v. Wade*, 1967). System variables include biased lineup instruction, confirmatory feedback, multiple viewings, and foil representation, among others. When a lineup administrator fails to disclose to the eyewitness that the suspect may or may not be present in the lineup (i.e., biased lineup instructions), the eyewitness may feel compelled to make an identification (Cutler, Dexter, & Penrod, 1990). This might lead the eyewitness to pick the individual in the lineup who bears the most resemblance to the perpetrator (Sheehan, 2011). After an identification is made, it is important for administrators to avoid verbal contact with the eyewitness, for feedback, whether confirmatory or not, triggers witnesses to reconsider their certainty, confidence, and clarity (Wells & Bradfield, 1998). When choosing lineup fillers, it is important to match both the description of the culprit and the appearance of all suspects (Sheehan, 2011). Officials can further control other system variables, such as the multiple viewings variable, which was examined in this study. If an eyewitness views the same face in multiple lineups, it is much more likely that the suspect will be misidentified as present at the scene of the crime (Pezdek & Blandon-Gitlin). Foil bias manifests itself when the lineup fillers share little physical resemblance with the suspect, and this negatively affects the validity of identification, as the suspect stands out amongst the crowd (Sheehan, 2011). One method that has succeeded in minimizing the risk of influencing the witness is the implementation of a double-blind lineup procedure (Wilford & Wells, 2013). In this case, the lineup administrator is unaware of the suspect's identity; therefore, he cannot engage in suggestive questioning procedures (Sheehan, 2011).

Regardless of whether an eyewitness was positively or negatively influenced by these system and estimator variables, the jury should be informed of the potential effects of these variables. Many studies have supported the notion that eyewitness identifications are especially susceptible to error (Goldstein et al., 1989), which warrants increased juror awareness. When addressing the jury, they must be made aware of the effects of these system and estimator variables (expert testimony, judge's instructions, etc.). Written instructions have served as longstanding safeguards against eyewitness misidentification. Judges may use instructions to focus jurors' attention on eyewitness inaccuracies.

The tendency of eyewitness testimony to mislead the jury has necessitated certain protocols to sensitize jurors to its dangers. The *Telfaire* instructions, named so after the defendant in *United States v. Telfaire, 1972*, have been the most prevalent set of instructions in years since 1972; they aim to educate the jury in regards to the innate hazards of eyewitness identification, the requisite cautioned examination of such evidence, and the relatively high frequency of unreliable eyewitness identification. The *Telfaire* instructions highlight 13 factors key to eyewitness identification, "including the strength of the identification [used to determine witness confidence], the circumstances under which the identification was made, and the length of time between the crime and identification" (Greene, 1988). These instructions represented the legal standard since their implementation in 1972 (Greene, 1988). One study examined whether jurors' verdicts change if they were given a cautionary message (Katzev & Wishart, 1985). In Greene's study, the *Telfaire* instructions were administered to half of mock jurors, who were asked to deliberate to a unanimous verdict after viewing one of two trials; one trial had a strong eyewitness identification, while the other had a weak eyewitness identification. The results showed little to no effect of the instructions. The *Telfaire* instructions were again examined in 1990, where researchers found that the instructions resulted in "some skepticism [towards the eyewitness] but no sensitization" (Cutler, Dexter, & Penrod, 1990). These studies, among others, discounted the effectiveness of the *Telfaire* instructions; as a result, attempts at alternative, more efficient methods of instructing the jury were made. Finally, in 2012, after the decision in State v. Larry R. Henderson (2011), New Jersey issued an expanded set of instructions to replace the *Telfaire* instructions. The current study investigated these recently approved New Jersey *Henderson* instructions. While the Henderson instructions themselves have seldom been experimented with, prior research indicates that instructions provided to the jury by the judge are often misinterpreted (Greene 1988).

Jurors who preside over criminal trials often enter with their own misconceptions with respect to these estimator variables (Wells et al., 2006), which may prevent them from making unbiased decisions about an identification. Upon recent experimentation, the approved New Jersey Henderson instructions did not strengthen juror perceptions of eyewitness identification quality. For this reason, the instructions did not have a large effect on evidence evaluation. Rather than sensitizing jurors to the caliber of eyewitness identification, the instructions produced a general mistrust of eyewitnesses (Berman, et al., 2014). The instructions aimed to increase confidence-weighted guilty verdicts in cases of strong identifications, but regardless of the quality of the identification, guilty verdicts decreased among jurors who heard the Henderson instruction. The instructions were somewhat effective, however, as the data suggested that participants were able to apply their knowledge of the instructions towards their consideration of some system and estimator variables (Berman et al., 2014). But certain variables did not seem to require instructions; jurors were able to recognize the effects of variables such as duration and identification procedure and appropriately factor them into their evaluations of evidence (Berman et al., 2014). This prompted further consideration of how the Henderson instructions addressed variables jurors had trouble discerning the effects of.

The purpose of the current study was to examine whether the Henderson instruction are more effective when jurors are instructed to evaluate the evidence and the judicial instructions prior to rendering a verdict. The study consisted of 12 conditions that manipulated two system (multiple viewings and lineup instructions) and two estimator variables; three instruction types (Standard instructions [reasonable doubt, burden of proof], *Henderson*-case-specific, and *Henderson*-non-case-specific); and In the case of this study, the lighting and exposure duration estimator variables were manipulated in the procedure; nonetheless, jurors were exposed implicitly to many, if not all, of the variables discussed.

Objective

How can prompting jurors to evaluate the evidence prior to rendering a verdict help sensitize jurors to eyewitness identification accuracy?

Hypothesis

Jurors who are prompted to consider an eyewitness identification and non-case-specific judicial instruction prior to their evidence evaluation will be better sensitized to the quality of identification and thereby render more accurate rulings.

Method

Participants

295 community members took part in this study (55% Male, 44% Female; ranging in age from 18 to 71). Participants were of diverse backgrounds (73% White, non-Hispanic; 6% Hispanic; 9% Black, non-Hispanic; and 12% Other). Participants were recruited via Amazon Mechanical Turk (mTurk) to ensure a randomized, varied sample of the general population. As this study pertains to the criminal justice system, all participants were jury eligible (U.S. citizens, age 18 and over) so that the results of the study could generalize to the larger population. Participants received $1.50 compensation for participating in the thirty-minute study.

Design

The study was organized into a 3 (Instruction: Henderson bad v. Henderson both v. No Henderson) X 2 (System/Estimators: Good v. Poor) X 2 (Verdict Prompt: Pre v. Post) fully-crossed factorial design.

Trial Stimulus

The novel trial transcript was based on the fictitious armed robbery of Stone Jewelers, a neighborhood jewelry store. The trial included opening and closing statements from both attorneys, direct and cross-examinations of a police officer, the victim, and the defendant's girlfriend, and a closing instruction regarding burden of proof. In the trial, the defendant was charged with armed robbery, alleging that the defendant entered the victim's establishment and robbed her at gunpoint. The defendant's girlfriend gave credence to his alibi—she claimed they were attending a neighborhood barbecue—but she admitted to leaving the function for two hours to assist her mother, leaving the defendant unaccounted for.

Manipulations

Instructions. In all conditions, the judge provided standard instructions regarding burden of proof and proof beyond a reasonable doubt at the end of the trial. These were the only instructions jurors in the no instruction condition received. In the *Henderson*-case-specific conditions, the judge provided research-based issue-specific instructions that highlighted the effects of bad system and bad estimator variables. In addition, certain variables that were held constant were addressed. In the *Henderson*-non-case-specific conditions, the judge provided research-based instructions that were not case-specific. That is, participants in the *Henderson*-non-case-specific conditions received instruction on all of the constant and manipulated system and estimator variables, regardless of whether the quality of system and estimators was good or poor. The *Henderson* instructions used in this study were an abbreviated version of the instruction developed in *New Jersey v. Henderson* (2011;http://www.judiciary.state.nj.us/pressrel/2012/jury_instruction.pdf).

Estimator variables. The manipulated estimator variables were the lighting conditions during the crime and exposure duration. In the good estimator conditions, all of the store lights were turned on and the victim was able to get a good look at the robber's face for almost one minute. In the poor estimator conditions, most of the store lights were turned off and the victim was only able to view the robber's face for roughly ten seconds. The levels of each estimator variable were chosen based on previous research showing that these levels reflect conditions shown to produce differences in identification accuracy.

System variables. The manipulated system variables were the lineup instructions and the number of viewings. In the good system conditions, the police officer 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 showed Ms. Stone some mugshots of possible suspects prior to the lineup, one of which was the defendant's, and failed to mention that the perpetrator may or may not be present. The levels of each system variable were chosen to reflect conditions shown to produce differences in identification accuracy.

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

Evidence Evaluation. Evidence was evaluated in the form of a questionnaire, in which verdict was rendered pre-questionnaire by half of participants and rendered post-questionnaire by the other half of participants.

Measures

Voir dire questionnaire. Prior to reading the trial stimulus, all participants answered questions regarding their gender, political views, and ethnic background. To ensure that all participants were eligible for the study, they indicated whether they were U.S. citizens, if they were registered to vote, if they had a driver's license, and their age. Those who were not eligible were directed to the end of the study and thanked for their time.

Verdict: Participants indicated whether they believed the defendant was guilty or not guilty of armed robbery, their confidence in their verdict (1 = not at all confident; 9 = extremely confident), and the probability that the defendant was guilty (0% = not at all probable; 100% = completely probable).

Witness ratings. Participants indicated their impressions of each witness on a scale of 1 to 7 (1 = extremely…; 8 = not at all…) in regards to trustworthiness, honesty, convincingness, and certainty.

Manipulation checks. The study was conducted online using Amazon's Mechanical Turk and Qualtrics. In an effort to ensure reading and recollection of trial evidence, participants were asked to retype a message embedded in the trial, confirming that they had read each testimony. Furthermore, certain post-trial questions asked about facts pertaining to the manipulated variables in the study in order to gauge the attentiveness of each participant. Additionally, participants indicated which, if any, factors were mentioned by the judge in the instructions. Participants were instructed to check the box next to a factor if the topic was covered in the judge's instructions. Factors included each of the manipulated system and estimator variables as well as filler items for participants in the no instruction conditions (e.g., burden of proof reasonable doubt).

Procedure

Amazon mTurk workers logged onto the mTurk website and viewed our study as a Human Intelligence Task. They were asked to answer a series of Voir Dire questions to determine eligibility. Once participants were deemed eligible, they were asked to read and give their informed consent. After consenting to participate, participants read one of twelve different trial transcripts. Participants then completed the post-trial questionnaire. The study took approximately 30 minutes to complete.

Analysis

Data was collected via the Qualtrics survey website, and analyzed by SPSS. Two univariate ANOVAs were run to analyze the two-way interactions between quality of system and estimator variables and type of instruction in the cases of pre-evidence evaluation verdict and post-evidence evaluation verdict. A third univariate ANOVA was run against the entire data set to analyze main effects and all-encompassing interaction significance.

Results

The results of the univariate ANOVA revealed a significant main effect of quality of system and estimator variables on confidence-weighted verdict (F (1, 283) = 20.69, $p < .001$, $\eta^2 = .07$) (see Table 1). When the quality of the system and estimator variables was good, mock jurors rendered a significantly higher number of guilty verdicts (M = 1.82) compared to when the quality of the system and estimator variables was poor (M = -1.83).

Although the two-way interaction between instruction type and quality of system and estimator variables was not significant, there was statistical significance at the univariate level. When jurors did not receive *Henderson* instructions, they rendered a significantly higher number of guilty verdicts when the quality of the system and estimator variables was good (*M* = 2.64) compared to when the quality of the system and estimator variables was poor (*M* = -.56; *F* (1, 283) = 5.58, $p < .02$, $\eta^2 = .02$) (See Table 3 and Table 1). Similarly, when jurors received non case-specific *Henderson* instructions, they rendered a significantly higher number of guilty verdicts when the quality of the system and estimator variables was good (*M* = 2.15) compared to when the quality of the system and estimator variables was poor (*M* = -3.32; F (1, 283) = 14.07, $p < .001$, $\eta^2 = .05$) (See Table 3 and Table 1). There was no significant interaction between case-specific *Henderson* instructions and quality of system and estimator variables.

Table 1: Between-subjects ANOVA

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1454.968[a]	11	132.270	2.812	.002	.099
Intercept	.019	1	.019	.000	.984	.000
Instructions	164.165	2	82.083	1.745	.177	.012
Sys_Est	973.557	1	973.557	20.694	.000	.068
Verdict_prompt	14.944	1	14.944	.318	.573	.001
Instructions * Sys_Est	124.981	2	62.491	1.328	.267	.009
Instructions * Verdict_prompt	26.129	2	13.065	.278	.758	.002
Sys_Est * Verdict_prompt	2.930	1	2.930	.062	.803	.000
Instructions * Sys_Est * Verdict_prompt	174.092	2	87.046	1.850	.159	.013
Error	13313.933	283	47.046			
Total	14770.000	295				
Corrected Total	14768.902	294				

Table 2: Univariate Tests

Dependent Variable: WeightedConfVerdict

instruction condition	1=verdict first; 2=verdict last		Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
No instruction	1	Contrast	160.702	1	160.702	3.416	.066	.012	3.416	.453
		Error	13313.933	283	47.046					
	2	Contrast	104.924	1	104.924	2.230	.136	.008	2.230	.319
		Error	13313.933	283	47.046					
Henderson bad	1	Contrast	240.913	1	240.913	5.121	.024	.018	5.121	.616
		Error	13313.933	283	47.046					
	2	Contrast	.725	1	.725	.015	.901	.000	.015	.052
		Error	13313.933	283	47.046					
Henderson both	1	Contrast	159.936	1	159.936	3.400	.066	.012	3.400	.451
		Error	13313.933	283	47.046					
	2	Contrast	549.567	1	549.567	11.682	.001	.040	11.682	.926
		Error	13313.933	283	47.046					

Each F tests the simple effects of quality of system and estimator variables: 0=bad; 1=good within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Computed using alpha = .05

Table 3: Estimates

Dependent Variable: WeightedConfVerdict

instruction condition	quality of system and estimator variables: 0=bad; 1=good	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
No instruction	0	-.559	.952	-2.433	1.314
	1	2.635	.961	.744	4.526
Henderson bad	0	-1.620	.970	-3.529	.289
	1	.673	.951	-1.199	2.545
Henderson both	0	-3.323	.991	-5.273	-1.372
	1	2.146	1.069	.041	4.251

The univariate ANOVA was run again after splitting the dataset to compare results of those who gave their verdicts before evaluating the evidence versus those who gave their verdicts after evaluating the evidence. The results revealed a significant main effect of quality of system and estimator variables on confidence-weighted verdict, both when verdict occurred prior to evidence evaluation (F (1, 144) = 11.73, $p < .001$, $\eta^2 = .08$) and when verdict occurred after evidence evaluation (F (1, 139) = 9.08, $p < .003$, $\eta^2 = .06$) (See Table 4).

The results of this ANOVA revealed a significant two-way interaction between instruction type and evidence quality when jurors evaluated evidence before rendering a verdict (F (2, 139) = 3.04, $p < .051$, $\eta^2 = .04$) (See Table 4). Specifically, when jurors received non-case-specific *Henderson* instructions and examined the evidence prior to rendering a verdict, they rendered a significantly higher number of guilty verdicts when the quality of the system and estimator variables was good ($M = 3.17$) compared to when the quality of the system and estimator variables was poor ($M = -4.08$; F (1, 139) = 11.76, $p < .002$, $\eta^2 = .08$) (See Table 5). There was no significant difference in verdict based on evidence quality when jurors examined the evidence prior to rendering a verdict and received either the case-specific *Henderson* instructions ($p = .90$) or No instruction ($p = .14$) (See Table 6). This two-way interaction between instruction type and evidence quality was not significant when jurors examined evidence after rendering a verdict (F (2, 144) = .05, $p = .95$, $\eta^2 = .001$).

Table 4: Between-subjects ANOVA (After split)

Dependent Variable: WeightedConfVerdict

1=verdict first; 2=verdict last	Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
verdict first	Corrected Model	646.857[a]	5	129.371	2.733	.022	.087	13.667	.811
	Intercept	7.120	1	7.120	.150	.699	.001	.150	.067
	Instructions	95.866	2	47.933	1.013	.366	.014	2.025	.224
	Sys_Est	554.975	1	554.975	11.726	.001	.075	11.726	.925
	Verdict_prompt	.000	0	.	.	.	.000	.000	.
	Instructions * Sys_Est	4.882	2	2.441	.052	.950	.001	.103	.058
	Instructions * Verdict_prompt	.000	0	.	.	.	.000	.000	.
	Sys_Est * Verdict_prompt	.000	0	.	.	.	.000	.000	.
	Instructions * Sys_Est * Verdict_prompt	.000	0	.	.	.	.000	.000	.
	Error	6815.516	144	47.330					
	Total	7472.000	150						
	Corrected Total	7462.373	149						
verdict last	Corrected Model	777.955[c]	5	155.591	3.328	.007	.107	16.640	.890
	Intercept	7.826	1	7.826	.167	.683	.001	.167	.069
	Instructions	96.818	2	48.409	1.035	.358	.015	2.071	.228
	Sys_Est	424.641	1	424.641	9.083	.003	.061	9.083	.849
	Verdict_prompt	.000	0	.	.	.	.000	.000	.
	Instructions * Sys_Est	284.545	2	142.273	3.043	.051	.042	6.086	.581
	Instructions * Verdict_prompt	.000	0	.	.	.	.000	.000	.
	Sys_Est * Verdict_prompt	.000	0	.	.	.	.000	.000	.
	Instructions * Sys_Est * Verdict_prompt	.000	0	.	.	.	.000	.000	.
	Error	6498.417	139	46.751					
	Total	7298.000	145						
	Corrected Total	7276.372	144						

a. R Squared = .087 (Adjusted R Squared = .055)

b. Computed using alpha = .05

c. R Squared = .107 (Adjusted R Squared = .075)

Table 5: Estimates (After split)

Dependent Variable: WeightedConfVerdict

1=verdict first; 2=verdict last	instruction condition	quality of system and estimator variables: 0=bad; 1=good	Mean	Std. Error	95% Confidence Interval	
					Lower Bound	Upper Bound
verdict first	No instruction	0	-.519	1.324	-3.135	2.098
		1	3.000	1.376	.280	5.720
	Henderson bad	0	-2.040	1.376	-4.760	.680
		1	2.308	1.349	-.359	4.975
	Henderson both	0	-2.565	1.435	-5.401	.270
		1	1.125	1.404	-1.651	3.901
verdict last	No instruction	0	-.600	1.367	-3.304	2.104
		1	2.269	1.341	-.382	4.921
	Henderson bad	0	-1.200	1.367	-3.904	1.504
		1	-.962	1.341	-3.613	1.690
	Henderson both	0	-4.080	1.367	-6.784	-1.376
		1	3.167	1.612	-.020	6.353

Table 6: Univariate Tests (After split)

Dependent Variable: WeightedConfVerdict

1=verdict first; 2=verdict last	instruction condition		Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
verdict first	No instruction	Contrast	160.702	1	160.702	3.395	.067	.023	3.395	.449
		Error	6815.516	144	47.330					
	Henderson bad	Contrast	240.913	1	240.913	5.090	.026	.034	5.090	.611
		Error	6815.516	144	47.330					
	Henderson both	Contrast	159.936	1	159.936	3.379	.068	.023	3.379	.447
		Error	6815.516	144	47.330					
verdict last	No instruction	Contrast	104.924	1	104.924	2.244	.136	.016	2.244	.319
		Error	6498.417	139	46.751					
	Henderson bad	Contrast	.725	1	.725	.016	.901	.000	.016	.052
		Error	6498.417	139	46.751					
	Henderson both	Contrast	549.567	1	549.567	11.755	.001	.078	11.755	.926
		Error	6498.417	139	46.751					

Each F tests the simple effects of quality of system and estimator variables: 0=bad; 1=good within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Computed using alpha = .05

Discussion

It is of paramount importance that jurors are accurately sensitized to eyewitness identification accuracy. Past research has shown that case-specific *Henderson* instructions do not effectively sensitize jurors to the quality of identification evidence (Berman, et al., 2014). Rather than discount the *Henderson* instructions themselves, this study investigated possible amendments to the *Henderson* instructions. Specifically, it looked at the implementation of case-specific versus non-case specific *Henderson* instructions. The difference between case-specific and non-case-specific is as follows: case-specific instructions highlight system and estimator variables when they have a negative impact on eyewitness accuracy and do not mention these variables when they have a positive impact on eyewitness accuracy; non-case-specific instructions highlight all pertinent system and estimator variables, regardless of their effect on eyewitness accuracy. In conjunction with the amended instructions, this study hypothesized that the non-case-specific instructions would be more effective it jurors had the opportunity to evaluate the evidence and instructions before rendering a verdict. Consistent with the proposed hypothesis, when jurors heard the non-case-specific *Henderson* instructions and had the opportunity to evaluate evidence before rendering a verdict they were sensitive to whether the quality of eyewitness identification variables were good or poor. When case-specific instructions were administered, however, jurors were not significantly sensitized to the quality of the variables. As such, jurors rendered a significantly higher number of guilty verdicts when identification variable quality pointed towards guilt in conditions with non-case-specific *Henderson* instructions than in other conditions.

The dependent variable for this experiment was a confidence-weighted verdict. A participant's dichotomous verdict was multiplied by their confidence in verdict, which allowed for more variability in the dependent measure when running univariate ANOVAs. The significant main effect of variable quality observed across the entire dataset indicates that jurors were able to distinguish between good and poor system and estimator variable conditions when all other factors were held constant. Yet this main effect alone does not alone show the effectiveness of instructions. The two-way interaction between variable quality and instruction type was not significant in the between-subjects analysis; however, results of the pairwise comparisons indicated that instructions did in fact sensitize jurors to identification quality. In conditions with no *Henderson* instructions, as well as in conditions with non-case-specific *Henderson* instructions, more guilty verdicts were rendered when identification quality pointed to guilt (good system and estimators) compared to when variable quality pointed to innocence (poor system and estimators). While both instruction types significantly increased jurors' sensitization, the effect on juror sensitization was greater when non-case-specific *Henderson* instructions were administered ($\eta^2 = .05$), compared to no *Henderson* instructions were administered ($\eta^2 = .02$), which suggests that the non-case-specific *Henderson* instructions improve the accuracy of juror perceptions of eyewitness testimony. Case-specific *Henderson* instructions failed to achieve a significant effect, further indicating that non-case-specific *Henderson* instructions more sufficiently sensitize jurors to identification quality.

When the dataset was split to compare pre- and post-evidence evaluation verdicts, the significant main effect of variable quality persisted both when evidence was evaluated prior to rendering a verdict and when evidence was evaluated after rendering a verdict. The splitting of the dataset also revealed a significant two-way interaction between variable quality and type of instruction when evidence evaluation occurred before jurors gave their final verdict. When non-case-specific *Henderson* instructions were coupled with evidence evaluation before verdict rendering, jurors were significantly sensitized to variable quality and eyewitness accuracy. This interaction supports the study's hypothesis, for there is a significant difference in verdict based upon the quality of the system and estimator variables when *Henderson* instructions are non-case-specific and jurors are prompted to consider evidence before rendering a verdict. No significant difference in verdict was observed when case-specific *Henderson* instructions or no *Henderson* instructions were administered, and no significant two-way interaction occurred when evidence evaluation occurred after a verdict was rendered. This is concurrent with the study's hypothesis as well, for it demonstrates that both case-specific instructions and no instructions whatsoever are less effective than non-case-specific *Henderson* instructions.

While the total sample size consisted of 300 participants, participants were divided among 12 conditions. A greater number of participants assigned to each condition would help to better generalize the results of the study to the public. The goal of jury-decision making research is to better the legal system in the real world, and a larger sample of the population will yield more significant and more adaptable results, as a higher power is observed. Perhaps the most crucial limitation of this study stems from its online nature. As the study was conducted online using Amazon mTurk, participants could not be closely monitored to ensure accuracy or attention. The online survey did include manipulation checks during the trial reading period and within the questionnaires, but these few safeguards may not have been enough to combat distraction or lack of attention among participants. Regardless, studies have been conducted that attest to the reliability of Amazon mTurk studies (Mason W, Suri S, 2012). For example, when Mechanical Turk was used to conduct replications of perception and decision-making studies, results were merely marginally different when analyzed quantitatively and were qualitatively consistent (Paolacci et al., 2010). Lastly, the trial transcript created for this study had not been previously used. The trial transcript, which served as the medium for the study, is entirely fictitious, and is not based off of a specific trial. As such, it had the potential to be unbalanced, swaying jurors unequivocally towards innocence or guilt. However, the transcript was pilot tested to verify its authenticity and balance, and piloting was successful; there was an adequate split in verdict for the conditions expected to point towards guilt and those expected to point towards innocence.

This research aims to find and implement better legal standards and procedures that decrease the probability of a false conviction. The findings of this experiment can help to prevent jurors from convicting innocent suspects based on eyewitness misidentification. Non-case-specific instructions were shown to be more effective than other forms of instruction (case-specific / none). Prior research of the *Henderson* instructions has failed to show a widespread effect on juror sensitization to eyewitness accuracy. This study has demonstrated that *Henderson* instructions that are non-case-specific are significantly more effective when jurors are given the opportunity to evaluate evidence before rendering a verdict. The findings suggest the *Henderson* instructions may be able to aid jurors if utilized and administered as demonstrated in this study. The *Henderson* instructions have already been adopted by the New Jersey State Supreme Court (2012), but the positive implications of the results may lead to a wider implementation of non-case-specific *Henderson* instructions, whether in other states' or federal courts.

Evidence evaluation was used as the primary means of sensitization, but the current legal system employs jury deliberation rather than individual evaluations. Since evidence evaluation was used in lieu of deliberation, future research could directly compare results of questionnaires and deliberation sessions. Running the same study with live participants would allow for group deliberation sessions, which may allow us to examine how perceptions are influenced by other members of the jury. Furthermore, jury deliberation parallels more closely with real-world courtroom practices.

Until now, little evidence has suggested that the *Henderson* instructions function as intended. But by altering the presentation of variables from case-specific to non-case-specific, the instructions were found to be both statistically significant and effective on the qualitative level. Participants were gathered from a random sample of the population, and their results suggested that the *Henderson* instructions had a significant impact; the intention is to project the results of the study onto the general public, who makes up the jury pool. This can be accomplished through increased testing and closely relating research methods to those seen in US criminal trials. If further testing continues to support the hypotheses of this study, jurors who serve on criminal trials, whether novice or experienced, will be able to use the revamped *Henderson* instructions to evaluate relevant facts and misconceptions prior to ruling on a defendant's fate.

Acknowledgements

I'd like to express my sincerest gratitude to Ms. Marlee K. Berman for her guidance and for her assistance with my research. I'd also like to thank my research teachers at school, Mrs. Barbara Franklin, Mrs. Barbi Frank, Mr. Helmut Schleith, and Mr. Andy Bramante, for their instruction and unyielding support as I completed my research, even as moved from one high school to the next. Lastly, I would acknowledge my parents and family for their constant love and support and for the many wonderful opportunities they have provided me.

References

Berman, Yarbrough, Nicholson, Hui, & Penrod. (2014). *"Judicial instruction and eyewitness identification."* MS. John Jay College & the Graduate Center, CUNY, New York.

Cavoukian, A. (1980). Eyewitness testimony: The ineffectiveness of discrediting information. Paper presented at the *American Psychological Association* Meeting, Montreal.

Cutler, B. (2013). Expert testimony on the psychology of eyewitness identification. Paper presented at the annual meeting of the *American Psychology - Law Society*, TBA, San Antonio, 2013-12-13 from http://citation.allacademic.com/meta/p303480_index.html

Cutler, B. L., Dexter, H. R. and Penrod, S. D. (1990). 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

Deffenbacher, K. A. (2008). Estimating the impact of estimator variables on eyewitness identification: A fruitful marriage of practical problem solving and psychological theorizing. *Applied Cognitive Psychology, 22(6),* 815-826. doi:10.1002/acp.1485

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & Kiernan McGorty, E. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior, 28(6),* 687-706. doi:http://dx.doi.org/10.1007/s10979-004-0565-x

Dempsey & Pozzulo (2009). Witness factors and their influence on jurors' perceptions and verdicts. *Criminal Justice and Behavior, 36*, 923 originally published online 11 June 2009 DOI: 10.1177/0093854809338450

Devenport, Penrod, & Cutler (1997). Eyewitness identification evidence: Evaluating commonsense evaluations.*Psychology Pub. Politics & Law, 2*, 338, 347 (1997)

Fradella, (2007). Why judges should admit expert testimony on the unreliability of eyewitness testimony. *Federal Courts Law, Rev. 1, 9* (2007)

Goldstein, A.G., et al. (1989). Frequency of eyewitness identification in criminal cases: A survey of prosecutors. *Bulletin of Psychonomic Society 27, 71, 73* (1989).

Greene, E. (1988). Judge's instruction on eyewitness testimony: Evaluation and revision. *Journal of Applied Social Psychology, 18,* 252–276. doi: 10.1111/j.1559-1816.1988.tb00016.x

Kassin, S. M., Ellsworth, P. C., & Smith, V. L. (1989). The 'general acceptance' of psychological research on eyewitness testimony: A survey of the experts. *American Psychologist, 44(8),* 1089-1098. doi:10.1037/0003-066X.44.8.1089

Kerr, N. L., Atkin, R. S., Stasser, G., Meek, D., Holt, R. W., & Davis, J. H. (1976). Guilt beyond a reasonable doubt: Effects of concept definition and assigned decision rule on the judgments of mock jurors. *Journal Of Personality And Social Psychology, 34(2),* 282-294. doi:10.1037/0022-3514.34.2.282

Lampinen, J., Scott, J., Pratt, D., Leding, J. K., & Arnal, J. D. (2007). 'Good, you identified the suspect...but please ignore this feedback': Can warnings eliminate the effects of post-identification feedback?. *Applied Cognitive Psychology, 21(8),* 1037-1056. doi:10.1002/acp.1313

Lloyd-Bostock, S., & Clifford, B. R. (Eds.). (1983). Evaluating witness evidence. *Wiley, Chicbester*, England

Loftus, E. F. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard Press.

Magnussen, S., Eilertsen, D., Teigen, K., & Wessel, E. (2014). The probability of guilt in criminal cases: Are people aware of being 'beyond reasonable doubt'?. *Applied Cognitive Psychology, 28(2),* 196-203. doi:10.1002/acp.2978

Mansour, J. K.; Beaudry, J. L.; Bertrand, M. I.; Kalmet, N.; Melsom, E. I.; Lindsay, R. C. L. (2012). Impact of disguise on identification decisions and confidence with simultaneous and sequential lineups. *Law and Human Behavior, Vol 36(6)*, 513-526. Dec 2012, http://dx.doi.org/10.1037/h0093937

Mason W, Suri S, (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods 44*, 1–23, doi:10.3758/s13428-011-0124-6 [PubMed]

McGonigle & Emily, (2008). A Blind Faith in Eyewitnesses: 18 Dallas County Cases Overturned by DNA Relied Heavily on Eyewitness Testimony, *Dallas Morning News, Oct. 12, 2008*

Memon, A., Mastroberardino, S., & Fraser, J. (2008). Münsterberg's legacy: What does eyewitness research tell us about the reliability of eyewitness testimony?. *Applied Cognitive Psychology, 22(6)*, 841-851. doi:10.1002/acp.1487

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5,* 411–419.

Penrod, S. (2003). Eyewitness identification evidence: How well are witnesses and police performing? *Criminal Justice Magazine, Spring, 54,* 36-47.

Penrod, S. D., & Cutler, B. L. (1999). Preventing mistaken convictions in eyewitness identification trials: The case against traditional safeguards. In R. Roesch, S. D. Hart, and J. R. P. Ogloff (Eds.), *Psychology and law: The state of the discipline (pp. 89-118).* New York: Kluwer.

Pezdek, K., & Blandon-Gitlin, I. (2005). When is an intervening line-up most likely to affect eyewitness identification accuracy?. *Legal And Criminological Psychology, 10(2),* 247-263. doi:10.1348/135532505X49846

Pezdek, K., Avila-Mora, E., & Sperry, K. (2010). Does trial presentation medium matter in jury simulation research? Evaluating the effectiveness of eyewitness expert testimony. *Applied Cognitive Psychology, 24(5),* 673-690. doi:10.1002/acp.1578

Saunders, D. M., Vidmar, N., & Hewitt, E. C. (1983). Eyewitness testimony and the discrediting effect. In B. R. Clifford & S. M. A, Lloyd-Bostock (Eds.), *Evaluating witness evidence (pp. 57-78)*. Chicester, England: *Wiley*.

Shapiro, Barbara J. (1991). Beyond reasonable doubt and probable cause: Historical perspectives on the anglo-american law of evidence. *Berkeley: University of California Press, c1991* http://ark.cdlib.org/ark:/13030/ft409nb30v/

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

Sheehan, (2011). Making the jurors the "experts": The case for eyewitness identification jury Instructions. *52 B.C.L. Rev. 651.* http://lawdigitalcommons.bc.edu/bclr/vol52/iss2/10

Stovall v. Denno, *388 U.S. 293*, 301–02 (1967)

Supreme Court Releases Eyewitness Identification Criteria for Criminal Cases. (2012, July 19). Retrieved September 9, 2014, from http://www.judiciary.state.nj.us/pressrel/2012/pr120719a.htm

Tenney, E. R., Cleary, H. D., & Spellman, B. A. (2009). Unpacking the doubt in "beyond a reasonable doubt": Plausible alternative stories increase not guilty verdicts. *Basic And Applied Social Psychology, 31(1),* 1-8. doi:10.1080/01973530802659687

Thompson, Sandra Guerra (2008). Beyond a reasonable doubt?: Reconsidering uncorroborated eyewitness identification testimony. *UC Davis Law Review, Vol. 41,* 2008; *U of Houston Law Center No. 2008*-A-06.

U.S. Const*. amend. VII*

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

United States v. Wade, *388 U.S. 218, 228* (*1967*)

Walsh, D. (2013). The dangers of eyewitness identification: A call for greater state involvement to ensure fundamental fairness. *Boston College. Law School. Boston College Law Review, 54(3),* 1415-1453. Retrieved from http://search.proquest.com/docview/1528143576?accountid=37512

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal Of Personality And Social Psychology, 36(12)*, 1546-1557. doi:10.1037/0022-3514.36.12.1546

Wells, G. L., & Bradfield, A. L. (1998). 'Good, you identified the suspect': Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal Of Applied Psychology, 83(3)*, 360-376. doi:10.1037/0021-9010.83.3.360

Wells, G. L., Lindsay, R. C. L., & Tousignant, J. P. (1980). Effects of expert psychological advice on juror judgments in eyewitness testimony. *Law and Human Behavior, 4*, 275-286.

Wells, G. L., & Loftus, E. F. (Eds.). (1984). Eyewitness testimony: Psychological perspectives. *New York: Cambridge University Press.*

Wells, G. L. and Loftus, E. F. (2013). Eyewitness Memory for People and Events (Chapter 25) (January 16, 2013). *Handbook of Psychology, Vol. 11, 2013*, *Forensic Psychology, Chapter 25,* R.K. Otto and & I.B. Weiner (Eds), Hoboken, NJ: John Wiley & Sons, Inc.; *UC Irvine School of Law Research Paper No. 2013-88*. Available at SSRN: http://ssrn.com/abstract=2201954

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

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2014, June 16). Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior*. Advance online publication. http://dx.doi.org/10.1037/lhb0000096

Wilford, M. M., & Wells, G. L. (2013). Eyewitness system variables. In B. L. Cutler (Ed.) , *Reform of eyewitness identification procedures (pp. 23-43).* Washington, DC, US: *American Psychological Association*. doi:10.1037/14094-002

Winter, R. J., & Robicheaux, T. (2011). Questions about the jury: What trial consultants should know about jury decision making. In R. L. Wiener, B. H. Bornstein (Eds.) , *Handbook of trial consulting (pp. 63-91).* New York, NY, US: *Springer Science + Business Media*. doi:10.1007/978-1-4419-7569-0_4

Wright, D. B., & Hall, M. (2007). How a 'reasonable doubt' instruction affects decisions of guilt. *Basic And Applied Social Psychology, 29(1),* 91-98. doi:10.1080/01973530701331254