Spring 5-8-2023

Juror's Perceptions of Evidence-Based Suspicion

Nicholas Welter
CUNY John Jay College, nicholas.welter@jjay.cuny.edu

Juror Perceptions of Evidence-Based Suspicion

A Thesis Presented in Partial Fulfillment of the Requirement for the Degree of Master of Arts in Forensic Psychology

John Jay College of Criminal Justice

City University of New York

Nicholas Welter

May 2023

Juror's Perceptions of Evidence-Based Suspicion

Nicholas Welter

Thesis Committee

Thesis Advisor: Margaret Bull-Kovera

Second Reader: Steven Penrod

External Reader: Jacqueline Katzman

Abstract

Eyewitness misidentification is a leading cause of wrongful conviction. Although the prior probability of guilt (i.e., pre-identification evidence strength) is the most important factor for predicting a defendant's actual guilt status and the accuracy of any subsequent eyewitness identification, no study has examined whether it affects juror decisions. This oversight is problematic because when officers place suspects in lineups when there is little evidence connecting them to the crime, it falls on jurors to examine the probative value of identification evidence. Participants (N = 357) watched a mock trial depicting an armed robbery that varied pre-identification evidence (strong vs. weak), and whether the expert witness testified about the importance of base-rates of guilt (present, not present). I predicted that when there is expert testimony on the importance of the base rate of guilty suspects in identification procedures, participants will be more sensitive to variations in the strength of the evidentiary connection between the defendant and the crime prior to the positive identification. Expert testimony on the importance of base rates of guilt did not affect participant judgments, but it did cause participants to be more skeptical of lineup fairness regardless of the strength of the evidence connecting the suspect to the crime. Overall, participants were sensitive to variations in the strength of the pre-identification evidence without the benefit of expert testimony; however, their judgments underestimated the likely accuracy of identifications made when the pre-identification connection between the suspect and the crime was strong. Thus, as expert testimony is not sufficient for helping jurors evaluate the probability of a suspect's guilt post-identification, it will be important to implement procedures that will ensure that the base rate of guilty suspects in identification procedures is high enough to produce a sufficient probability of guilt after a positive identification, like a departmental requirement of evidence-based suspicion.

Juror Perceptions of Evidence-Based Suspicion

Eyewitness Identification

In July of 1984, Jennifer Thompson – a White woman from North Carolina – was considered the 'perfect eyewitness' by police (O'Neil, 2001). She was sexually assaulted and during the assault claimed to focus entirely on remembering her attacker's facial features so that she could assist in later identifying the culprit. Thompson aided the police in the development of a composite sketch shortly after the assault which led to an anonymous tip. The police received a call regarding a possible suspect named Ronald Cotton, who worked near the scene of the assault. Three days later, Ronald Cotton was identified from a police lineup as the culprit. The evidence connecting Cotton to the crime was weak: Cotton was placed in the lineup because a flashlight found in his home resembled one used by the assailant, and rubber from Cotton's tennis shoe resembled rubber found at the crime scene. Although countless other individuals' own flashlights and practically all shoes have rubber soles, the entire case relied on the eyewitness identification by Jennifer Thompson. In July of 1995 after serving 10 ½ years in prison, Ronald Cotton was pardoned by the governor of North Carolina because of new DNA evidence (Innocence Project, n.d). This case study provokes the question: How reliable was the evidence connecting Ronald Cotton to the rape of Jennifer Thompson *before* the police placed him in a lineup and risked his misidentification?

A Need for Evidence-Based Suspicion

The dangers of eyewitness misidentifications are not a new phenomenon and these perils have been documented by the Courts and scholars for decades. The United States Supreme Court wrote in 1967, "Identification evidence is peculiarly riddled with innumerable dangers and

variable factors which might seriously, even crucially, derogate from a fair trial. The vagaries of eyewitness identification are well-known; the annals of criminal law are rife with instances of mistaken identification" United States v. Wade, 388 U.S. 218, 228 (1967). Brigham and Bothwell (1983) devised a study where participants predicted the accuracy rates of eyewitness identifications from descriptions of actual research studies. The results showed that 83% of participants overestimated the accuracy rates of eyewitness identifications across three different scenarios (Brigham & Bothwell, 1983). Wells and colleagues (2020) analyzed data from 6,743 lineups to estimate actual eyewitness accuracy rates outside of the laboratory. Although I cannot know for sure how many witnesses identified innocent suspects (as it is unknown in archival and field studies whether identifications of suspects are correct or mistaken), I can examine whether witnesses are prone to memory errors by examining whether witnesses positively identify fillers (or those lineup members who are not the suspect and are known to be innocent). One of every four witnesses (23.7%) who were shown a lineup selected an innocent filler, which is over 1/3 of witnesses who made an identification (35.5%) of a filler or a suspect (as opposed to no identification). The field of psychology and law has made significant advances in accumulating knowledge regarding best practices for identification procedures, however, the leading cause of wrongful convictions remains eyewitness misidentifications (Innocence Project, 2023). One of the factors that influences the rates at which witnesses misidentify innocent suspects is the base rate at which guilty suspects appear in police identification procedures; as the ratio of innocent to guilty suspects increases, so does the ratio of mistaken to correct identifications (Katzman & Kovera, 2022; 2023; Wells et al., 2015). Law enforcement officers decide whether to subject a suspect to an identification procedure; thus, the police control the base rate of guilty suspects in identification procedures.

Yet there is little research on the amount of evidence police officers require to connect a suspect to a crime before they conduct an identification procedure. In a national survey of United States Law enforcement agencies, one third of officers reported that they required no evidence to place a suspect in a lineup or they would place an individual in a lineup with no evidence of guilt other than a "mere hunch" (Wise et al., 2011). Behrman and Richards (2005) examined lineups at a Northern California police department and found that 40% of the time an identification procedure is implemented, there is no extrinsic evidence against the suspect and for an additional 30% of the lineups they had minimal evidence. Furthermore, in one of the only studies analyzing actual lineups at a Houston, Texas Police Department, researchers estimated only 35% of suspects who were subjected to an identification procedure were guilty (Wixted et al., 2016).

In the latest scientific review paper on best practices for eyewitness procedures, Wells and colleagues (2020) argued that there should be "evidence-based grounds to suspect that an individual is guilty of the specific crime being investigated before including that individual in an identification procedure" (p. 11). A requirement for evidence-based suspicion refers to the notion that there should be articulable evidence that a particular person committed the crime in question prior to his placement in an identification procedure. Evidence-based suspicion should not be based on the grounds of a "hunch" from a police officer or from an individual matching an eyewitness's general description of a culprit (Wells et al., 2020). Wells and colleagues (2020) provided an example of sufficient evidence-based suspicion to place an individual in a lineup such as a "unique fit to a specific description that was given by the eyewitness (e.g., blue teardrop tattoo under left eye; moon shaped scar on chin" (p.12). Moreover, there are no laws or policies in place that would prevent police departments from conducting lineups with low base rates of guilt (Wells et al., 2020). If a police department places a suspect into a lineup on nothing

more than a hunch, the base rate of guilt is likely to be low and the probability of an accurate identification of the suspect is going to be low (Wixted & Wells, 2017). To increase the base rate of guilt and increase the likelihood of a correct identification, police departments need to require pre-identification evidence connecting a suspect to a crime before conducting an identification procedure. Therefore, educating jurors about the benefits of having evidence-based suspicion before conducting an identification procedure may minimize wrongful convictions resulting from mistaken eyewitness identifications.

Base-Rate Neglect

The phenomenon of base-rate neglect, also known as the base-rate fallacy, refers to the idea that individuals do not use or tend to ignore base-rate information when making decisions. This phenomenon has been well studied by psychologists and it is generally understood that people do not intuitively use base rate information when making predictions (Kahneman & Tversky, 1973; Katzman & Kovera, 2022). Kahneman and Tversky (1973) published one of the first studies based on the phenomenon of base-rate neglect; participants were instructed to read a description of a first-year graduate student and then predict the student's field of study from the brief description. Participants (95%) were more likely to predict that a male student was a computer science major than a humanities or education major, even though participants knew more students were enrolled in the latter majors. Thus, participants relied on the representativeness of the personality sketch to predict the students' field of specialization and ignored the reliability of the personality sketch or the prior probability of group membership (Kahneman & Tversky, 1973).

Researchers have repeatedly demonstrated the far-reaching effects of base-rate neglect. Participants ignored the reliability of medical diagnostic equipment when making predictions

regarding if a person has a disease (Hammerton, 1973). People rely on a witness's testimony about the color of a taxicab that was involved in an accident, rather than the base rate of taxicab colors on the road (Lyon & Slovic, 1976). Probation officers ignored base-rate information when discussing recidivism rates (Carroll, 1977). In addition, police officers fall victim to base-rate neglect when making decisions about whom to place in a lineup, outlining a need for evidence-based suspicion before an identification procedure (Wells et al., 2020). Although there is a paucity of research on the topic, there is some evidence that mock jurors fall victim to base-rate neglect (Kaye & Kohler, 1991).

Juror Comprehension

For cases that rely on eyewitness identification evidence, the burden of interpreting the probative value of an identification falls upon jurors. Empirical research on juror comprehension of eyewitness identifications has demonstrated that juries overestimate their and do not understand which factors influence their reliability (APA, 2016). Schmechel and colleagues (2006) published a study involving one-thousand mock jurors that were asked if they thought a violent crime would make eyewitness memory more reliable, less reliable, or have no effect. Only three out of ten participants were able to understand that violence has a tendency to make an eyewitness memory less reliable (Schmechel et. al, 2006). The question of whether jurors are sensitive to the strength of the identification evidence is critical because suspects are more likely to reject a plea offer and go to trial when there is a low prior probability of guilt (Gazal-Ayal & Tor, 2012).

Yet lay people fall victim to base rate neglect and sitting in a courtroom does not seem to improve laypeople's methodological reasoning skills. For example, mock jurors demonstrate a tendency to rely on a forensic scientist's background and experience instead of a study's

scientific validity when they make evidence strength judgments (Koehler et al., 2016). Mock jurors also struggle to identify threats to internal validity that arise during expert testimony (Jones & Kovera, 2015). Therefore, it is unlikely that jurors will be sensitive to the present base rate issue when left to their own devices. Although jurors often struggle to interpret complex statistical concepts when left unaided, whether jurors naturally attend to information about a suspect's probability of guilt prior to a positive identification has yet to be investigated experimentally.

Expert Witness Testimony

Expert testimony is one strategy that can help jurors better understand complex scientific issues in the courtroom. Specific empirically based strategies can make expert testimony even more effective on juror comprehension. One such strategy is an expert's use of a demonstrative during their testimony. Brain and Boderick (1992) define demonstrative evidence as "any display that is principally used to illustrate or explain other admissible testimonial, documentary, or real evidence, or judicially noticed fact." Scholars suggest that demonstratives are helpful because they can transmit more information to an audience than any other method of communication (Watts, 2000), and are an effective way to reduce jurors' cognitive load (Jones & Kovera, 2015). In a study that examined jurors' comprehension of jury instructions, Wiener and colleagues (2005) found that comprehension of complex instructions improved with the addition of a flowchart decision, which outlined the steps jurors should take to correctly arrive at a verdict.

Moreover, research suggests that demonstrative evidence is an effective tool for experts who testify regarding variations in scientific evidence quality (Jones & Kovera, 2015). In one study, when an opposing expert presented mock jurors with a visual aid that illustrated the steps that they needed to take to evaluate the internal validity of the research presented by the other

expert, participants' verdicts reflected the validity of the expert's evidence (Jones & Kovera, 2015). In another study, an expert who testified with a demonstrative regarding Bayes' Theorem in evaluating probabilistic evidence effectively sensitized jurors to the flaws in a defense attorney's argument (Smith et al., 1996).

The type and timing of expert testimony also influences jurors' use and comprehension of statistical evidence (Brekke & Borgida, 1988). Expert testimony is most impactful if it appears at the beginning of a case presentation and makes the connection between the testimony and case facts concrete (Kovera et al., 1997). For example, in a study that examined mock jurors' use of base rates in a simulated rape trial, participants made the most use of an expert's scientific testimony when it was linked explicitly to the case and was presented early in the trial. This finding was not a result of the length of the testimony, but was facilitated by the expert's concrete, rather than abstract, presentation of information (Brekke & Borgida, 1988).

One method to assist people in overcoming base-rate neglect is to present base-rate information as frequencies. People are more easily able to understand frequencies rather than percentages or probabilities (Gilovich et al, 2002, Cosmides & Tooby, 1996). Gilovich and colleagues (2002) conducted a study supporting this notion; frequency representation helps people visualize that one set is contained in another, which helps to overcome the heuristics and biases that lead to base-rate neglect. Furthermore, when a base-rate problem is presented as a frequency rather than a percentage, participants implement base-rate information in their predictions at a substantially higher rate (Cosmides & Tooby, 1996). Although base-rate neglect may pose a challenge for mock jurors to overcome, presenting the probabilities as frequencies is one effective strategy for doing so.

Ultimately, strategically incorporated expert testimony is an effective strategy that is frequently used to educate jurors regarding factors that influence the reliability of eyewitness identifications (State v. Henderson, 2011; State v. Lawson, 2012). Although base rate neglect is notoriously difficult to overcome, perhaps empirically informed expert testimony could sensitize jurors to variations in lineup quality based on the probative value of an identification.

Present Study

The current study will investigate whether expert testimony on the importance of base rates will sensitize participants viewing a mock trial to variations in the evidence connecting the defendant to the crime prior to the positive identification. I expected to find that when there is expert testimony on the importance of base rates of guilt, participants will be more sensitive to variations in the evidence that would influence the base rate of guilt. Specifically, participants who view the trial in which there is base rate testimony will be less likely to render a guilty verdict when the pre-identification evidence connection is weak rather than strong; verdicts will not significantly differ for participants who did not hear expert testimony about the importance of base rates.

Method

Participants

I collected data from a total of 515 participants (357 participants after accounting for manipulation check failures) from Amazon's Mechanical Turk (MTurk) – an online crowdsourcing marketplace allowing researchers to outsource and distribute their materials (surveys, questionnaires, polls) to a virtual workforce. MTurk maintains demographic information about its participants, making it possible to only include participants who meet the eligibility criteria. In addition, I used the data quality controls available in CloudResarch. Doing

so produces data quality that is comparable to other crowdsourcing platforms (e.g., Prolific). An *a priori* power analysis using the superpower package (Lakens & Caldwell, 2021; R 3.6.3) indicated that this sample size was sufficient to detect medium effects ($d = 0.3$) with power = .80 and $p < .05$ for my analyses (i.e., two-way ANOVAs) (Lakens, 2021).

Participants were limited to United States citizens who were 18 and older and have working knowledge of English. Participants' gender was evenly represented with males accounting for (51%) and females (47%), racially and ethnically diverse (76% White, 10% Black, 9% Asian, 5% Hispanic), and an average age of 43 years old ($SD = 12.35$). All participants were jury-eligible.

Upon choosing to participate in the study, participants first viewed the informed consent where they read the purposes and procedures of the study and agreed to participate or not participate. They completed the study from the location of their choosing (as long as it is within the US). As a means of protecting privacy, participants had no interaction with the researchers associated with the project, except if they needed to resolve any technical difficulties they experienced during the study.

Participants were assigned a random number for purposes of compensation and their MTurk worker ID was not recorded. Participants were compensated for completing the survey in the amount of $3.50. Compensation occurred within three days of completion of the study. Participants were fully compensated, regardless of whether they finished the survey.

Design

The study has a 2 (Pre-Identification Evidence: weak vs. strong) × 2 (Base Rate Expert Testimony: present vs. absent between-subjects factorial design. I randomly assigned participants

to a condition. The research protocol was approved by the Institutional Review Board at the John Jay College of Criminal Justice.

Procedure

Participants were screened for juror eligibility and completed the study online in Qualtrics survey software. They viewed the advertisement for the study on MTurk and followed a link to participate in Qualtrics. If they consented, participants viewed the video-recorded mock trial. The survey software randomly assigned them to one of the four versions of the mock trial. They then completed the post-trial questionnaire and a demographics survey. The study took participants approximately thirty minutes.

Materials

Participants viewed a video-recorded mock trial that was uploaded to Youtube prior to being included in my Qualtrics survey. The video depicted a defendant accused of an armed robbery based on an eyewitness identification. The mock trial consists of six different components: opening jury instructions, opening statements, testimony from the eyewitness and a police officer, expert testimony, closing arguments, and closing jury instructions. There were four different versions of the trial based on expert testimony about base rates of guilt (present vs. absent) and pre-identification evidence (weak vs. strong). The trial lasted approximately thirty minutes depending on the conditions. The mock trial was transcribed using Otter.ai and the transcript was uploaded to Open Science Framework (OSF).

Pre-Identification Evidence (Weak vs. Strong)

In all conditions, participants heard testimony in which an officer identified a potential suspect while patrolling the area of the crime. The potential had general identifying characteristics (a Black man in his mid-twenties) from the eyewitness account in the weak pre-

identification evidence condition. Whereas, in the strong pre-identification evidence connection the potential suspect matched specific characteristics (a Black man in his mid-twenties, with an average build and a scar on his chin, who is wearing jeans and a black hoodie with orange and white lettering) from the eyewitness report.

Base Rate Expert Testimony (Present, Not Present)

In all conditions, a defense expert testified regarding the relationship between stress and accuracy (e.g., that high stress impedes encoding and produces less accurate identifications). For half of the participants the expert also discussed base rate issues via a PowerPoint presentation that visually displayed how base rates influence posterior probabilities, based on research suggesting that jurors most easily understand complex statistical issues when an expert uses a demonstrative in their testimony (Jones & Kovera, 2015). Finally, to maximize the effects of the expert testimony, the expert was the first and only witness for the defense, as experts are more impactful if they appear at the beginning of the case presentation and make the connection between the testimony and the case facts concrete (Brekke & Borgida, 1988; Kovera et al., 1997).

Dependent Measures

Participants first answered a series of questions about their trial judgments, then answered a series of questions about the strength of the prosecution's case, their perceptions of the eyewitness credibility, their perceptions of expert credibility, and their perceptions of lineup fairness. Finally, participants answered questions designed to check the effectiveness of the manipulations.

Verdict

Participants first indicated whether they found the defendant guilty or not guilty.

Strength of the Prosecution's Case

Participants then rated their agreement with statements that measured the strength of the prosecution's case using seven-point scales (1 = *strongly disagree*, 7 = *strongly agree*): (a) The prosecution's case was strong, (b) The defense's case was strong (R). For all scales, items followed by an (R) were reverse coded. A test for internal reliability showed that the items created a sufficient scale (Cronbach's $\alpha = .69$) that could not be meaningfully improved by omission of any items. Thus, I created a composite variable representing participants' perceptions of the relative strength of the prosecution's case by averaging across the two items.

Credibility of the Eyewitness

Participants also provided their ratings of the credibility of the eyewitness using seven-point Likert scales (1 = *not at all,* 7 = *extremely*): credible, honest, confident, certain, believable, and accurate. I subjected the seven items to an exploratory factor analysis using principal axis factoring and varimax rotation. A test for internal reliability with the items that I retained showed that they created a strong scale (Cronbach's $\alpha = .93$) that was meaningfully improved by the omission of "intelligence". Thus, I created a composite variable representing participants' perceptions of the credibility of the eyewitness by averaging across all seven items.

Credibility of the Expert

Participants also provided their ratings of the credibility of the expert witness using seven-point Likert scales (1 = not at all, 7 = extremely): credible, competent, intelligent, trustworthy, skilled, capable, and fair. A test for internal reliability showed that the items created a strong scale (Cronbach's $\alpha = .95$) that could not be meaningfully improved by omission of any items. Thus, I created a composite variable representing participants' perceptions of the credibility of the expert witness by averaging across the seven items.

Lineup Fairness

Participants then rated their agreement with a statement that measured line up fairness using a seven-point scale (1 = strongly disagree, 7 = strongly agree): (a) The identification procedure was fair (b) The identification procedure was unfair (R). For all scales, items followed by an (R) were reverse coded. A test for internal reliability showed that the items created a strong scale (Cronbach's $\alpha = .81$) that could not be meaningfully improved by omission of any items. Thus, I created a composite variable representing participants' perceptions of lineup fairness by averaging across the seven items.

Additional Items

Participants also answered questions about the case itself, including their memory for the case facts, which served as manipulation checks (e.g., on which descriptors did the defendant match the witness' description when he was found during the officers' canvas of the neighborhood, whether the participant heard the expert witness testify about the importance of the base rate of guilt or not). They concluded the study by providing demographics (e.g., age, gender, race/ethnicity, jury experience).

Results

Data Analytic Plan

I conducted binary logistic regressions for dichotomous dependent variables or analyses of variance (ANOVAs) for continuous dependent variables, in each case testing the full 2 (Pre-Identification Evidence: weak vs. strong) × 2 (Base Rate Expert Testimony: present vs. absent) model. When the omnibus test suggested that there were significant interactions, I conducted follow-up analyses.

Manipulation Checks

Pre-Identification Evidence Manipulation

Participants who viewed the trial in which the pre-identification evidence was weak were less likely to indicate that the defendant matched on race/age/jeans/letters on hoodie/scar on chin (25%) than were participants who viewed the trial in which the pre-identification evidence was strong (91%), χ^2 (1 N = 515) = 164.16, $p < .001$.

Expert Witness Manipulation

Participants who viewed the trial in which the expert witness testified about the importance of base rates of guilt were more likely to indicate that they heard testimony on base-rates (91%) than participants who viewed the trial in which the expert witness did not testify about the importance of base rates of guilt (26%), χ^2 (1, N = 515) = 159.36, $p < .001$. Consistent with my pre-registered analysis plan, the data from all participants who failed either manipulation check item were excluded in subsequent analyses, leaving us with a sample of 357 participants.

Verdict

I conducted a binary logistic regression with participants' verdict decision as the dependent variable and pre-identification evidence, presence of expert testimony, and the interaction term as the independent variables. Participants were significantly more likely to vote guilty when the pre-identification evidence was strong (54%) than weak (14%), χ^2 (1, N = 357) = 31.53, $p < .001$, OR = 8.80, 95% CI [4.12, 18.79]. Participants did not make significantly different verdict decisions when expert testified about the importance of base rates of guilt (34%) as compared to when they did not (34%), χ^2 (1, N = 357) = 0.17, p = .68, OR = 1.20, 95% CI [0.51, 2.84]. Last, participants were equally likely to convict when the pre-identification evidence was strong or weak, irrespective of the presence of expert testimony on the importance of base rates of guilt, (13% $_{\text{weak evidence connection, no expert}}$ vs. 15% $_{\text{weak evidence connection, expert}}$) or not

($56\%_{\text{strong evidence connection, no expert}}$ vs. $52\%_{\text{strong evidence connection, expert present}}$), as no significant interaction emerged, $\chi^2(1, N = 357) = 0.36$, $p = .55$, $OR = 0.73$, 95% CI [0.26, 2.06] (see Table 1 for full results).

Strength of Prosecution Case

I also conducted a 2 (Pre-Identification Evidence: weak vs. strong) × 2 (Base Rate Expert Testimony: present vs. absent) analysis of variance (ANOVA) on participants' evaluations of the strength of the prosecution's case. The strength of pre-identification evidence had a significant effect on participants' evaluations of the prosecution's case, $F(1, 353) = 122.28$, $p < .001$, $d = 1.18$, 95% CI [1.03, 1.32]. Participants found the prosecution's case to be weaker when the pre-identification evidence was weak ($M = 2.86$, $SD = 1.40$) than when it was strong ($M = 4.49$, $SD = 1.40$). The presence of expert testimony on base rates did not affect participant judgments, $F(1, 353) = 0.20$, $p = .66$, $d = 0.05$, 95% CI [- 0.10, 0.19]. Participants evaluated the strength of the prosecution's case as equally strong when the expert testified about the importance of base-rates of guilt ($M = 3.70$, $SD = 1.40$) and when the expert just testified about the effects of stress on memory accuracy ($M = 3.64$, $SD = 1.40$). No significant interaction emerged, $F(1, 353) = 0.35$, $p = .55$, $\eta^2 = 0.001$ (see Table 2 for full results).

Eyewitness Credibility

I also conducted a 2 (Pre-Identification Evidence: weak vs. strong) × 2 (Base Rate Expert Testimony: present vs. absent) analysis of variance (ANOVA) on participants' evaluations of eyewitness credibility. The strength of pre-identification evidence had a significant effect on participants' evaluations of eyewitness credibility, $F(1, 353) = 41.66$, $p < .001$, $d = 0.68$, 95% CI [0.56, 0.81]. Participants rated the eyewitness as more credible when the pre-identification evidence was strong ($M = 5.28$, $SD = 1.22$) than when it was weak ($M = 4.44$, $SD = 1.22$). The

presence of base rate expert testimony did not affect participants' evaluations of eyewitness credibility, $F(1, 353) = 2.63$, $p = .11$, $d = 0.17$, 95% CI [0.05, 0.30]. Participants rated the eyewitness as equally credible when the expert testified about the importance of base rates ($M = 4.75$, $SD = 1.22$) and when they did not ($M = 4.96$, $SD = 1.22$). No significant interaction emerged, $F(1, 353) = 0.34$, $p = .56$, $\eta^2 = 0.001$ (see Table 3 for full results).

Expert Witness Credibility

I also conducted a 2 (Pre-Identification Evidence: weak vs. strong) × 2 (Base Rate Expert Testimony: present vs. absent) analysis of variance (ANOVA) on participants' evaluations of expert witness credibility. The strength of pre-identification evidence strength did not have a significant effect on participants' evaluations of expert witness credibility**,** $F(1, 353) = 0.24$, $p = .63$ $d = 0.04$, 95% CI [-0.06, 0.14]. Participants rated the expert witness as equally credible when pre-identification evidence was weak ($M = 5.93$, $SD = .94$) and strong ($M = 5.89$, $SD = .94$). The presence of base rate expert testimony did not affect participants' evaluations of expert witness credibility, $F(1, 353) = 0.44$, $p = .51$ $d = 0.06$, 95% CI [-0.03, 0.16]. Participants rated the expert witness as equally credible when they testified about base-rates of guilt ($M = 5.89$, $SD = .94$) and or not ($M = 5.94$, $SD = .94$). No significant interaction emerged, $F(1, 353) = 0.91$, $p = .34$, $\eta^2 = 0.003$ (see Table 4 for full results).

Lineup Fairness

I also conducted a 2 (Pre-Identification Evidence: weak vs. strong) × 2 (Base Rate Expert Testimony: present vs. absent) analysis of variance (ANOVA) on participants' evaluations of lineup fairness. The strength of pre-identification evidence had a significant effect on participants' evaluations of lineup fairness**,** $F(1, 353) = 98.36$, $p < .001$, $d = 1.05$, 95% CI [0.90, 1.21]. Participants found the lineup to be more fair when the pre-identification evidence was

strong (M = 4.67, SD = 1.48) than when it was weak (M = 3.14, SD = .1.48). The presence of base-rate expert testimony had a significant effect on participants' evaluations of lineup fairness, $F(1, 353) = 7.49$, $p = .007$, $d = 0.29$, 95% CI [0.14 0.44]. Participants found the lineup to be more fair when the expert did not testify about the base-rates of guilt (M = 4.13, SD = 1.48) than when she did (M = 3.70, SD = 1.48). No significant interaction emerged, $F(1, 353) = 0.12$, $p = .73$, $\eta^2 = 0.001$ (see Table 5 for full results).

Discussion

Eyewitness misidentification is one of the leading contributors to wrongful convictions (Innocence Project, 2023). The probability of guilt before a defendant is placed into a lineup (prior probability of guilt), or strength of the evidence against a suspect before being placed into an identification procedure (pre-identification evidence strength), is one of the most important factors relating to a defendant's actual status of guilt. However, no previous study has analyzed whether pre-identification evidence strength would influence juror decision making. Thus, this study is an important first step in examining juror sensitivity to evidence-based suspicion. This study also addressed whether expert testimony about the role of the base-rate of guilty suspects in identification accuracy sensitizes jurors to the strength of the suspect- crime connection *before* the identification procedure.

Pre-Identification Evidence Connection Effects

Participants were relatively sensitive to pre-identification evidence strength when the evidence connection was strong versus weak. Participants were more likely to vote guilty when the pre-identification evidence was strong rather than weak, demonstrating that participants could appropriately distinguish between a strong versus weak pre-identification evidence connection. Moreover, the strength of pre-identification evidence influenced participants'

evaluations of the strength of the prosecution's case. Participants were more likely to indicate the prosecution's case was strong when the pre-identification evidence was strong rather than weak. The strength of pre-identification evidence had similar effects on participants evaluations of eyewitness credibility. Participants rated the eyewitness as more credible when the evidence connection was strong and less credible when the pre-identification evidence was weak. When the evidence connection was high the participants found the line-up to be more fair versus when the evidence connection was low they perceived the line-up to be less fair. Thus, mock jurors were already sensitive to pre-identification evidence strength on their own. Taken together, these results suggest that participants were sensitive to the strength of the pre-identification evidence.

Expert Testimony Effects

Expert testimony on the importance of the base rate of guilt had little effect on participants' judgments**.** I hypothesized that participants would be more likely to vote guilty when the expert testified about the importance of base rates of guilt. However, participants' verdict did not differ when the expert testified about baste rates of guilt versus when the expert merely testified about the effects of stress on memory accuracy, nor did expert testimony content affect participants' ratings of case strength. Participants also found both the expert witness and eyewitness as equally credible irrespective of whether the expert testimony addressed the importance of base rates of guilt. However, participants found the lineup to be more fair when the expert did not testify about the base-rates of guilt. Thus, at least on this measure, base rate expert testimony caused participants to be more skeptical rather than more attuned to the fairness of the lineup. Overall, these (primarily null) effects underscore the point that expert testimony on base rates of guilt may be an ineffective safeguard in this context.

Limitations and Future Research

There are various limitations of this research that warrant discussion. The first limitation of this study is I used an MTurk community sample as my participant pool. Crowdsourcing platforms will never be able to fully represent a jury, however prior research has indicated that the data collected on these online platforms are highly reliable, highly similar to subjects on other platforms, and produce findings similar to in-person research (Irvine et al., 2018). Therefore, in future research there would be a possible benefit of replicating my study using in-person data collection or other online platforms to test for differing results.

Second, the trial summary used in my mock trial, while detailed and realistic, does not mirror the stimulus provided in the courtroom in a real case. I filmed the video in a mock courtroom at the John Jay College of Criminal Justice and used actors. Therefore, the trial lacks audience-related cues (reactions of other jurors) that can have the potential to influence jurors' decisions. Further, the participants only heard testimony from one eyewitness, one police officer, and one expert witness. In the base-rates of guilt condition the expert testified for approximately fifteen minutes and in the condition with no testimony on the prior base-rates of guilt the testimony lasted approximately eight minutes. Additionally, the entire mock trial video lasted for approximately thirty minutes. Whereas, in an actual jury trial the jurors would hear from numerous witnesses and experts with varying points of view and the trial would take place over multiple days or weeks.

Third, the mock jurors in this study rendered their verdicts and deliberated independent of one another. Ellsworth (1989) found that jurors who deliberated as a jury did better at understanding and remembering the evidence presented at trial. Salerno and McCauley (2009) showed that deliberating can increase jurors' confidence levels in a verdict that closely aligns

with an accurate expert. The perceived skepticism that participants experienced due to the expert witness may have been exacerbated or neutralized with the addition of deliberation. Thus, for future research the addition of juror deliberation would be a point of interest.

Implications

My findings have various practical implications, as the goal of my research was to explore whether jurors are sensitive to variations in pre-identification evidence strength and if expert testimony would increase their sensitivity. Participants were able to identify when the pre-identification evidence was strong or weak and this variation was reflected in their verdicts. However, jurors' evaluations did not fully reflect the strength of the eyewitness identification evidence when the evidence connection was strong, suggesting that further education of jurors is necessary to help them understand how the probability of the suspect's guilt before the eyewitness makes an identification affects the probability of the suspect's guilt after the identification. That is, all identifications are not equal.

Expert testimony was not sensitizing jurors to the fairness of the lineup but was rather creating skepticism. Based on these findings, strategically incorporated expert testimony may still be an effective solution to remedying jurors base-rate neglect. However, expert testimony on the importance of base rates of guilt by itself is not enough to serve as a safeguard for correcting eyewitness misidentifications. Instead, the police need to make policy changes that require evidence connecting suspects to a crime before conducting lineup procedures (Wells et al., 2020). If the police adopted a policy such as evidence-based suspicion it would increase the base-rates of guilt and therefore, increase the likelihood of a correct identification. Thus, rather than placing suspects into lineups based on a "hunch", the police need to have specific evidence

connecting a suspect to a crime prior to an identification procedure to lower the rates of eyewitness misidentifications.

References

American Psychological Association. (2016). Brief for Amicus Curiae, Payne v. Commonwealth of Virginia. Retrieved from http://www.apa.org/about/offices/ogc/amicus/payne.pdf

Behrman, B. W., & Richards, R. E. (2005). Suspect/foil identification in actual crimes and in the laboratory: A reality monitoring analysis. *Law and Human Behavior*, *29*, 279–301.

Brain, R. D., & Broderick, D. J. (1991). The derivative relevance of demonstrative evidence: Charting its proper evidentiary status. *UC Davis Law Review*, *25*, 957–1027.

Brekke, N., & Borgida, E. (1988). Expert psychological testimony in rape trials: A social-cognitive analysis. *Journal of Personality and Social Psychology*, *55*(3), 372–386. https://doi.org/10.1037/0022-3514.55.3.372.

Brigham, J. C., & Bothwell, R. K. (1983). The ability of prospective jurors to estimate the accuracy of eyewitness identifications. *Law and Human Behavior*, *7*(1), 19–30. https://doi.org/10.1007/BF01045284

Carroll, J. S. (1977). Judgments of recidivism risk: Conflicts between clinical strategies and base-rate information. *Law and Human Behavior*, *1*(2), 191–198.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58*(1), 1–73.

Epley, N., Savitsky, K., & Gilovich, T. (2002). Empathy neglect: Reconciling the spotlight effect and the correspondence bias. *Journal of Personality and Social Psychology*, *83*(2), 300–312. https://doi.org/10.1037/0022-3514.83.2.300.

Faigman, D. L., & Baglioni Jr., A. J. (1988). Bayes' theorem in the trial process: Instructing jurors on the value of statistical evidence. *Law and Human Behavior*, *12*(1), 1–17.

Innocence Project (2023). https://www.innocenceproject.org/dna-exonerations-in-the-united-states/

Innocence Project. (n.d.). *Ronald Cotton*. Innocence Project.

Hammerton, M. (1973). A case of radical probability estimation. *Journal of Experimental Psychology*, *101*(2), 252–254. https://doi.org/10.1037/h0035224.

Jones, A. M., & Kovera, M. B. (2015). A demonstrative helps opposing expert testimony sensitize jurors to the validity of scientific evidence. *Journal of Forensic Psychology Practice*, *15*(5), 401–422.

Katzman, J., & Kovera, M. B. (2022). Evidence strength (insufficiently) affects police officers' decisions to place a suspect in a lineup. *Law and Human Behavior*, *46*(1), 30.

Kaye, D. H., & Koehler, J. J. (1991). Can jurors understand probabilistic evidence? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *154*(1), 75–81.

Lyon, D., & Slovic, P. (1976). Dominance of accuracy information and neglect of base rates in probability estimation. *Acta Psychologica*, *40*(4), 287–298.

McAuliff, B. D., & Kovera, M. B. (2008). Juror Need for Cognition and sensitivity to methodological flaws in expert evidence. *Journal of Applied Social Psychology*, *38*(2), 385–408.

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

Schmechel, R. S., O'Toole, T. P., Easterly, C., & Loftus, E. F. (2005). Beyond the ken? Testing jurors' understanding of eyewitness reliability evidence. *Jurimetrics*, *46*, 177–214.

Salerno, J. M., & McCauley, M. R. (2009). Mock jurors' judgments about opposing scientific experts: Do cross-examination, deliberation and need for cognition matter? *American Journal of Forensic Psychology, 27*(3), 37–60.

Schuller, R. A., & Cripps, J. (1998). Expert evidence pertaining to battered women: The impact of gender of expert and timing of testimony. *Law and Human Behavior*, *22*, 17–31.

Smith, V. L., Kassin, S. M., & Ellsworth, P. C. (1989). Eyewitness accuracy and confidence: Within-versus between-subjects correlations. *Journal of Applied Psychology*, *74*(2), 356 – 359. https://doi.org/10.1037/0021-9010.74.2.356.

Wells, G. L., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C. A., & Wixted, J. T. (2020). Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior*, *44*(1), 3–36. https://doi.org/10.1037/lhb0000359

Wixted, J. T., Mickes, L., Dunn, J. C., Clark, S. E., & Wells, W. (2016). Estimating the reliability of eyewitness identifications from police lineups. *Proceedings of the National Academy of Sciences*, *113*(2), 304–309.

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, *18*(1), 10–65. https://doi.org/10.1177/1529100616686966

Wise, R. A., Safer, M. A., & Maro, C. M. (2011). What US law enforcement officers know and believe about eyewitness factors, eyewitness interviews and identification procedures. *Applied Cognitive Psychology*, *25*(3), 488–500.

Table 1

Binary Logistic Regression Results of Manipulated Variables and their Interaction on Verdict

Variables in the equation	B	*SE*	Wald	*p*	*OR*	95% CI for *OR*	
						Lower	Upper
Expert Testimony	0.18	0.44	0.17	.68	1.20	0.51	2.84
Pre-Identification Evidence	2.17	0.39	31.53	.001	0.73	4.12	18.79
Expert Testimony × Pre-ID Evidence	-0.32	0.53	0.36	.55	0.73	0.26	2.06
Constant	-1.95	0.32	36.45	<.001	0.14		

Note. CI = confidence interval; *OR* = odds ratio. All variables were simultaneously entered into the logistic regression equation. All $df = 1$, $N = 357$.

Table 2

Results of ANOVA Examining the Effects of Manipulated Variables on Perceptions of Prosecution Case Strength

Variables in the equation	*df*	*F*	η_p^2	*p*
Expert Testimony	1	0.20	.001	.66
Pre-ID Evidence	1	122.28	.26	.001
Expert Testimony × Pre-ID Evidence	1	0.12	.001	.73

Table 3

Results of ANOVA Examining the Effects of Manipulated Variables on Perceptions of Eyewitness Credibility

Variables in the equation	*df*	*F*	η_p^2	*p*
Expert Testimony	1	2.63	.01	.11
Pre-ID Evidence	1	41.66	.11	.001
Expert Testimony × Pre-ID Evidence	1	0.34	.001	.56

Table 4

Results of ANOVA Examining the Effects of Manipulated Variables on Perceptions of Expert Credibility

Variables in the equation	*df*	*F*	η_p^2	*p*
Expert Testimony	1	0.39	.001	.51
Pre-ID Evidence	1	0.21	.001	.63
Expert Testimony × Pre-ID Evidence	1	1.08	.003	.27

Table 5

Results of ANOVA Examining the Effects of Manipulated Variables on Perceptions of Lineup Fairness

Variables in the equation	*df*	*F*	η_p^2	*p*
Expert Testimony	1	7.49	.02	.007
Pre-ID Evidence	1	93.36	.22	.001
Expert Testimony × Pre-ID Evidence	1	0.14	.001	.73