JUROR ASSESSMENT OF CERTAINTY ABOUT FIREARMS IDENTIFICATION EVIDENCE

Dr. Sarah L. Cooper & Dr. Páraic Scanlon***

I. INTRODUCTION

Many firearms examiners believe they can reliably engage in individualization, i.e., conclude that a particular gun fired a particular bullet to the "exclusion of all other[s] . . . ."[1] Over the last century, to aid the determination of criminal liability, such conclusions have been routinely admitted into American courtrooms as expert evidence.[2]

However, criticism about the ability of crime-solving forensic disciplines, including firearms/tool-mark identification, to engage in reliable individualization has grown throughout the 1990s and new millennium.[3] This is largely due to the increasing sophistication and ability of DNA technology to identify both criminals and persons wrongly convicted.[4] In particular, the National Academy of Sciences ("NAS") in its landmark 2009 report – *Strengthening Forensic Science in the United States: A Path Forward* ("*Strengthening*") - concluded that "no forensic method [including firearms identification] has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source"[5] except for DNA analysis.[6]

---

* Reader in Law; Director, Centre for Law, Science and Policy at Birmingham City University, School of Law, UK.
** Lecturer in Developmental Psychology, Birmingham City University, School of Social Sciences, Department of Psychology, UK.
The authors would like to thank undergraduate psychology student, Lily Edwards, for her excellent research assistance.

1. Adina Schwartz, *Challenging Firearms and Toolmark Identification—Part One*, 32 CHAMPION 14, Oct. 2008.
2. Paul C. Giannelli, Edward J. Imwinkelried, & Joseph L. Peterson, *Reference Guide on Forensic Identification Expertise*, *in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 55, 91 (3rd ed. 2011). ("The technique subsequently [after the 1920's] gained widespread judicial acceptance and was not seriously challenged until recently.").
3. Eadaoin O'Brien et al., *Science in the Court: Pitfalls, Challenges and Solutions,* 370 Philosophical Transactions of the Royal Society B: Biological Sciences 1674, 1 (2015).
4. COMM. ON IDENTIFYING THE NEEDS OF THE FORENSIC SCI. CMTY., NAT'L RESEARCH COUNCIL OF THE NAT'L ACADEMIES, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 8 (2009) [hereinafter STRENGTHENING].
5. *Id.* at 7.
6. *Id.*

Although challenged by some stakeholders,[7] concerns raised in *Strengthening* have been respectively echoed and followed up by the President's Council of Advisors on Science and Technology[8] ("PCAST") and the National Commission on Forensic Science ("NCFS").[9]

Some American courts have acknowledged the existence of scientific uncertainty in the context of firearms individualization evidence since 2005.[10] To address this uncertainty, these courts have restricted expert testimony; instructing experts to not testify in absolute terms, such as "there is an exact match," but rather in allegedly more diluted terms, such as a match can be made "more likely than not" and "to a reasonable degree of certainty."[11] Jurors must determine the weight of such evidence in criminal trials.[12] Research indicates various challenges can arise when jurors assess scientific evidence, including that jurors often lack scientific expertise; find comfort in alleged expert certainty; and can be confused by phrases such as "to a reasonable degree of certainty."[13]

It is therefore important to further investigate the levels of certainty jurors attach to common expert phrases. To do this, the authors conducted a study examining the effect of twelve expert statements of certainty on potential jurors.[14] Results from a sample of 107 participants found a significant main effect for certainty, suggesting that participant certainty was influenced by expert testimony.[15] There was a general trend of significant findings, with increased expert certainty leading to increased participant certainty, with some notable exceptions.[16]

---

7. *See infra* Part I(B).

8. *See generally* 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].

9. *See generally National Commission on Forensic Science*, U.S. DEP'T OF JUST., http://www.justice.gov/ncfs (last visited May 17, 2017). Note the NCFS' charter expired on April 23, 2017.

10. *See, e.g., United States v. Green,* 405 F. Supp. 2d 104 (D. Mass. 2005); *United States v. Glynn,* 578 F. Supp. 2d 567 (S.D.N.Y. 2008); *United States v. Monteiro,* 407 F. Supp. 2d. 351 (2006).

11. *See infra* Part I(C).

12. Sarah Lucy Cooper, *Forensic Identification Evidence: Tensions Between Law and Science*, 16 J. PHIL. SCI. & L. 1, 21 (2016), http://jpsl.org/files/7814/6014/5245/ForensicScienceIdentification.pdf. ("Jurors play a pivotal role in assessing the reliability and weight of scientific evidence.").

13. *Id.* at 13–14. *See also* Bonnie Lanigan, *Firearms Identification: The Need for a Critical Approach to, and Possible Guidelines for, the Admissibility of "Ballistics" Evidence*, 17 SUFFOLK J. TRIAL & APP. ADVOC. 54, 71 (2012).

14. *See infra* Appendix A for t-tests data.

15. *Id.*

16. *Id.*

This article contextualizes and presents the method and findings of our study in order to contribute to multi-disciplinary efforts nationwide to generate more knowledge about juror interpretation of expert forensic science evidence. Part II provides an overview of the process of firearms identification; debates concerning its validity; and court approaches to addressing scientific uncertainty in the discipline.[17] Part III explores current literature concerning juror interpretation of expert evidence.[18] Part IV discusses the methods and findings of the authors' study; comments on how the study's findings might inform current judicial approaches; and outlines the authors' recommendations for furthering this research topic.[19]

II. CONTEXTUALIZING THE AUTHORS' STUDY

This section sets out the process of firearms identification, recent debates about the validity of the discipline, and court approaches to legal challenges to the admissibility of firearms identification evidence.

A. The Process of Firearms Identification

When the hard metal of an internal part of a firearm connects with the softer metal of the ammunition, it makes a tool-mark on the ammunition.[20] Two distinct types of tool-marks may be created: striations and impressions.[21] Striations are similar to small scratches, and are most often produced on the bullet as it passes through the gun barrel.[22] Impressions usually resemble dimples or craters, and are typically produced on the cartridge as it comes into contact with the various internal parts of the firing chamber (e.g. the firing pin, breach face, extractor, and ejector).[23] Tool-marks can be divided into class, subclass, and individual characteristics.[24] Class characteristics result from design factors and are determined prior to manufacture,[25] which means they are "distinctively designed features" and will be present on every tool in that class.[26] By contrast, individual characteristics are unique to a particular tool and consist of purportedly random, microscopic imperfections and irregularities present on the tool's surface.[27] Subclass

---

17. *See infra* Part II.
18. *See infra* Part III.
19. *See infra* Part IV.
20. Schwartz, *supra* note 1, at 10, 12.
21. *Id.* at 12.
22. *Id.*
23. *Id.*
24. *Id.*
25. *Id.*
26. Schwartz, *supra* note 1, at 12.
27. *Id*.

characteristics straddle the line between class and individual characteristics.[28] Subclass characteristics arise when manufacturing processes create batches of tools that are similar to each other but distinct from other tools of the same class.[29]

To evaluate whether a suspect firearm fired suspect ammunition, examiners visually compare tool-marks present on suspect ammunition to those present on test ammunition fired by the suspect weapon.[30] For more than eighty years, examiners have primarily used the comparison microscope to undertake this evaluation.[31] Nowadays, their work can also be supported by ballistics imaging technology and national databases, such as the National Integrated Ballistic Information Network ("NIBIN").[32]

Through applying these methods many examiners believe they can make reliable conclusions about identification, including individualization. In 1998, the Association of Firearms and Tool Mark Examiners ("AFTE"), the leading professional organization in this field, developed a protocol ("AFTE Protocol") detailing when an examiner may reach a certain conclusion.[33] Under the AFTE Protocol, an examiner may make one of four conclusions: (1) identification; (2) inconclusive; (3) elimination; or (4) unsuitable for comparison.[34]

To make an "identification" (i.e., a "match"), there must be "sufficient agreement" between the tool-marks subject to examination.[35] Under AFTE's Theory of Identification, which was adopted in 1992, "sufficient agreement" relates "to the significant duplication of random toolmarks"[36] as evidenced by "the correspondence of a pattern or combination of patterns of surface contours."[37] An agreement is considered significant "when the agreement in individual characteristics exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools and is consistent

---

28. *See Id.* (stating that subclass characteristics differ from individual characteristics because they are shared by more than one tool, but they cannot fall under class characteristics because every tool in that class does not share them).

29. *Id*.

30. STRENGTHENING, *supra* note 4, at 152–53.

31. ROBERT M. THOMPSON, FIREARMS IDENTIFICATION IN THE FORENSIC SCIENCE LABORATORY 8 (2010), http://ndaa.org/pdf/Firearms_identity_NDAAsm.pdf.

32. *Id.* at 29–30 (summarizing NIBIN). *See generally National Integrated Ballistic Information Network (NIBIN)*, BUREAU OF ALCOHOL, TOBACCO, FIREARMS, AND EXPLOSIVES, https://www.atf.gov/firearms/national-integrated-ballistic-information-network-nibin (last visited Apr. 15, 2018).

33. Ass'n of Firearms & Tool Mark Exam'rs, *Theory of Identification as It Relates to Toolmarks*, 30(1) ASS'N FIREARMS & TOOL MARK EXAM'RS J., 86 (1998).

34. *Id.* at 86–87.

35. *AFTE Theory of Identification*, ASS'N OF FIREARMS & TOOL MARK EXAM'RS, https://afte.org/about-us/what-is-afte/afte-theory-of-identification (last visited May 17, 2017).

36. *Id*.

37. *Id*.

with agreement demonstrated by toolmarks known to have been produced by the same tool."[38] A conclusion that there is "sufficient agreement" between two toolmarks "means that the agreement of individual characteristics is of a quantity and quality that the likelihood another tool could have made the mark is so remote as to be considered a practical impossibility."[39] The theory states that the interpretation of identifications is subjective but "founded on scientific principles and based on the examiner's training and experience."[40]

B. Current Debates About Validity

Over the last three decades in particular, forensic disciplines such as firearms identification have been subject to growing criticism.[41] Unreliable forensic science evidence is now a proven cause of wrongful conviction, with nearly half of the wrongful convictions associated with all 356 post-conviction DNA exonerations being attributable, in some way, to such evidence.[42] These cases have underscored criticism that "little systematic research has been conducted to validate the [forensic science] field's basic premises and techniques. . . ."[43]

Such criticism has underpinned national level concern about firearms identification evidence. In 2008, the National Research Council of the National Academies published its *Ballistic Imaging Report*,[44] and although the report was not intended to be an overall assessment of the firearms identification discipline, the report found the validity of the fundamental assumptions of uniqueness and reproducibility of firearms-related tool-marks had not yet been fully demonstrated.[45] The Committee took the view that a significant amount of research would be needed to scientifically determine the degree to which firearms-related tool-marks are unique, or even to qualitatively characterize the probability of uniqueness.[46]

Strengthening, which examined past, current, and the future use of forensic science in the United States, further cemented these views in 2009.

---

38. *Id*.
39. *Id*.
40. *Id*.
41. *See, e.g.,* O'Brien et al., *supra* note 3.
42. *See Misapplication of Forensic Science,* INNOCENCE PROJECT, https://www.innocenceproject.org/causes/misapplication-forensic-science/ (last visited May 04, 2018).
43. Paul C. Giannelli, Commentary, *Scientific Evidence in Civil and Criminal Cases*, 33 ARIZ. ST. L.J. 103, 112 (2001).
44. NAT'L RESEARCH COUNCIL, BALLISTIC IMAGING 3 (Daniel L. Cork et al. eds., 2008).
45. *See* United States v. Taylor, 663 F. Supp. 2d 1170, 1175–1176 (D.N.M. 2009) (discussing the focus and scope of *Ballistic Imaging*).
46. *Id*.

Strengthening concluded that class characteristics can be "helpful in narrowing the pool of tools that may have left a distinctive mark"[47] and that individual characteristics "might, in some cases, be distinctive enough to suggest one particular source."[48] It also commented, however, that the AFTE Protocol was not defined in a sufficiently precise way for examiners to follow, particularly in relation to when an examiner can conclude a "match."[49] The AFTE Protocol was limited because it "does not even consider, let alone address, questions regarding variability, reliability, repeatability, or the number of correlations needed to achieve a given degree of confidence."[50] Overall, *Strengthening* concluded "the scientific knowledge base for toolmark and firearms analysis is fairly limited,"[51] and in order to make the process of individualization more precise and repeatable "additional studies should be performed."[52] *Strengthening* did not comment specifically on the admissibility of such evidence.

In 2016, President Obama's PCAST published a report, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*, in response to questions raised by President Obama post-*Strengthening*. President Obama wanted to know if additional steps remained to be taken to help "ensure the validity of forensic evidence used in the Nation's legal system."[53] PCAST concluded that two important gaps remain, namely (1) a need for clarity about the scientific standards for the validity and reliability of forensic methods and (2) a need to evaluate specific forensic methods (including firearms identification) to determine whether their validity and reliability has been scientifically established.[54] With regard to firearms identification, PCAST stated its conclusions were "consistent" with those in *Strengthening*,[55] and that only one appropriate black box study had been undertaken.[56] On the question of admissibility, PCAST stated that the decision "belongs to the courts,"[57] but "[i]f firearms analysis *is* allowed in court, the scientific criteria for validity as applied should be understood to require clearly reporting the error rates seen in the one appropriately designed black-box study. Claims of higher accuracy are not scientifically

47. STRENGTHENING, *supra* note 4, at 154.
48. *Id*.
49. *Id.*
50. *Id.* at 155.
51. *Id*.
52. *Id.* at 154.
53. PCAST REPORT*, supra* note 8, at 1.
54. *Id*.
55. *Id.* at 11.
56. *Id.* (showing estimated error-rates could be as high as 1 in 46).
57. *Id* at 12.

justified at present."[58] PCAST made a number of recommendations to judges in this vein.[59]

Although stakeholders largely share the vision that there is a need to continually enhance forensic science, they do not equally share concerns about the validity of firearms identification evidence. For example, in 2009, post-*Strengthening*, FBI employees specializing in firearms and tool-marks published views that the discipline is both "highly valuable and highly reliable in its traditional methods."[60] Other groups have challenged the findings in *Strengthening*, including AFTE.[61] More recently, AFTE has responded critically to the PCAST Report, expressing its "disappointment" that the report ignored research indicating "firearm and toolmark identification is scientifically valid,"[62] and that if the AFTE Theory is applied properly, "examiners are able to conduct quality, accurate analysis."[63] Indeed, a number of studies have highlighted reliability in the field.[64] The FBI, Department of Justice, and the Bureau of Alcohol, Tobacco, Firearms and Explosives (ATF) have also expressed "disappointment" with the PCAST Report.[65] They strongly disagree with the report's recommendations "regarding the admission of forensic evidence in criminal trials, particularly with respect to firmly established firearm and tool mark forensic evidence."[66] ATF summarized,

---

58. *Id*.

59. PCAST REPORT, *supra* note 8, at 19.

60. *See* Stephen G. Bunch et al., *Is a Match Really a Match? A Primer on the Procedures and Validity of Firearm and Toolmark Identification*, FBI (July 2009), https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/july2009/review/2009_07_review01.htm.

61. *See* AFTE Committee for the Advancement of the Science of Firearm and Tool Mark Identification, *The Response of the Association of Firearm and Tool Mark Examiners to the February 2009 National Academy of Science Report "Strengthening Forensic Science in the United States: A Path Forward,"* 41 ASS'N FIREARMS & TOOL MARK EXAM'RS J. 204, 205 (2009), https://afte.org/uploads/documents/position-nas-2009.pdf ("There is an extensive body of research, extending back over one hundred years, which establishes the accuracy, reliability, and validity of conclusions rendered in the field of firearm and toolmark identification.").

62. *See Response to the PCAST Report on Forensic Science*, ASS'N OF FIREARMS & TOOL MARK EXAM'RS (Oct. 31, 2016), https://afte.org/uploads/documents/AFTE-PCAST-Response.pdf.

63. *Id*.

64. *See generally* studies referred to in Sarah Lucy Cooper, *Firearms Identification Evidence: Emerging Themes from Recent Criticism, Research and Case Law*, *in* FORENSIC REFORM: PROTECTING THE INNOCENT 174 (Wendy J. Kohen & C. Michael Bowers eds., 2017).

65. *See ATF's Response to the President's Council of Advisors on Science and Technology Report,* INT'L ASS'N FOR IDENTIFICATION 1 (Sept. 21, 2016), https://www.theiai.org/president/20160921_ATF_PCAST_Response.pdf.

66. *Id.*

> With respect to PCAST's recommendation that courts should restrict the admission of firearm and tool mark evidence, ATF strongly agrees with the DOJ decision not to adopt that recommendation as the existing legal standards regarding the admissibility of firearm and tool mark evidence are based on sound science and sound legal reasoning. Decades of legal precedent—and the underlying scientific research on which the courts have relied—establish that forensic firearm and tool mark evidence is both reliable and of substantial value to juries in determining the facts. Firearm and tool mark evidence not only aids prosecutors and defense attorneys in the courtroom, it also enhances public safety and protects the innocent by providing law enforcement with science-based tools to focus scarce investigative resources on actual perpetrators.[67]

Notably, the NCFS, established in 2013 following recommendations in *Strengthening,* has adopted a number of recommendations related to advancing scientific inquiry and research across forensic science, including how forensic science evidence is reported and testified about in court.[68] The Department of Justice is also in the process of developing guidance documents governing the testimony and reports of its forensic experts.[69] To date, no uniform testimony guidelines have been provided in relation to firearms identification.[70]

C. Court Responses

The disagreement amongst stakeholders described in Part II(B) naturally demonstrates that a level of uncertainty exists about firearms identification evidence. Although firearms identification evidence has routinely satisfied the leading legal standards for admissibility in America for nearly a century,[71] lawyers have recently used this uncertainty to challenge the admissibility of firearms identification evidence in criminal cases.[72] This section presents a cohort of such cases from which the authors selected expert statements to use in their study.

In response, some courts have restricted expert testimony to account for the alleged uncertainty. These courts have taken various approaches. Some courts have required examiners to describe their observations. For example, in *United States v. Green*,[73] the state sought to admit expert testi-

---

67. *Id*.

68. *See Reporting and Testimony,* U.S. DEP'T OF JUST., https://www.justice.gov/ncfs/reporting-and-testimony (last updated Nov. 6, 2017).

69. *See Forensic Science,* U.S. DEP'T OF JUST., https://www.justice.gov/archives/dag/forensic-science (last visited May 17, 2017).

70. *Id*.

71. *See generally* Cooper, *supra* note 64.

72. *Id.*

73. United States v. Green, 405 F. Supp. 2d 104 (D. Mass. 2005).

mony that Green's pistol could be matched, "to the exclusion of every other firearm in the world."[74] The court prevented this conclusion, permitting the examiner to only "describe and explain the ways in which the earlier casings are similar to the shell casings test-fired from the . . . pistol."[75]

Other courts have required examiners to testify to degrees of certainty. In *United States v. Diaz*,[76] the court allowed the firearms examiner to testify to "a reasonable degree of ballistic certainty."[77] This language was also permitted in *Commonwealth v. Pytou Heang*.[78] Similarly, in *United States v. Taylor*,[79] the court restricted the examiner to testifying that the ammunition came from Taylor's rifle to within a "reasonable degree of certainty in the firearms examination field."[80] In *United States v. Otero,*[81] the court found testimony to a "reasonable degree of professional certainty" was permissible.[82] A slightly different approach emerged in *United States v. Glynn*,[83] where the court limited the examiner, *inter alia*, to testifying that "a firearms match was 'more likely than not.'"[84]

On the other hand, some courts have continued to allow testimony that conveys absolute conclusions. For instance, in *United States v. Natson,*[85] a court sanctioned testimony from a firearms examiner that the tool-marks present on a suspect cartridge "was fired" by Natson's gun. In *United States v. Melcher,*[86] the trial court ordered that the expert should not testify that he was "one hundred percent" sure, but the expert did, in fact, state that the "'chances of another firearm creating [the] exact same pattern are so remote to be considered practically impossible.'"[87] The appellate court acknowledged that the expert had come "close to the line" of expressing one hundred percent certainty, but nonetheless rejected the appeal.[88] In *United States v. Mouzone*,[89] the expert was ordered not to testify that it was a practical impossibility for different firearms to have fired the suspect casings or that he

---

74. *Id.* at 107.
75. *Id.* at 108–09.
76. United States v. Diaz, No. CR 05-00167 WHA, 2007 WL 485967 at *11–14.
77. *Id.* at *11.
78. Commonwealth v. Pytou Heang, 458 Mass. 827, 848–851, 942 N.E.2d 927, 945–948 (Mass. 2011).
79. United States v. Taylor, 663 F. Supp. 2d 1170, 1180 (D.N.M. 2009).
80. *Id.*
81. United States v. Otero, 849 F. Supp. 2d 425, 429 (D.N.J. 2012).
82. *Id.*
83. United States v. Glynn, 578 F. Supp. 2d 567 (S.D.N.Y. 2008).
84. *Id.* at 575.
85. United States v. Natson, 469 F. Supp. 2d 1253 (M.D. Ga. 2007).
86. United States v. Melcher, No. 12-0544 WHO (PR), 2014 WL 31359, at *12 (N.D. Cal. Jan. 3, 2014).
87. *Id.* at *12.
88. *Id.* at *13.
89. United States v. Mouzone, 687 F.3d 207, 216 (4th Cir. 2012).

was certain about his conclusions.[90] At trial, however, the expert repeatedly testified that the casings found at two different murder scenes were "'fired from the same firearm.'"[91] Mouzone appealed on this point.[92] Like in *Melcher*, however, the appellate court rejected the appeal.[93]

The authors used these cases (and the AFTE Theory of Identification) to generate and select expert statements for their study. The practical impact of such statements in a courtroom is important to investigate. This is because jurors must determine the weight of such evidence in criminal trials, and research indicates that various challenges can arise when jurors assess scientific evidence. Part III discusses associated literature and the authors' study.

III. EXPERT TESTIMONY AND JUROR ASSESSMENT OF CERTAINTY: EXISTING LITERATURE AND RATIONALE FOR THE STUDY

Jurors in criminal cases can encounter various obstacles, both of a personal and institutional nature, when assessing scientific evidence.[94] Given the crucial role jurors and forensic science evidence play in the criminal justice system, research examining this issue is important. As the NAS stated in *Strengthening*,

> Jurors' use and comprehension of forensic evidence is not well studied. Better understanding is needed in this area, and recommendations are needed for programs or methods that will better prepare juries in appropriate, unbiased ways for trials in which scientific evidence is expected to play a large or pivotal role.[95]

This section explores existing literature on this topic and presents the method and findings of the authors' study against this backdrop.

A. Existing Literature

Examining juror certainty based on expert testimony is a complex process because of the context-dependant nature of the field.[96] Participant assessment of the credibility of experts has been extensively studied, both

---

90. *Id.*
91. *Id.*
92. *Id.*
93. *Id.* at 217.
94. *See generally* Cooper, *supra* note 12; Cooper, *supra* note 64.
95. STRENGTHENING, *supra* note 4, at 237.
96. Saks, M.J., Risinger, D.M., Rosenthal, R. & Thompson, W.C., *Context Effects in Forensic Science: A Review and Application of the Science of Science to Crime Laboratory Practice in the United States,* SCI. & JUST., 43*,* 77–90 (2003).

generally[97] and specifically in courtrooms.[98] Recent studies continue to reflect the complexity of the area. In 2015, Blackwell and Seymour concluded that jurors rank relevant professional experience, lack of bias, and clarity of evidence in order of importance.[99] Confidence, eye-contact and academic qualifications are ranked as less important.[100] In 2015, Thompson and Newman found that perceptions of both DNA and shoeprint evidence are modified by prior expectation and belief as well as the content of the evidence itself.[101] Studies have also shown that some expert witnesses are prepared for testimony by trial consultants, to increase the credibility and believability of their evidence.[102] This level of context makes examining the content of any expert statements, including statements about firearms identification evidence, challenging.

Koehler and Ritchie attempted to remove much of the context while examining expert statements (about DNA evidence) of numerical certainty using simplified experimental designs.[103] These studies suggest that exclusion percentages[104] are more likely to result in conviction than if an expert were to testify in terms of frequency ratios.[105] However, the case law described in Part II(C) shows that forensic experts, specifically firearms examiners, can and often do testify using solely linguistic or ordinal category-based evidence.[106] Naturally, any verbal certainty judgements, made without explicit statistical information, are a balance between the meaning of a

---

97. *See generally* Chanthika Pornpitakpan, *The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence*, 34 J. APPLIED SOC. PSYCHOL. 243 (2004).

98. *See generally* Stanley L. Brodsky, Michael P. Griffin, & Robert J. Cramer, *The Witness Credibility Scale: An Outcome Measure for Expert Witness Research,* 28 BEHAV. SCI. & L., 892 (2010).

99. Suzanne Blackwell & Fred Seymour, *Expert Evidence and Jurors' Views on Expert Witnesses*, 22 PSYCHIATRY, PSYCHOL. & L. 673, 678 (2015).

100. *Id*.

101. 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, 332 (2015).

102. Robert J. Cramer, et al., *A Confidence-Credibility Model of Expert Witness Persuasion: Mediating Effects & Implications for Trial Consultation,* CONSULTING PSYCHOL. J: PRAC. & RES. 129, 130 (2011).

103. *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).

104. 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%."

105. 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."

106. Evidence that suggests a fixed hierarchy, but the numerical difference between the categories is not fixed, for example, statements such as "likely," "very likely," and "extremely likely."

phrase given by the expert and the interpreter's own subjective understanding of the evidence.[107]

The inherent problem with these ordinal statements has been the focus of some examination. Krauss and Sales[108] manipulated the type of expert testimony (clinical versus actuarial), finding that participants were more influenced by the qualitative, linguistic evidence.[109] Martire *et al* found that inculpatory evidence was significantly more likely to be seen as weak if it was presented in terms of linguistic descriptions than in terms of numerical likelihood.[110] Despite the Association of Forensic Science Providers guidelines suggesting forensic experts use likelihood ratios, uptake has varied between disciplines and jurisdictions.[111]

In 2008, McQuiston-Surrett and Saks asked participants to rate an odontology expert's intended certainty, on a scale from zero to one hundred, from four phrases taken from American Board of Forensic Odontology (ABFO) guidelines.[112] Responses showed that participant estimates did not mirror the intended hierarchy.[113] For example, the use of the term "*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 "*consistent with*" (75) another statement assumed to be uncertain, and significantly above both more certain phrases – "*probable*" (57) and "*reasonable scientific certainty*" (70).[114] The researchers concluded that expert witnesses cannot merely create a definition for a term and expect judges and juries to understand what they mean.[115]

107. Thomas S. Wallsten, Samuel Fillenbaum, & James A. Cox, *Base Rate Effects on the Interpretations of Probability and Frequency Expressions*, J. MEMORY & LANGUAGE 571 (1986).

108. Daniel A. Kraus & Bruce D. Sales, *The Effects of Clinical and Scientific Expert Testimony on Juror Decision Making in Capital Sentencing,* 7 PSYCHOL., PUB. POL'Y & L. 267, 276–77 (2001).

109. *Id.* Clinical meaning expert testimony based on personal experience of the subject; and actuarial meaning a numerically-based risk factor taken from research on large groups of people.

110. Kristy A. Martire et al., *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).

111. *Id*.

112. Dawn McQuiston-Surrett, & Michael J. Saks, *Communicating Opinion Evidence in the Forensic Identification Sciences: Accuracy and Impact*, 59 HASTINGS L. J. 1159 (2008).

113. *Id.* at 1162–63.

114. *Id.*

115. *Id.* at 1163.

McQuiston-Surrett and Saks followed up this study with an examination of certainty statements in microscopic hair evidence.[116] They examined both potential jurors and judicial participants, including 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*."[117] They found that the qualitative statements were deemed significantly more certain than the subjective probability or objective multiple-frequency statements.[118] Non-judicial participants were particularly susceptible to this effect.[119] Unlike in their 2008 study, they did not find a significant difference between the two qualitative statements, with both scoring around seventy-nine on a zero to one hundred scale.[120] Based on these studies, they concluded that empirical testing of the responses to the relevant words and phrases is needed, and that a simple approach would be best.[121]

Understanding more about juror interpretation of qualitative, language-based evidence is particularly important. This is because studies suggest that expert testimony couched in absolute terms may sensitize or inure potential jury members to evidence, resulting in higher certainty judgments because such testimony is considered to clarify an otherwise 'grey' situation.[122] Jurors have been found to show a lack of skill in distinguishing between evidence-based findings and flawed science.[123] Jurors' understanding of internal validity difficulties in expert evidence, such as the robustness of methods used by experts, has also been shown to be flawed.[124] The conclusion was that there is a need to examine the effectiveness of traditional legal safeguards against unfounded, seemingly-scientific testimony.[125] In fact, in general, it has been found that participants in a range of experimental situa-

---

116. Dawn McQuiston-Surrett, & Michael J. Saks, *The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Factfinders Hear,* 33 L. & HUM. BEHAV, 436–453 (2009).

117. *Id.* at 437–438.

118. *Id.* at 444.

119. *Id.* at 445.

120. *Id.* at 443.

121. *Id*; McQuiston-Surrett & Saks, *supra* note 112, at 1163.

122. Nancy Brekke & Eugene Borgida, *Expert Psychological Testimony in Rape Trials: A Social-Cognitive Analysis*, 55(3) J. PERSONALITY & SOC. PSYCHOL. 372, 383 (1988); Margaret Bell Kovera et al., *Does Expert Psychological Testimony Inform or Influence Juror Decision Making? A Social Cognitive Analysis,* 82(1) J. APPLIED PSYCHOL. 178, 188 (1997).

123. Bradley D. McAuliff & Margaret Bell Kovera, *Juror Need for Cognition and Sensitivity to Methodological Flaws in Expert Evidence,* 38 J. APPLIED SOC. PSYCHOL. 385 (2008).

124. *Id.*

125. Bradley D. McAuliff & Tejah D. Duckworth, *I Spy with My Little Eye: Jurors' Detection of Internal Validity Threats in Expert Evidence,* 34(6) L. & HUM. BEHAV, 489, 499 (2010).

tions are often poor at assessing their own understanding of information and their own decision-making processes.[126]

This existing literature suggests that qualitative, language-based expert testimony, including that outlined in Part II(C),[127] is poorly understood by the juror pool, but simultaneously viewed as powerful and compelling by those same participants. This is a concerning combination of findings, meaning further understanding of these statements and their hierarchies is vital moving forward. The authors' study aimed to contribute to this need in the context of firearms identification evidence. Sub-section B outlines our rationale, methods, and findings.

B. Rationale

While individualization of firearms identification evidence has the support of some scientific findings,[128] a number of studies contest whether these underpinnings amount to sufficiently powerful evidence for definitive statements of certainty by experts.[129] Some United States courts have acknowledged this issue, as outlined in Part II(C).[130] These courts have restricted expert testimony; instructing experts to not testify in absolute terms, such as "there is an exact match," but rather in allegedly more diluted terms, such as a match can be made "more likely than not" and "to a reasonable degree of certainty." The reason the authors chose to focus on firearms identification evidence is because this pattern in judicial decision-making is particularly pronounced.[131] Through such cases, an ordinal hierarchy of expert testimony has, in effect, been created, presuming that such statements would be ranked as less certain [than an exact match] by jurors. What is not known, however, are the numerical differences between these statements when jurors assess certainty. This is the area the authors' study aimed to examine further.

---

126. *See, e.g.,* Timothy D. Wilson & Elizabeth W. Dunn, *Self-Knowledge: Its Limits, Value, and Potential for Improvement*, 55 ANN. REV. OF PSYCHOL. 493, 504 (2004).

127. *See supra* Part II(C).

128. *See, e.g.,* Ronald G. Nichols, *Firearm and Toolmark Identification Criteria: A Review of the Literature, Part II,* 48(2) J. FORENSIC SCI. 318 (2003); Ronald G. Nichols, *Defending the Scientific Foundations of the Firearms and Tool Mark Identification Discipline: Responding to Recent Challenges*, 52(3) J. FORENSIC SCI. 586–93 (2007).

129. *See* Mark Page, Jane Taylor & Matt Blenkin, *Forensic Identification Science Evidence Since Daubert: Part II - Judicial Reasoning in Decisions to Exclude Forensic Identification Evidence on Grounds of Reliability*, 56(4) J. FORENSIC SCI. 913, 917 (2011).

130. *See supra* Part II(C).

131. *See, e.g.,* Sarah Lucy Cooper, *The Collision of Law and Science: American Court Responses to Developments in Forensic Science*, 33 PACE L. REV. 234, 287 (2013); Cooper, *supra* note 62, at 470

IV. METHODS, RESULTS, AND FINDINGS

A. Methods

Using an online questionnaire, participants were recruited through academic, professional, and community links across the United States. In total, 107 participants were placed in the role of a juror in a criminal trial. Participants were told that a qualified firearms examiner, testifying for the state, had been asked to testify as to whether tool-marks produced on ammunition test-fired from the Defendant's gun matched tool-marks present on suspect ammunition found at the crime scene. While appreciating the wider context in which jury decision-making occurs, the authors purposely designed such a basic scenario in order to remove as much context as possible and to therefore encourage participants to focus solely on the content of the expert statements. The authors also deliberately sought participants from across the general population. The only eligibility requirement was that participants be eligible for federal jury service, which was determined by participants answering a number of closed-answer questions.

Participants were provided twelve different expert statements in pseudo-random order. They were asked to rank their level of certainty, based on each statement, on a scale of zero to one hundred.[132] The expert statements were chosen from statements referenced in case law, such as that outlined in Part II(C),[133] and with reference to the AFTE Theory. The authors ordered the statements into a generally ordinal hierarchy of certainty. Three statements were hypothesised to attract high-certainty and three low-certainty. The remaining seven statements were hypothesised to attract moderate levels of certainty.

The study simply sought to determine the statistical validity of the specific hierarchy of certainty.

B. Findings

The ordinal hierarchy was found to show a significant general trend in the expected direction. The following sub-sections present means and standard deviations for each statement; results that showed significant differences between statements; and a summary about the meaning and implications of our findings.

---

132. Zero being the least certain and one hundred being the most.
133. *See supra* Part II(C).

1. Means and Standard Deviations

Table A shows the mean certainties, standard deviations, and significant differences found for each expert statement. The Mean is the average score of certainty for each statement across participants. The Standard Deviation is the average amount each individual participant varied away from that Mean. The larger the Standard Deviation, the wider spread the participants' scores for certainty were for the relevant statement. A smaller Standard Deviation shows participants' scores for the relevant statement are more clustered together. The size of the Standard Deviation is important because it impacts 'significance' as explained in sub-section (b)2 below.

Table A: Mean Certainties, Standard Deviations, and Significant Differences

Number	Expert Statement	Mean Certainty	Standard Deviation	Significantly less certain statements
STATEMENTS HYPOTHESISED TO ATTRACT HIGH LEVELS OF CERTAINTY				
1	"There is a match to the exclusion of every other firearm in the world."	68.94	35.12	9, 10, 11, 12
2	"There is an exact match between the suspect ammunition and the ammunition test fired from the Defendant's gun."	92.05	15.92	1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
3	"The chances of another firearm creating the exact same tool-marks are so remote as to be considered practically impossible."	77.96	28.95	1, 8, 9, 10, 11, 12
STATEMENTS HYPOTHESISED TO ATTRACT MODERATE LEVELS OF CERTAINTY				
4	"A match between suspect ammunition and ammunition test-fired from the defendant's gun can be made to a practical certainty."	72.50	20.83	8, 9 10, 11,12
5	"There is a match between suspect ammunition and ammunition test-fired from the defendant's gun to a reasonable degree of ballistic certainty."	73.80	19.97	8, 9, 10, 11, 12
6	"A match between suspect ammunition and ammunition test-fired from the defendant's gun can be made to a reasonable degree of certainty in the ballistics field."	74.79	17.91	8, 9, 10, 11, 12
7	"A match between suspect ammunition	72.64	18.25	8, 9, 10, 11, 12

	and ammunition test-fired from the defendant's gun can be made to a reasonable degree of professional certainty."			
8	"A match between suspect ammunition and ammunition test-fired from the defendant's gun can be made to a reasonable degree of certainty in the firearms examination field."	66.59	21.57	9, 10, 11, 12
9	"A match between suspect ammunition and ammunition test-fired from the defendant's gun is more likely than not."	56.83	21.42	10, 11, 12
STATEMENTS HYPOTHESISED TO ATTRACT LOW LEVELS OF CERTAINTY				
10	"The results of my examination of the ammunition test-fired from the Defendant's gun and suspect ammunition are inconclusive."	22.75	30.90	1-9
11	"There is significant disagreement in discernible class characteristics and individual characteristics between the suspect ammunition and those test-fired by the Defendant's gun."	28.43	29.93	1-9
12	"The suspect ammunition is unsuitable for comparison with ammunition test-fired from the Defendant's gun."	22.29	29.18	1-9

Figure A: Means and Standard Deviation Ranges

Figure A shows the Mean certainty for statements 1-12 (0-100) and the Standard Deviation range for each statement.

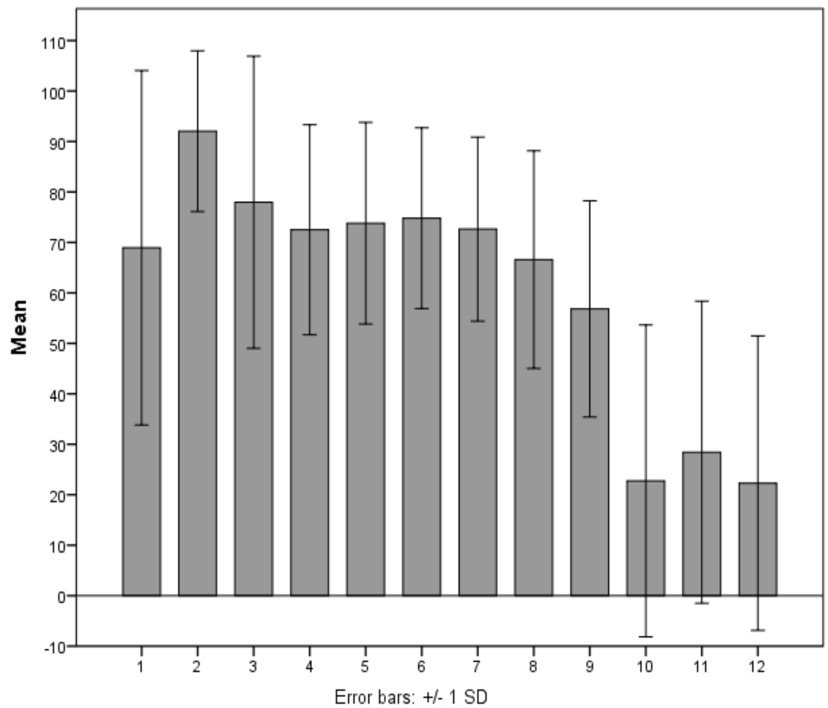

Figure A: Means and Standard Deviations for each of the statements

The information from Table A and Figure A was statistically analyzed to consider whether significant differences existed between the statements and, therefore, whether the general ordinal hierarchy was valid.

2. Significant Differences

The authors undertook a statistical analysis that examined the Mean and Standard Deviation variance and compared it between statements. In short, if the Means (for each statement) are different and the Standard Deviations (within each statement) show there is a low level of variability in scores, the statements are significantly different. The further apart the Means are between each statement, the more likely it is that a significant difference will be produced. However, as Standard Deviation increases, significance is less likely, as the scores are spread out more. A 'significant difference' means we can generalize our findings to the population our participant sample came from, in this case, the United States federal juror pool.

We used an ANOVA (statistical Analysis of Variance) to see if there was an overall significant difference between the statements and found there was. We then used t-tests[134] to compare each pair of statements separately for differences, one-on-one. This produced the results in Table B.

Table B: Significant Differences Between Expert Statements[135]

Number	Expert Statement	Expert Statement was found to elicit significantly more certainty than statements
1	"There is a match to the exclusion of every other firearm in the world."	9, 10, 11 and 12.
2	"There is an exact match between the suspect ammunition and the ammunition test fired from the Defendant's gun."	1, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12.
3	"The chances of another firearm creating the exact same tool-marks are so remote as to be considered practically impossible."	1, 8, 9, 10, 11 and 12.
4	"A match between suspect ammunition and ammunition test-fired from the defendant's gun can be made to a practical certainty."	8, 9 10, 11, and 12.
5	"There is a match between suspect ammunition and ammunition test-fired from the defendant's gun to a reasonable degree of ballistic certainty."	8, 9, 10, 11, and 12.
6	"A match between suspect ammunition and ammunition test-fired from the defendant's gun can be made to a reasonable degree of certainty in the ballistics field."	8, 9, 10, 11, and 12.
7	"A match between suspect ammunition and ammunition test-fired from the defendant's gun can be made to a reasonable degree of professional certainty."	8, 9, 10, 11, and 12.
8	"A match between suspect ammunition and ammunition test-fired from the defendant's gun can be made to a reasonable degree of certainty in the firearms examination field."	9, 10, 11, and 12.

134. Sixty-six t tests were used.
135. *See infra* Appendix A for t-tests data.

9	"A match between suspect ammunition and ammunition test-fired from the defendant's gun is more likely than not."	10, 11, and 12.
10	"The results of my examination of the ammunition test-fired from the Defendant's gun and suspect ammunition are inconclusive."	N/A.
11	"There is significant disagreement in discernible class characteristics and individual characteristics between the suspect ammunition and those test-fired by the Defendant's gun."	N/A.
12	"The suspect ammunition is unsuitable for comparison with ammunition test-fired from the Defendant's gun."	N/A.

C. Summary and Implications of Findings

The results show a general trend that the statements hypothesized as being high-certainty elicit significantly higher levels of certainty than those hypothesized as moderate-certainty statements and low-certainty statements. Moderate-certainty statements elicited higher levels of certainty than low-certainty statements. In other words, the general ordinal hierarchy was found to show a significant general trend in the expected direction.

The only exception to this trend was Statement 1 ("to the exclusion of every other firearm in the world"), which was found to elicit significantly less certainty than the other two high-certainty phrases, and no difference with all of the moderate-certainty phrases, except for Statement 9 ("more likely than not") This is interesting given that the precise language of Statement 1 suggests it was one of, if not the most, certain statement included in the study. As such, this finding warrants further investigation. Notably, Statement 3 ("practically impossible") was only found to elicit more certainty than two of the moderate-certainty statements, namely Statement 8 ("a reasonable degree of certainty in the firearms examination field) and Statement 9, and showed no significant difference with Statements 4, 5, 6, and 7. Statement 2 ("an exact match") was found to elicit the highest level of certainty from participants, suggesting such language is particularly persuasive to jurors. This finding suggests that judicial findings that the admissibility of such evidence is harmless error or is not prejudicial are overlooking juror perceptions of such evidence; a point Cooper has raised previously.[136]

---

136. Cooper, *supra* note 64, at 460 ("In relation to preventing frivolous claims from flooding the system, courts often conclude that the admission of such evidence was non-prejudicial in light of other evidence against the defendant. In other words, courts are terming

With regards to the moderate-certainty statements, Statements 4 ("practical certainty"), 5 ("reasonable degree of ballistic certainty"), 6 ("reasonable degree of certainty in the ballistics field"), and 7 ("a reasonable degree of professional certainty") did not show any inter-statement differences. However, these statements did elicit more certainty from participants than Statements 8 ("a reasonable degree of certainty in the firearms examination field") and 9 ("more likely than not"). This suggests that when experts convey their conclusions in terms of "practical," "professional," and "ballistic" certainty, they are seen by participants as more certainty-inducing than when they convey their conclusions in terms of "firearms-related" certainty. Notably, Statement 8 was found to induce more certainty than Statement 9. Statement 9 only elicited more certainty than the low-certainty statements. As expected, all high-certainty and moderate-certainty statements induced more certainty in participants than the low certainty statements, namely Statements 10 ("inconclusive"), 11 ("significant disagreement"), and 12 ("unsuitable for comparisons"). No significant differences were found between the three low-certainty phrases. All low-certainty statements generated some certainty in participants, including Statement 12, even though the statement conveyed that examined comparison could not take place.

The authors' findings with regard to the moderate-certainty statements are particularly important because these sorts of statements have been employed by the judiciary to address current uncertainty about firearms identification evidence. Our results suggest that the judiciary's assumption that these moderate-certainty phrases convey less certainty (and will be interpreted as such by jurors) than language akin to individualization is valid.

However, there are a few points to note. First, our study suggests that moderate-certainty statements do still attract notable certainty scores (72.5–74.8%) – they are not immaterial. Our results suggest that moderate-certainty statements do still persuade jurors more towards determining that the relevant evidence is inculpatory. As such, stakeholders should consider that such statements will not necessarily detract jurors from findings of guilt. Again, our findings suggest that the judiciary should be mindful of rejecting challenges to firearms identification evidence on the basis that admitting such evidence (presented in moderate-certainty formats, as well as high-certainty formats) was non-prejudicial or harmless in light of other evidence against the defendant.

Second, our results show that not all moderate-certainty statements are interpreted to convey similar levels of certainty by potential jurors. Statement 8 (66.6%) and Statement 9 (56.8%) induced significantly lower levels

---

the legally sound or unsound admission of firearms-identification evidence as immaterial. However, this rationale arguably overlooks the high impact scientific evidence has on jurors and the difficulty they have in accurately evaluating scientific evidence.").

of certainty in participants than all other moderate-certainty phrases, which attracted certainty scores of 72.5–74.8%. As such, the judiciary should be mindful not to use all moderate-certainty statements as if they are synonymous in meaning. Consider the case of *Melcher*, for example. In rejecting Melcher's appeal, the court commented that the difference between "practical certainty" and "considered practically impossible" versus "reasonable degree of certainty" or "more likely than not" would not tip the outcome of the case.[137] Although our results support the notion that there is no significant difference between the phrases "practical certainty," "practical impossibility," and "reasonable degree of certainty," they do suggest there is a significant difference between an expert using these three phrases and the phrase "more likely than not," which, our study suggests, is viewed as significantly less certain by jurors.

V. CONCLUSION

Criticism about the ability of crime-solving forensic disciplines, including firearms identification, to engage in valid individualization has grown throughout the 1990s and new millennium. Some American courts have acknowledged the existence of scientific uncertainty in the context of firearms individualization evidence since 2005. To address this uncertainty, these courts have restricted expert testimony; instructing experts to not testify in absolute terms, such as "there is an exact match," but rather in allegedly more diluted terms, such as a match can be made "more likely than not" and "to a reasonable degree of certainty."[138]

In criminal trials, jurors must determine the weight of expert evidence, and, thus, must assess how much certainty to attach to these various expert phrases. Given that research suggests that jurors might find this task challenging, and the existence of a particularly pronounced pattern of such judicial decision-making in firearms identification evidence cases, the authors conducted a study examining the effect of twelve expert statements of certainty, by a qualified firearms examiner, on potential jurors. Results from a sample of 107 participants found a significant main effect for certainty, suggesting that participant certainty was influenced by expert testimony in an expected direction along an ordinal hierarchy.

In the context of firearms identification evidence, the findings of the authors' study add to the body of literature that has found jurors to be influenced by qualitative, linguistic evidence, and that such evidence can elicit notable scores of certainty from jurors. Our results also affirm the findings

---

137. United States v. Melcher, No. 12-0544 WHO (PR), 2014 WL 31359, at *13 (N.D. Cal. Jan. 3, 2014).

138. *See supra* Part II(C)

of previous studies that expert testimony couched in absolute terms may particularly inure jurors to such evidence. In so doing, the authors' study responds to calls from researchers and national organizations to further investigate jurors' use and comprehension of certain words and phrases involved in forensic expert testimony, and the study provides some evidence with which to move forward.

On the basis of our findings, the authors echo the call for further research in this area. In particular, the authors suggest further investigation of (1) ordinal hierarchies of expert language used in relation to other forensic science disciplines subject to similar criticisms as firearms identification evidence; (2) whether a more significant ordinal hierarchy exists in relation to the moderate-certainty statements identified in the context of firearms identification evidence; (3) further examination of very high certainty phrases; (4) whether adding context influences juror interpretation of certainty, and, thus, the established hierarchy, *e.g.* by highlighting the alleged limitations of firearms identification evidence; and (5) on what basis jurors attach certainty to low, moderate, and high-certainty expert statements.

Appendix A

Pair 11: Statement 1 was found to elicit significantly more certainty than statement 12 (t[106] = 11.051, p = .000).

Pair 12: Statement 2 was found to elicit significantly more certainty than statement 3 (t[106] = 5.040, p = .000).

Pair 13: Statement 2 was found to elicit significantly more certainty than statement 4 (t[106] = 11.090, p = .000).

Pair 14: Statement 2 was found to elicit significantly more certainty than statement 5 (t[106] = 10.210, p = .000).

Pair 15: Statement 2 was found to elicit significantly more certainty than statement 6 (t[106] = 8.638, p = .000).

Pair 16: Statement 2 was found to elicit significantly more certainty than statement 7 (t[106] = 11.425, p = .000).

Pair 17: Statement 2 was found to elicit significantly more certainty than statement 8 (t[106] = 11.549, p = .000).

Pair 18: Statement 2 was found to elicit significantly more certainty than statement 9 (t[106] = 13.005, p = .000).

Pair 19: Statement 2 was found to elicit significantly more certainty than statement 10 (t[106] = 19.05, p = .000).

Pair 20: Statement 2 was found to elicit significantly more certainty than statement 11 (t[106] = 17.943, p = .000).

Pair 21: Statement 2 was found to elicit significantly more certainty than statement 12 t(106) = 11.051, p<.000.

Pair 26: Statement 3 was found to elicit significantly more certainty than statement 8 ($t[106] = 3.528$, $p = .001$).

Pair 27: Statement 3 was found to elicit significantly more certainty than statement 9 ($t[106] = 5.669$, $p = .000$).

Pair 28: Statement 3 was found to elicit significantly more certainty than statement 10 ($t[106] = 12.870$, $p = .000$).

Pair 29: Statement 3 was found to elicit significantly more certainty than statement 11 ($t[106] = 11.770$, $p = .000$).

Pair 30: Statement 3 was found to elicit significantly more certainty than statement 12 ($t[106] = 14.091$, $p = .000$).

Pair 34: Statement 4 was found to elicit significantly more certainty than statement 8 ($t[106] = 2.610$, $p = .010$).

Pair 35: Statement 4 was found to elicit significantly more certainty than statement 9 ($t[106] = 6.099$, $p = .000$).

Pair 36: Statement 4 was found to elicit significantly more certainty than statement 10 ($t[106] = 13.200$, $p = .000$).

Pair 37: Statement 4 was found to elicit significantly more certainty than statement 11 ($t[106] = 13.179$, $p = .000$).

Pair 38: Statement 4 was found to elicit significantly more certainty than statement 12 ($t[106] = 14.875$, $p = .000$).

Pair 41: Statement 5 was found to elicit significantly more certainty than statement 8 ($t[106] = 3.271$, $p < 0.001$).

Pair 42: Statement 5 was found to elicit significantly more certainty than statement 9 ($t[106] = 6.913$, $p < 0.001$).

Pair 43: Statement 5 was found to elicit significantly more certainty than statement 10 ($t[106] = 14.296$, $p < 0.001$).

Pair 44: Statement 5 was found to elicit significantly more certainty than statement 11 ($t[106] = 13.626$, $p < 0.001$).

Pair 45: Statement 5 was found to elicit significantly more certainty than statement 12 ($t[106] = 15.641$, $p < 0.001$).

Pair 47: Statement 6 was found to elicit significantly more certainty than statement 8 ($t[106] = 3.587$, $p = 0.001$).

Pair 48: Statement 6 was found to elicit significantly more certainty than statement 9 ($t[106] = 7.873$, $p < 0.001$).

Pair 49: Statement 6 was found to elicit significantly more certainty than statement 10 ($t[106] = 16.196$, $p < 0.001$).

Pair 50: Statement 6 was found to elicit significantly more certainty than statement 11 ($t[106] = 14.793$, $p < 0.001$).

Pair 51: Statement 6 was found to elicit significantly more certainty than statement 12 ($t[106] = 17.066$, $p < 0.001$).

Pair 52: Statement 7 was found to elicit significantly more certainty than statement 8 ($t[106] = 3.279$, $p = 0.001$).

Pair 53: Statement 7 was found to elicit significantly more certainty than statement 9 ($t[106] = 6.690$, $p < 0.001$).

Pair 54: Statement 7 was found to elicit significantly more certainty than statement 10 ($t[106] = 13.740$, $p < 0.001$).

Pair 55: Statement 7 was found to elicit significantly more certainty than statement 11 ($t[106] = 13.368$, $p < 0.001$).

Pair 56: Statement 7 was found to elicit significantly more certainty than statement 12 ($t[106] = 16.138$, $p < 0.001$).

Pair 57: Statement 8 was found to elicit significantly more certainty than statement 9 ($t[106] = 3.915$, $p < 0.001$).

Pair 58: Statement 8 was found to elicit significantly more certainty than statement 10 ($t[106] = 12.609$, $p < 0.001$).

Pair 59: Statement 8 was found to elicit significantly more certainty than statement 11 ($t[106] = 10.804$, $p < 0.001$).

Pair 60: Statement 8 was found to elicit significantly more certainty than statement 12 ($t[106] = 13.724$, $p < 0.001$).

Pair 61: Statement 9 was found to elicit significantly more certainty than statement 10 ($t[106] = 10.146$, $p < 0.001$).

Pair 62: Statement 9 was found to elicit significantly more certainty than statement 11 ($t[106] = 8.346$, $p < 0.001$).

Pair 63: Statement 9 was found to elicit significantly more certainty than statement 12 ($t[106] = 10.877$, $p < 0.001$).