JUROR CERTAINTY ABOUT EXPERT FIREARMS IDENTIFICATION EVIDENCE AND THE IMPACT OF CROSS-EXAMINATION

PARAIC SCANLON, BOGLARKA BANYAI, ELLIS HART, & SARAH L. COOPER[1]

ABSTRACT

Firearms examiners compare toolmarks present on suspect ammunition to those present on ammunition test-fired by a suspect weapon to evaluate association between the two. Examiners' conclusions are generally admissible in U.S. courts, yet the scientific underpinnings of the toolmark discipline have been subject to considerable criticism. Cross-examination can be used to bring such criticism to the attention of jurors in order to highlight the flaws inherent in toolmark analysis to those who determine the weight of expert evidence.We investigated the effect of such cross-examination on juror certainty about expert firearms evidence using online vignettes. A community sample of 437 U.S. participants was asked to rate its certainty, on a scale of 0 to 100, of a forensic match between a suspect weapon and suspect ammunition for each of four expert statements of certainty, in two groups; either with or without a cross-examination highlighting limitations of the toolmark discipline's scientific underpinnings. We analyzed differences between both groups and between the statements given to each group. Results suggest that cross-examination can have a strong influence on juror decision-making, particularly when experts express their conclusions in certain terms.

[1]Dr. Paraic Scanlon, Senior Lecturer in Developmental Psychology, Birmingham City University, School of Social Sciences, Department of Psychology; Boglarka Banyai (Bsc) Hons, MSc; Ellis Hart LL.B Hons; Dr. Sarah L. Cooper, Reader in Law, Birmingham City University, School of Law, UK. This research was supported by the Small Development Grant scheme ran by Birmingham City University's Faculty of Business, Law and Social Sciences. The authors would like to thank participants at the 2018 and 2019 European Association of Psychology and Law Annual Conferences for feedback on this series of work, and Turan Avkesen for his editorial assistance.

INTRODUCTION

Firearms examiners compare toolmarks present on suspect ammunition to those present on ammunition test-fired by a suspect weapon to evaluate association between the two. To evaluate association, examiners compare toolmarks present on suspect ammunition to those present on ammunition test-fired by a suspect weapon.[2] The Association of Firearms and Tool Mark Examiners ("AFTE") has developed a protocol ("AFTE Protocol") to guide examiners, under which they can come to one of four conclusions: (1) identification; (2) inconclusive; (3) elimination; or (4) unsuitable for comparison.[3] Examiner conclusions have been routinely admitted into U.S. courts as expert evidence for around a century.[4] However, in recent decades, concerns about the reliability of toolmarks have been repeatedly raised, with reports by the National Research Council[5] and President's Council of Advisors on Science and Technology[6] querying, for example, the validity of assumptions about uniqueness and reproducibility in the context of tool-marks made by firearms;[7] the precision of the AFTE Protocol;[8] and the discipline's scientific knowledge base.[9] Such concerns are not shared equally across all stakeholders, which

[2] NAT'L RSCH. COUNCIL OF THE NAT'L ACADS., STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 152–53 (Nat'l Academies Press 2009) [hereinafter STRENGTHENING].
[3] Ass'n of Forensic Sci. Providers, *Theory of Identification as it Relates to Toolmarks*, 30 ASS'N FIREARMS & TOOL MARK EXAM'RS J., 86–87 (1998).
[4] Paul C. Giannelli, Edward J. Imwinkelried, & Joseph L. Peterson, *Reference Guide on Forensic Identification Expertise*, *in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 55, 91 (3d ed., Nat'l Academies Press 2011) ("The technique subsequently [after the 1920s] gained widespread judicial acceptance and was not seriously challenged until recently.").
[5] *See generally* NAT'L RSCH. COUNCIL, BALLISTIC IMAGING (Daniel L. Cork et al. eds., Nat'l Academies Press 2008) [hereinafter BALLISTIC IMAGING].
[6] PRESIDENT'S COUNCIL OF ADVISORS ON SCI. AND TECH., FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS 11 (2016) [hereinafter PCAST REPORT].
[7] BALLISTIC IMAGING, *supra* note 5, at 3.
[8] STRENGTHENING, *supra* note 2, at 155.
[9] *Id.* at 155 (concluding that "the scientific knowledge base for toolmark and firearms analysis is fairly limited . . . . "). Note that, with regard to firearms identification, PCAST stated its conclusions were "consistent" with those in STRENGTHENING. *See* PCAST REPORT, *supra* note 6, at 11.

include the Federal Bureau of Investigation, AFTE, and the Department of Justice, yet a shared vision for the continual enhancement of forensic science generally is.[10]

Some U.S. courts have responded to the debate by restricting the degree of certainty that experts may express in their conclusions. For example, phrases such as "there is an exact match" would in some courts have to be replaced with phrases such as "a match can be made," "more likely than not," or "to a reasonable degree of certainty."[11] Jurors must determine the weight of these phrases, and they present various challenges.[12] Noting this, two of the Article's authors, Paraic Scanlon and Sarah Cooper, using online vignettes – short text descriptions of courtroom interactions – investigated juror certainty of the association between a suspect weapon and suspect ammunition when presented with a variety of phrases by a qualified firearms examiner in a trial setting (n=107). They found a significant main effect for certainty, with increased expert certainty generally leading to increased participant certainty.[13] Cooper and Scanlon suggested, *inter alia*,

[10] Sarah L. Cooper & Paraic Scanlon*, Juror Assessment of Certainty about Firearms Identification Evidence,* 40 U. Ark. Little Rock L. Rev. 95, 101–02 (2017).

[11] *Id.* at 102–04 (summarizing numerous court responses between 2005 and 2012). Notably, more recent case law reflects the ongoing use of/reference to debate around these phrases. *See, e.g.*, People v. Ross, 129 N.Y.S.3d 629, 639–40 (Sup. Ct. 2020) ("Some courts still permit a testifying ballistics examiner to recite the 'reasonable degree of ballistics certainty' standard . . . . But other courts have found that testimony too misleading . . . . The NYPD laboratory itself has now turned away from the 'reasonable degree of scientific certainty' standard in drawing its conclusions about ballistics and to the 'sufficient agreement' language consistent with AFTE guidelines . . . . Consequently, the scope of permissible expert toolmark testimony is narrowing overall.") (citations omitted); U.S. v. Harris, 502 F. Supp. 3d 28, 44–45 (D.D.C. 2020) ("Limitations restricting the degree of certainty that may be expressed on firearm and toolmark expert testimony are not uncommon. *See, e.g., Romero-Lobato*, 379 F. Supp. 3d at 1117 (noting the "general consensus" of the courts "is that firearm examiners should not testify that their conclusions are infallible or not subject to any rate of error, nor should they arbitrarily give a statistical probability for the accuracy of their conclusions"); *Ashburn*, 88 F. Supp. 3d at 249 (limiting expressions of an expert's conclusions to that of a 'reasonable degree of ballistics certainty' or a 'reasonable degree of certainty in the ballistics field.'); *Diaz*, 2007 WL 485967 at *1 (same).").

[12] *See infra* Part II.

[13] Cooper & Scanlon, *supra* note 10 at 114.

further investigation into whether adding context about the considered limitations of firearms evidence would influence juror certainty.[14]

This Article reports on the authors' study of the influence of this additional context, incorporated through cross-examination. Part I explores current literature concerning the intersection of juror decision-making, expert evidence, and cross-examination. Part II describes the authors' study design and results, which suggest that cross-examination can have a strong influence on juror decision-making, particularly when experts express their conclusions in certain terms. Part III discusses their findings in the context of current understandings of the effect of cross-examination, highlighting particular connections to reforms focused on the standardization of expert evidence and education and training for jurors. It concludes by noting study limitations and suggesting areas for further research, including on the style, length, and content of cross-examination and the effect of witness rehabilitation.

I. JUROR DECISION-MAKING, EXPERT EVIDENCE & CROSS-EXAMINATION

To position the authors' study, this Part broadly summarizes some relevant findings across the literature at the intersection of juror decision-making, expert evidence, and cross-examination.

Various factors come into play when jurors engage with expert evidence. Paul Roberts and Adrian Zukerman have explained that jurors can be confused by complex evidence and prefer to base their decisions on an expert's conclusions.[15] Suzanne Blackwell and Fred Seymour concluded that, when engaging with expert witnesses, jurors rank relevant professional experience, lack of

[14] *Id.* at 117.
[15] PAUL ROBERTS & ADRIAN ZUCKERMAN, CRIMINAL EVIDENCE 294–95 (2d ed., 2010).

bias, and clarity of evidence in order of importance.[16] Studies show jurors experience difficulties in recognizing bias, however. For instance, Bradley McAuliff and Tajeh Duckworth's study attempting to educate jurors on the reliability of forensic evidence concluded that while some jurors are able to recognise methodological flaws presented in expert testimony, they remained unable to identify bias.[17] Further, by introducing an opposing expert in a mock trial, Lora Levett and Margaret Bull Kovera attempted to sensitize participant jurors to flaws in forensic evidence and examinined ratings given by participating jurors to both experts' credibility, trustworthiness, research quality, and verdicts.[18] Results indicated that jurors did not become sensitized to flaws in the testimony, but, rather, the presence of opposing experts made jurors skeptical of both the experts and their trustworthiness.[19]

The content and expression of expert forensic evidence have also been explored. For instance, William Thompson and Eryn Newman found that perceptions of both DNA and shoe-print evidence are modified by prior expectation and belief as well, as the content of the evidence itself.[20] Johnathan Koehler and J. Ritchie[21] removed much of the context while examining expert statements of numerical certainty about DNA evidence, and they found that expert testimony that

[16] Suzanne Blackwell & Fred Seymour, *Expert Evidence and Jurors' Views on Expert Witnesses*, 22 PSYCHIATRY PSYCH. & L. 673, 678–79 (2015).
[17] Bradley D. McAuliff & Tejah D. Duckworth, *I Spy with My Little Eye: Jurors' Detection of Internal Validity Threats in Expert Evidence*, 34 LAW & HUM. BEHAV. 489, 499 (2010).
[18] Lora M. Levett & Margaret Bull Kovera, *The Effectiveness of Opposing Expert Witnesses for Educating Jurors About Unreliable Expert Evidence*, 32 LAW & HUM. BEHAV. 363, 363 (2008).
[19] *Id.* at 370.
[20] William C. Thompson & Eryn J. Newman, *Lay Understanding of Forensic Statistics: Evaluation of Random Match Probabilities, Likelihood Ratios, and Verbal Equivalents*, 39 LAW & HUM. BEHAV. 332, 347 (2015).
[21] *See generally* Jonathan J. Koehler, *The Psychology of Numbers in the Courtroom: How to Make DNA Match Statistics Seem Impressive or Insufficient*, 74 S. CAL. L. REV. 1275 (2001); J. Ritchie, *Probabilistic DNA Evidence: The Layperson's Interpretation*, 47 AUSTL. J. FORENSIC SCI. 440 (2015).

formulates DNA evidence in terms of exclusion percentages[22] is more likely to result in a conviction than if an expert were to testify in terms of frequency ratios.[23] Case law suggests, however, that forensic experts (including firearms examiners[24]) routinely testify using linguistic or ordinal category-based evidence[25]—that is, evidence that suggests a fixed hierarchy-—but the numerical difference between the categories is not fixed. Any verbal certainty judgements made without explicit statistical information necessarily balance the meaning of a phrase provided by the expert with the interpreter's own subjective understanding of the evidence.[26] Kristy Martire et al. found that inculpatory evidence was significantly more likely to be judged as weak if it was presented in terms of linguistic descriptions than if it was expressed in terms of statistical or numerical likelihood.[27] The Association of Forensic Science Providers has released guidelines suggesting forensic experts use likelihood ratios,[28] yet "uptake of these standards has varied considerably across jurisdictions and disciplines."[29]

Studies have specifically explored juror interpretation of certainty expressed by different expert linguitsic phrases. Dawn McQuiston-Surrett and Michael Saks asked participants in a study to rate an odontology expert's intended certainty (on a scale of one to one hundred) from four

[22] For example, a statement such as: "The probability that the suspect would match the blood specimen if he wasn't the source is 0.1%."
[23] For example, a statement such as: "The frequency that the suspect would match the blood specimen if he wasn't the source is 1 in 1000."
[24] Cooper & Scanlon, *supra* note 10, at 102–04.
[25] For example, phrases such as "likely," "very likely," and "extremely likely."
[26] Thomas S. Wallsten, Samuel Fillenbaum & James A Cox, *Base Rate Effects on the Interpretations of Probability and Frequency Expressions*, 25 J. MEMORY & LANGUAGE 571, 577 (1986).
[27] Kristy A. Martire, Richard I. Kemp, Ian Watkins, Malindi A. Sayle & Ben R. Newell, *The Expression and Interpretation of Uncertain Forensic Science Evidence: Verbal Equivalence, Evidence Strength, and the Weak Evidence Effect*, 37 LAW & HUM. BEHAV. 197, 200 (2013).
[28] *Id.* at 197; Ass'n of Forensic Sci. Providers, *Standards for the Formulation of Evaluative Forensic Science Expert Opinion*, 49 SCI. & JUST. 161, 161–62 (2009).
[29] Martire et al., *supra* note 27, at 197.

phrases taken from American Board of Forensic Odontology ("ABFO") guidelines.[30] Responses showed that participant estimates did not mirror the ABFO's intended hierarchy.[31] "*Match*" was assumed by ABFO to mean "*no expression of specificity intended; generally similar but true for large percentage of population,*" but was rated as the most certain statement (86/100) by participants, ahead of "*Consistent with*" (75/1000) another statement assumed to be less certain and significantly above the two assumedly more certain phrases: "*Probable*" (57/100) and "*Reasonable Scientific Certainty*" (70/100).[32] The researchers concluded that "[t]hese findings suggest a straightforward lesson. Forensic expert witnesses cannot simply adopt a term, define for themselves what they wish it to mean, and expect judges and juries to understand what they mean by it."[33] McQuiston-Surrett and Saks also examined certainty statements in the context of microscopic hair analysis.[34] The study involved both potential juror and judicial participants and included a comparison between two subjective qualitative statements—"*match*" and "*similar in all microscopic characteristics*"—and three quantitative statements—"*objective single-probability*," "*subjective probability*," and "*objective multiple-frequency*."[35] The researchers found that participants deemed qualitative statements significantly more certain than the subjective probability or objective multiple-frequency statements.[36] Non-judicial participants were particularly susceptible to this effect. [37] More recent research on qualitative firearms examiner

[30] Dawn McQuiston-Surrett & Michael J. Saks, *Communicating Opinion Evidence in the Forensic Identification Sciences: Accuracy and Impact*, 59 HASTINGS L.J. 1159, 1161 (2008).
[31] *Id.* at 1162–63.
[32] *Id.* (emphasis added).
[33] *Id.* at 1163.
[34] Dawn McQuiston-Surrett & Michael J. Saks, *The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Factfinders Hear*, 33 LAW & HUM. BEHAV. 436, 437 (2009).
[35] *Id.* at 437–38.
[36] *Id.* at 444.
[37] *Id.* at 445.

testimony has found that it also "constitutes powerful evidence" in comparison to an inconclusive phrase.[38]

These findings of varying interpretations fit with general concerns about the challenges that can emerge when legal actors engage with scientific evidence. Legal education is considered a "black hole"[39] for STEM education, meaning lawyers can be "ill-equipped to speak the language of science."[40] This educational deficiency "often places lawyers [and judges] at a disadvantage when confronted with scientific evidence . . . . [L]awyers . . . often fail to ask the right questions and uncritically accept scientific assertions."[41] The National Academy of Sciences has recognized these challenges, as well.[42] Equally, most jurors lack scientific expertise.[43] Research has shown both that jury and bench trial verdicts are consistent regardless of the scientific complexity involved, and that outcomes are generally justified in cases where jurors have expressed incomplete or flawed understanding of scientific or technical evidence.[44] However, jurors may struggle to engage with scientific and technical evidence,[45] although a better understanding of juror comprehension of forensic evidence is needed.[46] Based on their work, McQuiston-Surrett and Saks

[38] Brandon L. Garrett, Nicholas Scurich & William E. Crozier, *Mock Jurors' Evaluation of Firearm Examiner Testimony*, 44 LAW HUM. BEHAV. 412 (2020) (emphasis added). (In Study 1, the researchers hypothesized that jurors would accord significant weight to testimony declaring a match between two catridge cases over testimony with only an inconclusive statement *Id.* at 414. Their findings corroborated this, leading to a conclusion that firearms expert testimony is powerful evidence to jurors. *Id.* at 416.

[39] Jessica D. Gabel, *Forensiphilia: Is Public Fascination with Forensic Science A Love Affair or Fatal Attraction?*, 36 NEW ENG. J. ON CRIM. & CIV. CONFINEMENT 233, 255 (2010).

[40] *Id.* at 258.

[41] Fredric I. Lederer, *Scientific Evidence—An Introduction*, 25 WM. & MARY L. REV. 517, 519–20 (1984).

[42] STRENGTHENING, *supra* note 2, at 27 ("In addition, lawyers and judges often have insufficient training and background in scientific methodology, and they often fail to fully comprehend the approaches employed by different forensic science disciplines and the reliability of forensic science evidence that is offered in trial.").

[43] DAVID L. FAIGMAN, LEGAL ALCHEMY: THE USE AND MISUSE OF SCIENCE IN THE LAW 53 (W.H. Freeman & Co. 1999).

[44] STRENGTHENING, *supra* note 2, at 236.

[45] *See* Symposium, *Judges, Juries, and Scientific Evidence*, 16 J.L. & POL'Y 19, 24 (2007) (stating that jurors can have difficulty understanding scientific and technical evidence, particularly DNA evidence).

[46] STRENGTHENING, *supra* note 2, at 237.

conclude that a better understanding can be developed through "empirical testing of the responses to the words"[47] and that such experimentation "need not be difficult."[48]

Noting this, Cooper and Scanlon took a simplified approach to assessing juror certainty about expert firearms evidence, removing as much context as possible in order to encourage participants to focus on the linguistic content of the phrases from exper testimony.[49] Participants were asked to rate their level of certainty (on a scale from zero to one hundred) of a match between a defendant's firearm and suspect ammunition based on various expert statements, with each expert statement hypothesized to attract high, moderate, or low certainty.[50] The study found a general trend towards high-certainty expert statements, and also suggested that when experts convey their conclusions in terms of "practical," "professional," and "ballistic" certainty, participants see them as more certainty-inducing while more absolute statements of certainty or uncertainty are less appealing.[51] Based on their findings, Cooper and Scanlon suggested, *inter alia*, investigating whether adding more context, for example "by highlighting the alleged limitations of firearms identification evidence" would influence juror certainty. [52]

One way to integrate such context would be to add a cross-examination statement to the study design. Cross-examination is the process by which lawyers question opposing witnesses, including expert witnesses, with the aim of testing the reliability and relevance of admissible testimony. Cross-examination of expert witnesses is particularly important because expert

[47] McQuiston-Surrett & Saks, *supra* note 30, at 1163.
[48] *Id.*
[49] Cooper & Scanlon, *supra* note 10, at 109.
[50] *Id.* For example, researchers hypothesized that an expert statement containing the words "exact match" would attract high certainty, "match to a reasonable degree of professional certainty" would attract moderate certainty, and a statement that the results of the expert's examination were "inconclusive" would attract low certainty. *Id.* at 110–114.
[51] *Id.* at 109–15.
[52] *Id.* at 117.

witnesses provide testimony about technical and scientific matters beyond the knowledge of the fact-finder (i.e., jurors) whose role it is to weigh such evidence. Cross-examination is a characteristic feature of adversarial justice systems, such as that of the United States. In fact, the U.S. justice system relies on cross-examination to resolve fallibilities associated with expert evidence, as underscored by the U.S. Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), with the Court stating, "Vigorous cross-examination, presentation of contrary evidence . . . are the traditional and appropriate means of attacking shaky but admissible evidence."[53] This is so, despite some studies producing findings that should "give pause to anyone who believes that the traditional tools of the adversarial process [e.g., cross-examination] will always undo the adverse effects of weak expert testimony."[54]

The effect of cross-examination in the general context of this study has been explored elsewhere. Margeret Bull Kovera, Bradley McAuliff, and Kellye Hebert found that cross-examination failed to highlight issues with the validity of scientific evidence.[55] Brandon Garrett, Nicholas Scurich, and William Crozier found comparable evidence using a transcript of a cross-examination. They focused on three expert equivocal statements—based on an inconclusive result, a "simple identification," and an inability to exclude on the level of evidence—and found that cross-examination had no significant effect on potential jurors' decisions.[56] However, some research has begun to identify circumstances in which jurors were able to determine the validity of scientific measures. For example, Jacqueline Austin and Margaret Kovera found that jurors

[53] Daubert v. Merrell Dow Pharms. Inc., 509 U.S. 579, 596 (1993).
[54] McQuiston-Surrett & Saks, *supra* note 30, at 1188; McQuiston-Surrett & Saks, *supra* note 34, at 451.
[55] Margaret Bull Kovera, Bradley D. McAuliff & Kellye S. Hebert, *Reasoning About Scientific Evidence: Effects of Juror Gender and Evidence Quality on Juror Decisions in a Hostile Work Environment Case*, 84 J. APPLIED PSYCH. 362, 372 (1999).
[56] Garrett et al., *supra* note 38 at 420–21.

informed about methodological flaws or validity standards in DNA evidence through cross-examination were more discerning about the "quality" of the evidence.[57] Joel Lieberman et al. also looked at DNA and lab testing, "finding that jurors introduced to scientifically informed cross-examination were more likely to convict based on evidence tested in reliable labs than unreliable labs and reported less certainty in the defendant's guilt where evidence was tested in an unreliable lab."[58]

Contextual bias—where contextual factors, environmental factors, and attributes of a stimulus or situation influence the perception and interpretation of an object or event—has also been explored.[59] Concerns regarding contextual bias are widespread, as a context effect or bias can occur without an individual being aware of it; even well-trained experts are susceptible to this type bias.[60] The influence of contextual bias tends to be stronger when the data interpreted are ambiguous and weaker when the correct interpretation is more obvious.[61] In the context of jury decision-making following cross-examination, the risk of contextual bias is lower if the forensic evidence is built upon objective standards that produce standardized interpretable results.[62] However, the degree to which jurors understand standardized measures of forensic scientific testing is limited. William Thompson and Nicholas Scurich examined the use of blinding

[57] Jacqueline L. Austin & Margaret Bull Kovera, *Cross-Examination Educates Jurors About Missing Control Groups in Scientific Evidence*, 21 PSYCH. PUB. POL'Y. & L. 252, 261 (2015).
[58] Joel D. Lieberman, Courtney A. Carrell, Terance D. Miethe & Daniel A. Krauss, *Gold Versus Platinum: Do Jurors Recognize the Superiority and Limitations of DNA Evidence Compared to Other Types of Forensic Evidence?*, 14 PSYCH. PUB. POL'Y. & L. 27, 27 (2008).
[59] *See generally* Saul M. Kassin, Itiel E. Dror, & Jeff Kukucka, *The Forensic Confirmation Bias: Problems, Perspectives, and Proposed Solutions*, 2 J. APPLIED RSCH. MEMORY & COGNITION 42 (2013).
[60] *Id.* at 43.
[61] *See* Gary Edmond, Jason M. Tangen, Rachel A. Searston & Itiel E. Dror, *Contextual Bias and Cross-Contamination in the Forensic Sciences: The Corrosive Implications for Investigations, Plea Bargains, Trials and Appeals*, 14 LAW PROBABILITY & RISK 1, 11 (2015).
[62] *Id.* at 20–21.

procedures to prevent contextual bias in expert witness testimony, addressed during cross-examination, on juror perception of expert credibility.[63] Their mock jury perceived experts who used blinding procedures to prevent possible contextual bias as more credible than experts exposed to extraneous information.[64] Some forensic experts oppose the use of blinding procedures, however, "deny that being influenced by contextual information constitutes a bias."[65] This potentially allows forensic experts to be influenced by matters beyond their scientific expertise,[66] meaning experts might be able to testify on non-scientific matters, potentially exposing the jury to biased information. Such exposure to non-scientific information threatens the value of the expert's opinion and interpretation.[67] As a countermeasure, the now-disbanded U.S. National Commission on Forensic Science issued a statement urging forensic scientists to ensure that analysis is based solely upon task-relevant information.[68] Recent work has confirmed the efficacy of this statement, with a study by Thompson and Scurich suggesting that "jurors will view the examiners as less credible if the opposing lawyer can show through cross-examination either that the expert's interpretation relied on subjective judgment rather than objective standards or that the expert was exposed to potentially biasing task-irrelevant information."[69]

Evidently, the intersection of juror decision-making, expert evidence, and cross-examination is rich for exploration. Using firearms evidence as a context, the study by Cooper and

[63] William C. Thompson & Nicholas Scurich, *How Cross-Examination on Subjectivity and Bias Affects Jurors' Evaluations of Forensic Science Evidence*, 64 J. FORENSIC SCI. 1379, 1380 (2019).
[64] *Id.* at 1385–86.
[65] *Id.* at 1381 (citing to William C. Thompson, *What role should investigative facts play in the evaluation of scientific evidence?*, 43 AUSTL. J. FORENSIC SCI., 2-3, 123–134 (2011)).
[66] Thompson & Scurich, *supra* note 63 at 1381.
[67] *Id.*
[68] NAT'L. COMM'N ON FORENSIC SCI., ENSURING THAT FORENSIC ANALYSIS IS BASED UPON TASK-RELEVANT INFORMATION 1, https://www.justice.gov/archives/ncfs/page/file/641676/download [https://perma.cc/85CW-M64W] (last visited Sept. 20, 2021).
[69] Thompson & Scurich, *supra* note 63, at 1387.

Scanlon aimed to examine one particular aspect: the effect of cross-examination that questioned scientific rigor on juror certainty about association when presented with expert evidence in linguistic phrases. Part II outlines the study design and results.

II. STUDY DESIGN AND RESULTS[70]

Our study builds on the work described in Part I (particularly, the methods employed by McQuiston-Surrett and Saks and Cooper and Scanlon) by asking its U.S. participants to judge their certainty in expert statements that have had their strength challenged through cross-examination versus a control condition in which the statements were not. We predicted that

1. The phrases with more certain wording by the expert would be judged as more certainty-inducing by participants.

2. The group subjected to the cross-examination condition would score expert statements at lower certainty compared to the control group.

3. The cross-examination effect would be stronger for the more certain expert phrases.

A. METHOD

1. Design

We used a mixed quasi-experimental design. The within-subjects independent variable was the expert witness statements, which expressed varying levels of expert certainty in an ammunition-weapon association: very certain, certain (within the scope of the firearms examination field), certain (within the scope of the ballistics field), and uncertain. The between-

[70] All study materials and results data are on file with the authors.

subjects independent variable was cross-examination, on two levels, present or absent. The effects of this independent variable was measured through the dependent variable, the level of of participant certainty in an ammuition-weapon association, based on each expert statement, rated on a zero to one hundred scale.

2. Participants

The sample consisted of 437 members of the U.S. public[71] who volunteered to participate in the study online and were randomly assigned to one of two groups: a control group of 218 members and a group subjected to the cross-examination, consisting of 219 members. This was a national sample with participants from thirty-eight U.S. states, who were recruited through snowball sampling via professional networks, social networks, and online forums. No incentive was offered for taking part in the study. The eligibility criteria were all those that rendered participants eligible to sit on a federal jury in the Unted States: be a U.S. citizen; be at least 18 years of age; reside primarily in the judicial district for one year; be adequately proficient in English to satisfactorily complete the juror qualification form; have no disqualifying mental or physical condition; not currently be subject to felony charges punishable by imprisonment for more than one year; and never have been convicted of a felony. These criteria were presented in a screening questionnaire prior to testing, and all participants who took part responded to all exclusion questions.

[71] Mean age = 31.9, range = 19-85, 56m, 381f.

3. Materials and Procedure

Following the screening questionnaire, participants were asked to imagine that they were serving on a jury in a criminal trial. They were serially presented with four randomly ordered vignettes of a case involving a firearm owned by the defendant. They were told a qualified firearms examiner would testify for the state as to whether toolmarks produced on ammunition test-fired from the defendant's gun matched toolmarks present on suspect ammunition found at the crime scene. The expert's conclusions varied between the four otherwise identical vignettes for each of the two groups. Conclusions were based on those in Cooper and Scanlon's previous study, which used statements based on U.S. case law and were hypothesized to be either very certain (conclusion 1); certain (conclusions 2 and 3), and uncertain (conclusion 4).[72]

The conclusions used were as follows:

1. There is an exact match between the suspect ammunition and the Defendant's gun

2. A match can be made to a reasonable degree of certainty in the ballistics field

3. A match can be made to a reasonable degree of certainty in the firearms examination field

4. The suspect ammunition is unsuitable for comparison with ammunition test-fired from the Defendant's gun

The conclusion on each vignette was followed by a closing statement which differed between the two participant groups.

> For the Control Group: *The Judge says it is your role to determine the weight of this evidence, and you can give it as much or as little weight as you think it warrants.*

[72] Cooper & Scanlon, *supra* note 10, at 96, 108, 111, 113.

For the Experimental Group: *During cross-examination, the Firearms Examiner concedes that the scientific rigor of the methods they used to compare the tool-marks has been subject to significant criticism.*

On completion of each vignette, participants were asked to rate their certainty that the defendant's gun fired the suspect ammunition on a scale of zero to one hundred (least certain to most certain), based on the evidence presented.

B. RESULTS

1. Data Analysis

A 2(Cross examination) x 4(Expert statement) mixed analysis of variance ("ANOVA") was carried out to examine different effects of the between- and within-participant variables.[73] Bootstrapped Bonferroni-corrected *t*-tests were used to examine specific differences in the hypotheses—namely, that the more certainly-worded expert statements would elicit higher certainty in participants, and that cross-examination would have a negative effect on participant certainty across all expert statements. [74]

2. Descriptive Statistics

Table 1 and *Figure 1* provide the means and standard deviations of participant certanty for each of the conditions. For both the control and cross-examination groups, participants reported the highest level of certainty in an ammunition-weapon match based on the very certain conclusion, and reported similar levels of certainty based on both of the professionally certain

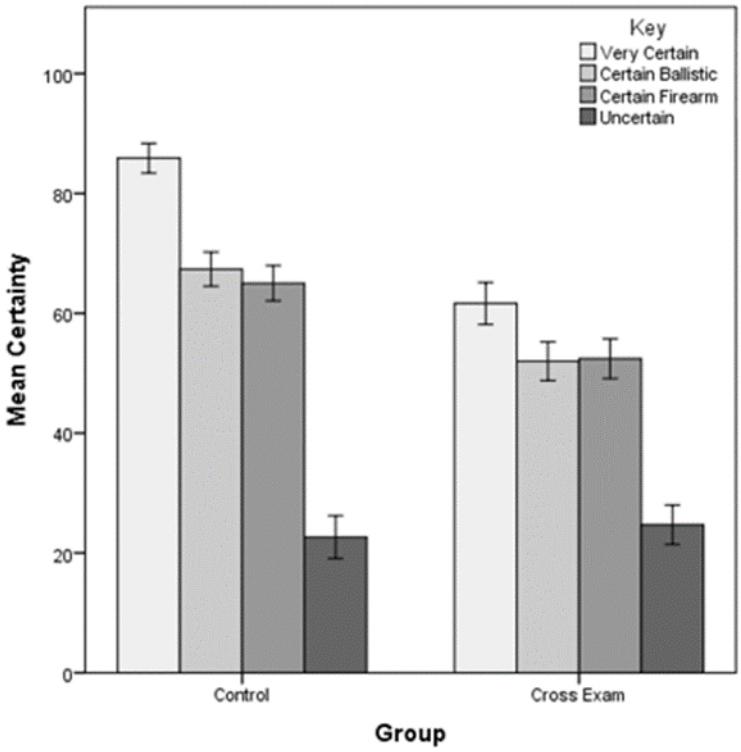
[73] This allows between- or within-group differences to be explained in terms of statistical significance – that the variation between statements can be generalised beyond this sample of participants.
[74] The t-tests examine each separate comparison individually, with the Bonferroni correction and bootstrapping in place, which require more difference between groups to be present for a significant finding to be reported, to reduce the likelihood of statistical error due to the large number of individual comparisons.

phrases judged similarly. The uncertain conclusion—that the evidence could not be compared—resulted in the lowest certainty scores. Overall, the control group reported higher certainty than the cross-examination group, apart from the uncertain conclusion, which reported similar means from both groups. Standard deviations were similar across conditions.

Table 1: Mean and standard deviations differences for each condition for participant certainty

Statement	**Control Means (SD)**	**Cross Exam Means (SD)**
Exact Match	85.88 (18.31)	61.63 (25.84)
Certain (Ballistic)	67.36 (21.05)	52.01 (24.01)
Certain (Firearms)	65.02 (21.7)	52.42 (24.46)
Unsuitable to compare	22.64 (26.34)	24.69 (24.11)

Figure 1: Mean and Standard Deviations differences for each condition for participant certainty



3. Inferential Statistics and Hypothesis Findings

The mixed ANOVA showed a main effect for group ($F[1,435]=61.43$, $p<0.001$, $\eta 2p=0.12$) and for statement ($F[2.13, 925.84]=559.07$, $p<0.001$, $\eta 2p=0.56$), and an interaction effect ($F[2.13, 925.84]=36.65$, $p<0.001$, $\eta 2p=0.08$). A Greenhouse-Geisser correction was implemented as the assumption of sphericity was violated.

The accompanying *t*-tests showed that within the control group, the Very Certain phrase elicited significantly more participant certainty than the Ballistic Certainty phrase ($t[222]=12.60$, $p<0.001$), the Firearms Certainty phrase ($t[219]=14.18$, $p<0.001$), and the Unsuitable for

Comparison phrase ($t[220]=27.81$, $p<0.001$). The Ballistic Certainty phrase ($t[220]=20.51$, $p<0.001$), and the Firearms Certainty phrase ($t[218]=19.23$, $p<0.001$) elicited significantly higher participant certainty than the Unsuitable for Comparison phrase but did not elicit significantly different levels of participant certainty between the three groups. This pattern is as predicted in Hypothesis 1, with qualitatively more certain phrases eliciting more certain responses.

The same pattern was evident within the cross-examination group. Again, the Very Certain phrase elicited significantly higher participant certainty than the Ballistic Certainty phrase ($t[222]=7.42$, $p<0.001$), the Firearms Certainty phrase ($t[221]=5.95$, $p<0.001$), and the Unsuitable for Comparison phrase ($t[219]=16.71$, $p<0.001$). The Ballistic Certainty phrase ($t[220]=13.70$, $p<0.001$) and the Firearms Certainty phrase ($t[220]=13.15$, $p<0.001$) elicited significantly higher participant certainty than the Unsuitable for Comparison phrase but did not elicit significantly different levels of participant certainty between the three groups. Again, this result is as predicted in Hypothesis 1, with the more certain phrases eliciting more certain responses.

Between groups, the Very Certain phrase elicited significantly higher participant certainty in the control group than in the cross-examination group ($t[435]=11.31$, $p<0.001$), as did both the Firearms Certainty phrase ($t[435]=5.69$, $p<0.001$) and the Ballistic Certainty phrase ($t[435]=7.10$, $p<0.001$). No significant difference was found between groups for the Unsuitable for Comparison phrases. This generally matches Hypothesis 2—that cross-examination will have a negative effect on participant certainty—though only for the Very Certain and Certain phrases.

The inferential findings mirror the descriptive statistics, showing that cross-examination had a significant detrimental effect on participants' certainty judgments for those cases in which the expert professed certainty, but not when the expert professed a lack of certainty. This generally

supports Hypothesis 3—that cross-examination would have a stronger effect on more certain phrases. A more certain expert also induced higher levels of certainty across participants regardless of cross-examination.

III. DISCUSSION

Overall, as described in Part II, the results followed our expectations that cross-examination can impact juror certainty, particularly when experts present their conclusions in terms of varying degrees of certainty. Two studies by Margaret Bull Kovera and colleagues have suggested that the impact of cross-examination on jury decision-making is negligible for forensic and scientific evidence.[75] One of these studies found that even a scientific-validity-specific cross-examination did not significantly affect jurors' ability to discern differences in construct validity, or differences in the degree to which experts' tests of the evidence measured what the tests purportedly measured, among different expert testimonies of accuracy in the original evidence.[76] More recently, Brandon Garrett et al. also found no significant effect for cross-examination on equivocal expert statements.[77] As in our study, however, others have noted that cross-examination can cause a reduction in juror certainty. Dawn McQuiston-Surrett and Michael Saks state that cross-examination can highlight the subjectivity of forensic testimony to jurors, thus leading to more conservative certainty judgments.[78] Jacqueline Austin and Kovera found that cross-examination using validity measures decreased certainty in jurors,[79] Joel Lieberman et al. showed

[75] *See generally* Margaret Bull Kovera, Robert J. Levy, Eugene Borgida & Steven D. Penrod, *Expert Testimony in Child Sexual Abuse Cases: Effects of Expert Evidence Type and Cross-Examination*, 18 LAW & HUM. BEHAV. 653 (1994); Kovera et al., *supra* note 55.
[76] *See generally* Kovera et al., *supra* note 75; Kovera et al., *supra* note 55.
[77] Garrett et al., *supra* note 38, at 420.
[78] McQuiston-Surrett & Saks, *supra* note 34, at 1170.
[79] Austin & Kovera, *supra* note 57, at 260.

strong evidence of the same result using DNA evidence,[80] and William Thompson and Nicholas Scurich's examination from the point of view of the expert admitting subjectivity under cross-examination found a similar pattern of results—admitting to methodological or reporting issues under cross-examination resulted in lower juror certainty.[81]

The decrease in juror certainty has been attributed to a number of causes. The study by McQuiston-Surrett and Saks suggests that decreases in certainty may be due to jurors feeling they better understand the issues around forensic testimony when presented with evidence of subjectivity, leading to conservatism in their own judgment, in an attempt to mitigate the effects of overzealous experts.[82] A study by Austin and Kovera provides some support for the theory that education can motivate a shift toward lower certainty scores, although they do not frame it in terms of conservatism.[83] Thompson and Scurich point out that reduced juror certainty is due to reduced credibility for the expert following damaging cross-examination.[84] Our design focuses on that credibility and shows the same pattern. However, both reasons are likely to contribute to the change in certainty. Pointing out methodological issues with a forensic science—be it DNA, odontology, microscopic hair examination, or firearms—through cross-examination increases jurors' focus on the scientific rigor needed for good forensic evidence, meaning that an expert who showed high certainty in testimony based on that now less powerful evidence may lose credibility.

The standardization of expert evidence is one potential way to assist jurors in their evaluation of forensic evidence. Noting that terms used to "describe findings, conclusions, and

[80] Lieberman et al., *supra* note 58, at 40.
[81] Thompson & Scurich, *supra* note 63, at 1385.
[82] McQuiston-Surrett & Saks, *supra* note 30, at 1170.
[83] Austin & Kovera, *supra* note 57, at 260.
[84] *See* Thompson & Scurich, *supra* note 63, at 1385.

degrees of association . . . can and do have a profound effect on how the trier of fact . . . perceives and evaluates scientific evidence,"[85] the National Academy of Sciences has recommended that "[t]he terminology used in reporting and testifying about the results of forensic science investigations must be standardized,"[86] Gary Edmond et al. suggests that "ambiguous information that is consistent with what is expected tends to be sharpened, whereas ambiguous information that is inconsistent with what is expected tends to be levelled."[87] However, if the information is unambiguous "it is almost impossible for sharpening or levelling to change the perception and interpretation."[88] The use of unequivocal and equivocal expert statements in our study supports this, with unequivocal statements losing the most credibility under even straightforward cross-examination. Edmond et al. also concluded that the risk of contextual bias is lower if the forensic evidence has objective standards that produce standardized, interpretable results.[89] This has been shown to be mitigated for numeracy; jurors who are numerate are not unbiased, but they tend to rely on issues around subjectivity of data rather than anecdotal information or the vividness of evidence descriptions when evaluating the possibility that forensic scientific evidence is fallible, and, in turn, rendering a verdict.[90]

Education and training for jurors has also been suggested as an answer. It has been noted that jurors' comprehension of forensic evidence is limited and not well studied; there is a need to better prepare jurors for their role.[91] The limitations of lawyers and judges can also exacerbate the

[85] STRENGTHENING, *supra* note 2, at 21.
[86] *Id.*
[87] Edmond et al., *supra* note 61, at 11.
[88] *Id.*
[89] *See id.* at 23.
[90] Nicholas Scurich, The Effect of Numeracy and Anecdotes on the Perceived Fallibility of Forensic Science, 22 PSYCHIATRY PSYCH. & L. 616, 621 (2015).
[91] STRENGTHENING, *supra* note 2, at 237.

challenges experienced by jurors, with some research suggesting that jurors' errors in interpreting evidentiary information can be traced to errors by lawyers and judges.[92] It has been proposed that lawyers, judges and jurors likely require higher levels of science literacy.[93] Despite this, "[t]he value of science literacy in societal systems such as the . . . justice system . . . as well as the opportunities that these systems provide to develop science literacy[] have not been studied in sufficient detail."[94] Specifically, it is "important to know what fields of science are most frequently referenced . . . and what level of understanding of scientific principles, methodologies, and habits of mind are needed for the proper and equitable operation of the justice system."[95]

CONCLUSION

Our study attempted to create a simple yet realistic mock case extract. However, as an online study, it obviously did not capture all the complexities of a real trial. This rich context is important for full understanding, but answering the fundamental research question required us to limit that context for increased experimental control.[96] A closer facsimile to trial settings may improve validity, perhaps through presenting information as a trial transcript, or even organizing a full mock trial. Further, the participants in our study may not represent a true community sample, despite the careful sampling method and sample size. Engagement with online information is significantly lower in older adults even now,[97] meaning that participatants will likely skew

[92] *Id.* at 236.
[93] NAT'L ACADS. OF SCI., ENGINEERING & MEDICINE, SCIENCE LITERACY: CONCEPTS, CONTEXTS, AND CONSEQUENCES 110–11 (2016).
[94] *Id.* at 110.
[95] *Id.* at 111.
[96] *See generally* Herman Aguinis & Kyle J. Bradley, *Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies*, 17 ORG. RSCH. METHODS 351 (2014).
[97] Eszter Hargittai, Anne Marie Piper, & Meredith Ringel Morris, *From Internet Access to Internet Skills: Digital Inequality Among Older Adults*, 18 UNIVERSAL ACCESS INFO. SOC'Y 881, 882 (2019).

younger, as seen by our mean age despite the age range. This may not be reflective of mean jury ages in the United States, and it has been found that juror age can affect decision-making.[98]

In the context of our study, developing further understanding of how jurors interpret expert firearms evidence could take multiple routes. Further investigation into the influence of cross-examination would be valuable and could be conducted by changing the style, length, and content of lawyers' questioning in relation to actual scientific content of the expert evidence. The theories that cross-examination might make defense lawyers seem more knowledgeable, or make experts appear defensive or weak are also deserving of more focus.[99] Further research on the sequence of interactions between lawyers and witnesses would also be valuable. A natural next step in the sequence presented in this study, for instance, following the cross-examination of an expert, would be for an opposing lawyer to "rehabilitate" a witness by asking them to reconfirm the validity and/or reliability of their initial testimony after a defense lawyer has cross-examined the expert. Designing methods for examining these areas of interest is key to achieving a deeper understanding of how jurors interpret the weight of forensic expert testimony within the adversarial system.

[98] *See* Christine L. Ruva & Elizabeth M. Hudak, *Pretrial Publicity and Juror Age Affect Mock-Juror Decision Making*, 19 PSYCH. CRIME & L. 179, 194–95 (2013).
[99] Thompson & Scurich, *supra* note 63, at 1385.