ACCEPTED MANUSCRIPT *forthcoming in the Southern California Interdisciplinary Law Journal Vol. 31*
Not to be cited without authors' permission.

Juror Certainty about Expert Firearms Identification Evidence and the Impact of Cross-examination

Paraic Scanlon, Boglarka Banyai, Ellis Hart, Sarah L. Cooper[1]

Abstract

To evaluate association, firearms examiners compare tool-marks present on suspect ammunition to those present on ammunition test-fired by a suspect weapon. Examiners' conclusions are generally admissible in US courts, yet the scientific underpinnings of the discipline have been subject to considerable criticism. Cross-examination can be used to bring such criticism to the attention of jurors, who determine the weight of expert evidence. The authors investigated the effect of such cross-examination on juror certainty about expert firearms evidence using online vignettes. A community sample of US participants (n=437) were asked to rate their certainty (0-100) of a forensic match for each of 4 expert statements of certainty, in 2 groups; either with or without a cross-examination highlighting limitations of the tool-mark discipline's scientific underpinnings. Analysis was undertaken both between 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.

INTRODUCTION

To evaluate association, examiners compare tool-marks 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 make one of four conclusions: (1) identification; (2) inconclusive; (3) elimination; or (4) unsuitable for comparison.[3] Examiner conclusions have been routinely admitted into US courts as expert evidence for around a century.[4] However, in recent decades, concerns about reliability have been repeatedly raised, with reports by the National Research Council[5]

[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.

[2] NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 152-53 (2009) [hereinafter *Strengthening*].

[3] Ass'n of Firearms & Tool Mark Exam'rs, *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. 2011). ("The technique subsequently [after the 1920's] gained widespread judicial acceptance and was not seriously challenged until recently.")

[5] STRENGTHENING, *supra* note 2, at 27; and NAT'L RESEARCH COUNCIL, BALLISTIC IMAGING (Daniel L. Cork et al. eds., 2008) [hereinafter *Ballistics Imaging*].

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, including the FBI, AFTE, and Department of Justice,[10] yet a shared vision for the continual enhancement of forensic science generally does.[11]

Some US courts have responded to the debate by restricting the degree of certainty that experts may express in their conclusions. For example by requiring that phrases like "there is an exact match" be replaced with phrases such as a match can be made "more likely than not" or "to a reasonable degree of certainty."[12] Jurors must determine the weight of these phrases, a task that can comprise various challenges.[13] Noting this, Cooper and Scanlon, using online vignettes, investigated juror certainty of 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. They suggested, *inter alia*, 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 into 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

---

[6] PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY, FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE- COMPARISON METHODS (2016) [hereinafter *PCAST Report*].

[7] *Ballistics Imaging*, *supra* note 5, at 3; United States v. Taylor, 663 F. Supp. 2d 1170, 1176 (D.N.M. 2009) (discussing the focus and scope of the Ballistic Imaging Report).

[8] STRENGTHENING, *supra* note 2, at 154.

[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.

[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 101.

[12] *Id.* at 102-04 (summarising numerous court responses between 2005 and 2012). Notably, more recent case law reflects the ongoing use of/reference to/debate around these phrases. See, for example, People v. Ross, 68 Misc. 3d 899, 914, 129 N.Y.S.3d 629, 640 (N.Y. 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."); United States v. Harris, No. CR 19-358 (RC), 2020 WL 6488714, at *11 (D.D.C. Nov. 4, 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).").

[13] *See infra* Part II.

[14] *See generally* Cooper & Scanlon, *supra* note 10.

in the context of some current understandings about 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."

PART I: JUROR DECISION-MAKING, EXPERT EVIDENCE & CROSS-EXAMINATION

To position the authors' study, this section broadly summarises 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. Robert and Zukerman have explained that jurors can be confused by complex evidence and prefer to base their decisions on the expert's conclusions.[15] Blackwell and Seymour concluded that, in the context of expert witnesses, jurors rank relevant professional experience, lack of bias, and clarity of evidence in order of importance.[16] Studies show jurors can be challenged in recognising bias, however. For instance, McAuliff and 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, using opposing experts in a mock trial, Levett and Kovera attempted to sensitise participants to flaws and bias in forensic evidence through examining both experts' credibility, trustworthiness, research quality and verdicts.[18] Results indicated that jurors were not sensitive to flaws and bias, but rather were made skeptical of the expert and their trustworthiness.[19]

The content and expression of expert forensic evidence has been explored. For instance, Thompson and 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] Koehler and Ritchie[21] removed much of the context while examining expert statements of numerical certainty about DNA evidence, and found exclusion percentages[22] are more likely to result in conviction than if an expert were to testify in terms of frequency ratios.[23] Case law suggests, however, that forensic experts (including

---

[15] PAUL ROBERT & ADRIAN ZUCKERMAN, CRIMINAL EVIDENCE (2d ed. 2010).
[16] Suzanne Blackwell & Fred Seymour, *Expert Evidence and Jurors' Views on Expert Witnesses*, 22 PSYCHIATRY PSYCHOL. & L. 673 (2015).
[17] Bradley D. McAuliff & Tejah D. Duckworth, *I Spy with My Little Eye: Jurors' Detection of Internal Validity Threats in Expert Evidence*, 34 L. & HUM. BEHAV. 489 (2010).
[18] Lora M. Levett & Margaret Bull Kovera, *The Effectiveness of Opposing Expert Witnesses for Educating Jurors about Unreliable Expert Evidence*, 32 L. & HUM. BEHAV. 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 L. & HUM. BEHAV. 332 (2015).
[21] *See generally* Johnathan 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); *See generally* J. Ritchie, *Probabilistic DNA Evidence: The Layperson's Interpretation*, 47 AUSTL. J. FORENSIC SCI. 440 (2015).
[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."

firearms examiners[24]) routinely testify using linguistic or ordinal category-based evidence i.e., evidence that suggests a fixed hierarchy, but the numerical difference between the categories is not fixed.[25] Any verbal certainty judgements, made without explicit statistical information, are by necessity a balance between the meaning of a phrase provided by the expert and the interpreter's own subjective understanding of the evidence.[26] Martire, Kemp, Watkins, Sayle, & Newell found that inculpatory evidence was significantly more likely to be judged as weak if it was presented in terms of linguistic descriptions than in terms of statistical or numerical likelihood.[27] The Association of Forensic Science Providers' have released guidelines suggesting forensic experts use likelihood ratios,[28] yet uptake "uptake of these standards has varied considerably across jurisdictions and disciplines."[29]

Studies have specifically explored juror interpretation of certainty when presented with expert linguistic phrases. McQuiston-Surrett and Saks asked participants to rate an odontology expert's intended certainty (on a scale of 1-100) from four 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] "A 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 "c*onsistent with*" (75) another statement assumed to be uncertain, and significantly above both more assumedly certain phrases – "*Probable*" (57) and "*Reasonable Scientific Certainty*" (70).[32] The researchers concluded "These 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] Recent firearms-based research corroborates this to some extent, finding that juror assessment of the weight of expert evidence does not always follow researcher expectations, again using case law-based phrases including *"Ballistic Certainty"* (67.9), *"More likely than not"* (69.5) and *"Complete agreement"* (65.5).[34] McQuiston-Surrett and Saks also examined certainty statements in the context of microscopic hair analysis.[35] 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*

[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 (1986).
[27] Kristy A. Martire et al., *The Expression and Interpretation of Uncertain Forensic Science Evidence: Verbal Equivalence, Evidence Strength, and the Weak Evidence E*ffect, 37 L. & HUM. BEHAV. 197, 200 (2013).
[28] *Id.* at 197; Association of Forensic Science Providers, *Standards for the Formulation of Evaluative Forensic Science Expert Opinion*, 49 SCI. & JUST. 161 (2009).
[29] Martire et al, *supra* note 27, at 197.
[30] Dawn McQuiston-Surrett, & Michael J. Saks, *Communicating Opinion Evidence in the Forensic Identification Sciences: Accuracy and Impact*, 59 HASTINGS L. J. 1159 (2008).
[31] *Id.* at 1162-63.
[32] *Id.*
[33] *Id.* at 1163.
[34] Brandon L. Garrett, Nicholas Scurich & William E. Crozier, *Mock Jurors' Evaluation of Firearm Examiner Testimony*, 44 L. & HUM. BEHAV. 412, 417 (2020).
[35] Dawn McQuiston-Surrett, & Michael J. Saks, *The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Factfinders Hear*, 33 L. & HUM. BEHAV. 436 (2009).

multiple- frequency."[36] The researchers found that the qualitative statements were deemed significantly more certain than the subjective probability or objective multiple-frequency statements.[37] Non-judicial participants were particularly susceptible to this effect. [38]

These findings about varying interpretations fit with general concerns about the challenges that can emerge when legal agents 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 recognised these challenges too.[42] Equally, most jurors lack scientific expertise.[43] And, although research has shown consistency between jury and bench trial verdicts regardless of the scientific complexity involved, and generally justified outcomes in cases where jurors have expressed incomplete or flawed understanding of scientific or technical evidence,[44] it is known jurors can have difficulty engaging with scientific and technical evidence,[45] and a better understanding of juror comprehension of forensic evidence is needed.[46] Based on their work, McQuiston-Surrett & Saks have concluded better understanding can be developed through "empirical testing of the responses to the words"[47] and 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 expert phrases.[49] Participants were asked to rate their certainty (on a scale of 0 - 100) of a match between a defendant's firearm and suspect ammunition based on various expert statements, hypothesised to attract high (e.g., "*exact match*"), moderate (e.g. "*match to a reasonable degree of professional certainty*") or low-certainty (e.g "*inconclusive*").[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

[36] *Id.* at 437-38.
[37] *Id.* at 444.
[38] *Id.* at 445.
[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, 255-56 (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 and Company 1999).
[44] STRENGTHENING, *supra* note 2, at 236-37.
[45] *See* Valeria P. Hans, *Judges, Juries, and Scientific Evidence*, 16 J.L. & POL'Y 19, 23 (2007) (stating that jurors can have difficulty understanding scientific and technical evidence, particularly DNA evidence).
[46] STRENGTHENING, *supra* note 2, at 237.
[47] McQuiston-Surrett & Saks, *supra* note 30, at 1163.
[48] *Id.*
[49] Cooper & Scanlon, *supra* note 10, at 109.
[50] *Id.*

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"[52] would influence juror certainty.

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 they 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 characterising practice of adversarial justice systems, like that operated in the US. In fact, the US justice system relies on cross-examination to resolve fallibilities associated with expert evidence, as underscored by the US Supreme Court in *Daubert*, 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 readily undo the adverse effects of misleading expert testimony."[54]

The effect of cross-examination in the general context of this study has been explored. Kovera, McAuliff, and Hebert found cross-examination failed to highlight issues with validity of scientific evidence.[55] Garrett, Scurich and Crozier found comparable evidence using a cross-exam transcript. 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 cases in which jurors were able to distinguish between validity in scientific measures. Austin and Kovera found, for example, that jurors informed about methodological flaws or validity standards in DNA evidence through cross-examination were able to distinguish between 'quality' of evidence.[57] Lieberman *et al* also looked at DNA and lab testing, finding that the introduction of cross-examination lowered the participants' certainty scores, particularly when the scientific method was the focus.[58]

Contextual bias - where contextual factors, environmental factors, attributes of a stimulus or situation influence perception and interpretation of an object or event — has been explored. [59] Concerns regarding contextual bias are widespread, as a context effect

[51] *Id.* at 109-14.
[52] *Id.* at 117.
[53] Daubert v. Merrell Dow Pharm. Inc., 509 U.S. 579, 596 (1993).
[54] McQuiston-Surrett & Saks, *supra* note 30, at 1188; McQuiston-Surrett & Saks, *supra* note 35, at 451.
[55] Margaret Bull Kovera, Bradley D. McMuliff & 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 PSYCHOL. 362, 372 (1999).
[56] Garrett et al, *supra* note 34, at 420-21.
[57] Jacqueline Austin & Margaret Bull Kovera, *Cross-examination educates jurors about missing control groups in scientific evidence*, 21 PSYCHOL. PUB. POL'Y. L. 252 (2015).
[58] Joel D. Lieberman et al., *Gold versus platinum: Do jurors recognize the superiority and limitations of DNA evidence compared to other types of forensic evidence?*, 14 PSYCHOL. PUB. POL'Y. & L. 27 (2008).
[59] Saul M. Kassin, Itiel E. Dror & Jeff Kukucka, *The Forensic Confirmation Bias: Problems, Perspectives, and Proposed Solutions*, 2 J. APPLIED RES. MEMORY & COGNITION 42 (2013).

or bias can occur without an individual being aware of it, and even well-trained experts are susceptible to bias.[60] The influence of context bias tends to be stronger when the data interpreted are ambiguous and weaker than when the correct interpretation is more obvious.[61] In the context of jury decision-making following cross-examination, contextual bias is lower if the forensic evidence has objective standards that produce standardized interpretable results.[62] However, the degree to which jurors understand standardized measures of forensic scientific testing is limited. Thompson & Scurich examined the use of blinding procedures to remove contextual bias within expert witness testimony during cross-examination.[63] Their mock jury perceived blind procedures that removed possible contextual bias as more 'valid.'[64] Some forensic expert witnesses, however, resist blind procedures, believing that evidence without context is less powerful.[65] This potentially allows forensic experts to be influenced by matters beyond their scientific expertise,[66] meaning they could testify to non-scientific matters, which potentially exposes the jury to biased information.

Forensic experts are routinely exposed to task-irrelevant information to their processing and interpretation of evidence.[67] This exposure to non-expert information threatens the value of the expert's opinion and interpretation of evidence.[68] As a countermeasure, the now-disbanded US National Commission on Forensic Science issued a statement urging forensic scientists to ensure that analysis was based solely upon task-relevant information.[69] Recent work has confirmed the efficacy of these findings, 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."[70]

Evidently, the intersection of juror decision-making, expert evidence and cross-examination is rich for exploration. Using firearms evidence as a context, the authors' study aimed to examine one particular aspect: the effect of cross-examination questioning scientific rigour on juror certainty about association when presented with expert evidence in linguistic phrases. Part II outlines our study design and results.

PART II: STUDY DESIGN AND RESULTS[71]

[60] *Id.* at 43.
[61] Gary Edmond et al., *Contextual bias and cross-contamination in the forensic sciences: the corrosive implications for investigations, plea bargains, trials and appeals*, 14 L. PROBABILITY & RISK 1 (2015).
[62] *Id.* at 20-21.
[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 (2019).
[64] *Id.* at 1385.
[65] Nicholas Scurich, *The Effect of Numeracy and Anecdotes on the Perceived Fallibility of Forensic Science*, 22 PSYCHIATRY PSYCHOL. & L. 616 (2015).
[66] *Id.* at 620.
[67] McAuliff, *supra* note 17.
[68] *Id.* at 499.
[69] NAT'L. COMMISSION ON FORENSIC SCI., *ENSURING THAT FORENSIC ANALYSIS IS BASED UPON TASK-RELEVANT INFORMATION*, https://www.justice.gov/archives/ncfs/page/file/641676/download.
[70] William C. Thompson & Nicholas Scurich, *How Cross-Examination on Subjectivity and Bias Affects Jurors' Evaluations of Forensic Science Evidence*, 64 J. FORENSIC SCI. 1379 (2019).
[71] All study materials and results data are on file with the authors.

Our study builds on the work described in Part I (particularly methods employed by McQuiston-Surrett and Saks and Cooper and Scanlon), by asking US participants to judge the certainty of expert statements that have had their strength challenged through cross-examination vs a control condition. 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.*

Method

Design

We used a mixed quasi-experimental design. The within-subjects independent variable was the expert witness statement, which had four levels: (1) very certain, (2) certain [firearms], (3) certain [ballistics], (4) uncertain. The between-subjects independent variable was cross-examination, on two levels (present or absent). These were measured through a dependent variable of participant certainty judgments based on the expert statement on a 0-100 scale.

Participants

The sample consisted of n=437 members of the US public (mean age = 31.9, range = 19-85, 56m, 381f) who volunteered to participate in the study online and were randomly assigned to one of two groups (control, n=218; cross-examination, n=219). This was a national sample with participants from 38 US states included, recruited through snowball sampling via professional networks, social networks and online forums. No incentive was offered for taking part. The eligibility criteria were all those that rendered participants eligible to sit on a federal jury in the United States, as follows: Be a United States 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; 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.

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 and randomly presented with four vignettes of a case involving a firearm owned by the defendant. They were told a qualified Firearms Examiner testifies for the state as to whether tool-marks produced on ammunition test-fired from the Defendant's gun match tool-marks present on suspect ammunition found at the crime scene. The expert's conclusions varied between the otherwise identical vignettes for each of the two groups. Conclusions were based on Cooper and Scanlon's previous study that used statements based on US case law and were

hypothesised to be very certain (conclusion 1); certain (conclusions 2 and 3), and uncertain (conclusion 4).[72]

These conclusions were:

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

This 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.*

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 0 (Least Certain) to 100 (Most Certain) based on the evidence presented.

Results

Data Analysis

A 2(Cross examination)x4(Expert statement) mixed ANOVA was carried out to examine the between- and within-participant effects. Bootstrapped Bonferroni-corrected t-tests were used to examine specific differences in the hypotheses; 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 expert statements.

Descriptive Statistics

Table 1 and *Figure 1* provide the means and Standard Deviations for each of the conditions. For both groups, participants judged the more certain expert phrases highest, with both of the professionally certain phrases judged similarly. The expert's judgment that the evidence could not be compared resulted in the lowest scores. Overall, the control group scored higher certainty than the cross-examination group, apart from for the unsuitable for comparison phrase, where means were similar. Standard Deviations were similar across conditions.

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

[72] Cooper & Scanlon, *supra* note 10.

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)

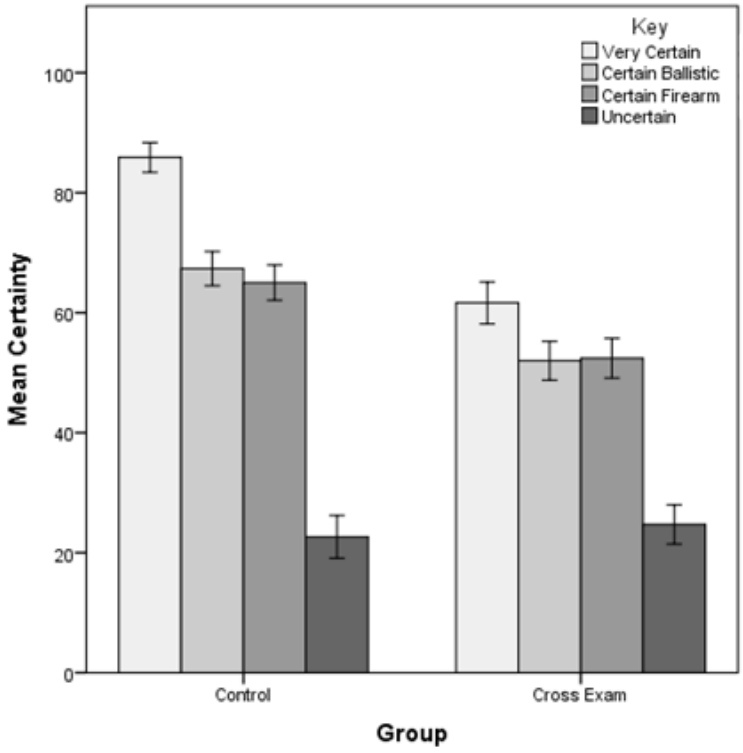

Figure 1: Mean and Standard Deviations differences for each condition for participant certainty

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 scored significantly more 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 ($t[218]=19.23$, $p<0.001$) were significantly higher than the Unsuitable for Comparison phrase, but not

significantly different than each other. 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 was judged significantly more certain 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 ($t[220]=13.15$, $p<0.001$) scored as significantly more certain than the Unsuitable for Comparison phrase, but were not significantly different from each other. Again, this is as predicted in Hypothesis 1, with the more certain phrases eliciting more certain responses.

Between groups, the Very Certain phrase scored significantly higher for the control group over the cross exam 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 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 the participant's certainty judgments for those cases in which the expert professed expertise, but not when the expert professed a lack of analysis. This generally follows 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.

PART III: DISCUSSION

Overall, as described in Part II, the results followed our expectations i.e., that cross-examination can impact juror certainty, particularly when experts present in terms of certainty. Two studies by Kovera and colleagues have suggested the impact of cross-examination on jury decision-making is negligible for forensic and scientific evidence.[73] They found that even a scientific validity-specific cross-examination did not significantly affect jurors' ratings of accuracy in the original evidence.[74] More recently, Garrett *et al* also found no significant effect for cross-examination on equivocal expert statements.[75] Others, however, have found, like the authors, that cross-examination can cause a reduction in certainty. McQuiston-Surrett and Saks found that cross-examination can highlight to jurors the subjectivity of forensic testimony - thus leading to more conservative certainty judgments.[76] Austin and Kovera found that cross-examination using validity measures decreased certainty in jurors,[77] Lieberman *et al* showed strong evidence of the same result using DNA evidence,[78] and Thompson and Scurich's examination from the point of view of the expert admitting subjectivity under cross-

[73] Margaret Bull Kovera et al., *Expert Testimony in Child Sexual Abuse Cases: Effects of Expert Evidence Type and Cross-Examination*, 18 L. & HUM. BEHAV. 653 (1994); Kovera, *supra* note 55.
[74] *Id.*
[75] Garrett et al, *supra* note 34, at 420.
[76] McQuiston-Surrett & Saks, *supra* note 30, at 1170.
[77] Austin & Kovera, *supra* note 57, at 260.
[78] Lieberman et al, *supra* note 58, at 40.

examination found a similar pattern of results – admitting to methodological or reporting issues under cross-examination negatively effects juror certainty.[79]

The decrease in certainty has been attributed to a number of causes. McQuiston-Surrett and Saks suggest it may be due to jurors feeling they better understand the issues around forensic testimony if they are presented with evidence of subjectivity, leading to conservatism in their own judgment, in an attempt to mitigate overzealous experts.[80] Austin and Kovera also theorize education motivates this shift toward lower certainty scores, although they do not frame it in terms of conservatism.[81] Thompson and Scurich broadly agree, but point out that reduced juror certainty may be due to reduced credibility for the expert following damaging cross-examination.[82] Our design focuses on that credibility, and shows the same pattern. However, both reasons are likely to contribute to the change in certainty. By pointing out methodological issues with a forensic science - be it DNA, odontology, microscopic hair examination, or firearms – jurors' focus on the scientific rigour needed for good forensic evidence seems to be increased through this cross-examination, meaning that an expert that showed high certainty in testimony based on that now less-powerful evidence will lose credibility.

The standardisation of expert evidence is one potential way to support jurors in their evaluation of forensic evidence. Noting that terms used to "describe findings, conclusions, and degrees of association… can and do[es] have a profound effect on how the trier of fact …. perceives and evaluates scientific evidence…."[83], the National Academy of Sciences has recommended that "The terminology used in reporting and testifying about the results of forensic science investigations must be standardized…"[84] Edmond *et al* reported that the influence of subjective evidence tends to be stronger when the data presented is ambiguous and weaker when the correct interpretation is more obvious.[85] The use of unequivocal and equivocal expert statements in our study supports this, with those unequivocal statements being most effected in terms of losing credibility under even straightforward cross-examination. Edmond *et al* concluded that contextual bias is lower if the forensic evidence has objective standards that produce standardized interpretable results.[86] This has been shown to be mitigated for numeracy – those jurors who are numerate are not unbiased, but tend to rely on issues around subjectivity of data rather than anecdote or vividness of evidence description when evaluating the possibility that forensic scientific evidence is fallible.[87]

Education and training for jurors has also been suggested as an answer for this issue of understanding. It has been noted that jurors' comprehension of forensic evidence is limited, not well studied, and there is a need to better prepare jurors for their role.[88] The limitations of lawyers and judges can also exacerbate the challenges experienced by jurors, with some research suggesting that jurors' errors in interpreting evidentiary

[79] Thompson, *supra* note 63.
[80] McQuiston-Surrett & Saks, *supra* note 30, at 1170.
[81] Austin & Kovera, *supra* note 57, at 260.
[82] Thompson & Scurich, *supra* note 79, at 1385.
[83] STRENGTHENING, *supra* note 2, at 21.
[84] *Id.*
[85] Edmond et al, *supra* note 61, at 20-21.
[86] *Id.* at 23.
[87] Scurich, *supra* note 65, at 620.
[88] STRENGTHENING, *supra* note 2, at 237.

information can be traced to errors by lawyers and judges.[89] It has been proposed that lawyers, judges and jurors likely require higher levels of science literacy.[90] Despite this, "The 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."[91] 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."[92]

CONCLUSION

Our study attempted to create a simple yet realistic mock case extract. However, as an online study with streamlined context it obviously did not capture all the complexities of a real trial. This rich context is important to full understanding, but answering the fundamental question required us to limit that context for increased experimental control.[93] A closer facsimile to trial settings may improve validity, perhaps through presenting information as a trial transcript, or even 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. Access to online information is significantly lower in older adults even now,[94] meaning a skew toward younger participants is likely, as seen by our mean age despite the age range. This may not be reflective of mean jury ages in the US, and it has been found that juror age can affect decision-making.[95]

In the context of our study, developing further understanding of how jurors interpret expert firearms evidence could take multiple routes. Further investigation of the influence of cross-examination would be valuable, for example, on the style, length, and content of lawyers' questioning in relation to actual scientific content of the expert evidence. The theory that cross-examination might have made the defense lawyer seem more knowledgeable, or made the expert appear defensive or weak is also deserving of more focus.[96] Further research on the sequence of interaction between lawyer and witness is also valuable. A natural next step in the sequence presented in this study, for instance, following the cross-examination of the expert, would be for the opposing lawyer to 'rehabilitate' the witness by asking them to reconfirm the validity and/or reliability of their initial testimony. Designing methods for examining these areas of interest are key to a deeper understanding.

---

[89] *Id.* at 236.
[90] NATIONAL ACADEMIES OF SCIENCES, ENGINEERING AND MEDICINE, SCIENCE LITERACY: CONCEPTS, CONTEXTS, AND CONSEQUENCES 110-11 (2016).
[91] *Id.* at 110.
[92] *Id.* at 111.
[93] Herman Anguinis & Kyle J. Bradley, *Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies*, 17 ORG. RES. METHODS 351 (2014).
[94] Eszter Hargittai et al., *From Internet Access to Internet Skills: Digital Inequality Among Older Adults*, 18 UNIVERSAL ACCESS INFO. SOC'Y 881 (2019).
[95] Christine L. Ruva & Elizabeth M. Hudak, *Pretrial Publicity and Juror Age Affect Mock-Juror Decision Making*, 19 PSYCHOL. CRIME & L. 179 (2013).
[96] Thompson & Scurich, *supra* note 82, at 1385.