

A167548

**IN THE COURT OF APPEAL
OF THE STATE OF CALIFORNIA**

FIRST APPELLATE DISTRICT, DIVISION THREE

PEOPLE OF THE STATE OF CALIFORNIA,
Plaintiff and Respondent,

v.

RAYMOND TIDD,
Defendant and Appellant.

APPEAL FROM THE SUPERIOR COURT FOR SAN FRANCISCO COUNTY
HON. STEPHEN MURPHY, JUDGE
CASE NOS. CRI21008732 & SCN234453

**APPLICATION TO FILE AMICI CURIAE
BRIEF AND BRIEF OF AMICI CURIAE THE
WILSON CENTER FOR SCIENCE AND
JUSTICE AT DUKE LAW SCHOOL AND A
GROUP OF ESTEEMED SCIENTISTS AND
SCHOLARS IN SUPPORT OF DEFENDANT**

COMPLEX APPELLATE LITIGATION GROUP LLP

***Kelly A. Woodruff** (No. 160235)
kelly.woodruff@calg.com

Jennifer Teaford (No. 228635)
jennifer.teaford@calg.com

96 Jessie Street
San Francisco, CA 94105
(415) 649-6700 • FAX: (415) 726-2527

ATTORNEYS FOR AMICI CURIAE

Table of Contents

Table of Authorities	4
Application to File Brief of Amici Curiae the Wilson Center for Science and Justice at Duke Law School and a Group of Esteemed Scientists and Scholars in Support of Defendant	8
Brief of Amici Curiae the Wilson Center for Science and Justice at Duke Law School and a Group of Esteemed Scientists and Scholars in Support of Defendant	12
Introduction.....	12
Discussion.....	14
I. The scientific community has rejected the notion that FA/TM identification evidence has been scientifically validated.	14
A. Reports authored by committees of experts from the scientific community have concluded that FA/TM identification evidence has not been scientifically validated.....	15
B. Since the publication of the reports, the scientific community has continued to cast doubt upon FA/TM identification as a discipline.....	18
C. Existing FA/TM studies do not validate FA/TM identification as a discipline....	22
1. Methodological Failures of Set- Based Studies.....	25
2. Methodological Failures of Sample-to-Sample Black-Box Studies	28
3. Alarming Error Rates.....	35
D. Just because there may be a number of validation studies concluding that FA/TM identification is valid does not make it so when the studies themselves are fatally flawed.....	39

II.	The <i>Kelly</i> rule should govern the admissibility of FA/TM evidence.....	43
III.	This Court Should Preclude FA/TM examiners from testifying about the presence of individual or unique characteristics of spent ammunition that imply a “match.”	49
A.	Allowing any expert testimony about the similarity or presence of individual or unique characteristics should be prohibited.	50
B.	There is a notable movement toward limiting the presentation of FA/TM evidence, but it does not go far enough.	54
C.	Even limiting FA/TM testimony to anodyne “consistent with” or “similar” language does not sufficiently protect against the overstatement of the probative value of FA/TM evidence.	56
	Conclusion	59
	Certificate of Word Count.....	60

Table of Authorities

Cases

<i>People v. Azcona</i> (2020) 58 Cal.App.5th 504.....	43, 46, 47, 48
<i>People v. Cowan</i> (2010) 50 Cal.4th 401.....	45, 46
<i>People v. Jackson</i> (2016) 1 Cal.5th 269.....	44
<i>People v. Kelly</i> (1976) 17 Cal.3d 24	<i>passim</i>
<i>People v. Venegas</i> (1998) 18 Cal.4th 47.....	15
<i>United States v. Adams</i> (D. Or. 2020) 444 F. Supp. 3d 1248.....	28, 55
<i>United States v. Shipp</i> (E.D.N.Y. 2019) 422 F. Supp. 3d 762.....	28
<i>United States v. Tibbs</i> (D.C. Super. Ct., Sept. 5, 2019, No. 2016-CF1-19431) 2019 WL 4359486	55
<i>Williams v. United States</i> (D.C.Ct.App. 2019) 210 A.3d 734	54

Rules

Cal. Rules of Court, rule 8.200.....	8, 10
Cal. Rules of Court, rule 8.360.....	8

Other Authorities

Albright, <i>How to Make Better Forensic Decisions</i> (2022) 119 Proceedings of the Nat. Academy of Sciences, No. 38.....	19, 32, 51
American Statistical Assn., Position on Statistical Statements for Forensic Evidence (Jan. 2, 2019).....	50, 51, 52, 53
Bajic et al., <i>Report: A Validation Study of the Accuracy, Repeatability, and Reproducibility of Firearms Comparison</i> (Oct. 7, 2020)	35

Baldwin et al., A Study of False-Positive and False-Negative Error Rates in Cartridge Case Comparisons (Apr. 7, 2014) 35, 36

Cole et al., *Microscopic Hair Comparison Analysis and Convicting the Innocent*, The Nat. Registry of Exonerations (2023)*passim*

Cuellar et al., *Methodological Problems in Every Black-Box Study of Forensic Firearm Comparisons* (2024).....*passim*

Dept. of Justice and FBI Joint Statement on Microscopic Hair Analysis (Apr. 19, 2015) <<https://www.fbi.gov/news/press-releases/department-of-justice-and-fbi-joint-statement-on-microscopic-hair-analysis>> [as of May 10, 2024] 42

Dorfman & Valliant, *A Re-analysis of Repeatability and Reproducibility in the Ames-USDOE-FBI Study* (2022) 9 Stats. & Pub. Policy 175..... 37, 38

Dorfman & Valliant, *Inconclusives, Errors, and Error Rates in Forensic Firearms Analysis: Three Statistical Perspectives* (2022) 5 Forensic Sci. Internat.: Synergy*passim*

Dror & Scurich, *(Mis)use of Scientific Measurements in Forensic Science* (2020) 2 Forensic Sci. Internat.: Synergy 333.. 12, 20, 30, 31

Ebersole, *Old-School Hair Analysis is Junk Science. But it Still Keeps People Behind Bars, The Marshall Project*, The Marshall Project (Dec. 15, 2023), <<https://www.themarshallproject.org/2023/12/15/florida-death-penalty-hair-analysis-junk-science>> 43

Epps & Todorow, *Refried Forensics: Screening Expert Testimony in Criminal Cases Through Frye Plus Reliability* (2018) 48 Seton Hall. L.Rev. 1161 45

Faigman et al., *The Field of Firearms Forensics Is Flawed*, Scientific American (May 25, 2022) <<https://www.scientificamerican.com/article/the-field-of-firearms-forensics-is-flawed/>> 19, 30, 37

Garrett et al., <i>Mock Jurors’ Evaluation of Firearm Examiner Testimony</i> (2020) 44 Law & Hum. Behav. 412.....	51, 57, 58
Hofmann et al., <i>Treatment of Inconclusives in the AFTE Range of Conclusions</i> (2020) 19 Law, Prob. & Risk 317	<i>passim</i>
Kafadar, <i>The Critical Role of Statistics in Demonstrating the Reliability of Expert Evidence</i> (2018) 86 Fordham L.Rev. 1617.....	29, 54
Krauss & Sales, <i>The Effects of Clinical and Scientific Expert Testimony on Juror Decision Making in Capital Sentencing</i> (2001) 7 Psych., Pub. Policy & Law 267	53
Levett & Kovera, <i>The Effectiveness of Opposing Expert Witnesses for Educating Jurors About Unreliable Expert Evidence</i> (2008) 32 Law & Hum. Behav. 363.....	58
Nat. Assn. of Criminal Defense Lawyers, <i>Microscopic Hair Comparison Analysis</i> < https://www.nacd.org/Landing/Microscopic-Hair-Comparison-Analysis > [as of May 10, 2024]	41, 42
Nat. Research Council of the Nat. Academies, <i>Ballistic Imaging</i> (2008)	15, 16, 23, 47
Nat. Research Council of the Nat.Academies, <i>Strengthening Forensic Science in the United States: A Path Forward</i> (2009)	<i>passim</i>
<i>People v. Auimatagi</i> , Order After Hearing on Motion to Exclude Testimony of Prosecution Witness Alex Taflya (Super. Ct. Yolo County, June 7, 2021, No. 19-4995)	48
President’s Council of Advisors on Science and Technology, <i>Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods</i> (2016)	<i>passim</i>
Schweitzer & Saks, <i>The Gatekeeper Effect: The Impact of Judges’ Admissibility Decisions on the Persuasiveness of Expert Testimony</i> (2009) 15 Psych., Pub. Policy & Law 1	53

Scurich et al., *Scientific Guidelines for Evaluating the Validity of Forensic Feature-Comparison Methods* (2023) 120 Proceedings of the Nat. Academy of Sciences, No. 41.....*passim*

Spiegelman & Tobin, *Analysis of Experiments in Forensic Firearms/Toolmarks Practice Offered as Support for Low Rates of Practice Error and Claims of Inferential Certainty* (2013) 12 Law, Prob. & Risk 115*passim*

Thompson et al., *Forensic Science Assessments: A Quality and Gap Analysis: Latent Fingerprint Examination*, American Assn. for Advancement of Science (2017)..... 39

Tobin & Blau, *Hypothesis Testing of the Critical Underlying Premise of Discernible Uniqueness in Firearms-Toolmarks Forensic Practice* (2013) 53 *Jurimetrics J.* 121 21, 22

Tobin et al., *Absence of Statistical and Scientific Ethos: The Common Denominator in Deficient Forensic Practices* (2017) 4 *Statistics & Pub. Policy* 1.....*passim*

United States v. Sutton, Order (D.C. Super. Ct., May 9, 2022, No. 2018 CF1 009709) 56

VanderPlas et al., *Firearms and Toolmark Error Rates. Submitted by the defense in Illinois v. Winfield* (2022) 15 CR 14066-01, <<https://digitalcommons.unl.edu/statisticsfacpub/159/>>*passim*

A167548

**IN THE COURT OF APPEAL
OF THE STATE OF CALIFORNIA**

FIRST APPELLATE DISTRICT, DIVISION THREE

PEOPLE OF THE STATE OF CALIFORNIA,
Plaintiff and Respondent,

v.

RAYMOND TIDD,
Defendant and Appellant.

**APPLICATION TO FILE BRIEF OF AMICI CURIAE
THE WILSON CENTER FOR SCIENCE AND JUSTICE
AT DUKE LAW SCHOOL AND A GROUP OF
ESTEEMED SCIENTISTS AND SCHOLARS
IN SUPPORT OF DEFENDANT**

The Wilson Center for Science and Justice at Duke Law School and an esteemed group of concerned scientists and scholars respectfully apply for leave to file the accompanying amici curiae brief in support of defendant and appellant Raymond Tidd pursuant to rules 8.200(c) and 8.360(f) of the California Rules of Court.

The Wilson Center for Science and Justice at Duke Law brings together faculty and students at Duke University in law, medicine, public policy, and arts and sciences to pursue research, policy, and education to improve criminal justice outcomes. Their work is non-partisan and evidence-informed. The Wilson Center

is devoted to identifying better ways for law enforcement to collect eyewitness, confession, forensic, and other evidence, and to enhance the ability of judges, lawyers, and jurors to understand evidence presented in court. The Center was founded and is led by Faculty Director, Professor Brandon L. Garrett, the L. Neil Williams, Jr., Professor of Law, where he has taught since 2018. Garrett was previously the Justice Thurgood Marshall Distinguished Professor of Law and White Burkett Miller Professor of Law and Public Affairs at the University of Virginia School of Law. His research and teaching interests include criminal procedure, wrongful convictions, habeas corpus, corporate crime, scientific evidence, civil rights, and constitutional law. One overriding concern of his work is to safeguard the accuracy and integrity of the criminal system, including through the use of reliable scientific and expert evidence.

The individual amici are law professors, scientists, and statisticians at some of America's leading universities who have devoted a substantial part of their teaching, work, research, and/or writing to criminal law and procedure, including issues pertaining to the accuracy and reliability of evidence and equity in criminal outcomes. Their work has been published by major university presses and in leading scientific and law journals. The amici consist of:

- **David L. Faigman**, Chancellor, Dean, and John F. Digardi Distinguished Professor of Law, University of California College of the Law, San Francisco;

- **Maria Cuellar**, Assistant Professor of Criminology and Statistics and Data Science, University of Pennsylvania;
- **Valena Elizabeth Beety**, Robert H. McKinney Professor of Law, Indiana University Maurer School of Law, and Senior Research Fellow, Academy for Justice, Arizona State University Sandra Day O'Connor College of Law;
- **Arturo Casadevall, MD, PhD**, Chair, of the Department of Molecular Microbiology & Immunology, Bloomberg Distinguished Professor, and Alfred and Jill Sommer Professor and Chair, Johns Hopkins School of Public Health;
- **M. Bonner Denton**, Galileo Professor of Chemistry, Professor of Geosciences, University of Arizona; and
- **Karen Kafadar**, Commonwealth Professor, Department of Statistics, University of Virginia.

No party or counsel for a party authored the proposed amici brief in whole or in part, or made any monetary contribution intended to fund the preparation or submission of the brief. No person or entity other than the amici curiae, their members, or their counsel in the pending appeal funded the preparation and submission of the proposed amici brief. (Cal. Rules of Court, rule 8.200(c)(3).)

May 14, 2024

Respectfully Submitted,

Complex Appellate Litigation Group LLP

Kelly A. Woodruff

Jennifer Teaford

By /s/ Kelly A. Woodruff

Kelly A. Woodruff

Attorneys for Amici Curiae

Document received by the CA 1st District Court of Appeal.

**BRIEF OF AMICI CURIAE THE WILSON CENTER FOR
SCIENCE AND JUSTICE AT DUKE LAW SCHOOL AND A
GROUP OF ESTEEMED SCIENTISTS AND SCHOLARS
IN SUPPORT OF DEFENDANT**

Introduction

“The fair administration of justice requires that science is accurately and effectively communicated to the fact finders” in judicial proceedings. (Dror & Scurich, *(Mis)use of Scientific Measurements in Forensic Science* (2020) 2 Forensic Sci. Internat.: Synergy 333, 333.) Firearm and toolmark (“FA/TM”) examiners purport to “match” spent ammunition to *one particular firearm* – not a type of firearm, make of firearm, or model of firearm – by visually comparing spent ammunition from an unknown source – typically recovered from a crime scene – with spent ammunition from a known source – typically a test fire from the firearm suspected to have been used in the crime. Expert testimony from FA/TM examiners is premised on the *unproven assumption* that each firearm leaves unique, accidental, individualized markings on spent ammunition.

Though routinely admitted by courts for decades, “[l]engthy judicial admissibility does not equate to scientific validity or reliability.” (Tobin et al., *Absence of Statistical and Scientific Ethos: The Common Denominator in Deficient Forensic Practices* (2017) 4 Statistics & Pub. Policy 1, 7, <<https://www.tandfonline.com/doi/pdf/10.1080/2330443X.2016.1270175>>.) And the scientific community at large has now sounded the alarm that the administration of justice is severely threatened by continued admission of FA/TM identification evidence. In particular, the

current scientific consensus is that the validity of FA/TM identification as a discipline lacks scientific support from well-designed, empirical studies, and the overwhelming majority of studies purporting to demonstrate the validity and reliability of FA/TM identification evidence do not in fact do so. As a result, courts are beginning to re-examine whether such evidence should be admissible in criminal trials and, if so, what the outer limitations on the testimony should be.

Given the now well-documented and widely recognized issues with FA/TM evidence, amici respectfully urge this court to take the opportunity presented by this case to provide urgently needed guidance to trial courts confronted with challenges to FA/TM evidence. To that end, amici respectfully urge this court to analyze the admissibility of FA/TM identification evidence under the test set forth in *People v. Kelly* (1976) 17 Cal.3d 24, superseded by statute on other grounds. Applying *Kelly*, this court should hold that, in criminal courts in the State of California, FA/TM examiners must be limited to making general group-level statements, without implying any similarity of individual or unique characteristics between the spent ammunition, because such individualized statements are not scientifically valid and are inherently unreliable.

Discussion

I. **The scientific community has rejected the notion that FA/TM identification evidence has been scientifically validated.**

The discipline of FA/TM identification “is based on the idea that the toolmarks produced by different firearms vary substantially enough (owing to variations in manufacture and use) to allow components of fired cartridges to be identified with *particular* firearms.” (President’s Council of Advisors on Science and Technology, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods* (2016) p. 104 (“PCAST Report”).) There are two steps to the comparison. First, the examiner compares the “class characteristics of the bullets and casings, which are features that are permanent and predetermined before manufacture.” (*Ibid.*) If the class characteristics (such as caliber) are different, the examiner can safely exclude or eliminate the items as coming from the same source. (*Ibid.*) But if the class characteristics are similar, they can safely be said to have come from the same *class* of guns. (*Ibid.*) Examiners go further, however, and examine whether there are “sufficient” similarities of individual characteristics between the spent ammunition to conclude that the exemplars came from the same source. (*Ibid.*)

It is this latter type of FA/TM identification evidence that has never been established to be scientifically valid. At the very least, in the last 10 to 15 years, there has been a seismic change in the scientific consensus on the validity of FA/TM identification as a forensic science, warranting consideration anew of whether

such evidence should be admissible in criminal trials. (*Kelly, supra*, 17 Cal.3d at pp. 30, 32 [if the admission of scientific evidence is affirmed in a published opinion based on a finding that the technique is generally accepted by the relevant scientific community, the precedent may control the admissibility of such evidence in future trials, “at least until new evidence is presented reflecting a *change in the attitude of the scientific community.*”], italics added; *People v. Venegas* (1998) 18 Cal.4th 47, 53 [previously admitted scientific evidence is no longer admissible in the face of evidence that the “prevailing scientific opinion *has materially changed.*”], italics added.)

A. Reports authored by committees of experts from the scientific community have concluded that FA/TM identification evidence has not been scientifically validated.

Beginning in 2008 and continuing through 2016, three separate panels of distinguished independent experts from the broader scientific and academic community (not limited to FA/TM) – convened by the National Academy of Sciences and the President’s Council of Advisors on Science and Technology (“PCAST”) – authored reports raising grave concerns about the scientific validity and reliability of FA/TM identification methods (as well as other “pattern-matching” fields).

- National Research Council of the National Academies, *Ballistic Imaging* (2008) (“2008 NRC Report”)
- National Research Council of the National Academies, *Strengthening Forensic Science in the United States: A Path Forward* (2009) (“2009 NRC Report”)

- PCAST Report, *supra*

Importantly, the committees authoring these reports consisted of independent scientists and scholars with expertise in physics, chemistry, biology, materials science, engineering, biostatistics, statistics, scientific methodology and study design, and medicine, as well as judges and lawyers – rather than toolmark examiners, whose financial and professional stake in the continued embrace of their discipline is apparent. Each of these committees heard testimony from forensic scientists, reviewed available journal articles and studies involving toolmark examination, and read every article or study submitted by members of the forensic community. As such, these bodies were uniquely qualified to determine whether this field is based on valid, reliable scientific principles or methodologies.

The conclusions of these committees were uniform and devastating in their assessment of FA/TM identification evidence. The 2008 NRC Report found “[t]he validity of the fundamental assumptions of uniqueness and reproducibility of firearms-related toolmarks has not yet been fully demonstrated” and that “an examiner’s assessment of the quality and quantity of resulting toolmarks and the decision of what does or does not constitute a match comes down to a subjective determination based on intuition and experience.” (2008 NRC Report, *supra*, at pp. 81, 55.) The 2009 NRC Report reiterated these conclusions and found that FA/TM evidence “is introduced in criminal trials without any meaningful scientific validation, determination of

error rates, or reliability testing to explain the limits of the discipline.” (2009 NRC Report, *supra*, at pp. 107-108.)

The PCAST Report, issued seven years later, concluded that “work to date has not addressed the 2009 NRC report’s call to examine the fundamental scientific validity and reliability of many forensic methods used every day in courts.” (PCAST Report, *supra*, at p. 39.) With regard to FA/TM analysis, the PCAST Report found that “the current evidence [concerning FA/TM identification] falls short of the scientific criteria for foundational validity.” (*Id.* at p. 111.) “Foundational validity requires that a method has been subjected to *empirical* testing by multiple groups, under conditions appropriate to its intended use . . . [that] (a) demonstrate that the method is repeatable and reproducible and (b) provide valid estimates of the method’s accuracy.” (*Id.* at p. 5.)

The PCAST Report highlighted that toolmark comparison methods are “subjective” and therefore “require particularly careful scrutiny because their heavy reliance on human judgment means they are especially vulnerable to human error, inconsistency across examiners, and cognitive bias.” (PCAST Report, *supra*, at p. 5.) But, as the PCAST Report emphasizes, “neither experience, nor judgment, nor good professional practices (such as certification programs and accreditation programs, standardized protocols, proficiency testing, and codes of ethics) can substitute for actual *evidence* of foundational validity and reliability.” (PCAST Report, *supra*, at p. 6, italics added.) Further, “an expert’s expression of *confidence* based on

personal professional experience or expressions of *consensus* among practitioners about the accuracy of their field is no substitute for error rates estimated from relevant studies.” (*Ibid.*) It is essential that, for FA/TM methods, “establishing foundational validity based on empirical evidence is thus a *sine qua non*. Nothing can substitute for it.” (*Id.* at p. 6.) And the PCAST Report’s central finding was that “firearms analysis currently falls short of the criteria for foundational validity.” (*Id.* at p. 112.)

In short, these major reports show that FA/TM examination consists of applying a subjective methodology to an unvalidated assumption, and that the practice has not been subjected to *empirical* studies necessary to demonstrate that it produces reproducible, repeatable, and valid results.

B. Since the publication of the reports, the scientific community has continued to cast doubt upon FA/TM identification as a discipline.

Since publication of the reports, the scientific community has elaborated upon and amplified the criticisms of FA/TM identification methods. A litany of publications by authors spanning multiple disciplines have continued to call FA/TM analysis into question, including but not limited to:

- Cuellar et al., *Methodological Problems in Every Black-Box Study of Forensic Firearm Comparisons* (2024) p. 3, <<https://doi.org/10.48550/arXiv.2403.17248>> [“all studies in our literature search have methodological flaws that are so grave that they render the studies invalid, that is,

incapable of establishing scientific validity of the field of firearms examination.”];

- Faigman et al., *The Field of Firearms Forensics Is Flawed*, Scientific American (May 25, 2022) <<https://www.scientificamerican.com/article/the-field-of-firearms-forensics-is-flawed/>> [“Few studies of firearms exist and those that do indicate that examiners cannot reliably determine whether bullets or cartridges were fired by a particular gun.”];
- Dorfman & Valliant, *Inconclusives, Errors, and Error Rates in Forensic Firearms Analysis: Three Statistical Perspectives* (2022) 5 Forensic Sci. Internat.: Synergy, at p. 7, <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9254335/>> [“sound estimates of error rates are elusive and, in light of the over-abundance of inconclusives, potential error rates must be considered large”];
- Albright, *How to Make Better Forensic Decisions* (2022) 119 Proceedings of the Nat. Academy of Sciences, No. 38, p. 1, <<https://doi.org/10.1073/pnas.2206567119>> [“Although forensic pattern comparison disciplines have for decades played a valued role in criminal investigation and prosecution, the extremely high personal and societal costs of failure – the conviction of innocent people – has elicited calls for caution and for the development of better practices.”];

- Hofmann et al., *Treatment of Inconclusives in the AFTE Range of Conclusions* (2020) 19 Law, Prob. & Risk 317, 342-343 [“It seems clear from our assessment of the currently available studies that there is significant work to be done before we can confidently state an error rate associated with different components of firearms and toolmark analysis.”];
- Dror & Scurich, *(Mis)use of Scientific Measurements in Forensic Science*, *supra*, 2 Forensic Sci. Internat. at p. 337 [“error rate studies fall short, and produce inaccurate and misleading error rate estimates]; and
- Tobin et al., *Absence of Statistical and Scientific Ethos*, *supra*, 4 Statistics & Pub. Policy at pp. 7, 8 [“[T]he foundational validity and reliability of the [FA/TM] practice has not been established and currently has no probative value . . . ,” and “[c]ritics have characterized the [FA/TM] practice as ‘pathological science’ and some insightful judges are in agreement, resulting in embryonic stages of judicial paradigm shift.”].

Most fundamentally, there is a notable lack of standard operating procedures and analysis protocol for FA/TM comparison methods. (Spiegelman & Tobin, *Analysis of Experiments in Forensic Firearms/Toolmarks Practice Offered as Support for Low Rates of Practice Error and Claims of Inferential Certainty* (2013) 12 Law, Prob. & Risk 115, 119, 124, 127.) “The forensic practice of firearms/toolmarks examination, with attendant opinions of specific source attributions, is totally

subjective The best guidance available for examiners is the ‘AFTE Theory of Identification’, but that guideline does not constitute a scientific protocol.” (*Id.* at p. 119; PCAST Report, *supra*, at p. 104 [the AFTE theory “is circular . . . stat[ing] that an examiner may conclude that two items have a common origin if their marks are in ‘sufficient agreement,’ where ‘sufficient agreement’ is defined as the examiner being convinced that the items are extremely unlikely to have a different origin.”]; Tobin et al., *Absence of Statistical and Scientific Ethos*, *supra*, 4 Statistics & Pub. Policy at p. 8 [“The guideline is circular (tautological), and is so vague that, once a target sample pool is narrowed by class characteristics (e.g., caliber), it basically allows opinions based on no objective criteria at all, but rather on ‘training and experience,’ a 100% subjective criterion.”].)

Indeed, “because discernible uniqueness has not been scientifically or forensically established, there exists no articulated protocol providing the foundation for assessing repeatability and reproducibility of experimental results in either purported validation studies or in actual casework.” (Tobin & Blau, *Hypothesis Testing of the Critical Underlying Premise of Discernible Uniqueness in Firearms-Toolmarks Forensic Practice* (2013) 53 *Jurimetrics J.* 121, 140, <<https://www.researchgate.net/publication/256040993>>.)

With no objective protocol, by what indicia or criteria do examiners assess uniqueness? And even if such metrics existed, by what rules should those criteria be applied? In other words, even if items are unique, what is a proper protocol for examiners to use to determine “same” or “different” and what

experiments should be performed to ascertain error rates as foundation for probabilistic expressions of certainty?

(Spiegelman & Tobin, *Analysis of Experiments, supra*, 12 Law, Prob. & Risk at p. 120; see also Cuellar et al., *Methodological Problems, supra*, at p. 29 [“This lack of conformance to a well-defined, clearly interpretable, firearms examination protocol in practice is a major roadblock to any research program that aims to evaluate scientific validity of a forensic method.”].)

This problem is fatal to most studies. “Without a detailed SOP, the practice of matching bullets and/or cartridge cases in studies claiming to support low rates of practice error do not have a sound scientific basis, in part because there are no measurable assurances that the experimental procedures in the reviewed papers mimic casework at all; in fact, some are starkly different from ‘real-world’ scenarios.” (Spiegelman & Tobin, *Analysis of Experiments, supra*, 12 Law, Prob. & Risk at p. 131.)

C. Existing FA/TM studies do not validate FA/TM identification as a discipline.

FA/TM evidence “is presented as ostensibly ‘scientific’ by innumerable purported ‘validation studies’ and practitioner claims based on ‘training and experience,’ and sold to courts without true scientific validation or scrutiny.” (Tobin & Blau, *Hypothesis Testing, supra*, 53 Jurimetrics J. at p. 123; see also Scurich et al., *Scientific Guidelines for Evaluating the Validity of Forensic Feature-Comparison Methods* (2023) 120 Proceedings of the Nat. Academy of Sciences, No. 41, at p. 7, <<https://www.pnas.org/doi/full/10.1073/pnas.2301843120>>.)

However, it cannot be overemphasized that existing FA/TM studies, including those conducted after the 2008 and 2009 NRC Reports and the PCAST Report, do *not* validate FA/TM identification as a discipline. (Cuellar et al., *Methodological Problems, supra*, at p. 3.)

For a scientific method to be valid, it must be:

(1) repeatable, i.e., in the context of FA/TM evidence, the examiner reaches the same conclusion when presented with the same evidence; (2) reproducible, i.e., different examiners reach the same conclusion when presented with the same evidence; and (3) accurate, i.e., the conclusion is correct. (PCAST Report, *supra*, at pp. 47-48.) Validation studies are intended to understand the range of conditions under which the method works as required, to understand how well it performs, and to identify conditions under which it is likely to fail. (*Ibid.*)

To date, none of the existing FA/TM studies, whether using a set-to-set design so harshly criticized in the PCAST Report (PCAST Report, *supra*, at p. 106) or sample-to-sample black-box studies, are capable of establishing the scientific validity of FA/TM identification. (Cuellar et al., *Methodological Problems, supra*, at p. 3.) Rather, though they purport to demonstrate the validity and reliability of FA/TM evidence, existing FA/TM studies have been poorly designed and were “developed by insular communities of nonscientist practitioners” – members of the Association of Firearm and Toolmark Examiners (“AFTE”), i.e., practicing FA/TM examiners rather than scientists – “who did not incorporate effective statistical methods.” (Tobin et al.,

Absence of Statistical and Scientific Ethos, *supra*, at pp. 1, 8, fn. 12; see also Scurich et al., *Scientific Guidelines*, *supra*, at p. 4 [existing studies of FA/TM identification “were exclusively created and conducted by FATM examiners with no specialized training in research design, statistics, and measurement”].) As a result, nearly all FA/TM studies suffer from fatal design flaws that render them incapable of providing meaningful, generalized results applicable to FA/TM evidence as a whole. (Hofmann et al., *Treatment of Inconclusives*, *supra*, 19 Law, Prob. & Risk at p. 333 [“We could not identify any studies that assess the error rates of bullet or toolmark examination in a manner that would produce reliable error rate estimates.”]; Spiegelman & Tobin, *Analysis of Experiments*, *supra*, 12 Law, Prob. & Risk 115, 131 [“From our review of the available studies and literature, we conclude that a statistical foundation supporting inferences of specific source attribution for firearms/toolmarks *does not exist.*”], italics added; VanderPlas et al., *Firearms and Toolmark Error Rates*. Submitted by the defense in *Illinois v. Winfield* (2022) 15 CR 14066-01, <<https://digitalcommons.unl.edu/statisticsfacpub/159/>> [“[A]ll of the studies we are aware of which are applicable to the state of firearms and toolmark examination as practiced in the United States at this time suffer from sampling and non-response bias that renders them *unreliable* for the purposes of establishing the science of firearms and toolmark examination as a reliable discipline.”], italics added; Cuellar et al., *Methodological Problems*, *supra*, at pp. 8-10 [identifying study design and analysis flaws the authors found in every one of the

28 existing validation studies they analyzed, some “so consequential that having even one such flaw renders a validation study scientifically unsound.”].)

The overwhelming opinion of the scientific community is that many of the existing studies of FA/TM identification methods “are poorly designed, with problems ranging from a complete inability to characterize the full error rate [citations] to the acknowledged inability of examiners to follow the instructions set out by the researchers [citation].” (VanderPlas et al., *Firearms and Toolmark Error Rates, supra*, 15 CR 14066-01, at p. 4.) “While these studies have been presented in court by FATM examiners as precisely the empirical support that science demands [citation], [the studies’] fundamental design flaws are now widely recognized as precluding their ability to measure a false positive error rate.” (Scurich et al., *Scientific Guidelines, supra*, at p. 8.)

1. *Methodological Failures of Set-Based Studies*

The scientific community agrees that existing “set-based” studies, where “examiners are asked to perform all pairwise comparisons within or between small sample sets” (PCAST Report, *supra*, at p. 106), suffer from one or more of the following design defects.

Multiple Known Exemplars. By including multiple exemplars from known sources, the examiner may use logical reasoning to reduce the number of comparisons made, rather than actually comparing an unknown source item with a known source item. (Hofmann et al., *Treatment of Inconclusives, supra*,

19 Law, Prob. & Risk at p. 333.) “[B]ecause all of the unknowns have a corresponding known, participants could use a deductive process to reduce the number of possible matches for subsequent comparisons.” (Scurich et al., *Scientific Guidelines, supra*, at p. 4.) For example, if the examiner is tasked with comparing an unknown source to 10 different known source exemplars, and the examiner matches the unknown source exemplar to the second source exemplar, there is no need to compare the unknown source to the remaining eight sources. (Hofmann et al., *Treatment of Inconclusives, supra*, 19 Law, Prob. & Risk at p. 333.) “This design ensures that it is not possible to count up the total number of different-source comparisons performed,” and an accurate error rate cannot be computed. (*Ibid.*)

Dependent Comparisons. In most existing studies, examiners are provided groups, or sets, of items from different sources and asked to make comparisons. But because the comparisons are not independent, it makes “statistical calculations of performance difficult, if not impossible.” (Scurich et al., *Scientific Guidelines, supra*, at p. 4.) By using set-to-set comparisons, there are no “true different-source comparisons (i.e., examiners were directly asked to evaluate two items fired by different guns), which is where a false positive error could theoretically happen.” (*Ibid.*) Without providing the examiner with some opportunities to directly compare two items from different sources, there is simply no way to assess the true false positive rate of error. (*Ibid.*)

To be able to accurately measure the false positive rate, the studies should contain true direct source comparisons, including comparisons of items from different sources. (Scurich et al., *Scientific Guidelines, supra*, at pp. 4-5.) A sample-to-sample study “gives the participant one ‘known’ item and one ‘unknown’ item and asks the participant to determine whether the unknown item came from the same source as the known item.” (*Id.* at p. 4.) The participant makes a judgment and then puts those items away. She is then presented with additional items to compare in the same fashion. In this way, each comparison is independent, which makes calculating performance metrics relatively straightforward.” (*Id.* at pp. 4-5.)

Closed-Set Studies. An additional problem with existing FA/TM studies is that they have largely used closed rather than open sets. In a closed-set study there is a “match” for every test sample in a set. (PCAST Report, *supra*, at pp. 108-109.) In contrast, open-set studies “remove the knowledge an examiner might have about whether a comparison is guaranteed to match one of the provided knowns.” (Hofmann et al., *Treatment of Inconclusives, supra*, 19 Law, Prob. & Risk at p. 321.) As such, a closed-set study is easier for the examiner than one in which a test sample might have no “match,” because in a closed set the examiner can merely look for the closest match in the set. Moreover, a closed set allows an examiner to come to correct conclusions simply by process of elimination, akin to a Sudoku puzzle. (PCAST Report, *supra*, at p. 106.)

Indeed, an examiner in a closed-set study will be able to come to at least some conclusions without having to perform any analysis. (See *United States v. Adams* (D. Or. 2020) 444 F. Supp. 3d 1248, 1264-1265 [error rates in tests using “partly closed sets” were “‘nearly 100-fold higher’ than from the closed-set tests”], quoting *United States v. Shipp* (E.D.N.Y. 2019) 422 F. Supp. 3d 762, 777-778, and citing PCAST Report, *supra*, at p. 109.) For example, in one study, participants were asked to evaluate spent ammunition and match it to 10 potential barrels; as a result, random guesses have a 1-in-10 chance of being correct. (Spiegelman & Tobin, *Analysis of Experiments*, *supra*, 12 Law, Prob. & Risk at p. 127.) Error rates in a closed-set study would not correlate with casework error rates, since in casework, there is by definition no guarantee that any of the evidence in fact “matches” the test samples. (PCAST Report, *supra*, at p. 108.) In short, “the closed-set design . . . is not appropriate for assessing scientific validity and measuring reliability.” (*Id.* at p. 109.)

2. *Methodological Failures of Sample-to-Sample Black-Box Studies*

PCAST called for additional black-box studies of FA/TM analysis to be conducted. (PCAST Report, *supra*, at pp. 5, 49, 106.) “In black-box studies, many examiners are presented with many independent comparison problems – typically, involving ‘questioned’ samples and one or more ‘known’ samples – and asked to declare whether the questioned samples came from the same source as one of the known samples.” (*Id.* at p. 49.) While the experimenter would know whether a comparison sample is a

match or not, the participant operates as a “black box” in which a subjective assessment of match or non-match is made, but the reasoning is unknown to the experimenter. Although a few black-box studies have been conducted, they too, suffer from numerous methodological flaws rendering them incapable of establishing the scientific validity of FA/TM comparison methods. (See, e.g., Cuellar et al., *Methodological Problems, supra*, at p. 3.)

Defining error. Proponents of FA/TM identification have asserted that it is impossible or impractical to conduct studies establishing a discipline-wide error rate in support of the validity of FA/TM methods and that criticism of existing studies and FA/TM evidence is thus overblown. (Spiegelman & Tobin, *Analysis of Experiments, supra*, 12 Law, Prob. & Risk at p. 117.) Not so. While it may be complex to establish a discipline-wide known or potential rate of error using well-designed empirical studies (*id.* at pp. 130-131), such difficulty should not prevent the FA/TM and scientific communities from attempting to do so – life and liberty are at stake. (See Kafadar, *The Critical Role of Statistics in Demonstrating the Reliability of Expert Evidence* (2018) 86 Fordham L.Rev. 1617, 1620 [“ ‘Scientific validity and reliability require that a method has been subjected to empirical testing . . . that provides valid estimates of how often the method reaches an incorrect conclusion.’ ”], quoting PCAST Report, *supra*, at p. 143.)

As discussed above, with FA/TM identification, the studies that claim low error rates have dubious and flawed methods, allowing purported FA/TM experts to make specious claims of

infallibility to courts and juries. (Faigman et al., *The Field of Firearms Forensics Is Flawed*, *supra*.) Most glaringly, FA/TM studies are fundamentally flawed in how they treat inconclusive results. Where an examiner cannot identify a match or definitively rule out a match, AFTE's theory of identification permits the examiner to report results as "inconclusive." (Dorfman & Valliant, *Inconclusives, Errors, and Error Rates*, *supra*, 5 Forensic Sci. Internat.: Synergy at p. 1.) Existing FA/TM studies "are not able to accurately estimate error rates because they fail to address inconclusive evidence and decisions." (Dror & Scurich, *(Mis)use of Scientific Measurements in Forensic Science*, *supra*, 2 Forensic Sci. Internat. at p. 336; see also Scurich et al., *Scientific Guidelines*, *supra*, at p. 5 ["[I]mproperly designed studies have masked a significant category of responses that could have real implications when the technique is applied in forensic contexts."].) Most existing studies count answers of "inconclusive" as correct or just exclude them. (*Ibid.*; Faigman et al., *The Field of Firearms Forensics Is Flawed*, *supra*.) However, "[s]cientifically, an inconclusive result has to be automatically *incorrect*: a comparison is either from a same-source or a different-source." (VanderPlas et al., *Firearms and Toolmark Error Rates*, *supra*, 15 CR 14066-01, at p. 9, italics added; see Dorfman & Valliant, 5 Forensic Sci. Internat.: Synergy at p. 2.) Although there is disagreement about how inconclusive results should be treated, what is indisputable is allowing an inconclusive result to be counted as correct grossly underestimates the reported error rates. (*Ibid.*)

Notably, the existing studies are not “test-blind,” and therefore the participants know they are being tested for a study. (Dorfman & Valliant, *Inconclusives, Errors, and Error Rates*, *supra*, 5 Forensic Sci. Internat.: Synergy at p. 6.) Consequently, there is a well-recognized risk of bias, and a tendency of the participants to call a result inconclusive than to risk either a false-negative or a false-positive. (*Ibid.*) This presents the serious risk that examiners doing a comparison as part of actual casework would conclude an identification, where in the non-blind study they reached an inconclusive result. (*Ibid.*)

Moreover, existing studies reveal that examiners are less likely to call the result “inconclusive” for inculpatory evidence (that is, they are more likely to claim a match than inconclusive) than they are for exculpatory evidence (that is, they are more likely to claim inconclusive than an exclusion). (Scurich et al., *Scientific Guidelines*, *supra*, at p. 5.) “This asymmetry is not contemplated in the AFTE protocol and therefore reveals a significant discrepancy between the AFTE theory and how it is effectuated by practicing FATM examiners.” (*Ibid.*)

In addition, many existing FA/TM error rate studies were designed such that test items are prescreened and removed from the study if they appear to be inconclusive in nature. (Dror & Scurich, *(Mis)use of Scientific Measurements in Forensic Science*, *supra*, 2 Forensic Sci. Internat. at p. 336.) Thus, test takers are effectively allowed to skip difficult questions – which are, by definition, more likely to yield wrong answers – by simply answering “inconclusive.” This method of scoring unquestionably

“misrepresents the reality of evidence in casework” and inherently, artificially depresses the true error rate. (*Id.* at p. 336; *id.* at p. 334 [“A priori presuming that inconclusive decisions can never be an error is problematic. If some examiners conclude an identification (or exclusion) whereas other examiners conclude as inconclusive, then at least some of the examiners are mistaken. . . . [I]t is obvious they cannot all be correct when examiners reach different conclusions on identical comparisons.”].)

Other studies simply exclude inconclusives from their analysis altogether, which also artificially deflates potential error rates and renders reported error rates uninformative. (Dorfman & Valliant, *Inconclusives, Errors, and Error Rates*, *supra*, 5 *Forensic Sci. Internat.: Synergy* at p. 2; Albright, *How to Make Better Forensic Decisions*, *supra*, 119 *Proceedings of the National Academy of Sciences*, No. 38, at p. 5.) At the extreme, these systems would allow an examiner to answer “inconclusive” on every test question and nevertheless receive a perfect score.

Participant Sampling. In order for studies to be generalizable to FA/TM analysis on the whole, the participants must be a representative sample from the population at issue – in the case of FA/TM, all qualified examiners in the United States. (VanderPlas et al., *Firearms and Toolmark Error Rates*, *supra*, 15 CR 14066-01, at pp. 5-6.) Existing FA/TM studies, however, have not randomly selected participants from the population, but have instead relied upon self-selected volunteers. (*Id.* at p. 6; see also Dorfman & Valliant, *Inconclusives, Errors, and Error Rates*,

supra, 5 Forensic Sci. Internat.: Synergy at p. 5.) This leads to inherent biases in the study population; for example, experienced examiners who may have lower error rates than the population of examiners on the whole may be more likely to volunteer “out of a sense of duty to the discipline.” (VanderPlas et al., *Firearms and Toolmark Error Rates*, *supra*, 15 CR 14066-01, at p. 5.) Without a representative sample of participants, a study can speak only to the error rate of those participants, not to the discipline as a whole.

Drop-Out or Attrition Rate. Research has found that if less than 5 percent of participants drop out, “there is little threat to the statistical validity of the study, but if more than 20% of participants drop out, the study’s validity is severely compromised.” (VanderPlas et al., *Firearms and Toolmark Error Rates*, *supra*, 15 CR 14066-01, at p. 4.) Studies with a high attrition rate exacerbate the bias from having the participants come primarily from the pool of examiners who are better at the task, with those finding the tasks too difficult to drop out. (*Ibid.*) “Analyzing only the results of the participants who chose to remain in the study can lead to biased statistical estimates and faulty conclusions.” (Scurich et al., *Scientific Guidelines*, *supra*, at p. 6.) Indeed, in the Ames II study (see, *post*, Section C.3), 69 percent of the participants who started the study dropped out without completing it. (*Ibid.*) Most existing FA/TM studies, however, do not report the drop-out rate for participants. (*Ibid.*)

Material Sampling. Like participant sampling, well-designed studies should also have a representative sample of the

ammunition and firearms that an examiner could encounter in a real-world setting. Existing FA/TM studies, however, largely concern a single type of firearm and/or a single type of ammunition. (VanderPlas et al., *Firearms and Toolmark Error Rates, supra*, 15 CR 14066-01, at p. 6.) In many cases, existing studies examine firearms of the same make and model manufactured closely in time. (*Ibid.*; see also Spiegelman & Tobin, *Analysis of Experiments, supra*, 12 Law, Prob. & Risk at p. 124 [noting that in one study oft-cited to courts by proponents of FA/TM, only three types of weapons were used, “two with sample size 1”], p. 127 [noting that in another oft-cited study, “one type of weapon” and “possibly two types of ammunition” were used].) Existing FA/TM studies are thus not representative of the discipline on the whole, and findings and error rates may not generalize well to FA/TM analysis on the whole.

* * *

In sum, existing “‘validation studies’” concerning FA/TM identification “typically result from,” among other things “statistical . . . deficiencies in the design and conduct of the experiments, and frequently lead to unjustified inferential extrapolation to universal” application to FA/TM. (Spiegelman & Tobin, *Analysis of Experiments, supra*, 12 Law, Prob. & Risk at p. 115.) In other words, “[t]he various ‘validation studies’ may be skilled experiments as forensic proficiency tests for specific examiners (test respondents) in controlled circumstances, but the same studies as currently exist are inappropriate for extrapolation to universal assumption or otherwise

representative of rates of error for the field of firearms/toolmarks examinations.” (*Ibid.*; see also VanderPlas et al., *Firearms and Toolmark Error Rates*, *supra*, 15 CR 14066-01, at p. 10.)

As a result, existing studies have “conclusions [that] far exceed statistically sound inferences from the experimental evidence.” (Spiegelman & Tobin, *Analysis of Experiments*, *supra*, 12 Law, Prob. & Risk at p. 130; see also *id.* at p. 118 [“Because most of the studies reviewed by the authors stray to varying degrees from the true scientific method, they frequently contain another characteristic of ‘pathological science’: wishful data interpretation.”].) Leading statisticians (and the scientific community on the whole) have therefore reached one inescapable conclusion: Multiple well-designed studies are still badly needed to demonstrate the general scientific validity and reliability of FA/TM evidence.

3. *Alarming Error Rates*

Two large black-box studies that are often cited by proponents of FA/TM evidence were undertaken by the Ames Laboratory, a Department of Energy national laboratory affiliated with Iowa State University, one in connection with the U.S. Department of Defense (“Ames I”) and the other in conjunction with the FBI (“Ames II”). (Baldwin et al., A Study of False-Positive and False-Negative Error Rates in Cartridge Case Comparisons (Apr. 7, 2014) <<https://www.ojp.gov/pdffiles1/nij/249874.pdf>> [Ames I]; Bajic et al., Report: A Validation Study of the Accuracy, Repeatability, and Reproducibility of Firearms Comparison (Oct. 7, 2020) [Ames II].) The studies, however,

demonstrate alarming error rates and do not support the validity of FA/TM evidence.

In Ames I, the research suffered from several of the same methodological issues discussed above. Even more troubling, though, is that the error rate was strikingly high. Although the researchers claimed a positive error rate of just over 1 percent, they did not include inconclusive responses or non-responses. (Baldwin et al., *A Study of False-Positive and False-Negative Error Rates*, *supra*, at p. 16.) Notably, examiners labeled almost 34 percent of the true different-source comparisons as inconclusive; that is, they failed to correctly label them as eliminations and thus exculpatory. (*Ibid.*)¹ In actuality, the study contained 2,180 true different-source comparisons, but only 1,421 (65 percent) were correctly identified as eliminations. (*Ibid.*) This is a far cry from the near certainty firearms examiners portray in their testimony.

The results of the second Ames study are even more troubling. Participants concluded the results of comparisons of bullets were inconclusive in 51 percent of all the comparisons and 65 percent of the true different-source comparisons. (Dorfman & Valliant, *Inconclusives, Errors, and Error Rates*, *supra*, 5 *Forensic Sci. Internat.: Synergy* at p. 4.) With respect to casings,

¹ In contrast, of the true same-source comparisons, only 1 percent were labeled inconclusive. (Baldwin et al., *A Study of False-Positive and False-Negative Error Rates*, *supra*, at p. 15.) This is consistent with the observation that examiners are far less likely to call inconclusive for same-source comparisons than they are for different-source comparisons. (Scurich et al., *Scientific Guidelines*, *supra*, at p. 5.)

participants reached inconclusive results in 42 percent of all comparisons and 51 percent of different-source comparisons. (*Ibid.*) The authors of Ames II reported inconclusives as “neutral non-errors,” which allowed them to report error rates of less than 0.8 percent. (*Id.* at pp. 4-5.) But if inconclusives are regarded as potential errors – and logically, they are (see, *ante*, Section I.C.2.) – the potential error rate for different-source bullets rises to more than 66 percent. (Dorfman & Valliant, 5 *Forensic Sci. Internat.: Synergy* at p. 5.)

Aside from the alarming potential error rate, Ames II highlighted the utter subjectivity and inability of FA/TM to demonstrate the repeatability and reproducibility of its methodology. Researchers sent the same exemplars back to the same examiner to re-examine and compare, and then to different examiners to conduct the same comparisons. (Faigman et al., *The Field of Firearms Forensics Is Flawed, supra.*) With respect to bullets, examiners were unable to repeat their own conclusions 21 percent of the time for known matches and 35.3 percent of known non-matches; the results for cartridge casings were similar, with examiners disagreeing with their own conclusions 24.4 percent of the time for known matches and 37.8 percent of the time for known non-matches. (Dorfman & Valliant, *Inconclusives, Errors, and Error Rates, supra*, 5 *Forensic Sci. Internat.: Synergy* at p. 6 [examiners re-examining the same exemplars a second time “disagree[d] with themselves between 20 and 40% of the time.”]; Dorfman & Valliant, *A Re-analysis of Repeatability and Reproducibility in the Ames-USDOE-FBI*

Study (2022) 9 Stats. & Pub. Policy 175, 177, <<https://doi.org/10.1080/2330443X.2022.2120137>>.) This amount of disagreement with an examiner’s own conclusions is substantial “and certainly gives us the impression of not very strong repeatability.” (Dorfman & Valliant, *A Re-analysis, supra*, 9 Stats. & Pub. Policy at p. 177.)

“Not surprisingly, reproducibility – the tendency of *different* examiners to come to the same conclusion on a given set of bullets – is lower than that for repeatability.” (Dorfman & Valliant, *A Re-analysis, supra*, 9 Stats. & Pub. Policy at p. 177.) For bullets, examiners were unable to repeat the conclusions of other examiners 32.2 percent of the time for known matches and 69.1 percent of the time for known non-matches; for cartridges, examiners disagreed with other examiners 36.4 percent of the time for known matches and 59.7 percent of the time for known non-matches. (*Ibid.*) These are astounding levels of disagreement and seriously undermine the reproducibility of the FA/TM analysis methods.

Indeed, “the results of the [Ames II] study suggest that, at least for the examiners represented in the study and for the guns and ammunition examined, repeatability and reproducibility are at best mediocre. The level of repeatability and reproducibility . . . would not appear to support the reliability of firearms examination.” (Dorfman & Valliant, *A Re-analysis, supra*, 9 Stats. & Pub. Policy at p. 178.)

D. Just because there may be a number of validation studies concluding that FA/TM identification is valid does not make it so when the studies themselves are fatally flawed.

Nor does the fact that there are several studies purporting to validate the practice of FA/TM comparison render the studies valid or justify the introduction of FA/TM evidence in legal proceedings. It is true there is a concept of “convergent validity,” which is “the possibility that various publications, each with distinct limitations when considered by itself, can reinforce each other and collectively support conclusions that would not be warranted on the basis of a single article.” (Thompson et al., *Forensic Science Assessments: A Quality and Gap Analysis: Latent Fingerprint Examination*, American Assn. for Advancement of Science (2017) at p. 94, <https://www.aaas.org/sites/default/files/s3fs-public/reports/Latent%2520Fingerprint%2520Report%2520FINAL%25209_14.pdf>.)

Convergent validity, however, is not properly applied to existing studies on FA/TM evidence because the literature does not consist of well-designed studies that, together, analyze FA/TM over a representative sample of examiners, firearms, ammunitions, and conditions. Rather, as discussed above, existing studies universally suffer from methodological flaws so grave that they are incapable of establishing the scientific validity of FA/TM examination of individual characteristics of ballistics. (Cuellar et al., *Methodological Problems*, *supra*, at p. 3.) Indeed, “because these flaws are shared across studies, it is not possible to combine results from the studies to overcome the flaws. The only way to demonstrate scientific validity is to

conduct future, adequately designed and analyzed studies.” (*Id.* at p. 9.)

As a result, no matter how many there are, these studies cannot support one another: Instead of filling in the gaps resulting from the limitations or flaws of each other, existing FA/TM studies accentuate and widen those same gaps. “A key lesson is that the sheer quantity of empirical studies is not a substitute for quality and that a bevy of improperly designed studies does little to provide empirical proof that the technique works.” (Scurich et al., *Scientific Guidelines, supra*, at p. 4.)

Worse, “[t]he adverse effect of such poor experimental design is exacerbated when the studies, such as those in the forensic domain, are insulated from public (scientific) scrutiny and become the basis for decisions that become entrenched in jurisprudence.” (Spiegelman & Tobin, *Analysis of Experiments, supra*, 12 Law, Prob. & Risk at p. 117.) A frightening example is the purported phenomenon of “polywater.” (*Ibid.*) During the 1960s and 1970s, the scientific community embraced the existence of water with much lower freezing points and much higher boiling points, which became known as polywater. (*Id.* at p. 118.) Several studies, including papers in prestigious scientific journals, purported to confirm the existence of polywater. (*Ibid.*) In fact, the studies were poorly designed and failed to account for critical factors, rendering them invalid. (*Ibid.*) It was not until the 1980s that it was established that “polywater” did not exist. (*Ibid.*) Spiegelman and Tobin emphasize how dangerous this example is:

In addition to the poor experimental design, it is notable that only through many good experiments of proper experimental design by skeptics was widespread acceptance of polywater overcome. Sometimes unsupported claims based on what has become known as ‘pathological science’ take on a life of their own and become entrenched in the popular culture or judicial community, requiring many years to dispel, even when widespread segments of the scientific community are involved. Such is the case of the public’s ‘CSI’ perception of exaggerated expressions of certainty for firearms/toolmarks ‘matches’ and the belief that forensic examiners can employ forensic techniques using combinations of non-unique geometry (lines) to identify specific sources of bullets and cartridge cases to the exclusion of all other possible sources in the world, including the vast pool of possible sources that the examiner(s) has never seen, with ‘near-zero error’.

(Ibid.)

Another frightening example is the acceptance by courts, for decades, of microscopic hair comparison analysis, “where an examiner uses a high powered microscope to view hair from a crime scene and compare it to a known hair sample.” (Nat. Assn. of Criminal Defense Lawyers, *Microscopic Hair Comparison Analysis* <<https://www.nacdl.org/Landing/Microscopic-Hair-Comparison-Analysis>> [as of May 10, 2024].) As with FA/TM identification, the examiner looks for similarities between the samples and, if there are “enough” characteristics that are the same, concludes the hairs are a “match.” *(Ibid.)* With the development and acceptance of DNA evidence, the FBI began to limit the use of hair microscopy only in conjunction with DNA evidence. *(Ibid.)* It was not until several defendants who had been convicted on the basis of MHCA were exonerated based on

DNA evidence, that the FBI undertook a massive review of prior convictions in collaboration with the Innocence Project and National Association of Criminal Defense Lawyers to determine whether any convictions were tainted by the flawed evidence. (*Ibid.*; see also *Dept. of Justice and FBI Joint Statement on Microscopic Hair Analysis* (Apr. 19, 2015) <<https://www.fbi.gov/news/press-releases/department-of-justice-and-fbi-joint-statement-on-microscopic-hair-analysis>> [as of May 10, 2024].)

In a report issued in December 2023 by the National Registry of Exonerations, the authors concluded that, as of the time of publication, “[a]t least 129 people have been falsely convicted based at least in part on MHCA.” (Cole et al., *Microscopic Hair Comparison Analysis and Convicting the Innocent*, The Nat. Registry of Exonerations (2023), at p. 7, <<https://n2t.net/ark:/88112/x2d909>>.) Fifteen of the exonerees had been sentenced to death! (*Ibid.*) And while the FBI attributed the wrongful convictions to “erroneous testimony” (such as testifying “that defendants *were* the source of hairs, when they should only have said they *might be* sources of hairs, cited baseless statistics, and misleadingly implied that their experience examining hair was a measure of the accuracy of their conclusions”), the authors concluded that “appropriate testimony” (such as the hairs are “similar” or “consistent with” the defendant’s hair or the defendant “could be” the source of the hair) “contributed to more false convictions than erroneous testimony did.” (*Id.* at pp. 8-10.)

At bottom, while proponents of FA/TM evidence may wish to invoke “convergent validity” to prop it up, it simply cannot do so. No matter how many studies purportedly validate FA/TM identification methods, it cannot be refuted that the *scientific* community does not agree it is a valid discipline. And simply because FA/TM evidence has been accepted by courts for decades does not mean courts should turn a blind eye to this new overwhelming scientific consensus.

II. The *Kelly* rule should govern the admissibility of FA/TM evidence.

FA/TM evidence should not be immune from scrutiny in California just because historically it has been admitted in the criminal courts. (*People v. Azcona* (2020) 58 Cal.App.5th 504, 523-529 (conc. opn. of Greenwood, J.)) “Lengthy judicial admissibility does not equate to scientific validity or reliability.” (Tobin et al., *Absence of Statistical and Scientific Ethos*, *supra*, at p. 7.) “In courtrooms across America, ‘scientific evidence’ used to imprison people for heinous crimes has been increasingly discredited. Blood-spatter patterns, arson analysis, bite-mark comparisons, even some fingerprint evidence have all turned out to be unreliable.” (Ebersole, *Old-School Hair Analysis is Junk Science. But it Still Keeps People Behind Bars*, *The Marshall Project*, The Marshall Project (Dec. 15, 2023), <<https://www.themarshallproject.org/2023/12/15/florida-death-penalty-hair-analysis-junk-science>>.) Indeed, according to the National Registry of Exonerations, a quarter of the 3,439 exonerations of convicted defendants “involved false or misleading forensic evidence.” (*Ibid.*)

In the interest of justice and in light of the current understanding of FA/TM evidence, amici respectfully request that this court hold that the admissibility of FA/TM evidence is subject to the *Kelly* test. (*People v. Jackson* (2016) 1 Cal.5th 269, 316 [appellate courts review de novo whether the admissibility of a technique is subject to *Kelly*].) And in light of the abundant new evidence “reflecting a change in the attitude of the scientific community,” amici request that the court conclude that FA/TM evidence suggesting individualization of ballistics markings is not admissible. (*Kelly, supra*, 17 Cal.3d at p. 32.)

Under *Kelly*, expert testimony based on the application of a new scientific technique is admissible in California if the technique is generally acceptable in the pertinent scientific community. (*Kelly, supra*, 17 Cal.3d at pp. 31-32.) Thus, “the task of determining reliability of the evolving technique” is assigned “to members of the scientific community from which the new method emerges.” (*Id.* at p. 31.) An advantage to this approach is to “promote a degree of uniformity of decision” with guidance from “substantial agreement and consensus in the scientific community.” (*Ibid.*) But the “primary advantage” of the *Kelly* approach is “its essentially conservative nature.” (*Ibid.*) Judicial restraint in admitting scientific evidence is “especially warranted when the identification technique is offered to identify the perpetrator of a crime” as FA/TM evidence routinely is. (*Id.* at p. 32.) Moreover, judicial caution is warranted given that “[l]ay jurors tend to give considerable weight to ‘scientific’ evidence when presented by ‘experts’ with impressive

credentials” and that there is “ ‘misleading aura of certainty which often envelops a new scientific process, obscuring its currently experimental nature.’ ” (*Id.* at pp. 31-32.)

As noted in the 2009 NRC Report, “[i]n a number of forensic science disciplines [including FA/TM analysis], forensic science professionals have yet to establish either the validity of their approach or the accuracy of their conclusions.” (2009 NRC Report, *supra*, at p. 53.) Although FA/TM analysis may not be a “new” scientific method under *Kelly*, to “hew more closely to the underlying goals of forensic science, namely identifying the guilty and exonerating the innocent, [citation] courts must take a more proactive approach to preventing the admission of flawed science.” (Epps & Todorow, *Refried Forensics: Screening Expert Testimony in Criminal Cases Through Frye Plus Reliability* (2018) 48 Seton Hall. L.Rev. 1161, 1164.) Applying *Kelly*, this court should conclude that FA/TM evidence is not generally accepted in the scientific community and restrict its admission, because “once an expert is able to present scientifically unsound evidence to the jury, it is often too late for the defendant to recover, even if that evidence is ‘palpably wrong.’ ” (*Id.* at p. 1168; see also PCAST Report, *supra*, at p. 19 [FA/TM evidence should not be admitted in courts until there are “adequate empirical studies and/or statistical models to provide meaningful information about the accuracy of [the FA/TM] comparison method.”].)

In *People v. Cowan* (2010) 50 Cal.4th 401, the California Supreme Court held that *Kelly* did not govern the admissibility of

a specific FA/TM technique (dynamic ballistics comparison). (*Id.* at pp. 468-471.) The court, however, did not determine whether dynamic ballistics comparison was a new scientific technique because defendant had not made that argument at trial. (*Id.* at p. 471.) Instead, the court held that *Kelly* did not govern because dynamic ballistics comparison was not a matter so beyond common understanding that lay jurors could not give it proper weight. (*Ibid.*)

Cowan should not prevent this court from holding that *Kelly* governs the admissibility of FA/TM evidence. In *Cowan*, the court did not consider the current state of FA/TM science (most notably, the landmark PCAST Report was published after *Cowan*). Accordingly, the court did not have the opportunity to consider the current attitude of the scientific community with respect to the lack of validity and reliability of FA/TM evidence. Indeed, absent this context and given the historical acceptance of FA/TM evidence, the court fairly assumed the soundness of the discipline. (See *Kelly, supra*, 17 Cal.3d at pp. 31-32 [precedent established by acceptance of evidence controls until evidence is presented that reflects a change in the attitude of the scientific community].)

Further, the court in *Cowan* did not have an opportunity to consider that the current scientific literature demonstrates that FA/TM evidence is *not* within the common understanding of lay jurors. While a lay juror is “capable of visually comparing the similarity of the marks,” as the literature shows, FA/TM expert testimony goes “far beyond that.” (*Azcona, supra*, 58 Cal.App.5th

at p. 526 (conc. opn. of Greenwood, J.).) Indeed, three panels of experts and multiple independent scientists have reviewed all of the studies up to the most recent and detail all their flaws and the commonality of those flaws. Nevertheless, FA/TM examiners continue to make statements expressing certainty or high probability, unaware of or unbowed by the current literature. FA/TM evidence should not continue to escape *Kelly* scrutiny given that it has been revealed to be essentially “experimental” in nature despite its longstanding and “misleading aura of certainty.” (*Kelly, supra*, 17 Cal.3d at pp. 31-32.)

Justice Mary J. Greenwood’s sharply rendered concurrence in *Azcona* shows the urgent need to consider the admissibility of FA/TM evidence anew. Assuming *Kelly* scrutiny applied, Justice Greenwood discussed the 2008 NRC Report, the 2009 NRC Report, and the PCAST Report at length. (*Azcona, supra*, 58 Cal.App.5th at pp. 519-529.) She concluded that “[g]iven the provenance of the NRC and PCAST reports and the detailed analyses they provide on the current state of the science, I would conclude they reflect a material change in the attitude of the scientific community regarding the validity of firearms toolmark analysis.” (*Id.* at p. 527.) And she found the trial court erred because it appeared “the trial court never even considered whether the attitude of the scientific community had changed over time . . . [and] relied solely on the claim that such evidence has historically been admitted.” (*Id.* at p. 528.)

Justice Greenwood also pointed out that “[t]here is no question that trial courts will be faced with many challenges to

previously accepted expert testimony on forensic techniques that may prove, with the swift advance of science, to be less sound than previously supposed.” (*Azcona, supra*, 58 Cal.App.5th at p. 529; see *ante*, Section I.C.0 [discussion of “polywater” and microscopic hair comparison analysis].)

It is unsurprising then that in 2021, a trial court in Yolo County granted a motion to limit FA/TM testimony regarding the similarities of the markings on bullets. (*People v. Auimatagi*, Order After Hearing on Motion to Exclude Testimony of Prosecution Witness Alex Taflya (Super. Ct. Yolo County, June 7, 2021, No. 19-4995) p. 1.) The trial court allowed the expert witness “to describe the class characteristics he observed on the bullets” and “the theory of toolmark analysis and how firearms leave markings on bullets and shell casings.” (*Id.* at pp. 1-2.) The expert was not, however, allowed to testify that he could not “exclude or eliminate the bullets as coming from different guns” or “permitted to describe any greater level of scientific certainty than the bullets may or may not have come from the same gun.” (*Id.* at p. 2.) Citing *Azcona*, the court found that “the methodology[] and the questionable foundational validity and analysis of error rates in the studies[] do not support permitting opinions of a *specific* level of scientific certainty that bullets were fired from the same gun.” (*Ibid.*)

Given the seismic shift reflected in the literature and the need to provide judicial guidance and to promote uniformity (as between defendants charged in San Francisco County and those charged in Yolo County), the court should take this opportunity

to apply *Kelly*, review the pertinent scientific literature, and conclude that FA/TM identification evidence is no longer admissible as reliable evidence.

III. This Court Should Preclude FA/TM examiners from testifying about the presence of individual or unique characteristics of spent ammunition that imply a “match.”

In light of the complete absence of studies “establishing scientific validity of the field of firearms examination . . . statements about the common origin of bullets or cartridge cases that are based on examination of ‘individual’ characteristics” should not be permitted in criminal trials. (Cuellar et al., *Methodological Problems, supra*, at p. 3.) Rather, FA/TM evidence should be limited to that for which there are *objective* measurable standards, namely the consistency of class characteristics between items of interest.

An example of such testimony might be “the bullet that killed the victim is consistent with having been shot from a 38 caliber Smith & Wesson, and there are approximately 10,000 such guns in circulation in the Southwest United States. Any one of those 10,000 guns could have left similar striae found on the bullet.”

(Scurich et al., *Scientific Guidelines, supra*, at p. 8.)

Any further testimony that includes, in any form, statements about the similarity or presence of “individual” characteristics or “unique” characteristics is not scientifically defensible and should be excluded.

A. Allowing any expert testimony about the similarity or presence of individual or unique characteristics should be prohibited.

In addition to the overwhelming evidence that existing FA/TM studies do *not* validate FA/TM identification as a discipline, statisticians criticize many common pattern matching forensic disciplines such as FA/TM and fingerprint comparison as allowing examiners to testify to statistically inappropriate conclusions. (E.g., American Statistical Assn., Position on Statistical Statements for Forensic Evidence (Jan. 2, 2019) p. 2, <<https://www.amstat.org/asa/files/pdfs/POL-ForensicScience.pdf>> (“ASA Report”); Spiegelman & Tobin, *Analysis of Experiments*, *supra*, 12 Law, Prob. & Risk at p. 115.) Opining that two pieces of ammunition were fired from the same gun, for example, “requires knowledge of how common or rare the association is, based on empirical data linked to the case at hand.” (*Id.* at p. 2.)

The American Statistical Association recommends that forensic witnesses in comparative disciplines who are allowed to couch their conclusions with some level of statistical certainty – or implied certainty – should also be required to include in their testimony explanations of the features compared, the process used to determine the level of similarity/dissimilarity, relative frequencies of the features compared, relative frequencies of combinations of features compared, quantitative statements of confidence, and the sensitivity of the methodology. (ASA Report, *supra*, at pp. 4-5.) *But no such data currently exists with regard to FA/TM evidence.*

Despite these failings, FA/TM evidence is often presented in broad – and seriously misleading – categorical strokes: there is a definite identification (i.e., a “match”); it is “more likely than not” the spent ammunition came from the suspect firearm; the markings on the spent ammunition from defendant’s firearm were “consistent with” or “similar to” the unknown item; there is some agreement between individual characteristics of the exemplars, suggesting they came from the same firearm. (Garrett et al., *Mock Jurors’ Evaluation of Firearm Examiner Testimony* (2020) 44 Law & Hum. Behav. 412, 413, 415.) On the surface, “is consistent with” being fired or “likely” fired from the suspect firearm may seem to address the serious concerns with examiners’ categorical but statistically unsound statements of a “match.” But such statements inherently imply certainty – or at least a very high probability. (See ASA Report at p. 2; Albright, *How to Make Better Forensic Decisions*, *supra*, 119 Proceedings of the Nat. Academy of Sciences, No. 38, at p. 9 [examiner statements “foster an illusion of certainty”].) And, as described above, there is no basis in existing data to suggest that a FA/TM examiner’s conclusion can be made with any degree of probability or certainty.

Moreover, opinions about consistency or similarity are both misleading to jurors and insufficient to prevent confusion and overemphasis by jurors of FA/TM evidence. Indeed, in the National Registry of Exonerations’ analysis of MHCA cases, the authors found that “[s]eemingly innocuous statements like ‘similar’ and ‘consistent with’ contributed to false convictions at

least as often as did egregious misstatements” of certainty. (Cole et al., *Microscopic Hair Comparison Analysis*, *supra*, at p. 9, 14-16.)

First, “is consistent with” being fired or “likely” fired from the suspect firearm is inherently ambiguous: Different people may understand it differently. Some people may understand the phrases to connote that the ammunition at issue definitely came from the same source as the test ammunition. Others may understand it to mean that the two pieces of ammunition are indistinguishable. Yet others may understand it to mean that the two pieces of ammunition have some similarities – to varying extents. As the American Statistical Association noted, “[t]o evaluate the weight of any set of observations made on questioned and control samples,” jurors need to be able “to relate the probability of making these observations if the samples came from the same source to the probability of making these observations if the questioned sample came from another” potential source. (ASA Report, *supra*, at p. 2.) Language like “is consistent with” or “likely” gives jurors no ability to make that evaluation.

Second, this type of phrase fails to acknowledge that the conclusion is subjective and heavily reliant on the “personal impressions” of the FA/TM examiner. Even if cross-examined and asked to justify their level of certainty, FA/TM examiners will likely cite their years of experience and their professional judgment. But “although training and experience are important in applying valid techniques, practitioners’ subjective opinions

are not sufficient for establishing the uncertainty in measurements or inferences.” (ASA Report, *supra*, at p. 2.) This is particularly problematic because jurors are persuaded by, for example, clinicians who testify as “experts,” even when their testimony is based solely on personal experience and has no support by reliable scientific studies. (See Krauss & Sales, *The Effects of Clinical and Scientific Expert Testimony on Juror Decision Making in Capital Sentencing* (2001) 7 Psych., Pub. Policy & Law 267, 272.) This is because jurors generally presume that “scientific” evidence presented to them via “expert” testimony has been thoroughly vetted and screened by the court before its presentation. (See Schweitzer & Saks, *The Gatekeeper Effect: The Impact of Judges’ Admissibility Decisions on the Persuasiveness of Expert Testimony* (2009) 15 Psych., Pub. Policy & Law 1, 4.)

Third, conclusions like “is consistent with” or “likely” connote a statistical basis for the conclusion. The phrases inherently suggest that the FA/TM examiner knows how common certain markings are, how rare other markings are, etc.: He or she could not render such an opinion and make such a comparison without information about the relative frequencies of the characteristics at issue. But no such relative frequency data has been established for FA/TM evidence. (See, *ante*, Section I.B.) Compounding this problem, examiners often lack a sufficient understanding of statistics, statistical models, and the statistical issues inherent in the evaluation of forensic evidence.

(Kafadar, *The Critical Role of Statistics, supra*, 86 Fordham L.Rev. at pp. 1634-1635.)

At a minimum, then, the court should hold that FA/TM examiners must be limited to making general group-level statements about the class characteristics of the compared items. (Scurich et al., *Scientific Guidelines, supra*, at p. 8.) Allowing experts to testify to group data of this sort is akin to allowing someone to testify that the perpetrator drove a red Ford Mustang. We do not need to know exactly how many such cars are in the general area, but we know that there are many. The fact that the defendant drives a red Ford Mustang is thus relevant, but it is not determinative.

B. There is a notable movement toward limiting the presentation of FA/TM evidence, but it does not go far enough.

There is a nascent nationwide shift toward limiting the admissibility and presentation of FA/TM evidence. In *Williams v. United States* (D.C.Ct.App. 2019) 210 A.3d 734, the D.C. Court of Appeals found that the trial court committed plain error when it allowed a FA/TM examiner's testimony that the ammunition in question "had all been fired by the same gun," and "fired from" the specific gun recovered in connection with the case. (*Id.* at p. 738.) The court unequivocally noted that "the empirical foundation does not currently exist to permit these examiners to opine with certainty that a specific bullet can be matched to a specific gun," and that "these conclusions are simply unreliable." (*Id.* at p. 742.)

In *United States v. Adams*, *supra*, 444 F. Supp. 3d 1248, the district court found that the results of FA/TM comparison analysis “cannot be random” but that “it is not clear that those results are the product of a *scientific* inquiry.” (*Id.* at p. 1266.) The court noted that the expert did not explain how or why he reached his comparison conclusion in “any quantifiable, replicable way.” (*Ibid.*) Accordingly, the court excluded testimony about “conclusions relating to whether the shell casings [seized from the shooting scene] matched” the gun recovered from defendant’s home. (*Id.* at p. 1267.)

And in *United States v. Tibbs* (D.C. Super. Ct., Sept. 5, 2019, No. 2016-CF1-19431) 2019 WL 4359486, the district court judge held that the government’s FA/TM examiner

may testify that based on his examination, the recovered firearm *cannot be excluded as the source of the cartridge casing found* on the scene of the alleged shooting. . . . Any statements by the expert involving more certainty regarding the relationship between a casing and a firearm would stray into territory not presently supported by reliable principles and methods.

(*Id.* at *23.) Among other things, the judge recognized that “threshold design issues” with existing FA/TM studies “surely impact the validity of these studies’ conclusions and limit their utility to some extent.” (*Id.* at *14.)

This court should take this opportunity to hold that, at most, FA/TM examiners may opine about consistent class characteristics between the ammunition at issue, but not any individualized statements that would imply the exemplars are a match.

C. Even limiting FA/TM testimony to anodyne “consistent with” or “similar” language does not sufficiently protect against the overstatement of the probative value of FA/TM evidence.

Despite the lack of scientific or statistical validity to FA/TM identification evidence, proponents of FA/TM techniques have continued to successfully push courts to allow FA/TM testimony to go beyond the consistency of class characteristics between the ballistic evidence. For example, in an order in *United States v. Sutton* (D.C. Super. Ct., May 9, 2022, No. 2018 CF1 009709) (attached as Exhibit 1), the district court held that the government’s FA/TM examiner could not state “without any qualifications or limitations that the ammunition at issue was fired from the same firearm,” but allowed the examiner to opine “that the ammunition at issue is consistent with being fired from the same firearm.” (*Id.* at p. 5.)

But a recent analysis of wrongful convictions from microscopic hair comparisons established that even “anodyne ‘consistent with’ statements” “appear[] to have done an equally good job at contributing to false convictions” as did egregious overstatements about the evidence. (Cole et al., *Microscopic Hair Comparison Analysis, supra*, at pp. 79-80.) Such “statements are appealing to expert disciplines that lack foundational data because they mean almost nothing while appearing to mean at least something and perhaps anything. The problems exposed with MHCA should be a warning about these kinds of statements in any expert domain.” (*Id.* at p. 79.) Notably, from this study, the authors concluded that limiting expert forensic testimony to

modest conclusions does not change the outcome. “It is difficult to avoid the conclusion that what actually mattered to juries was the mere fact that the evidence was called ‘forensic’ and that it appeared to tend to implicate the defendant in some way.” (*Id.* at p. 80.)

Another recent study highlights how FA/TM identification conclusions connoting any degree of statistical certainty – i.e., anything other than inconclusive – is confusing to jurors and fails to adequately protect against the overstatement of the probative value of FA/TM evidence. (Garrett et al., *Mock Jurors’ Evaluation, supra*, 44 Law & Hum. Behav. 412.) Mock jurors were presented with seven different potential conclusions by a FA/TM examiner. A conclusion of “inconclusive” was defined as lacking sufficient agreement to determine the bullets/casings came from the same gun. (*Id.* at p. 415.) On the positive identification side, the conclusions ranged from “cannot be excluded” to absolute certainty and included a response using “consistent with” language. (*Ibid.*) The study found that jurors were significantly more likely to convict on the basis of the FA/TM evidence alone under any of the six positive opinions than under the inconclusive opinion. (*Id.* at pp. 416-417.) Indeed, “[r]egardless of the specific language used, the odds of conviction increased by 5 to 6 when a match was declared relative to when an inconclusive was declared.” (*Id.* at p. 416.) Even the “cannot be excluded” result, “which had the smallest relative effect size, increased the odds of conviction by 2.5.” (*Ibid.*) There was little practical change in results when participants were shown cross-

examination of the FA/TM examiner. (*Id.* at pp. 419-422; see also Levett & Kovera, *The Effectiveness of Opposing Expert Witnesses for Educating Jurors About Unreliable Expert Evidence* (2008) 32 *Law & Hum. Behav.* 363 [even when cross-examination is well-constructed and exposes considerable flaws, it has little to no effect on juries].)

At bottom, “adopting types of modified conclusion language . . . did not affect guilty verdicts [and] . . . many judicial and prosecution-driven interventions to limit conclusion language for firearms testimony are not likely to be effective.” (Garrett et al., *Mock Jurors’ Evaluation*, *supra*, 44 *Law & Hum. Behav.* at p. 422; see also Cole et al., *Microscopic Hair Comparison Analysis*, *supra*, at pp. 9, 66, 79-80 [testimony that hairs are “consistent with,” similar, or “could be” a match may result in wrongful convictions as often as erroneous testimony of an actual match].)

In the present case, the expert testified the firearm he examined was “likely” the one used in the shooting. (RB 14.) He also asserted that “[m]y expectation to find another firearm with the same signature or same fingerprint is remote or really small.” (RB 15.)

This court should advance the interests of justice and hold that FA/TM evidence concerning individual characteristics is inadmissible because it is unreliable. At most, FA/TM evidence should go no farther than what the data actually tend to support: testimony that the items shared similar group-level or class characteristics. Here, the testimony went well beyond that presentation.

Conclusion

For the foregoing reasons, amici respectfully urge this court to hold that FA/TM identification testimony is inadmissible because it is not scientifically validated and thus not generally accepted by the scientific community. At most, FA/TM examiners should be permitted to testify that the bullets or casings have consistent class characteristics with the spent ammunition at issue, without implying any similarity of individual characteristics between the items.

May 14, 2024

Respectfully Submitted,

Complex Appellate Litigation Group LLP

Kelly A. Woodruff

Jennifer Teaford

By /s/ Kelly A. Woodruff

Kelly A. Woodruff

Attorneys for Amici Curiae

Document received by the CA 1st District Court of Appeal.

Certificate of Word Count

(Cal. Rules of Court, rules 8.204(c), 8.360(b)(1))

The text of this brief consists of 11,354 words as counted by the Microsoft Word program used to generate this brief.

Dated: May 14, 2024

/s/ Kelly A. Woodruff
Kelly A. Woodruff

Proof of Service

I, Stacey Schiager, declare as follows:

I am employed in the County of San Francisco, State of California, am over the age of eighteen years, and am not a party to this action. My business address is 96 Jessie Street, San Francisco, CA 94105. On May 14, 2024, I mailed the following document:

Application to File Amici Curiae Brief and Brief of Amici Curiae the Wilson Center for Science and Justice at Duke Law School and a Group of Esteemed Scientists and Scholars in Support of Defendant

I enclosed a copy of the document identified above in an envelope and deposited the sealed envelope with the U.S. Postal Service, with the postage fully prepaid. The envelope was addressed as follows:

Clerk for Hon. Stephen Murphy
San Francisco County Superior Court
400 McAllister Street, Dept. 204
San Francisco, CA 94102
Trial Judge

Additionally, on May 14, 2024, I caused the above-identified document to be electronically served on all parties, pursuant to the parties’ agreement to electronic service, and the California Supreme Court via TrueFiling, which will submit a separate proof of service.

I declare under penalty of perjury under the laws of the State of California that the foregoing is true and correct. Executed on May 14, 2024.

/s/ Stacey Schiager
Stacey Schiager

Document received by the CA 1st District Court of Appeal.

Exhibit 1

SUPERIOR COURT OF THE DISTRICT OF COLUMBIA

Criminal Division

UNITED STATES OF AMERICA	:	
	:	Case No. 2018 CF1 009709
v.	:	
	:	
KAEVON SUTTON	:	Hon. Robert Okun

ORDER

Defendant’s Motion to Exclude Expert Testimony in Firearms Identification (“Defendant’s Motion to Exclude”) is pending before the Court. For the reasons set forth below, Defendant’s Motion to Exclude will be granted in part.

RELEVANT PROCEDURAL HISTORY

On February 2, 2021, Defendant filed his Motion to Exclude, requesting the following limitations on the Government’s presentation of ballistics comparison evidence in this case: 1) the examiner be precluded from testifying to the source attribution opinion that, based on pattern-matching, casings and bullets were identified as having been fired from the same weapon; and 2) the examiner’s discussion of markings on the ammunition be limited to class characteristics. In support of this request, Defendant asserted that there was a lack of scientific support for the Government’s claim that a firearms examiner can reliably make a source determination – i.e., that the examiner can match markings on expelled casings and bullets to a single firearm. Defendant cited the findings of three reports – 1) the 2008 Ballistics Imaging Report by the National Research Council (“2008 NRC Report”); 2) the 2009 Forensic Science Report by the National Research Council (“2009 NRC Report”); and 3) the 2016 Forensic Science Report by the President’s Council of Advisors on Science and Technology (“2016 PCAST Report”) – in support of his argument, as well as Judge Edelman’s opinion in *United States v. Tibbs*, 2016 CF1 19431, cited at 2019 D.C. Super. LEXIS 9 (Sept. 5, 2019).

On March 12, 2021, the Government filed its Opposition to Defendant's Motion to Exclude. In its Opposition, the Government asserted that studies issued subsequent to the 2016 PCAST Report undercut the conclusion reached by Judge Edelman in *Tibbs*, and cited Judge Contreras's opinion in *United States v. Harris*, 502 F.Supp.3d 28 (D.D.C. 2020), in support of its argument that a firearms examiner could reliably testify that ammunition was fired from the same firearm if there were sufficient levels of agreement among the individual characteristics of the firearm. The Government noted several limitations in any expert testimony it would offer – namely, the expert would not use unqualified terms such as “match,” would not state his expert opinion with any level of scientific certainty, and would not render his opinion “to the exclusion of all other firearms” or use the phrase “to a reasonable degree of scientific certainty.”

On March 31, 2022, Defendant filed his Amended Reply to the Government's Opposition. In his Amended Reply, Defendant argued that the Government ignored the growing number of judicial opinions, including those from this jurisdiction, that placed greater limits on firearm testimony than those proposed by the Government. Defendant also asserted that the scientific landscape had not changed in the Government's favor since the 2016 PCAST report was issued and that the weight of the evidence supported Defendant's proposed limitations. In addition, Defendant claimed that the Government's firearms expert (Jay Stuart) did not reliably apply firearms identification methodology in this case.

Finally, on May 3, 2022, the Government filed its Sur-Reply to Defendant's Motion to Exclude. In its Sur-Reply, the Government noted that another firearms expert (Rick Wyant) had examined the relevant ammunition and had reached almost an almost identical conclusion as Mr. Stuart, and that another examiner (Aaron Brundenell) verified Mr. Wyant's conclusions. In

addition, the Government argued that Defendant’s proposed limitations exceeded those imposed by the Court of Appeals or by Judge Edelman in *Tibbs*.

RELEVANT LEGAL STANDARDS

The case law concerning the admission of testimony from a firearm or toolmark expert has evolved over the past several years, but the Court of Appeals’ two most recent cases are very informative. In *Gardner v. United States*, 140 A.3d 1172 (D.C. 2016), the Court of Appeals held that a firearm and toolmark expert “may not give an unqualified opinion, or testify with absolute or 100% certainty, that based on ballistics pattern comparison matching a fatal shot was fired from one firearm, to the exclusion of all other firearms.” *Id.* at 1177. The Court further noted that its holding was limited “in that it allows toolmark experts to offer an opinion that a bullet or shell casing was fired by a particular firearm, but it does not permit them to do so with absolute certainty,” and noted that it had doubts as to whether toolmark experts should be allowed to state their opinions “with a reasonable degree of certainty.” *Id.* at 1184, n.19.

In *Williams v. United States*, 210 A.3d 734 (D.C. 2019), the Court of Appeals reiterated that it is error to allow a firearm and toolmark examiner to “provide unqualified opinion testimony that purports to identify a specific bullet as having been fired by a specific gun via toolmark pattern matching.” *Id.* at 743. The Court of Appeals did not resolve the Government’s argument that *Gardner* only prohibited certainty statements and otherwise continued to authorize opinion testimony identifying a specific bullet as having been fired by a specific gun, because the examiner in that case had given a certainty statement. *Id.* at 741-42. However, the Court noted that the Government’s argument was “difficult to square” with the Court’s holding in *Gardner* that the trial court had erred by admitting the examiner’s unqualified opinion that a specific gun was the murder weapon. *Id.* at 739. The Court also noted that its opinion did “not limit firearms and

toolmark examiners from making other observations about the ballistics evidence recovered in a particular case,” because those observations were not at issue in the case. *Id.* at 743, n.19.

In addition, two relatively recent trial court decisions are informative. In *Tibbs*, Judge Edelman, after conducting an extensive evidentiary hearing, precluded the government from eliciting testimony identifying the recovered firearm “as the source of the recovered cartridge casing,” and instead ruled that the government’s expert must limit his testimony to a conclusion that “based on his examination of the evidence and the consistency of the class characteristics and microscopic toolmarks, the firearm cannot be excluded as the source of the casing.” 2019 D.C. Super LEXIS 9 at *3.

By contrast, in *Harris*, Judge Contreras disagreed with Judge Edelman’s analysis, after conducting an evidentiary hearing, and held that the Government’s firearms and toolmark expert could testify that casings were fired from the same firearm when all class characteristics were in agreement and “the quality and quantity of corresponding individual characteristics is such that the examiner would not expect to find that same combination of individual characteristics repeated in another source and has found insufficient disagreement of individual characteristics to conclude that they originated from different sources.” 502 F.Supp.3d at 45. Judge Contreras also noted with approval that the Government had agreed that its expert would not use terms such as “match” or state his opinion with any level of statistical or scientific certainty or to the “exclusion of all other firearms.” *Id.* at 44.

APPLICATION OF THE RELEVANT LEGAL STANDARDS

Although the issue is not free from doubt, the Court finds that the legal standards set forth above preclude the Government’s firearm expert from conclusively stating that the various pieces of ammunition that are at issue in this case were fired from the same firearm. While the Court

acknowledges that the Government's expert will not testify as to any level of scientific certainty or use unqualified terms such as "match," the Court nonetheless believes that opinion testimony that ammunition was fired from a particular firearm, without any qualifications or limitations on that opinion, is inconsistent with *Williams*, where the Court noted that the Government's argument on appeal was "difficult to square" with the Court's holding in *Gardner* that the trial court had erred by admitting the examiner's unqualified opinion that a specific gun was the murder weapon. 210 A.3d at 739. Thus, the Court will preclude the Government's expert from stating without any qualifications or limitations that the ammunition at issue was fired from the same firearm, and instead will limit the examiner's opinion to a conclusion that the ammunition at issue is consistent with being fired from the same firearm.

However, the Court does not agree with the other limitation proposed by Defendant – namely, that the examiner's testimony about the markings on the ammunition be limited to class characteristics. Such a limitation goes beyond the limitations set forth in *Williams* and *Gardner*, and even goes beyond the limitations imposed by Judge Edelman in *Tibbs*, because the Government's firearms expert in *Tibbs* was allowed to testify about his observations concerning individual ammunition characteristics such as the similarity in striations.

Therefore, it is this 9th day of May, 2022, hereby

ORDERED that Defendant's Motion to Exclude Expert Testimony in Firearms Identification is **GRANTED IN PART AND DENIED IN PART**; and it is

ORDERED that the Government may present expert testimony on firearms identification in a manner that is consistent with the limitations described above.



Judge Robert Okun

Copies to:

Terrance Austin, Joseph Wong
Defendant's Counsel

Jack Korba, Marybeth Manfred
United States Attorney's Office

Document received by the CA 1st District Court of Appeal.