SUPERIOR COURT OF THE DISTRICT OF COLUMBIA
Family Court—Juvenile Branch

In the Matter of	)	Docket No.
	)	
M W	)	Trial Date:
	)	
Respondent	)	Judge Andrea Hertzfeld

GOVERNMENT'S OPPOSITION TO RESPONDENT'S MOTION TO EXCLUDE OR IN THE ALTERNATIVE LIMIT FIREARM AND TOOLMARK EXAMINER TESTIMONY

The District of Columbia, by and through its attorney, the Office of the Attorney General, responds to the Respondent's Motion to Exclude or, in the Alternative, Limit Firearm and Toolmark Examiner Testimony and states the following:

PROCEDURAL BACKGROUND

On or about October 21, 2021, the respondent was charged with one count of First Degree Murder While Armed in violation of D.C. Code §§ 22-2101, 22-4502; one count Second Degree Murder While Armed in violation of D.C. Code § 22-2103, 22-4502; one count of Assault with Intent to Commit Murder While Armed in violation of D.C. Code § 22-401, 22-4502; one count of Assault with the Intent to Kill While Armed in violation of D.C. Code §§ 22-401, 22-4502; one count of Conspiracy to commit First Degree Murder While Armed in violation of D.C. Code § 22-1805(a); one count of Carrying a Pistol Without a License in violation of D.C. Code § 22-4504; one count of Possess of a Large Capacity Ammunition Feeder in violation of D.C. Code § 7-2506.01(b); one count of Possession of a Firearm without a Registration in violation of D.C. Code § 7-2506.01 (a); one count of Possession of a Ammunition without Valid Registration in violation of D.C. Code § 7-2506 (a)(3); one count of Tampering with Physical Evidence in violation of D.C. Code § 22-723; and one count of Unlawful

Discharge of a Firearm in violation of DC Code 22-4503.01 as part of the August 9, 2020 murder of Richard Bangura.

Probable cause was found on the charges of Second Degree Murder While Armed, Assault with Intent to Commit Murder While Armed, Assault with the Intent to Kill While Armed, Possession of a Firearm without a Registration, Possession of a Ammunition without Valid Registration, and Unlawful Discharge of a Firearm by the Honorable Tyrona DeWitt on October 21, 2021.

On March 1, 2022, the respondent filed a t Motion to Exclude or, in the Alternative, Limit Firearm and Toolmark Examiner Testimony. The government files this response to that motion.

STATEMENT OF FACTS

Upon information and belief, on Sunday, August 9, 2020, at around 7:00 p.m., MPD responded to a shooting at the intersection of [redacted], D.C. Officers located victim Richard Bangura ("decedent") in the driver's seat of a black four-door Lexus suffering from a gunshot wound to the head. Decedent was unconscious but still breathing. Decedent was transferred to Washington Hospital Center and admitted in critical condition. On August 16, 2020, decedent was pronounced dead at 8:42 a.m. by Dr. Jason Chang.

On August 17, 2020, Dr. Kristinza Geise, Office of the Chief Medical Examiner, conducted decedent's autopsy and reported that he suffered one gunshot wound to the head, and that the bullet entered and remained in his skull. The cause of death was ruled to be a gunshot to the head, manner of death homicide.

Detectives responded to the scene at the intersection of [redacted] and observed decedent's 2010 black sedan-style Lexus ("decedent's vehicle") where it came to

rest on [redacted], facing southbound towards [redacted] Street. Decedent's vehicle had multiple gunshot defects to the driver's side door, roof, and windshield and blood on the driver's seat. Several shell casings were recovered from the eastern portion of the intersection. ShotSpotter also recorded five rounds of gunshots at 6:58:26 pm at [redacted]

Detectives interviewed several witnesses and recovered surveillance footage in the vicinity of the offense and determined that just before the shooting, decedent was standing next to his vehicle in his driveway at the [redacted] A white four-door Toyota Camry with tinted windows drove by decedent. Moments later, decedent got in his vehicle and drove west towards [redacted].

The Department of Forensic Sciences (DFS) responded to the scene at 2[redacted]. and recovered six 9mm shell casings, latent print lifts from the decedent's vehicle exterior, and an apple iPhone from the front passenger compartment. DFS later entered the shell casings recovered into the National Integrated Ballistic Information Network database (NIBIN), which produced links to several Central Complaint Numbers (CCN), two of which are outlined below:

- August 4, 2020 – NIBIN LINK ADW Gun incident ([redacted]) Shots fired. [redacted] executed on Respondent [redacted].
- August 9, 2020 – NIBIN LINK Homicide [redacted]

On October 20, 2020, Prince George's County Police Department (PGPD) recovered a 9mm Polymer 80 PF940C "Ghost Gun" from an unoccupied overturned vehicle at Brinkley Road and Kildare Court in Fort Washington, Maryland. The firearm was test-fired by PGPD and OAG coordinated with PGPD, MPD, and the Alcohol, Tobacco, and Firearms Bureau (ATF) to have this firearm retested to confirm the NIBIN hits. The NIBIN connection to the firearm from these shell casings was confirmed in relevant part below:

> *The Exhibit 1 firearm was examined, found to be in operable condition, and test fired using the submitted ammunition magazine. Two (2) cartridges from laboratory stock ammunition were used for test firing. The test fires were retained and designated Exhibit1.1.*
>
> *The Exhibit 3 and 4 cartridge cases were microscopically compared to the Exhibit 1.1 test fired cartridge cases. Based on agreement of all discernible class characteristics and sufficient correspondence of individual characteristics, the Exhibit 3 and 4 cartridge cases were identified as having been fired in the Exhibit 1 firearm. An identification conclusion indicates the probability that the Exhibit 3 and 4 cartridge cases were fired in a different firearm is so small that it is negligible.*

[redacted]

[redacted]

[redacted]

The government retained the services of Therese C. Moynihan[1], a Firearm & Toolmark examiner employed by the Bureau of Alcohol, Tobacco, Firearms & Explosives to examine the shell casings recovered in this case on August 9, 2020, at the intersection of [redacted], as well as those recovered from the shooting at [redacted] on August 4, 2020. Ms. Moynihan issued two reports in connection with those examinations. On July 8, 2021, Ms. Moynihan issued a report stating the six shell casings recovered from 2[redacted] were fired from the same weapon. (Govt Ex 1b)

On February 9, 2022, Ms. Moynihan issued a report stating the four shell casing recovered from in front of [redacted] were all fired from the same firearm. (Govt Ex 1c.) In Ms. Moynihan's expert opinion, the firearm that discharged those four shelling casings was the same firearm as one that discharged the six shelling casings recovered on August 9, 2020, at th intersection of [redacted] She is of the opinion that the probability that the two toolmarks were made by different sources is so small that it is negligible.

After that information was provided to defense counsel, the respondent files his motion to exclude.

FIREARMS AND TOOLMARK IDENTIFICATION TESTIMONY

The respondent filed a Motion to Exclude or, in the Alternative, Limit Firearm and

[1] Therese Moynihan's CV is attached as Govt Ex 1

Toolmark Examiner Testimony pursuant to Rule 702 and *Daubert v. Merrell Dow Pharm. Inc.,* 509 U.S. 579 (1993). Respondent's Motion at 2. The Court should deny the respondent's Motion to Exclude or, in the Alternative, Limit Firearm and Toolmark Examiner Testimony in this case. The scientific evidence supports admission of the expert firearm and toolmark identification testimony of Therese Moynihan, which will be within the limitations set forth in the Uniform Language for Testimony and Reports for the Forensic Firearms/Toolmarks Discipline – Pattern Matching Examination ("DOJ ULTR") (Govt. Ex. 3a.) Nothing in the respondent's Motion undermines the scientific data supporting firearm and toolmark identification.

The respondent presents three main arguments in asking the Court to exclude the government's expert: 1) Firearm and Toolmark Analysis is not reliable enough to assert conclusive matches; 2) Other courts across the country are limiting the scope of Firearm and Toolmark testimony; and 3) If the Court allows any testimony at all, the Court should bar the examiner from opining that the cartridge cases came from the same gun. The Court should deny the defense's motion because, as the government will set out in arguments below, Firearm and Toolmark Analysis is reliable, courts across the country overwhelming admit Firearm and Toolmark testimony and those that do not are outliers, and the respondent has not provided the Court with any basis to limit the testimony of the examiner.

Underpinning the respondent's arguments are the findings of three outdated policy reports – 1) the 2008 Ballistic Imaging Report (2008 NAS Report); 2) the 2009 Report by the National Academy of Sciences (2009 NAS Report); and 3) the 2016 President's Council of Advisors on Science and Technology (PCAST Report). Paradoxically, respondent spends much time echoing PCAST's emphasis on and call for black box, open-set studies, but ignores entirely

the post-PCAST body of such studies published in scientific journals (or in pre-publication review) that provide unrefuted scientific data demonstrating the reliability of firearms and toolmark identification, including establishing a low false positive error rate (accompanied by high "sensitivity"[2]).

Although these post-PCAST studies have satisfied PCAST's standard for establishing what it termed "foundational validity," it is worth noting that not only is PCAST outdated, the PCAST report itself is flawed in numerous ways. PCAST chose to ignore dozens of firearm and toolmark validation studies because the committee members were under the mistaken impression that there was a significant difference in error rates when comparing sample-to-sample/open with set-to-set/closed validation studies. As described herein, subsequent scientific research does not support this theory. While well intended, there is a simple explanation for why PCAST members missed the mark: there was not a single trained firearms examiner or scientist with firearm and toolmark research experience on the PCAST committee. Conversely, the post-PCAST scientific data were produced collaboratively by trained firearms and toolmark examiners and applied scientists. It is one thing for scientists (untrained in the discipline in question) to postulate theories, but trained examiners and applied scientists have now put these theories to the test. This ground truth scientific data, relied upon herein, has been peer-reviewed and published in scientific journals, or is in the process of peer review, and unambiguously demonstrates the reliability of firearms and toolmark analysis.

The cases relied upon by respondent fail to address any of the post-PCAST scientific research. Rather, the respondent relies upon a handful of outlier legal decisions, including *United*

---

[2] "Sensitivity is the portion of times examiners reported identifications when the ground truth is same source (examiners correctly reported an identification)." Govt. Ex. 2, 1/12/21 Declaration of Todd Weller, ¶ C2 hereinafter "Weller Decl.").

States v. Tibbs, 2016 CF1 19432, 2019 WL 4359486 (D.C. S. Ct. September 5, 2019), a nonbinding and unpersuasive decision that ignored one post-PCAST scientific study altogether and discarded another because researchers published the scientific data in the AFTE Journal.

Indeed, many judges who have had the benefit of and considered such new scientific research have declined to preclude or limit the admissibility of firearms and toolmark expert testimony as requested by the respondent in this case. One such judge, the Honorable United States District Court Judge Rudolph Contreras, had the benefit of addressing new scientific research that directly refutes the respondent's arguments. After a thorough review of the scientific data, Judge Contreras ruled that testimony consistent with DOJ guidelines for firearms and toolmark analysis is admissible under Rule 702:

> Setting aside for the moment the utility of this "black box" requirement – which goes beyond what is required by Rule 702 – the Government has provided to the Court three recent scientific studies that meet the PCAST's black-box model requirements and demonstrate the reliability of the firearm and toolmark identification method.

United States v. Harris, 502 F.Supp.3d 28, 38 (D.D.C. 2020).[3]

Similarly here, the government is confident that after a review of the scientific information presented in this pleading, which primarily focuses on post-PCAST studies, and any testimony at a hearing, this Court will decline to adopt the respondent's reliance upon a handful of outlier legal decisions and opinions provided by a psychologist (Dr. Scurich) and a law professor (Mr. Faigman) who have ***no*** firearm and toolmark training, have conducted ***no*** research in the field, and whose ideas have limited support in the scientific community. In sum,

---

[3] Unlike the hearing in *Tibbs*, where Judge Edelman held a "truncated" hearing in which witnesses did not give direct testimony, but rather were cross-examined by opposing counsel, Judge Contreras conducted a full hearing where he had the benefit of hearing Todd Weller discuss the three Post-PCAST studies. Unfortunately, the truncated nature of the hearing prevented Judge Edelman from considering a full evidentiary record, which may have prevented him from comprehending the full import of the two post-PCAST studies published at the time of the *Tibbs* hearing.

the Court should deny the respondent's motion and allow Ms Moynihan to testify consistent with the limitations laid out in the DOJ ULTR.

BACKGROUND

A. Firearms and Toolmark Identification[4]

Firearm identification has been a forensic discipline since the 1920s. See The History of Firearm and Toolmark Identification, AFTE Journal, 1999 Volume 31, Number 3 (Summer), pp. 266-284. Firearms identification is a subset of the broader forensic discipline known as toolmark identification. Toolmark examiners are trained to examine the marks left by tools on any variety of surfaces in an attempt to associate a toolmark to a particular tool that made the mark. Firearms are simply a specialized subset of tools that impart marks on bullets and cartridge cases. See AFTE.org/resources/swggun-ark, Foundational Overview of Firearm/Toolmark Identification.

A firearm imparts different types of marks on the various components of a cartridge. With respect to bullets, cuts within a gun barrel ("grooves") and raised surfaces ("lands") create corresponding depressed "land impressions" and raised "groove impressions" as bullets travel through a barrel. The twist imparted on a bullet may be either left or right, depending on the direction of the lands and grooves. With respect to cartridge casings, contact between the cartridge and the breech create "breech face marks," and the impact of the firing pin on the primer creates a "firing pin impression" on the primer itself. The working edges of tools, which include components of firearms that contact ammunition, generally consist of some type of hard material, such as steel, to ensure strength and durability of the tool while work pieces are

[4] The contents of this section is based on the work of Stephen Bunch (formerly the Unit Chief with the FBI Firearms/Toolmark Unit), FBI Firearms Analyst Douglas Murphey, ATF firearms examiner Greg Klees, and John Murdock.

generally made of softer materials. These tool surfaces contain random, microscopic irregularities that are produced during the tool's manufacture and/or subsequent wear through use and abuse. These irregularities, which are formed randomly, can individualize or distinguish one tool from another. Because these irregularities or individual characteristics are typically imparted by contact onto the work piece, the comparative study of the imparted markings allow the tool to be individually associated or identified as having produced the mark. The presence, observation, and comparison of these random toolmarks on tools form the hypothetical propositions upon which the discipline of Toolmark Identification is based.

Firearm and toolmark identification is based upon two testable propositions:

Proposition #1:

Toolmarks imparted to objects by different tools will rarely if ever display agreement sufficient to lead a qualified examiner to conclude the objects were marked by the same tool. That is, a qualified examiner will rarely if ever commit a false positive error (misidentification).

Proposition #2:

Most manufacturing processes involve the transfer of rapidly changing or random marks onto work pieces such as barrel bores, breech faces, firing pins, screwdriver blades, and the working surfaces of other common tools. This is caused principally by the phenomena of tool wear and chip formation or by electrical/chemical erosion. Microscopic marks on tools may then continue to change from further wear, corrosion, or abuse.

See Bunch S., Smith E, Grioux B., and Murphy D., Is a Match Really a Match? A Primer on the Procedures and Validity of Firearm and Toolmark Identification, Forensic Science Communications, July 2009, Vol 11, No. 3.

Examiners are trained to recognize and evaluate the following characteristics: (1) class

characteristics; (2) subclass characteristics; and (3) individual characteristics. Class characteristics, such as caliber, number of land and grooves, etc., are predetermined during the manufacturing process. For a fired bullet, class characteristics include the number of land and groove impressions, the direction of twist of the land and groove impressions, and the width of the land and groove impressions. For a fired cartridge case, class characteristics are typically limited to (1) the firing pin impression on the primer, which can appear in various shapes (including circular, rectangular, hemispherical, and elliptical); and (2) the shape of the firing pin aperture and the type of breach face impression, which can be in different shapes and orientations (e.g., arched, circular, parallel, etc.). These are measurable features of a specimen, which indicate a restricted group source. They result from design factors and are determined prior to manufacture. See ASSOCIATION OF FIREARM & TOOL MARK EXAMINERS, AFTE GLOSSARY 38 (6th ed. 2013).[5]

Subclass characteristics are more restrictive than class characteristics and are consistent among items manufactured by the same tool in the same approximate state of wear. See id. These characteristics can exist within a particular production run in the manufacturing process and occasionally arise from (1) imperfections in a machine tool that persist during the production of multiple firearm components; or (2) extreme hardness differences between the machine tool and the work pieces. Qualified examiners are trained to distinguish subclass characteristics from individual characteristics because identifications may not be made from subclass characteristics.

Individual characteristics, on the other hand, consist of microscopic, random imperfections in the barrel or firing mechanism created by the manufacturing process, wear, corrosion, or abuse. Individual characteristics are unintended microscopic features that occur due

---

[5] The AFTE Training Manual, Technical Procedures Manual and Glossary are too large to be attached to this pleading. They are available for download at www.AFTE.org.

to machining process and random chip formation during manufacturing. Individual characteristics typically fall into two categories: (1) striated marks made by movement of a fired bullet through a gun's barrel (typically appearing as scratches or striations), and (2) impressed marks that are pressed into a surface. A fired bullet usually has striated marks. A spent cartridge case, on the other hand, can have both impressed and striated marks. Before firing, the process of feeding the cartridge into the chamber can create striated marks. Once the firearm is fired, striated marks also can be imparted to the cartridge case wall (side), and impressed marks are imparted to the cartridge case by the gun's firing pin and breech. With semi-automatic weapons, additional marks can be made as the cartridge case is expelled from the gun. Marks produced by the random imperfections or irregularities of tool surfaces are produced incidental to manufacture and/or caused by use, corrosion, or damage. See id. In general, the tool working surfaces in a firearm can slowly change over time from wear and may leave different marks on bullets and casings. As microscopic similarities diminish, the likelihood of an inconclusive result increases, but the likelihood of a false positive should remain unchanged.

Since the inception of firearms and toolmark identification as a forensic discipline, firearms examiners have been using a method known as "pattern matching" to determine whether sufficient similarity exists between toolmarks to warrant a conclusion that two bullets or two cartridge cases came from the same firearm. In 1992, AFTE memorialized the Theory of Identification in an attempt to explain the basis of opinions of common origin in toolmark comparisons:

> 1. The theory of identification as it pertains to the comparison of toolmarks enables opinions of common origin to be made when the unique surface contours of two toolmarks are in "sufficient agreement."
>
> 2. This "sufficient agreement" is related to the significant duplication of random toolmarks as evidenced by a pattern or combination of patterns of surface

contours. Significance is determined by the comparative examination of two or more sets of surface contour patterns comprised of individual peaks, ridges and furrows. Specifically, the relative height or depth, width, curvature and spatial relationship of the individual peaks, ridges and furrows within one set of surface contours are defined and compared to the corresponding features in the second set of surface contours. **Agreement is significant when agreement in individual characteristics exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools and is consistent with agreement demonstrated by toolmarks known to have been produced by the same tool.** The statement that "sufficient agreement" exists between two toolmarks means that the agreement of individual characteristics is of a quantity and quality that the likelihood another tool could have made the mark is so remote as to be considered a practical impossibility.

3. Currently the interpretation of individualized/identification is subjective in nature, founded on scientific principles and based on the examiner's training and experience.

See AFTE Criteria for Identification Committee, *Theory of Identification, Range of Striae Comparison Reports and Modified Glossary Terms – An AFTE Criteria for Identification Committee Report*, 24 AFTE JOURNAL 336, 336–40 (1992) (emphasis supplied); AFTE GLOSSARY at 138.

The discipline of firearm and toolmarks identification follows a recognized examination methodology involving two phases of analysis. The first phase focuses on the objective evaluation of existing class characteristics, which if similar, the examination moves to the second phase involving the comparative microscopic evaluation by the examiner of individual characteristics that culminates with an opinion decision concerning sufficient agreement, which is subjective.

The following discussion outlines the levels of analysis firearm and toolmark examiners follow to determine a common source:

1. Evaluation: (Objective)

The initial examination phase evaluates evidence to determine if the observed class

characteristics are the same between two specimens (two unknown specimens, or an unknown and a known specimen). If the specimens are suitable for examination and the class characteristics are the same, then it is possible that the toolmarks were produced utilizing the same tool (such as a firearm). If they are different, then the two specimens can be eliminated as having been produced by the same tool.

2. Comparison: (Subjective – Pattern Matching)

If the class characteristics are consistent between two specimens, then a comparative examination is performed utilizing a comparison microscope. The methodology utilized in the examination process is pattern matching. This comparison is conducted to determine: 1) if any marks present are subclass characteristics and/or individual characteristics, and 2) the level of correspondence of any individual characteristics.

3. Conclusion:

If sufficient agreement of individual characteristics is observed between two specimens, an identification conclusion is rendered. If all of the discernible class characteristics are the same, but sufficient agreement of the individual characteristics is not observed, an inconclusive result is rendered. In some situations, an elimination conclusion may be rendered based on observed differences in individual characteristics.

4. Verification:

A verification process is employed to ensure proper conclusions are rendered. As outlined in a laboratory's quality assurance policy, a mechanism should be in place to determine which cases will require verification. Many laboratories require verification of all identifications. See afte.org/resources/swggun-ark/summary-of-the-examination-method.

Using this methodology for examining toolmarked surfaces, there are four conclusions

that examiners reach when conducting an examination: (1) identification, (2) inconclusive, (3) elimination, and (4) unsuitable for comparison. Examiners undergo standardized technical training designed to develop cognitive skills to recognize patterns of individual characteristics necessary to make an identification. However, there is no way to be *absolutely* (100%) certain of any identification without comparing a particular set of marks to marks created by every firearm produced since the invention of the modern-day firearm. This would be an impossible endeavor. Weller Decl. ¶ H4. Thus, an examiner cannot rule out with absolute certainty the highly unlikely event that two different firearms will produce indistinguishable individual characteristics.

As discussed below, the field of forensic firearm and toolmark identification continues to undergo testing in the form of (1) technical research; (2) validation studies; and (3) proficiency testing. Validation studies are the most comprehensive way to test and validate firearms and toolmark identification as a reliable forensic science. These tests involve "ground truth," so it is known with absolute certainty where each of the test components came from. Using the same methods and identification criteria as those in actual casework, qualified examiners have consistently reached correct conclusions with exceptionally low error rates. See Weller Decl. ¶¶ C1-C38. Even where researchers have studied bullets and cartridge cases fired from consecutively manufactured firearms – where the possibility of a false-positive conclusion is at its highest – trained examiners have been able to readily distinguish marks produced by the various firearms.

B. July 8, 2021, Firearms & Toolmark Report by Therese Moynihan

On July 8, 2021, firearms examiner Therese Moynihan issued a report comparing various fired cartridge casings to a recovered firearm. See Moynihan July 8 2021 Report (Govt. Ex. 2b).[6] Specifically,

[6] Ms. Moynihan's CV is attached as Govt. Ex. 2a.

Ms. Moynihan examined, test fired, and found operable, Item 1 the Polymer 80, model PF940SC, caliber 9mm Luger semiautomatic pistol with no serial number recovered in Prince George's County, Maryland. Ms. Moynihan used two cartridges from laboratory stock ammunition to test fire the weapon. She retained those test fires for comparison and Item 1.1. Ms. Moynihan microscopically compared cartridge casings test fired from this 9mm pistol (Item 1) to six (6) cartridge casings recovered from the intersection of [redacted] (Item 4). She reached the following conclusions:

Item 4 and Item 1.1 were microscopically compared and based on agreement of all discernable class characteristics and sufficient correspondence of individual characteristics; the six cartridge casings in Item 4 were identified as having been fired in Item 1 firearm. An identification conclusion indicates the probability that the cartridge casings in item 4 were fired in a different firearm is so small that it is negligible.

Arnold J. Esposito, a Firearm & Toolmark Examiner with Bureau of Alcohol, Tobacco, Firearms and Explosives, Forensic Science Laboratory – Washington, verified these conclusions.[7]

C. February 9, 2022, Firearms & Toolmark Report by Therese Moynihan

On February 9, 2022, firearms examiner Therese Moynihan issued a report comparing various fired cartridge casings to a recovered firearm. See Moynihan February 9 2021, Report (Govt. Ex. 2c). Specifically,

[7] Unlike the February 9, 2022, report, Mr. Arnold did not sign Ms. Moynihan's report to signify he had reviewed and verified. Per ATFE policy at the time, Mr. Esposito initial each page of the July 8, 2021, report to show he had reviewed and verified the report. Prior to the February 9, 2022, report, ATFE policy changed, and the reviewer was required to sign the document.

Ms. Moynihan examined, test fired, and found operable, Item 1 the Polymer 80, model PF940SC, caliber 9mm Luger semiautomatic pistol with no serial number recovered in Prince George's County, Maryland. Ms. Moynihan used two cartridges from laboratory stock ammunition to test fire the weapon. She retained those test fires for comparison as Item 1.1. Ms. Moynihan microscopically compared cartridge casings test fired from this 9mm pistol (Item 1) to four (4) cartridge casings recovered from in front of [redacted] (Item 2). She reached the following conclusions:

Item 2 and Item 1.1 were microscopically compared and based on agreement of all discernable class characteristics and sufficient correspondence of individual characteristics; the four cartridge casings in Item 2 were identified as having been fired in Item 1 firearm. An identification conclusion indicates the probability that the cartridge casings in item 2 were fired in a different firearm is so small that it is negligible.

Arnold J. Esposito, a Firearm & Toolmark Examiner with Bureau of Alcohol, Tobacco, Firearms and Explosives, Forensic Science Laboratory – Washington, verified these conclusions.

D. Testimony by Therese Moynihan

Here, firearms examiner Moynihan will opine based on her training and experience and the degree of agreement of individual characteristics observed under the comparison microscope. She will **not** use unqualified terms such as "match." She will **not** state her expert opinion with any level of statistical certainty, much less 100% or absolute certainty. She will **not** render her opinion "to the exclusion of all other firearms" or use the phrase "to a reasonable degree of scientific certainty." This is consistent with the DOJ ULTR, which defines "source identification" as "an examiner's conclusion that 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 they originated from different sources." DOJ ULTR, Section III (Govt. Ex. 1). The DOJ ULTR precludes examiners from associating a casing to a firearm "to the exclusion of all other sources," from asserting a "numerical degree of probability" without appropriate data, or from using the expression "reasonable degree of scientific certainty." Id.

ARGUMENT

In *Daubert*, the Supreme Court provided a non-exhaustive list of factors to consider when evaluating the admissibility of expert testimony under Rule 702, including whether a theory or technique (1) has been or can be tested; (2) has a known or potential rate of error; (3) has been subjected to peer review and publication; (4) has standards controlling the techniques operation; and (5) enjoys acceptance within the relevant scientific community. *Daubert*, 509 U.S. 579, at 593-94. Rule 702 also allows for expert testimony where the expert "reliably applied the principles and methods to the facts of the case." Fed. R. Evid. 702(d). Rule 702 embodies a more liberal standard of admissibility for expert opinions than did *Frye*. *United States v. Williams*, 506 F.3d 151, 161-62 (2d Cir. 2007). In exercising its gatekeeping function, courts must keep in mind the Supreme Court's admonition that "vigorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence." Daubert, 509 U.S. at 596.

Given the extensive testing, peer-review, low levels of error, and general acceptance throughout the world, courts have routinely admitted firearms evidence under Rule 702. *State v. Romero*, 341 P.3d 493, 498 (Ariz. App. Div. 2 2014) ("Several federal district courts have held

that firearms identification testimony is sufficiently reliable under *Daubert* and Federal Rule 702."); *United States v. Hicks*, 389 F.3d 514, 526 (5th Cir. 2004) (noting that defendant was unable to point the court to a single case in any circuit showing that the methodology was unreliable); *United States v. Santiago*, 199 F. Supp.2d 101, 111 (S.D.N.Y. 2002) ("The Court has not found a single case . . . that would suggest that the entire field of ballistics identification is unreliable."). Many federal courts have held extensive hearings before admitting firearms and toolmark identification evidence. *See e.g., United States v. Montiero*, 407 F. Supp. 2d 351, 355 (D.Mass. 2006) (six-day hearing); *United States v. Diaz*, 05-CR-00167, 2007 WL 485967, at *1 (N.D. Cal. Feb. 12, 2007) (four-day hearing); *United States v. Taylor*, 663 F. Supp. 2d 1170, 1171 (D.N.M. 2009) (two-day hearing); *United States v. Otero*, 849 F. Supp.2d 425, 429 (D.N.J. 2012) (three-day hearing).

While the issuance of the PCAST Report created a flurry of firearms litigation, it did not significantly alter the legal landscape:

> While no federal court (at least to the Court's knowledge) has found the AFTE method to be unreliable under *Daubert*, several have placed limitations on the manner in which the expert is allowed to testify. The general consensus is that firearm examiners should not testify that their conclusions are infallible or not subject to any rate of error, nor should they arbitrarily give a statistical probability for the accuracy of their conclusions. Several courts have also prohibited a firearm examiner from asserting that a particular bullet or shell casing could only have been discharged from a particular gun to the exclusion of all other guns in the world. These restrictions are in accord with guidelines issued by the Department of Justice for its own federal firearm examiners which went into effect in January 2019. But it is also important to note that the courts that imposed limitations on firearm and toolmark expert testimony were the exception rather than the rule. Many courts have continued to allow unfettered testimony from firearm examiners who have utilized the AFTE method.

United States v. Romero-Lobato, 379 F. Supp. 3d 1111, 1117 (D. Nev. 2019) (citations omitted, emphasis added). *See also United States v. Brown*, 973 F.3d 667, 704 (7th Cir. 2020) (no abuse of discretion in trial court's refusal to adhere to PCAST recommendations and its finding that firearms and

toolmark identification is tested, subjected to peer review and publication, has an overall low error rate (single digits), and is generally accepted in the specified scientific field); *United States v. Chavez*, ___ F.Supp.3d __, 15-CR-00285, 2021 WL 5882466 (N.D. Cal. Dec. 13, 2021) (denying request to limit prosecutions' ballistics evidence where defense submitted an affidavit from Dr. Scurich); Harris, 502 F.Supp.3d 28 (admitting firearms and toolmark expert testimony with the limitations identified in the DOJ ULTR); *United States v. Hunt*, 464 F. Supp. 3d 1252, 1259-62 (W.D. Okl. June 1, 2020) (admitting testimony following AFTE theory of identification and indicating DOJ ULTR is reasonable guidance scope of testimony); *United States v. Johnson*, 2019 WL 1130258, at *21-22 (S.D.N.Y. March 11, 2019) ("In the vast majority of cases in which courts have limited the opinions a firearms examiner may offer, the limitation has addressed whether the firearms examiner can state his or her opinion to a specific degree of scientific certainty. Often these limitations are imposed because of judicial or defense counsel concern that the firearms examiner intends to offer an opinion with absolute or 100% certainty.") (citations omitted); *United States v. Romero-Lobato*, 379 F. Supp. 3d 1111, 1117 (D. Nev. May 16, 2019); *United States v. White*, 17-CR-611, 2018 WL 4565140, at *2 (S.D.N.Y. Sept. 24, 2018) (finding no hearing necessary where the admissibility of firearms expert testimony has been repeatedly recognized by federal courts and noting the expert may not quantify or give an opinion to the exclusion of all firearms); *United States v. Gregory Chester, et. al.*, No 13-CR-00774, slip. op. at 2 (N.D. Ill. Oct. 7, 2016) ("In short, the PCAST report does not undermine the general reliability of firearm toolmark analysis or require exclusion of the proffered opinions in this case. Questions about the strength of the inferences to be drawn from the analysis of the examiners presented by the government may be addressed on cross-examination.") (attached as Govt. Ex. 30); *People v. Lozano-Membrano, et.al.*, No. 1501755, Oral Ruling, p.21 (Sup. Ct. Cal. July 16, 2020) (declining to limit testimony to class characteristics in a case where defendant's submitted an affidavit from Mr. Faigman

and ruling that "[t]he expert may testify as to exclusions and inclusions, but must identify the limitations of her opinion that, one, it does not exclude all firearms, two, that it is not presented as a scientific certainty, and three, that they will give no numerical or statistical calculation") (transcript of ruling attached as Govt Ex. 31)[8]; *Commonwealth v. Hernandez*, SUCR2014-10417 * 5, slip. op. (Super. Ct. Mass. Dec. 21, 2016) (denying defendant's motion to preclude firearms and toolmark identification evidence based on PCAST) (attached as Govt. Ex. 32); *Commonwealth v. Legore*, SUCR 2015-10363, slip op. at 2 (Superior Court Mass., Nov. 17, 2016) ("After a non-evidentiary hearing and argument, and upon review of the PCAST report (and in particular, pages 104-114), there is no basis to disturb settled law permitting a properly qualified firearms expert from offering opinion evidence under [*Daubert/Lanigan*] relating to a comparison and match between a bullet recovered from the alleged victim, and a bullet test-fired from a firearm allegedly associated with the defendant.") (attached as Govt. Ex. 33).

A. D.C. Court of Appeals Case Law Does Not Require the Limitations Requested by the Respondent.

The respondent requests that if the Court doesn't outright exclude Ms. Moynihan's testimony, that the Court limit Ms. Moynihan's conclusions to class characteristics. For a fired bullet these include the number, direction of twist, and measurement of land and groove impressions. For a fired cartridge casing these are typically limited to the firing pin impression

[8] The ruling and pleadings in *Lozano-Membrano* addressed a separate case *People v. Azcona*, 58 Cal. App. 5th 504, 510 (2020), as modified (Jan. 11, 2021). In that case, the court determined the defendant had failed to establish that firearm and toolmark identification was no longer generally accepted in the applicable scientific community. *Id.* at 512-13. The court further found, however, that it had erred by allowing the expert to testify that the matching marks on the relevant projectiles are "much more than can ever happen by random chance," and therefore the projectiles came from the same gun, "to the practical exclusion of all other guns." *Id.* at 513-14. The court went on to say that "[s]uch a purportedly infallible conclusion is a leap too far from what the underlying method allowed. There was support for the opinion that the projectiles likely came from the same gun, perhaps more likely than not, but there was no basis to present it as a scientific certainty." *Id.* The Azcona court only described the 2008 and 2009 NAS Reports, and 2016 PCAST Reports in support of its determination.

on the primer, shape of the firing pin aperture, and type of breach face impression. Limiting testimony to class characteristics leaves open a potentially wide array of firearms that could have fired the evidence in any particular case. The respondent seeks to exclude **any** testimony based on the firearms examiner's microscopic examination of individual characteristics, upon which firearms examiners may rely to make conclusions of identification, exclusion, or inconclusive.[9]

Such limitation goes even beyond those imposed in *Tibbs*, and finds no support from *United States v. Gardner*, 140 A.3d 1172 (D.C. 2016), *Unites States v. Williams*, 210 A.3d 734 (D.C. 2019) (*Williams II*), or the science. Neither *Gardner* nor *Williams II* had the benefit of any information related to Firearms and Toolmark Identification beyond the 2008 and 2009 NAS Reports and PCAST, which were either limited in their examination of the field or flawed as explained herein. More importantly, neither *Gardner* nor *Williams II* had the benefit of studies in the field, or any testimony or affidavits from well-respected members from the field or broader the scientific community. Neither of those cases had the benefit of post-PCAST studies that confirm the reliability of the field. This Court accordingly need not engage in an in-depth examination of the rulings in *Gardner* or *Williams II*, and notably neither opinion required the limitation requested by the respondent. Rather, each simply precluded the kind of unqualified opinion that the government does not intend to offer in this case.[10] Indeed, *Williams II* stated,

---

[9] Certainly, the respondent would not be seeking the requested limitation if the microscopic examination of individual marks established an exclusion.

[10] Specifically, the D.C.C.A. in *Williams II* confronted a record in which the examiner had testified, *inter alia*, "[t]hese three bullets were identified as being fired out of Exhibit No. 58. And it doesn't matter how many firearms Hi[-]Point made. Those markings are unique to that gun and that gun only." *Williams*, 210 A.3d at 738. Similarly, in *Gardner*, the Court's confronted a record where "the prosecutor specifically asked Mr. Watkins, 'Just to be clear, sir, your—your scientific—your opinion here is Government Exhibit Number 18, the bullet, [was] fired from Government 71[,] or was it consistent with being fired from Government Exhibit 71?" Mr. Watkins replied, "It was identified as having been fired from Government Exhibit 71." *Gardner*, 140 A.3d at 1182. The Court of Appeals concluded only that the examiner could not "give an unqualified opinion about the source of the bullet" and further held "that in this jurisdiction a firearms and toolmark expert may not give an unqualified opinion, or testify with absolute or 100%

"[W]e do not question the admissibility of the firearms and toolmark examiner's testimony generally." 210 A.3d at 743. *Williams II* went on to state,

> Following *Gardner*, we repeat that it is error to allow an examiner to provide this kind of unqualified opinion testimony, but we do not foreclose the possibility that the necessary data will exist at some point in the future to provide a foundation for opinion testimony that unqualifiedly connects a specific bullet to a specific gun.

Id. The scientific information provided herein and at any hearing in this case provides such a foundation. Even adopting the PCAST framework relied on by *Gardner* and *Williams II*, post-PCAST studies establish what PCAST termed "foundational validity." That said, the government does not intend to admit unqualified firearm and toolmark identification testimony. Rather, in ensuring Ms. Moynihan's testimony complies with the restrictions set out in the DOJ ULTR, the government asserts it is in compliance with Gardner and Williams II.

B. Firearms and Toolmark Identification Satisfies Reliability Under Rule 702

1. Testability[11]

Firearm and toolmark identification has been and continues to be tested and found reliable. This factor focuses on "whether the expert's theory can be challenged in some objective sense, or whether it is instead simply a subjective, conclusory approach that cannot reasonably be assessed for reliability." See Fed. R. Evid. 702 Advisory Committee's Note to 2000 Amendment. "[V]irtually every court that has evaluated the admissibility of firearms and toolmark identification has found the AFTE method to be testable and that the method has been repeatedly tested." *Tibbs*, 2019 WL 4359486, at *7 (listing cases); *see also Harris*, 502 F. Supp. 3d at 37

---

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 1184.

[11] Respondent addresses testability and error rate together. This pleading discusses them separately.

(stating that "[a] number of courts have examined this factor in depth to conclude that firearm toolmark identification can be tested and reproduced[,]" and compiling citations including Tibbs).

Such findings are supported by a long and continuous history of research. Since the "phenomenon behind Firearms and Toolmark Examination, namely that firearms can impart microscopic toolmarks on fired ammunition components, and that those toolmarks can be used for the purposes of source attribution (i.e., identification) and elimination was first documented over 100 years ago[,] . . . the profession has been engaged in observation, documentation and testing of firearms examination related topics." See Weller Decl. ¶ B1. Such foundational research appears in textbooks[12] and peer reviewed journal articles. *Id.* Articles often referred to as "Review Articles" serve as "encyclopedia-like sources of research that summarize the state of a topic or discipline." *Id.* Attached are two Review Articles related to Firearms and Toolmark Identification, one written by Ronald Nichols and published in the Journal of Forensic Sciences in May 2007,[13] which lists 65 references, and the second by Erwin J.A.T. Mattijssen and published in Forensic Science International: Synergy in 2020, with 189 references, addressing research occurring between 2016 and the end of 2018.[14] (Govt. Exs. 4a and 4b respectively). Summaries of foundational research also appear on the AFTE website, which lists over 100

[12] There are several textbooks published over a large span of time cited in Mr. Weller's declaration, including a recent text by RONALD NICHOLS, FIREARM AND TOOLMARK IDENTIFICATION: THE SCIENTIFIC RELIABILITY OF THE FORENSIC SCIENCE DISCIPLINE 1-159 (2018). This text "explains how past and recent research provide strong support for the science of firearm and toolmark examination." Weller Decl. ¶ B1.

[13] Ronald G. Nichols, *Defending the Scientific Foundations of the Firearms and Tool Mark Identification Discipline: Responding to Recent Challenges*, 52 J. FORENSIC SCIENCES 586, 586–94 (2007) (Govt. Ex. 4a.)

[14] Erwin J.A.T. Mattijssen, *Interpol review of forensic firearm examination 2016-2019*, 2 FOR. SCI INT. SYNERGY 389, 389–403 (2020) (Govt. Ex. 4b.)

citations with summaries of each article.[15] "These articles, which only represent a fraction of the total body of research, show the profession has published research papers that span over half a century." *See* Weller Decl. ¶ B1.

Firearms and Toolmark Identification research has been published in numerous peer-reviewed journals and conducted by experts in the field and experts in other fields such as physical scientists, statistics, and computer science. Mr. Weller's declaration cites to more than forty scientific studies involving firearms and toolmark identification analysis that were published in journals other than the AFTE Journal, including the following ten scientific journals: 1) Forensic Science International, 2) Journal of Forensic Sciences, 3) Science and Justice, 4) National Institute of Standards and Technology, 5) Surface Topography, 6) Measurement Science and Technology, 7) Scanning, 8) Three Dimensional Imaging, Processing and Applications, 9) The Annals of Applied Statistics, and 10) Journal of Physics. *See* Weller Decl. at 52-55, Appx. A. These studies alone were authored numerous scientists, many of whom hold PhDs in a wide range of the applied sciences, including statistics, engineering, quantum chemistry, mathematics, physics, computer science, and physical chemistry. *Id*. Many of these scientists, such as Dr. Max Morris, the former Chair of Statistics at Iowa State University, specialize in experimental design.

The respondent's challenge centers on three outdated policy papers, namely the 2008 and 2009 NAS Reports and the 2016 PCAST Report. The PCAST Report reviewed literature in the field and criticized the design of most of the studies that validated the accuracy of firearms and

---

[15] AFTE Website, https://afte.org/resources/swggun-ark/testability-of-the-scientific-principle, accessed 1/14/2021. Attached as Govt. Ex. 4c is a portion of the webpage reached when clicking on the link. Specifically, Govt. Ex. 4c is what appears when clicking on the hyperlink titled, "Firearm Identification – Bullets."

toolmark identification. PCAST Report, p.106. Nevertheless, such firearm and toolmark identification has a sound scientific basis, and a recently-released Department of Justice Report points out the faults with PCAST's criticism of firearms and toolmarks identification.[16] Mr. Weller's declaration provides further information related to the reliability of the PCAST Report and its criticism of firearm and toolmark identification. *See* Weller Decl. at 47-51. Among other concerns, PCAST did not include anyone from the firearm and toolmark community and therefore misunderstood fundamental aspects of the studies it discussed, miscounted or omitted data from several studies, and included multiple basic mathematical errors. Id. Contrary to conclusions in the PCAST Report, numerous pre-2016 studies demonstrate the validity of firearm and toolmark identification, many of which are contained in the resources cited above and discussed in Mr. Weller's declaration and herein.

It is unnecessary, however, for this Court to undertake a detailed examination of these three reports, because even accepting PCAST's concerns, recent studies confirm that the field meets and exceeds the benchmarks for scientific validity set by its critics. PCAST determined that two black-box, sample-to-sample/open studies, employing the use of independent comparisons would be necessary to establish the "foundational validity" of Firearms and Tool Marks Identification, and that one was already in existence, specifically the Ames I study.[17] *See* 2016 PCAST Report at 109-11. In a law review article, Dr. Eric Lander, co-Chair of the PCAST

---

[16] *See* U.S. Department of Justice Statement on the PCAST Report; Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods (January 2021) (Govt. Ex. 3b.)

[17] The full cite is David P. Baldwin et al., *A study of false positive and false negative error rates in cartridge case comparisons*, AMES LABORATORY, USDOE, TECHNICAL REPORT #IS-5207 (2014). (Govt. Ex. 5.) This study is referred to herein as Ames I. As noted in Mr. Weller's declaration, "While not published in peer reviewed journal, this study has undergone extensive review by both firearms and toolmark examiners as well as the PCAST commission. Given this extensive review by the general scientific community and the lack of any critique, it is unlikely a journal-based peer review would result in any substantial changes." Weller Decl. Appx. A, n.221.

Report reiterated that even under PCAST's standards (which exceed what is required under Rule 702) only one additional black box study was required to establish "foundational validity." *See* Eric S. Lander, *Fixing Rule 702: The PCAST Report and Steps to Ensure the Reliability of Forensic Feature-Comparison Methods in the Criminal Courts*, 86 FORDHAM L. REV. 1661, 1672 (2018).

PCAST defined a "black-box study" as an "empirical study that assesses a subjective method by having examiners analyze samples and render opinions about the origin or similarity of such samples." PCAST Report at 48. PCAST further stated, "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" PCAST Report at 49. PCAST stated the following as it relates to open and closed set designs in firearm and toolmark examinations:

> This closed-set design is simpler than the problem encountered in casework, because the correct answer is always present in the collection. In such studies, examiners can perform perfectly if they simply match each bullet to the standard that is *closest*. By contrast, in an open-set study (as in casework) there is no guarantee that the correct source is present – and thus no guarantee that the closest match is correct. Closed set comparisons would thus be expected to underestimate the false positive rate.

PCAST Report at 108. PCAST also indicated the study should include what it terms "independent" comparisons, which it defines as "examiners making a series of independent comparison decisions between questioned sample and one or more known samples that may or may not contain the source." *Id.* at. 110 (the second portion of the sentence reiterates the definition of "open"). This study design is referred to in Todd Weller's declaration as sample-to-sample, open, and will be referred to as such herein. *See* Weller Decl. ¶ C21.

The "single well-designed study" referenced by PCAST and Dr. Lander is Baldwin, et.al., *A study of false-positive and false-negative error rates in cartridge case comparisons,* Ames Laboratory, USDOE, Technical Report #IS-5207 (2014), a 2014 Study conducted by the Ames Laboratory of the Department of Energy ("Ames I") (Govt. Ex. 6). In Ames I, 218 tests were returned. The study calculated a false positive rate of 1.01%, false negative rate of 0.367%, sensitivity of 98.6%, and specificity of 65.2%.[18] See Ames I, Table III, p.17 (Govt. Ex. 6.); Weller Decl. ¶ C22. According to Dr. Lander, if one more study reproduced the results of Ames I, which Dr. Lander described as "well-designed," it would render firearms identification "scientifically valid." *Lander*, 86 FORDHAM L. REV. at 1672 ("A second study would solve this problem."). This is consistent with the position of the PCAST Report, which described Ames I as an "appropriately designed" black box study and explained that one additional, similar study would establish "foundational validity."[19] PCAST Report at 4, 109–111.

To the extent this specific PCAST threshold is required by Rule 702 or *Daubert*, which we do not concede, the field has responded with additional research and study. In 2018, the Keisler Study was published in the AFTE Journal. It adheres to the sample-to-sample/open, black-box requirements set by PCAST. See Mark A. Keisler et al., wrote a study, "Isolated Pairs Research Study", that was published in the AFTE Journal, 50 AFTE J 56-58 (2018); Mark A. Keisler et. al., *Letter to the Editor: Isolated Pairs Research Study*, 50 AFTE J 56-58 (2018)

---

[18] "Sensitivity is the portion of times examiners reported identifications when the ground truth is same source (examiners correctly reported an identification). Specificity is the portion of times examiners reported eliminations when ground truth is difference source (examiners correctly reported elimination)." Weller Decl. ¶ C2.

[19] PCAST stated that a showing of foundational validity "requires that a method has been subjected to empirical testing by multiple groups, under conditions appropriate to its intended use. The studies must (1) demonstrate that the method is repeatable and reproducible, and (2) provide valid estimates of the method's accuracy (that is, how often the method reaches an incorrect conclusion) that indicate the method is appropriate to the intended application." PCAST report at 5.

(Govt. Ex. 6a and 6b). In the Keisler Study, examiners completed 126 tests resulting in 1512 same-source and 1008 different-source comparisons. No false identifications or eliminations were reported. *See* Keisler, et. al, Winter 2018, p.57. The Keisler Study reported a sensitivity of 99.74% and specificity of 79.86%. *Id.*; *see also* Weller Decl. ¶ C24.

In 2020, the manuscript of a study referred to herein as Ames II was released. [20] Ames II was an open-set black-box study involving 173 examiners. The total number of comparisons performed was 20, 130, of which 8640 were used for accuracy. *See* Weller Decl. ¶ C29. The firearms (largely consecutively or closely manufactured) and ammunition (steel) were selected to represent a challenge not seen in typical casework and to introduce confusion. *See* Weller Decl. ¶ C29. Nonetheless, Ames II reported a false positive error rate of 0.66% for bullets and 0.93% for cartridge cases. *Id.*

Two additional post-PCAST studies using 3D image technology further support the field's reliability and establish "foundational validity." A 2018 study referred to herein as the Duez Study, [21] included a black-box (as well as sample-to-sample, open-set) and white-box [22] components and used Virtual Comparison Microscopy (VCM), which allows for side-by-side

---

[20] Stanley J. Bajic, *Report: Validation Study of the Accuracy, Repeatability, and Reproducibility of Firearm Comparisons*, AMES LABORATORY-USDOE TECHNICAL REPORT # ISTR-5220 (2020). It is the government's understanding this manuscript is currently undergoing pre-publication review (Govt. Ex.78).

[21] Pierre Duez et al., *Development and Validation of a Virtual Examination Tool for Firearm Forensics*, 63 J. FORENSIC SCI. 1069, 1069-1084 (2018). Both parts of this study survived double blind peer-review in the Journal of Forensic Sciences (JFS), demonstrating that the scientific community rejects the notion that only black box design is worthy of consideration. The involvement of two Ph.D. scientists – Drs. Marcus Brubaker and Ryan Lilien – further refute the argument that applied scientists are not involved in the validation of firearms and toolmark identification (Govt. Ex. 8).

[22] White-box studies are performed "to understand the factors that affect examiners' decisions." PCAST Report, p.9. The National Institute of Science and Technology (NIST) Glossary defines white box testing includes the following, "[A] method of testing software that tests internal structures or workings of an application as opposed to its functionality (i.e., black-box testing). See https://csrc.nist.gov/glossary/term/white_box_testing .

comparisons of toolmarks on fire ammunition components using computer software.[23] The empirical data illustrated the low rate of errors made by trained firearm examiners in identifying casings to a particular firearm. In the open-set black-box component, forty-six trained examiners "correctly reported 100% of the identifications (known matches) while reporting no false positives."[24] Weller Decl. ¶ C26 (citing Duez).[25] Among trained examiners, sensitivity was 100%, i.e., 276 identifications from 276 true same-source comparisons. Id. Specificity was 87%. i.e., 80 eliminations from 92 true different source comparisons. Id.

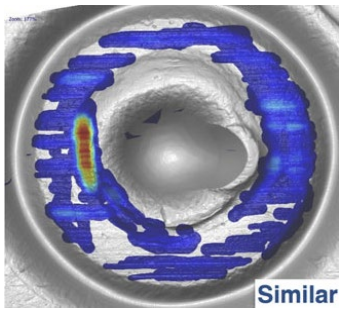
Duez also contained a white-box aspect. The software allowed each examiner to independently annotate the areas where he or she identified significant agreement. Weller Decl. ¶ C27. At the completion of all tests, each annotation was overlaid to compare where each examiner found significant agreement. Id. Following is a heatmap of breach face markings (from Duez) showing the combination of these annotations of significant agreement. The closer to red, the closer to 100% of examiners that marked that area. Id. The red and orange areas indicate a high degree of correspondence of marks by examiners. Id.

---

[23] The use of 3D images to perform firearms and toolmark identification is no different from the use of scanned latent and known prints to conduct ACE-V analysis in the latent fingerprint field – a practice commonly used by latent print examiners throughout the country and routinely admitted into evidence, including here. In other words, the method utilized is the same; examiners are merely evaluating 3D images rather than what appears under the comparison microscope.

[24] A trainee reported two false positive results. *See* Weller. Decl. ¶ C26.

[25] Importantly, the government admitted this study and the below image at the *Tibbs* hearing. Perhaps the government did not make it clear to the Court in the *truncated Tibbs* hearing that part of the heat map study comported with black box design. Either way, the import of this study was lost on the Court because it went unmentioned in the *Tibbs* ruling.



As Mr. Weller discusses in his declaration, the heat map shown above illustrates how "examiners from different laboratories (15), each working independently, are mostly using the same amount and same location of microscopic marks when concluding identification." Weller Decl. ¶ C27. Thus, not only did Duez reaffirm that trained examiners could accurately identify casings to a specific firearm, but it also showed that trained firearms examiners focus on the same location and amount of marks to make a conclusion, and that alternative forms of study design are valuable to the advancement of the field.

In a 2020 study, referred to herein as the Chapnick Study, firearm examiners used 3D technology that, according to the authors, will "potentially replace the light comparison microscope as the primary instrument used for firearm and toolmark examination."[26] The study involved 76 trained firearms examiners from the United States and Canada and 40 test sets of fired cartridge casings from firearms with a variety of makes, models and calibers. Weller Decl. ¶ C28. These 76 examiners completed a total of 1184 comparisons. *Id.* The overall error rate for this group was 0.253%. *Id.* Three false positives were reported from a total of 693 true different source samples, for a false positive error rate of 0.433%. Id. There were no false eliminations reported. *Id.* Sensitivity was 92.2% and specificity was 62.9%. Weller Decl. ¶ C28.

[26] Chad Chapnick et. al., *Results of the 3D Virtual Comparison Microscopy Error Rate (VCMER) Study for Firearm Forensics*, J. FORENSIC SCI. at 1 (Oct. 1, 2020) (Govt. Ex. 90.).

The use of 3D imaging has direct application to implementation of the AFTE theory of identification by humans:

> The fact that there are subjective elements to the firearm and toolmark identification methodology is not enough to show that the theory is not "testable." Indeed, studies have shown that "the AFTE theory is testable on the basis of achieving consistent and accurate results." *Otero*, 849 F. Supp.2d at 433; *see also* July 7, 2017 Decl. of Todd Weller at 2-6 (describing various studies that support the reproducibility of the AFTE identification theory). This conclusion has only been further strengthened in recent years due to advances in three-dimensional imaging technology, which has allowed the field to interrogate the process and sources of "subjectivity" behind firearm and toolmark examiners' conclusions. For example, Mr. Weller testified regarding a study which used 3D image technology to assess the process used by trained firearm examiners when identifying casings to a particular firearm. *See* Sept. 19, 2019 Decl. of Todd Weller at 15-16.

See Harris, 502 F. Supp. 3d at 37.

These four studies, each sample-to-sample/open black-box, confirm that trained examiners are able to make identifications with a very low rate of error and few inconclusive decisions where ground truth (correct answer) is an identification. Additionally, these studies refute the theory of the defense and the *Tibbs* court that examiners avoid difficult comparisons by applying an inconclusive determination. The studies specifically account for and report inconclusive results by calculating the specificity and sensitivity. *See* Weller Decl. ¶ E4. As pointed out by Mr. Weller in his declaration, the presence of inconclusive conclusions would certainly be a concern if sensitivity were extremely low, but Keisler, Amex II, Duez, and Chapnick refute this concern. In fact, the data from these and other studies confirms that "[w]hen examiners are faced with true same-source samples, the overwhelming amount of the time they will conclude identification, and also do so at very low error rates." See id.[27]

[27] Like the members of PCAST, the defense witnesses in Tibbs had no experience in study design, research or casework related to firearms and toolmark identification. Instead, the defense proffered a psychologist and a lawyer, neither with training or research experience in the field of firearm and toolmark. The *Tibbs*

2. Error Rates

a. Studies before and after PCAST have established low error rates.

While error rates are difficult to calculate for myriad reasons,[28] courts have addressed a wealth of validation studies ignored by PCAST and *Tibbs* and concluded that the validation research to date establishes a low error rate. *See e.g., Romero-Lobato*, 379 F. Supp. 3d at 1119-20 ("[T]he studies cited by [the firearms examiner] in his testimony and by other federal courts examining the issue universally report a low error rate for the AFTE method."); *United States v. Ashburn,* 88 F. Supp. 3d 239, 246 (E.D.N.Y. 2015) ("The court finds that due to the subjective nature of the inquiry, a definitive error rate is impossible to calculate, but also finds that the error rate, to the extent it can be measured, appears to be low, weighing in favor of admission of the expert testimony."); *Otero*, 849 F. Supp. 2d at 434 ("[I]nformation derived from [] proficiency testing is indicative of a low error rate[.]"); *Taylor*, 663 F. Supp.2d at 1177 (concluding that the error rate is "quite low"); *Diaz*, 2007 WL 485967, *8 (concluding that due to the subjective nature of the methodology, "it is not possible to calculate an absolute error rate for firearms identification," but that "the government has provided enough data to show that the error rates among trained firearms examiners are sufficiently low to counsel in favor of admitting the evidence.").

Court then largely disregarded nearly a century of validation research based on the opinions of a psychologist. It is worth noting that psychology itself is not free from experimental test design issues. *See generally* Estimating the Reproducibility of Psychological, *Science*, 28 August 2015 – Vol 349 Issue 6251 (reporting that over half of psychology studies fail the reproducibility test) (Gov. Ex. 28); Monya Baker, *Over half of psychology studies fail reproducibility test*, NATURE NEWS at 1 (Aug. 27, 2018) ("Don't trust everything you read in the psychology literature, in fact, two thirds of it should probably be distrusted.") (Gov. Ex. 29).

[28] *Tibbs* concluded that error rates cannot be calculated for most studies; however, that decision confused false positive error rates with overall error rates. Weller Decl. ¶ C9.

As stated previously, PCAST incorrectly concluded that studies that employ set-to-set/closed and set-to-set/partly-open designs will result in higher error rates than sample-to-sample/open designs. *See* PCAST report at 109. This perceived difference in study design is critical because it is the rationale PCAST, *Tibbs*, and the respondent here use to ignore numerous closed design validation studies establishing a low error rate for the field. PCAST report at106, 109, 111; *Tibbs*, 2019 WL 4359486 *14-15. This approach has been rejected by virtually all other courts to address the issue. *See e.g., Romero-Lobato*, 2019 WL 2150938, at *5 ("While the Court is cognizant of the PCAST Report's repeated criticisms regarding the lack of true black box tests, the Court declines to adopt such a strict requirement for which studies are proper and which are not. *Daubert* does not mandate such a prerequisite for a technique to satisfy its error rate element.").[29]

In a set-to-set/closed design test, the test takers are provided a set of questioned samples and set of known samples and asked to determine which, if any, of the questioned samples match one of the knowns. Weller Decl. ¶ C8. Examiners are not prescribed which comparisons to perform. *Id*. Ultimately each questioned sample has a ground-truth known match (closed). *Id*. Importantly, most of the designs in these listed studies are not one-question-to-one-known.[30] *Id.,* As Mr. Weller explains in his affidavit:

---

[29] The Mayland, et al., Study, (Govt. Ex. 14), has the highest reported overall error rate of all the studies listed in Mr. Weller's declaration. This is notable because PCAST concluded that the closed set-to-set design *underestimated* error rates. If this hypothesis was accurate, then the highest overall error rate would not be in a study using a closed, set-to-set design. Study design may have some effect on the rate of error, but so could other factors, including the samples used, the number of participants, and which examiners participated. Current data from studies using closed set-to-set and open designs provides evidence that study design is not a primary contributing factor. Both set-to-set and open study designs have some studies with false identifications and others with zero false identifications. *See* Weller Decl. ¶ C16 (citations omitted).

[30] In the Stroman Study, (*see* Govt. Ex. 15), each test was randomized, so it was more likely there was not a one-to-one relationship. *See* Weller Decl. at 12, n.65; Stroman, A. "Empirically Determined Frequency of Error in

> In other words, there are not 10 knowns and 10 questioned, each with one and only one correct match. The studies have unequal or even randomized numbers of true-matches. It is important to remember, as stated above, many of these studies use consecutively manufactured samples, and thus are attempting to test examiners with worst-case scenario samples. Finally, while the studies below do have a known match for each questioned sample, this fact was not disclosed to test takers.

Weller Decl. ¶ C8.

As it relates to set-to-set/closed designed studies, the *Tibbs* court identified two "significant problems" that made the studies "difficult to rely upon as evidence of the reliability of conclusions regarding toolmark evidence." 2019 WL 4359486, at *15. The first concern *Tibbs* expressed was an inability to calculate a "true error rate" because closed/set-to-set studies involve an unknown number of total comparisons. Id. As Mr. Weller's declaration explains, "this is not entirely correct." Weller Decl. ¶ C9. Rather, "[t]hese studies do allow for a calculation of an overall error rate by tabulating the total number of answers and the total number of incorrect answers." *Id*.

The second concern was that in set-to-set/closed designs, examiners have all of the questioned and known samples at once and can "employ inferences gained from looking at one of the individual problems . . . to solve other individual problems." 2019 WL4359486, at *16. "In other words, the participant can rely on other unrelated parts of the puzzle – or even the puzzle as a whole – to solve and individual part of the puzzle, and thus a match determination for each of the individual problems evaluated would depend not simply on one-to-one comparisons but also on information and inferences gleaned from other individual problems . . .". *Id*. This puzzle analogy is a bad fit given the study designs at issue:

Cartridge Case Examinations Using a Declared Double Blind Format." AFTE Journal, Vol. 46, No. 2 (2014) pp. 157-175.

> A puzzle analogy implies that as one fills in the answers, then the other possibilities become less and the puzzle/test becomes increasingly easier. However, when solving a puzzle, one knows the "rules" of the puzzle from the beginning. If a validation test were designed with the same number of questioned items and knowns, each known had only one match, and the test taker knew these parameters, then the puzzle analogy would be accurate. However, this assumption is not true. For example, in Phase 1 of the Fadul cartridge case study, test takers were provided with 15 questioned and 10 knowns. Therefore, when taking the test, each of the fifteen-questioned items has 10 possible answers. Test takers were not told whether each known had at least one matching questioned item. Using inference to try and deconstruct the test could result in an error that would then be propagated throughout the test resulting in multiple false identifications. It is also worth noting that this closed set-to-set design and resulting data has been used by non-practitioner researchers when testing 3D data and the performance of their algorithms

Weller Decl. ¶ C10 (citations omitted).

PCAST also declined to consider studies involving consecutively manufactured firearms on account of alleged flaws in the design of those studies, because PCAST equated such studies to completing a puzzle.[31] PCAST does not cite a single trained firearms examiner who subscribes to this theory. Rather, such studies are considered worst-case scenario design by practitioners and researchers in the field because such consecutively manufactured weapons are more likely to have subclass characteristics and toolmarks with little change or variation from one machined part to the next and thus may result in false positive errors. Weller Decl. ¶ C3. Although the import of these worst-case-scenario validation studies were lost on PCAST, courts have appreciated the significance of the studies on reliability under Rule 702. *Otero*, 849 F. Supp. 2d at 432 ("Some of these 'validation studies' seek to validate the theory that one can individualize tools, even when comparing marks made by tools of the greatest possible

[31] Researchers have produced objective data to support the proposition that consecutively manufactured firearms produce markings with distinguishable individual characteristics. See Todd Weller et al., *Confocal Microscopy Analysis of Breech Face Marks on Fired Cartridge Cases from 10 Consecutively Manufactured Pistol Slides*, 57 J. FORENSIC SCIENCE 912, 912-917 (2012) (Govt. Ex. 19).

similarity, such as those involved in the consecutive manufacture of various firearms of the same make.").[32]

Although not considered by PCAST and *Tibbs*, this Court should use these pre-PCAST studies to take a holistic view of the field that includes studies before and after PCAST, which establish low rates of error (overall and false positive).[33] Below are some studies dismissed by PCAST and Tibbs due to alleged design flaws, that in fact establish low overall error rates.

Pre-PCAST Studies	Error Rate (Overall Error Rate)
Brundage, D.J., *The Identification of Consecutively Rifomanled Gun Barrels*, AFTEJ 30(3) (Summer 1998) at 440. (Govt. Ex. 10.) Set-to-Set Closed, Weller Decl. ¶¶ C8 & C14	0%
DeFrance and Van Arsdale, *Validation study of electrochemical rifling*, AFTEJ 35(1) (Winter 2003) at 36. (Govt. Ex. 11.) Not Designed to Measure Overall Error Rates. Weller Decl. ¶ C6	0%

[32] The study with the highest reported overall error rate (1.7%) involved this type of "closed" study design. Weller Decl., ¶ C16 (*citing* B. Maryland et al., *Validation of Obturation Marks in Consecutively Reamed Chambers*, 44 AFTE J. 167-69 (2012). Both types of set design, whether sample-to-sample/open or closed contains studies with "some false identifications and others with zero false identifications." Id., pp.15-16.

[33] In his declaration, Mr. Weller describes numerous sources of information to establish an overall view of the error rates in the field. See Weller Decl. ¶¶ C1-C38. This includes proficiency test records. "While the error rates may be less reflective of qualified practitioners (anyone can buy a test, thus trainees or other non-examiners can submit data), the data can supplement other validation/black box studies." *Id.* ¶ C35.

Smith, E., *Cartridge case and bullet comparison validation study with firearms submitted in casework*, AFTEJ 37(2) (2005) at 132. (Govt. Ex. 12.) Weller Decl. ¶ C5	0%
Hamby, J.E., Brundage, D.J., and J.W. Thorpe, The identification of bullets fired from 10 consecutively rifled 9mm Ruger pistol barrels: a research project involving 507 participants from 20 countries, AFTE Journal, Vol. 41, No. 2 (2009): pp. 99-110. (Govt. Ex. 13.) Set-to-Set Closed, Weller Decl. ¶¶ C8 & C15	0%
Mayland B., Tucker C., *Validation of Obturation arks in Consecutively Reamed Chambers*, AFTEJ 44(2) (Spring 2012) at 167-169. (Govt. Ex. 14.) Set-to-Set Closed, Weller Decl. ¶¶ C8 & C16	1.7%
Fadul, T.G., Hernandez, G.A., Stoiloff, S., and S. Gulati, An Empirical Study to Improve the Scientific Foundation of Forensic Firearm and Tool Mark Identification Utilizing 10 Consecutively Manufactured Slides, AFTEJ 45(4) (Fall 2013) at 385-87. (Govt. Ex. 15.) Set-to-Set Closed, Weller Decl. ¶¶ C8 & C17	0.064% for Phase 0.18% for Phase 2

Fadul, T.G., Hernandez, G.A., Stoiloff, S., and S. Gulati, An empirical study to improve the scientific foundation of forensic firearm and tool mark identification utilizing consecutively manufactured Glock EBIS barrels with the same EBIS pattern, National Institute of Justice Grant #2010-DN-BX-K269 (2013) at 33. (Govt. Ex. 16.) Set-to-Set Open, Weller Decl. ¶¶ C18 & C19	0.7%
Smith, Smith, Snipes, J.B., A Validation Study of The Bullet and Cartridge Case Comparisons Using Samples Representative of Actual Casework, Journal of Forensic Sciences (2016) at 5. (Govt. Ex. 18.) With-in Set Open, Weller Decl.¶¶ C33 & C34	.303% Overall error rate incl. false negatives 0.144% false positive error rate-cartridge cases 0% false identification rate-bullets

Such studies are valuable in establishing low overall error rates and high sensitivity for the field. However, as stated above, this Court need not mediate any disagreement regarding the positions taken by PCAST, because the field has since completed several studies employing the sample-to-sample/open-set study-design preferred by PCAST. Below is a table listing Ames I, and four post-PCAST studies employing the same sample-to-sample/open design. Given their sample-to-sample/open set design, these studies allow for the calculation of a false positive error rate.

Importantly, the primary issue typically presented at trial is whether the identification rendered by the firearms examiner is, in fact, correct. That is precisely the question presented in the instant case. As discussed in greater detail in the next subsection, to answer that question, the Court should inquire how often examiners err when making a positive identification. *Harris*, 502 F. Supp. 3d at 434 ("[T]he critical inquiry under this factor is the rate of error in which an examiner makes a false positive identification, as this is the type of error that could lead to a conviction premised on faulty evidence."); *Otero*, 849 F. Supp. 2d at 433 ("Indeed, for the purposes of utilizing toolmark identification in legal proceedings, the critical validation analysis has to be the extent to which false positives occur."). The four studies listed below demonstrate a low rate of false positive identifications (ranging from zero to 1.01%). ***It is important to note that these error rates do not represent the percentage of time any particular examiner would make a false positive, or that the percentage of time a lab would report out a false positive***[34]

Study	False Positive & False Negative Error Rate
Ames I-Baldwin, Bajic, Morris, and Zamzow, *A study of false-positive and false-negative error rates in cartridge case comparisons*, Ames Laboratory,	1.01% False Positive Rate 0.367% False Negative Rate

[34] *See e.g.*, Ames I at 19 ("This finding does not mean that 1% of the time each examiner will make a false-positive error. Nor does it mean that 1% of the time laboratories or agencies would report false positives, since this study did not include standard or existing quality assurance procedures, such as peer review or blind reanalysis) (Govt. Ex. 5); Ames II ("Definitive false positive error rate estimates that take examiner heterogeneity [i.e., errors were clustered among a minority of examiners] into account are 0.66% for bullets and 0.933% for cartridge cases. False negative error rate estimates are 2.87% (bullets) and 1.87% (cartridge cases)". *See also* Weller Decl. ¶¶ C22 & C29 (citing to Monson K, Smith E, Stephenson L, Chumbley LS, Bajic S, Morris MD, Zamzow D, "A Validation Study of the Accuracy, Repeatability, and Reproducibility of Firearm Comparisons" Presented at 2021 American Academy of Forensic Sciences Annual meeting, February 2021.

USDOE, Technical Report #IS-5207 (2014). (Govt. Ex. 5.) Sample-to-Sample, Open, Weller Decl. ¶¶ C21 & C22-23	
Keisler-Mark A. Keisler, et al., *"Isolated Pairs Research Study,"* AFTE Journal, Vol 50 No 1 (Winter) 2018 at 56–58. (Govt. Ex. 6a & 6b.) Sample-to-Sample, Open, Weller Decl. ¶¶ *C21 & C24-25*	0% False Positive Rate 0% False Negative Rate
Duez-Duez, Weller, Brubaker, Hockensmith, Lilien, *Development and Validation of a Virtual Examination Tool for Firearm Forensics,* J Forensic Sci, Vol 63 No 4 (July 2018). Govt. Ex. 8. Sample-to-Sample, Open, Weller Decl. ¶¶ C21 &C26-27	0% False Positive Rate Trained Examiners 0% False Negative Rate Trained Examiners * 2 False positives reported with trainee examiner
Ames II- Bajic, Chumbley, Morris, and Zamzow, *Report: Validation Study of the Accuracy, Repeatability, and Reproducibility of Firearms Comparisons, Ames Laboratory-US DOE*, Technical Report #ISTR-5220 (2020). (Govt. Ex..7.) Sample-to-Sample, Open, Weller Decl. ¶¶ C21 & C29-30	0.66% False Positive Rate – bullets 0.93% False Positive Rate -- cartridge cases 2.87% False Negative Rate – bullets 1.87% False Negative Rate – cartridge cases

Chapnick- Chapnick, Weller, Duez, Mesche, Marshall, Lilien, *Results of the 3D Virtual Comparison Microscopy Error VCMER Study for Firearm Forensics*, J Forensic Sci October 2020. (Govt. Ex. 9.) Sample-to-Sample, Open, Weller Decl. ¶¶ C21 & C28	0.433% False Positive Error Rate, 0% False Negative Error Rate

PCAST opined that from a scientific perspective, an acceptable error rate should be less than 5%. PCAST Report, p.152. As shown above, these are far below that. As low as these error rates are, the error rate in casework is likely lower. Many of these studies were designed to be more difficult than casework in several respects. For example, the Ames II study designers chose to use consecutively or closely manufactured firearms and steel ammunition. Weller Decl. ¶¶ C29 & D2.

Most importantly, these studies overestimate the potential frequency of errors in that they lack a second-level review, an important step to ensure as close to zero errors in casework as possible.[35] Labs employed by the government employ a second examiner who re-examines the results of the first examiner. See Weller Decl. ¶ H4. This is likely to decrease the rate of errors. Additionally, a respondent may elect to have his or her own expert re-examine the evidence to ensure against the unlikely occurrence of a false identification.

[35] For example, the Ames II data related to reproducibility showed that none of the false identifications were reproduced by a second examiner. Weller Decl. ¶ C31. "This strongly supports the quality control measures forensic laboratories employ, such as verification. During verification, a second examiner will review the evidence to determine if the same result is obtained as the primary examiner. The Bajic [Ames II] reproducibility data provides evidence that the forensic laboratory practice of verification can catch false positive errors." Id.

Given the low overall error rates, and low false positive error rates associated with the firearms and toolmarks identification methodology, the mechanisms in place in casework to ensure that reports do not issue with erroneous results, and the potential for independent examination by the respondent, this factor weighs in favor of admission of the testimony as proposed by the government.

b. Inconclusive determinations are necessary and appropriate conclusions, are not errors, and are considered in much of the field's research.

Error rates are a fraction of two numbers. Weller Decl. ¶ C2. "The smaller the numerator (top number) in relation to the denominator (bottom number), the smaller the calculated percentage." *Id*. To calculate a *false positive error rate*, the number of false positive determinations is divided by the total number of comparisons where the ground truth is a different source. To calculate a false negative error rate, the number of false negative determinations is divided by the total number of comparisons where ground truth is same source. This is the calculation used by Ames I, Ames II, Keisler, Duez, and Chapnick, whose studies appear in the second chart above. *See infra*, pp. 33-34. In this calculation, the denominator includes inconclusive results. To calculate an *overall error rate*, the number of wrong answers (i.e., false positive and false negative determinations) is divided by the total number of answers in the study. This is the calculation used in the studies listed in the first chart above listing pre-PCAST studies. *See infra*, pp. 31-32.

In a modified calculation, PCAST removed inconclusive results from the denominator when calculating the false positive error rate.[36] *See* PCAST Report at 91, 153; Weller Decl. ¶¶

[36] Even under this approach the false positive rate only increased from 1.01% to 1.5% in Ames I. *See* PCAST Report at 111, Table 2. As noted by United States District Court Judge Paul G. Gardephe of the Southern District of New York, even accepting these rates – which are significantly higher than the false positive error rates calculated by the authors in studies ignored by PCAST, and post-PCAST studies – the error rate calculated under PCAST's method still satisfies this *Daubert* factor. *Johnson*, 2019 WL 1130258, at * 19 ("Finally, even accepting the PCAST Report's assertion that the error rate could be as high as 1 in 46, or close to 2.2%, such an error rate is not impermissibly high."). Of course, as shown

C2 & E2. PCAST noted that calculations including and excluding inconclusive determinations both had scientific merit. See PCAST Report at 91, 153; Weller Decl. ¶¶ C2 & E2. In the context of fingerprints, PCAST noted that it focused on the calculation excluding inconclusives because "fingerprint evidence used against a defendant in court will typically be the result of a conclusive determination." PCAST Report at 91-92; Weller Decl. ¶ E2. PCAST further validated the inclusion of inconclusive determinations in the denominator by calculating false positive error rates using all conclusions (with inconclusives in the denominator), as done in Ames I, Ames II, Keisler, and Chapnick. See PCAST Report at n.272, 276, 334; Weller Decl. n.135. Importantly, PCAST did ***not*** treat inconclusive results as errors. *See* Weller Decl. ¶ E & n.135.

"Until recently the only discussion was whether to include inconclusive results in the denominator of the false positive error rate calculation." Weller Decl., ¶ E2. The defense expert in *Tibbs*, Dr. Scurich, whom the respondent in this case continues to rely on, took the unreasonable and unsupported position in Tibbs that inconclusive decisions where the ground truth is known (thus it is known whether it is same source or different source) should be included with false positives in calculating the false positive error rate. *See Tibbs*, 2019 WL 4359486 at *17.

Since *Tibbs*, Dr. Scurich's opinion has shifted. He now advocates that study designs should account for inconclusives by identifying a process to determine which inconclusive calls are (and are not) errors. *See* Itiel E. Dror & Nicholas Scurich, *(Mis)use of scientific measurements in forensic science*, FOR. SCI. INT'L: SYNERGY 333-338 (2020) (Govt. Ex. 21a.) He proposes two designs to achieve such a result. In the first, a panel of experts would make

herein, the vast universe of validation data ignored by PCAST indicate that the error rate for the field is considerably lower.

such determination before the study begins. Id. In the second, the determination would be done after participating examiners completed their examinations on a majority rules basis. Id. Thus, if a majority of examiners reported inconclusive results, the evidence would be classified as such, and any other determination would be error (i.e., identification or exclusion). *Id*.

Dr. Scurich's opinions lack common sense or support in any scientific community. Even in a study with ground truth, a particular casing may not be well marked enough to make a reliable determination. In such cases, inconclusive may be the appropriate call. (Hence, this is the reason Dr. Scurich must fall back on a consensus-based approach to argue that inconclusives should be categorized as errors). Reaching an inconclusive conclusion is not suspect or unique to firearms examination. For example, a spectrometer or genetic analyzer may measure a detectable amount of controlled substance or DNA that falls below a diagnostic threshold for the reporting of reliable results. The ability of such a device to detect a pattern is analyzed as the device's sensitivity – not its error rate.

It is only by adopting the approach advocated by Dr. Scurich that one may characterize an inconclusive decision as a false positive or other type of error and inflate the false positive error rates in a select few studies to 20 or 30%. These theories are the conception of psychologists,[37] not applied scientists, statisticians, or anyone with training or knowledge of firearms and toolmark identification. Both of Dr. Scurich's approaches are addressed below. Although it appears Dr. Scurich has since abandoned the approach taken during his testimony in *Tibbs*, it is important to address the fault of that position because the court in *Tibbs* adopted it to inflate the error rate and discount the error rate studies presented by the government.

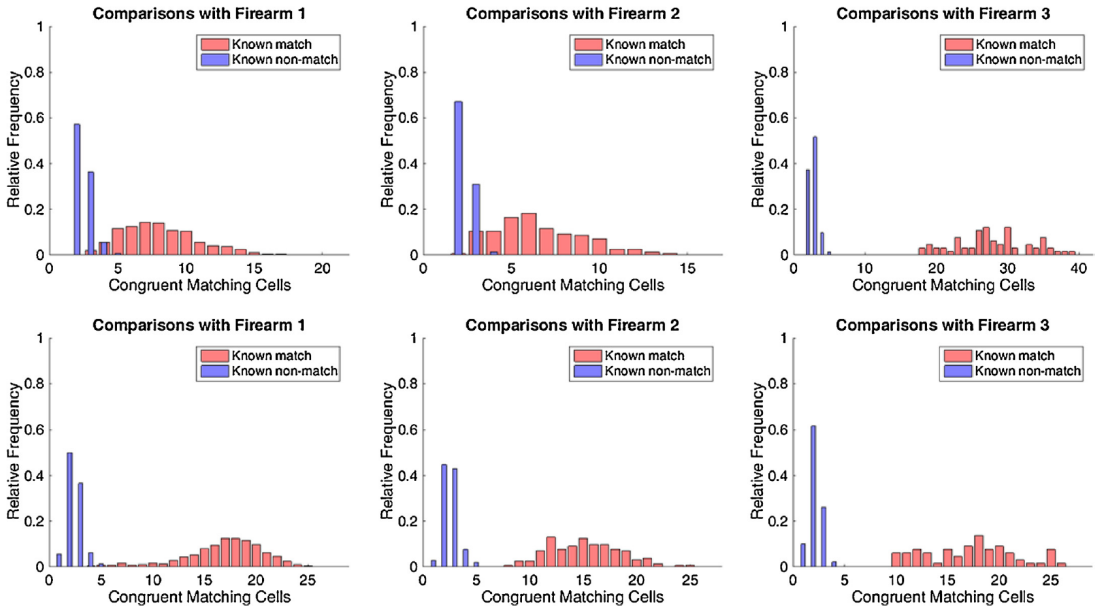
[37] Both Drs. Scurich and Dror are psychologists.

Part of why *Tibbs* strayed so far from a reasonable interpretation of the empirical data is explained by its acceptance of Dr. Scurich's theory that under "pristine" conditions of validation, all inconclusive determinations are a type of error and should be calculated in deciding the error rate for the field. *See Tibbs*, 2019 WL 4359486, at *18. This approach completely changes the playing field for firearms examiners whose training instructs that inconclusive is an appropriate response. In fact, under the range of acceptable AFTE conclusions, the "inconclusive" category recognizes three subcategories, i.e., some level of agreement in individual characteristics, lack of any discernable agreement or disagreement in individual characteristics, and some level of disagreement in individual characteristics, see AFTE GLOSSARY at 94; *see also*, Duez Study, p.1075 (listing same). It strains logic to change the playing field *after* examiners have participated in validation studies where they are instructed that inconclusive is an appropriate response.

More importantly, the theory that samples used in validation are somehow "pristine" rendering all "inconclusive" decisions errors is refuted by objective, 3D empirical data. The following series of graphs show histograms between same-source (red) and different source (blue) cartridge cases from the Ott et. al. manuscript. *See* Daniel Ott et al., *Applying 3D measurements and computer matching algorithms to two firearm examination proficiency tests*, FOR SCI. INT'L 98-106 Fig 6 (2017) (Govt. Ex. 20); Weller Decl. ¶ D4. The further to the right within each graph, the higher the CMC score (which is a statistical measure of toolmark similarity between two samples). Weller Decl. ¶ D4. The data shows that most very often known matches are separated from known non matches, i.e., red and blue are separated. *Id*. "However, a minority of known matches have a low number of matching features, i.e., a low CMC score."

Weller Decl. ¶ D4. The importance of the information in the graphs below is as follows, as explained by Mr. Weller, *Id.*:

> This is significant because the samples shown here were all collected under the controlled, non-casework environment that should have resulted in the supposed pristine samples. Additionally, the ground truth is known about these samples: researchers know which are same source and which are difference source. Despite these non-casework and controlled conditions, there are occasional known-matching samples that do not have sufficiently similar microscopic marks to be algorithmically separated from the non-matching samples. In other words, the comparison of these known-matches are not identifiable and are correctly categorized as inconclusive.



Knowledge of ground truth is simply not a guarantee that any given sample will have reproducible marks. As noted by Mr. Weller in his affidavit, in Ames II, "steel-based ammunition was purposely selected for the characteristic reproducing less toolmarks." Weller Decl. ¶ D2. Some but not all studies perform quality control on a percentage of samples, but this does not guarantee that "*all* the samples will have sufficient markings for source attribution." *Id.* ¶ D3. For example, the two Fadul studies considered this issue and noted: "The possibility exists that the questioned casings and known standards failed to mark clearly. Since every set

was not microscopically examined to ensure that the casings were comparable and identifiable, some sets may have contained casings that were not suitable for identification."[38] Ames I and Ames II did not prescreen samples at all. *See* Weller Decl. ¶ D3. Ames I noted that toolmarks are not always reproduced, despite knowledge of ground truth: "As is determined in this study, there are also a significant number of times that the firearm fails to make clear and reproducible marks (which very well might have happened for a questioned case)." Baldwin, et. al. (Ames I), p.6. (Govt. Ex. 5.) The Smith Study (2016) used firearms from crimes-related cases, see Tasha P. Smith et al., *A Validation Study of The Bullet and Cartridge Case Comparisons Using Samples Representative of Actual Casework*, J. FORENSIC SCI. 940 (2016), and the Keisler Study used confiscated and donated firearms, *see* Keisler et al., at 56 (Govt. Ex. 18.). The 2016 Smith Study noted that the firearms used in that study "were circulated in the general population and subjected to use, corrosion and abuse similar to that observed in a typical case." Tasha P. Smith T. et al., at 940 (Govt Ex. 18).

The passage below from the attached declaration of Dr. Max Morris further highlights the flaw in Dr. Scurich's approach. Unlike the defense expert in *Tibbs* (Dr. Scurich), Dr. Morris is a PhD statistician with a specialization in experimental design who has been involved in firearms and toolmark research, including Ames I, the Addendum to Keisler, and Ames II. He states:

> Using such a combined count as the numerator in an "error" proportion runs counter to any reasonable definition of "error." . . .
>
> In the referenced report of Baldwin et al. [Ames I] (of which I am a coauthor), our more straightforward approach was to see "exclusion", "identification" and "inconclusive" as three distinct outcomes that can result from different-source comparisons, and we used 22/2178 = 1.01% as the proportion of all such comparisons resulting [in] a (clearly) erroneous conclusion. As Scurich states in a

[38] Thomas G. Fadul Jr. et al., *An empirical study to improve the scientific foundation of forensic firearm and tool mark identification utilizing 10 consecutively manufactured slides*, 45 AFTE J. 376, 383 (2013) (Govt. Ex. 15.)

> footnote to his comments, an alternative approach would be to cite 22/1443 = 1.52% as the proportion of erroneous calls among only those comparisons judged to have sufficient probative value to support a conclusion regarding source, a complication we chose not to include in our report, but which might be appropriate under some circumstances if clearly explained where cited. Either of these approaches limits the characterization of "error" to what is clearly an incorrect assessment regarding source. Because an "inconclusive" determination cannot be regarded as a claim that contradicts fact, "error" proportions computed by pooling such counts with factual errors in the numerator (e.g., 22+735 in the
>
> Baldwin common-source samples) are **logically flawed and seriously misleading with respect to any reasonable interpretation of the word "error."**

See Dr. Max Morris Declaration 1 at 1-2 (Govt. Ex. 22a) (bolded emphasis added) (referred to herein as Morris Decl. 1). Dr. Morris furthermore explains why Tibbs (and the respondent) missed the mark in their evaluation of inconclusive decisions:

> [I]nconclusive calls are a common and necessary part of firearms examinations, both in controlled studies and in casework. In casework, such calls may be necessary due to material damage that can occur at the crime scene (which, if so severe as to obliterate the firearm-produced marks, may result in an "unsuitable for examination" evaluation rather than an inconclusive call). But even in controlled studies based on material produced in the laboratory, the random variation in tool marks left on bullets or cartridge cases can lead to comparison sets that cannot definitively be classified as exclusions or identifications. The type of firearms and ammunition used in these studies (along with other factors) influences the number of such calls. Characterizing the probative value of the tool marks left on any specific bullet or cartridge case, in the context of the required comparison(s), is a critical part of forensic examination. Classification of a comparison set as inconclusive, while frustrating from a legal perspective, is simply appropriate in some cases. To ignore this fact would be essentially equivalent to requiring an examiner to yield an exclusion or identification judgement when neither is physically justifiable. . . .
>
> To be sure, inconclusive calls are never the most *desirable* result of a forensic examination, just as it is disappointing to receive the result of an expensive medical diagnostic procedure that does not provide the hoped-for information regarding a patient's condition. But to suggest that inconclusive calls are never the most *appropriate* calls for any specific set of physical evidence, whether in casework or controlled studies, does not coincide with the process and goals of forensic examination. . . .

> On page 41, "While the [Tibbs] Court does not accept Dr. Scurich's inclusion of inconclusives in the false positive error rate, it agrees with his essential premise that such responses should represent an error by the examiner." While I certainly concur with the first part of this statement, I strongly disagree with the second part, as explained above. In fact, a more honest reporting of these studies is to refuse to combine the two conclusions of lesser desirability (as the Court agrees), but to understand that inconclusive calls are a distinctive and necessary (if frustrating) part of the evaluation process, and that proportions of conclusive errors, conclusive correct evaluations, and inconclusive calls are *all* important characterizations of what can be expected of firearms forensics. The frequency of inconclusive results tells us how often we should expect firearm examinations to be of limited or null probative value. **The relative frequencies of correct and incorrect definitive calls are estimates of the error probabilities we can expect among those examinations of material that are judged by the examiner to support such calls.**

See Dr. Max Morris 9/17/19 Declaration 2 at 1-2 (Govt. Ex. 22b) (emphasis added) (hereinafter Morris Decl. 2). No applied scientist, researcher or practitioner has ever advocated Dr. Scurich's use of inconclusive conclusions as false positive errors in the field of firearms and toolmark analysis, nor in any other forensic discipline for that matter.[39]

As it relates to Dr. Scurich's most recent approach to inconclusive calls, i.e., that studies should account for them by identifying a process to determine which inconclusive calls are (and are not) errors, it has been discredited in a published letter to the editor by Mr. Weller and Dr. Morris and in an article by Alex Biedermann and Kyriakos Kotsoglou. Dr. Morris and Mr. Weller note three overall issues with Dr. Scurich's current approach. First, it unmoors error rates from ground truth by proposing a study design that requires a ground truth regarding the sufficiency of information imparted to bullets or cartridge casings when such ground truth is not available. *See* Morris, M, Weller, T. *Commentary on I. Dror, N. Scurich: (Mis)use of scientific*

---

[39] Dr. Petraco, who teaches statistics to graduate students, testified that Dr. Scurich's approach is so out of touch with reality that if a student presented a thesis with Dr. Scurich's calculations of error rates, not only would he fail the student but he would suggest that the student consider another profession. See Defense Appendix C, Tibbs Transcripts, 5/19/19, p.30 (Transcript), 32/921 (Appendix C).

measurements in forensic science, Forensic Science Intl. 2 Synergy 2020, p.701. (Govt. Ex. 21b.) Second, the two proposed study designs "substitute the concept of reproducibility for … accuracy." *Id*. "That is, it considers as correct (an accuracy concept) calls that represent agreement among multiple examiners (the essence of reproducibility)." *Id*. Third, the proposed study design is circular in that it "assess[es] examiner accuracy via a process that assumes examiner accuracy." *Id*.

Biedermann and Kotsoglou systematically dismantle Scurich's and Dror's proposal in *Forensic science and the principle of excluded middle: "Inconclusive" decisions and the structure of error rate studies*, Forensic Science International: Synergy 3, March 2021. They conclude in part as follows:

> In all, our analysis does not leave much intact from recent attempts to label "inconclusives" as errors. To be clear, we do not argue that we should not focus on errors or error rates. Quite the contrary: our point is that recording errors is important, but errors of the suitable kind, and task-specific information regarding such errors. By this we mean genuine errors determined with respect to ground truth, rather than with respect to artificial category labels which lack a coherent conceptual basis and, thereby, lead to paradoxes in practice (e.g., penalising factually correct responses).
>
> Similarly, arguing against the scoring of "inconclusives" as potential errors does not mean to dismiss the focus on "inconclusive" decisions. They are an important category of decisions precisely because they allow "identifications" and "exclusions" to be used only when stringent requisite conditions are satisfied, not least because we value adverse outcomes of decisions such as 'identification', i.e. false positives, as highly undesirable. This assertion is not based on mere intuition but can be demonstrated through a decision-theoretic analysis.

See Alex Biedermann & Kyriakos N. Kotsoglou, *Forensic science and the principle of the excluded middle: "Inconclusive" decisions and the structure of error rate studies*, FORENSIC SCI. INT'L SYNERGY at 9 (2021) (Govt Ex. 21d).

Importantly, inconclusive results may be accounted for in the error rate studies by calculating metrics such as sensitivity and specificity. Id. "Sensitivity is the portion of true

same-source samples that were called as identifications. Specificity is the portion of true different-source samples that were called as eliminations." *Id.* For example, Keisler reported sensitivity and specificality rates of 99.74% and 79.86% respectively. *See* Keisler M et al. pp. 56-58 (Govt. Ex. 6a); Letter to the Editor "Isolated Pairs Research Study" AFTE Journal Vol 50 No 3 (Winter) 2018 p.131 (Govt. Ex. 6b). Such reports allow the consumer of the study to understand and appreciate the diagnostic sensitivity of human examiners without resorting to the inaccurate rubric that would label samples below that threshold as "errors".

The respondent and Dr. Scurich also argue that inconclusive rates are higher in studies than casework. They speculate that examiners in studies skip hard questions because they know only incorrect identifications and exclusions count as errors. Def. Mtn. p.20. As a result, they assert, this provides a misleading error rate. *Id.* This theory is speculative. The respondent provides no support other than circularly pointing to the article by Scurich and Dror. *Id.* Even if it were accurate that inconclusive calls increased in study examinations, the most reasonable explanation is not that proposed by the respondent and Dr. Scurich. Rather, a more reasonable explanation rests on the deliberate difficulty of the examinations in the studies. Many of the studies are designed to be more difficult than casework. For example, in Ames II the study designers almost exclusively used consecutively and closely manufactured firearms as well as steel ammunition. As a result, the barrels were likely to leave very similar markings. Additionally, steel ammunition, not commonly found in casework, does not mark as well as softer metals such as brass. In casework, it is exceedingly unlikely that one case would include steel ammunition at all, let alone fired from two consecutively or closely manufactured firearms.

Under these more challenging conditions, one should not only expect more inconclusive results, but also that examiners may be less accurate than in casework. The very low rates of overall and false

positive errors in these studies, under such circumstances, confirms the reliability of firearms and toolmark identification.

3. Peer-Review

Studies testing the foundational research of firearms and toolmark identification and examiners' ability to associate a cartridge case/bullet to a particular firearm have been authored by a variety of scientists, most of whom hold PhDs in a wide range of the applied sciences -- including statistics, engineering, quantum chemistry, mathematics, physics, computer science, and physical chemistry -- and have undergone peer-review in a variety of scientific journals. *See* Weller Decl. Appx. B (Non-Exhaustive List of Firearm and Toolmark Research by Individuals Associated with Non-Crime Lab Institutions.) In fact, nearly all of the citations in Mr. Weller's declaration are from peer-reviewed scientific journals.[40] Moreover, the field's research into objective means to quantify and validate firearms and toolmark identifications through the use of 3D topographical imaging, sophisticated computer algorithms, and the implementation of statistical tools through the establishment of firearms databases is conducted in peer-reviewed scientific journals as well.

Given the breadth of peer-reviewed publication, courts have consistently held that this factor weighs in favor of admissibility. *Harris*, F. Supp. 3d at 39.; *Johnson*, 2019 WL 1130258, at *16*; *Romero-Lobato*, 2019 WL 2150938, at *5 ("Several published federal decisions have also commented on the AFTE Journal, with all finding that it meets the *Daubert* peer review element."); *Ashburn*, 88 F.Supp.3d at 246 ("The court finds that the AFTE methodology has been published and subject to peer review, weighing in favor of admission . . ."); *Diaz*, 2007 WL 485967, at *8 ("The fact that the articles submitted to the AFTE Journal are subject to peer

[40] Mr. Weller has served as peer reviewer for both the Journal of Forensic Science and AFTE Journal. Although the process differs, the result is the same: articles are reviewed by subject matter experts who judge the paper on its scientific merit. Mr. Weller has accepted, revised, and rejected papers for both.

review weighs strongly in favor of admission."); *Otero*, 849 F. Supp. 2d at 433 (noting AFTE Journal's formal process for the submission of articles); *Taylor*, 633 F. Supp. 2d at 1176 (finding the peer review factor "clearly weighs in favor of admissibility.").[41]

PCAST and the outlier decisions cited by the defense seek to discard the vast body of peer-review and publication by taking the position that only studies that undergo what they deem to be "meaningful" peer review should be considered, and all other scientific data should be discarded altogether. There are several flaws in this approach. First, and foremost, as discussed at length in Mr. Weller's declaration, several studies have been peer reviewed and published in a variety of scientific journals. For example, both Duez and Chapnick in which Mr. Weller collaborated with Dr. Lilien and other scientists, were published post-PCAST in the Journal of Forensic Sciences. Not only do these and other post-PCAST data dispel the theory that open set studies result in higher error rates and lower sensitivity than closed set studies, but the data survived the very type of peer-review that the critics such as Judge Edelman in *Tibbs* deemed "meaningful." Yet, none of the decisions relied upon by the respondent cite a single post-PCAST scientific study published in a scientific journal.

Moreover, the idea that only one particular type of peer review is "meaningful" is simply wrong. Dr. Bruce Budowle is in a unique position to comment on the peer-review process as the most published forensic DNA scientist in the world who sits on the editorial boards of several scientific journals. Dr. Budowle Decl. (Gov. Ex. 23) at 1. According to Dr. Budowle, the *Tibbs'* opinion's critique of certain peer review processes is out of step with how the scientific

---

[41] Moreover, the defense may also elect to have its own expert examine and attempt to disprove the conclusions proffered by the prosecution's expert – an additional layer of peer-review. *John*, 597 F.3d at 276 ("[Appellant] had the opportunity to analyze the fingerprint evidence herself and question its validity."). To date, the defense has not even requested to have an expert independently examine the firearms evidence, much less proffer a contrary opinion by a qualified firearms expert.

community evaluates peer-review. Leaders in the scientific community, including Dr. Budowle, do not embrace the idea that only double-blind peer-review constitutes "meaningful" review; rather all peer review has value. *Id.*, at 2. Having reviewed Dr. Budowle's affidavit, Judge Contreras questioned "whether excluding certain journals from consideration based on the type of peer review the journal employs goes beyond a court's appropriate gatekeeping function under *Daubert*." *Harris*, 502 F.Supp.3d at 40.

Additionally, *Tibbs* was critical of the AFTE journal because it did not previously employ double-blind peer review. However, there is no consensus in the scientific community that double-blind peer review is the only meaningful kind; rather, there are three commonly used forms of peer review, one of which is the open peer-review used (at the time) by the AFTE journal. See Weller Decl. ¶¶ F1-F4. As noted by Mr. Weller, Wiley Publishers, which publishes more than 1600 academic and scientific journals, outlines the pros and cons of each type of peer-review and notes that the use of open peer review is growing. *Id.* ¶ F1. Moreover, for many prestigious journals, double-blind peer review is a recent phenomenon. For example, the prestigious journal Nature first instituted double-blind peer-review in 2015. Even now, such review is voluntary for authors, most of whom do not opt for it (from March 2015 to February 2017 only 12% of Nature authors opted for double-blind review). *Id.* ¶ F3. Does this mean that every scientific discipline that published in Nature prior to 2015, or studies in which the authors did not opt for double-blind review, were somehow deficient under *Daubert*? This would require reevaluating, e.g., whether DNA is a double helix. *See* Watson, J. D., & Crick, F.H. C. A Structure for deoxyribose nucleic acid, Nature 171, 737-738 (1953). Dr. Budowle echoes this sentiment: "[E]ach [form of peer-review] has some merit; however, there is no consensus at this

time on which approach is superior. Nonetheless, they all serve as part of the peer review process." Dr. Budowle Decl. at 1 (Govt. Ex. 23.)

Finally, the opinion that only double-blind peer-review amounts to a "meaningful" review is belied by studies that were initially published in the AFTE Journal and subsequently published in the Journal of Forensic Science with no alterations to the design study. See *Harris*, 502 F.Supp.3d at 40.[42] For example, Hamby et al. *A Worldwide Study of Bullets Fired From 10 Consecutively Rifled 9MM RUGER Pistol Barrels – Analysis of Examiner Error Rate*, 64 J. FORENSIC SCI (2019) (Gov. Ex. 24), is an update of a continuing study that has been ongoing for over twenty years. *Id.* at 551. The study used the consecutive barrel (worst-case scenario) design. *Id.* Although PCAST was critical of this design, the study survived the double-blind peer-review process in JFS. *Id.* The first four installments of the ongoing study consisted of two AFTE Journal publications, a dissertation, and a presentation at an AFTE training seminar. Id. at 557. According to PCAST (and *Tibbs*), the type of peer-review involved in the first four rounds of this study would be a basis to neglect the data altogether. However, the fifth installment of the study (which included all prior four installments) survived double-blind peer review at JFS. Notably, the study design remained unchanged from the earlier installment published in the AFTE Journal. *Id.* (referring readers to the earlier studies "for the complete design of the study"). This illustrates two points: 1) the data first published in the AFTE Journal was worthy of double-blind peer review later published in JFS and 2) the scientific community does not

[42] The Court in *Harris* stated, "Compellingly, the government also refuted the allegation by Judge Edelman in *Tibbs* that the AFTE Journal does not provide 'meaningful' review, by bringing to the Court's attention a study that was initially published in the AFTE Journal, and then was subsequently published in the Journal of Forensic Science with no further alterations."

subscribe to the PCAST theory (adopted by the defense) that only black box studies are worthy of scientific publication.

This Court should further reject claims that "the AFTE Journal's use of reviewers exclusively from within the field to review articles created for and by other practitioners in the field greatly reduces its value as a scientific publication, especially when considered in conjunction with the general lack of access to the journal for the broader academic and scientific community as well as its use of an open review process." *See Tibbs*, 2019 WL 4359486, at *10. The assumptions underlying this conclusion in *Tibbs* are simply inaccurate, rendering the related conclusions unreliable. In contrast to the understanding of the *Tibbs* Court, the AFTE Journal has a free, publicly accessible online searchable index.[43] Weller Decl. ¶ F5. The AFTE Journal is also listed in the SCOPUS database, "the largest abstract and citation database of peer-reviewed literature: scientific journals, books and conference proceedings."[44] *Id.* It is publicly accessible and non-AFTE-members can subscribe to it, and search for, find, and purchase individual articles. AFTE journal subscriptions have been available to non-AFTE members since at least

[43] https://afte.org/afte-journal/searchable-journal-index

[44] A list of SCOPUS "sources," available at https://www.scopus.com/sources.uri provides a list of "sources", showing the AFTE Journal is included in this database. "Scopus is the largest abstract and citation database of peer-reviewed literature: scientific journals, books and conference proceedings. Delivering a comprehensive overview of the world's research output in the fields of science, technology, medicine, social sciences, and arts and humanities, Scopus features smart tools to track, analyze and visualize research." https://service.elsevier.com/app/answers/detail/a_id/15534/supporthub/scopus/#tips, accessed 1/14/22.

October 1989.[45] Weller Decl. ¶ F5. The online, searchable index where pdf versions of articles can be purchased by non-members has been available since 2011.[46] *Id.*

In further contrast to the Tibbs Court's (mis)understanding, the AFTE Journal is available at academic institutions. *See id.* It may be freely searched on WorldCat.org and search results show the AFTE Journal is available at Cal State Sacramento, Nebraska Wesleyan University, University of Central Oklahoma, Truman State University, Grambling State University, University of Wisconsin-Platteville, Case Western Reserve University, Marshall University, Radford University, University of Toronto, West Virginia University, Stetson University College of Law, George Mason University, George Washington University, the Library of Congress, Syracuse University, Virginia Commonwealth University, Florida International University, John Jay College, The British Library (UK), University of Wolverhampton (UK), Cranfield University (UK), Bibliotheque de l'EPFL (Switzerland), ESR Mt Albert Science Center (New Zealand), ESR Kenepuru Science Center (New Zealand), NT Police, Fire and Emergency Services Library (Australia) and the University of Western Australia. Weller Decl. ¶ F7.

Tibbs wrongly concluded that AFTE and the AFTE Journal is an insular organization comprised solely of examiners in the field, who isolate themselves from the greater scientific community and academics, and who are motivated to confirm their own beliefs, "comparable to talk within congregations of true believers" rather than by a sincere desire to engage in critical scientific review and debate. *See Tibbs* 2019 WL 4359486, at *10. *Tibbs* ruled on this notion,

---

[45] See "Publication Information" from the AFTE Journal, October 1989, Volume 21, No 4: "Members of the Association of Firearm and Tool Mark Examiners receive a subscription to the AFTE Journal as a part of their dues…A subscription to the AFTE Journal is available by contacting the Editor: M James Kreiser, 2168 South 9th Street, Springfield Illinois, 62702, U.S.A. The subscription rate is $50.00 per calendar year."

[46] Clow, C., Letter to the Editor: *Message from the Chair of the AFTE Journal Index Committee*. The online AFTE Journal Index was launched on January 1, 2011 and is currently available at https://afte.org/afte-journal/searchable-journal-index. Individual articles can be purchased for $5.00 each.

along with the inaccurate assumption that only double-blind peer-review is valuable, to reduce the value of and ignore scientific data published in the AFTE Journal. *See id.* Given the faulty factual basis for this conclusion, this Court should disregard *Tibbs* and consider the many AFTE Journal articles, along with relevant articles in other publications in determining admission of the testimony at issue. Ignoring such articles and the data therein simply because of the journal of publication is not good science. It is more akin to ignoring data simply because it conflicts with the desired outcome or hypothesis. The Court should not rely on conclusions reached through the application of such unscientific principles.

4. Standards

Standards and controls for the firearms and toolmarks profession are published and maintained from several sources. AFTE has published the following standards for professional guidance and use:

- AFTE Training manual: 166-page document outlining all steps a new trainee should undertake prior to starting casework.
- AFTE Technical Procedures Manual: 116-page document providing technical procedures for typical examinations that may occur in firearms and toolmark identification laboratories.
- AFTE Glossary: 244-page document providing the profession with standardized terminology and definitions.[47]
- AFTE Theory of Identification:[48] First published in 1992, the Theory of Identification distills the essence of firearms identification into several paragraphs. The Theory of Identification describes the basis for an identification as well as the limitations of the profession (i.e., that identification/same-source conclusions are not absolute).

[47] The AFTE Training Manual, Technical Procedures Manual and Glossary are too large to be attached to this pleading. They are available for download at www.AFTE.org.

[48] Committee for the Advancement of the Science of Firearm & Toolmark Identification, *Theory of Identification as it Relates to Toolmarks: Revised*, 43 AFTE JOURNAL at 287 (2011).

Weller Decl. ¶ G1. Another source of standardized guidelines was established by the Scientific Working Group for Firearms and Toolmarks (SWGGUN). *Id.* SWGGUN was a committee of firearms and toolmark examiners whose responsibility was to publish guidelines for the firearm and toolmark community. It has been replaced by the Organization of Scientific Area Committees (OSAC) for which Mr. Weller currently serves as the Vice Chair, and previously served as the Chair. SWGGUN guidelines are on the OSAC website.[49] OSAC is in the process of revising and writing discipline-specific standards. *Id.* It has published one standard to the OSAC registry and 11 draft standards and is currently drafting additional best practices and standards. Id. National Institute for Standards and Technology also provides metrological standards. It currently offers a standard bullet and cartridge case and is working to produce a new set of reference standards for use in emerging 3D technology.[50]

This is one factor where courts have come to different conclusions. *Compare Johnson*, 2019 WL 1130258 at *17-18 (standards weighs in favor of admissibility) *with Harris*, 502 F. Supp. 3d at 41 ("While a close call, the Court finds that the lack of objective standards ultimately means this factor cannot be met.") However, disagreement between courts on this one factor has not prevented the vast majority of other courts from finding that firearms and toolmark identification satisfies *Daubert*. This is not surprising because the lack of objective standards or subjectivity of a methodology has little to do with whether the methodology has been tested, found to have low error rates in ground truth studies, and been subjected to peer review and scientific publication. *Harris*, 502 F. Supp. 3d at 42 ("It should be noted, however, that even if

[49] https://www.nist.gov/topics/forensic-science/firearms-and-toolmarks-subcommittee

[50] Stocker, M. et al., *Addressing Challenges in Quality Assurance of 3D Topography Measurements for Firearm and Toolmark Identification,* 50 AFTE JOURNAL 104-111 (2018).

this factor cannot be met, a partially subjective methodology is not inherently unreliable, or an immediate bar to admissibility. Rule 702 'does not impose a requirement that the expert must reach a conclusion via an objective set of criteria or that he be able to quantify his opinion with a statistical probability.'") (quoting *Romero-Lobato*, 379 F. Supp. 3d at 1120).

5. General Acceptance

Firearms and toolmark identification is practiced by accredited laboratories in the United States and throughout the world, including England (Scotland Yard), New Zealand, Canada, South Africa, Australia, Germany, Sweden, Greece, Turkey, China, Mexico, Singapore, Malaysia, Belgium, Netherlands, and Denmark. Weller Decl. ¶ H1. In the United States alone, there are 233 laboratories with firearm and toolmark accreditation certificates, all of which routinely identify bullets or casings to a particular firearm. *Id.* The firearms units do not exist in a vacuum, but rather are part of a greater scientific accreditation umbrella; virtually all of these accredited firearms units function within a larger forensic laboratory offering a multitude of accredited scientific units, *e.g.,* chemistry, DNA, latent fingerprint identification, etc. In this area alone, FBI and ATF maintain accredited firearms and toolmark units, along with a variety of accredited forensic disciplines that support local and federal investigations. Notably, PhD scientists in various applied scientific disciplines are in charge of many of these laboratories. In addition, as discussed supra, scientists who design, conduct, and publish validation studies in the area of firearms and toolmark identification accept the AFTE method of identification, as do leading scientific working groups such as the OSAC, which is comprised of a variety of applied scientists.

6. Therese Moynihan reliably applied the Firearms and Toolmark Identification methodology.

Ms. Moynihan applied the AFTE Theory of Identification reliably. Ms. Moynihan is an extremely experienced and qualified firearm and toolmark examiner who has practiced for more than fourteen years. From 2007 to 2020 she was a forensic scientist in the Firearm and Toolmark Section of the Virginia Department of Forensic Science. Since March of 2020, Ms. Moynihan has been a Firearm and Toolmark Examiner with the Firearm & Toolmark Section of the Bureau of Alcohol, Tobacco, Firearms & Explosives Agency. Both of those agencies are accredited. Ms. Moynihan has testified approximately 100 times in various circuit courts in Virginia, as well as Federal District Court in Virginia. She has always been qualified when offered as an expert.

Ms. Moynihan is certified in Firearm Evidence Examination and Identification and Toolmark Evidence and Identification by AFTE. To obtain such certification an examiner must pass written and practical examinations in the area of certification and recertify every five years. In addition, Ms. Moynihan has always passed her proficiency tests. In sum, she can be trusted to apply the principles of firearm and toolmark identification reliably.

In this particular case, she conducted the examination consistent with accepted practices and procedures in the field. As indicated above, for more than fourteen years she has worked and continues to work for accredited labs. She applies the procedures used in such labs to her work. Ms. Moynihan's training and experience provide a substantial basis to conclude he reliably applied the principles of firearm and toolmark identification in this case.

In addition, Ms. Moynihan documented her work in this case and all of her conclusions were verified by another qualified examiner, Arnold Esposito of the ATFE's Firearm & Toolmark Section.

In sum, the identification conclusions and other results to which Ms. Moynihan will testify in this case have been examined by two highly qualified, AFTE certified examiners, who have worked and continue to work in accredited labs. They have extensive knowledge of the procedures used at accredited labs and acted consistently in conducting the examinations in this case. Thus, the Court can have a degree of confidence that the examiners in this case reliably applied methodology underlying firearms and toolmarks identifications.

7. 2008 Ballistic Imaging Report & 2009 NAS Report are Outdated by Over a Decade

The 2008 Ballistic Imaging Report is outdated by over a decade. More to the point, Dr. John E. Rolph, the Chair of the Ballistic Imaging Report, put to rest (shortly after the report's issuance) any effort by litigants to use the report as a statement against the scientific validity of firearms and toolmark identification:

> The statement in the Report that the "validity of the fundamental assumptions of uniqueness and reproducibility of firearm-related toolmarks has not been fully demonstrated" (Report at 3-22) was not made in the context of assessing the admissibility of firearms-related evidence. Indeed, the Report states clearly that *"this study is neither a verdict on the uniqueness of firearm-related toolmarks generally nor an assessment of the validity of firearms identification as a discipline.* Our charge is to focus on 'the uniqueness of ballistic images' –that is, on the uniqueness and reproducibility of the markings (toolmarks) left on cartridge cases and bullets as they are recorded or measured by various technologies ...." Report at 1-5 (emphasis in the original)….
>
> The Committee's cautionary statement [about not casting conclusions in terms of absolute certainty, to the exclusion of all other firearms, or implying a zero error rate] is not a commentary on the admissibility of firearm-related toolmark evidence. In the Committee's view, "statements on toolmark matches (including legal testimony) should be supported by the work that was done in the laboratory, by the notes and documentation made by examiners, and by proficiency testing or established error rates for individual examiners in the field and in that particular laboratory." Report at 3-23 to 3-24.

See Sworn Statement of Dr. John E. Rolph (Gov. Ex. 23) ¶¶ 6 & 10 (quoting portions of the 2008 NAS Report). Courts have relied upon Dr. Rolph in placing the 2008 NAS Report in proper

context. *See United States v. Casey*, 928 F. Supp. 2d 397, 400 (D.P.R. 2013) (Dr. Rolph's statements greatly undermine the portions of the 2008 NAS report upon which defendant … relies]."); *State v. Langlois,* 2 N.E.3d 936, 945 (Ohio Ct. App. 2013) ("[T]he 2008 NRC report addressed the issue of establishing a nationwide database for the computer imaging of bullets. The report's primary focus was not firearms identification, comparative ballistics, or tool mark analysis."); *United States v. Taylor*, 663 F. Supp. 2d 1170, at 1175-76, 1179-80 (D.N.M.) (holding firearms evidence admissible after considering, *inter alia*, the 2008 Ballistic Imaging Report); *United States v. Otero*, 849 F. Supp. 2d 425, 430, 438 (D. N.J.) (same).[51]

Similarly, the 2009 NAS Report is only one year less outdated than the 2008 Ballistic Imaging Report. Like the earlier NAS report, defense attorneys did not gain any traction in attempts to use the report as a basis to exclude firearms and toolmark identification testimony in the manner respondent is attempting to do so here. *See United States v. Ashburn*, 88 F. Supp. 3d 239, 274 (E.D.N.Y. 2015) (holding that firearms and toolmark identification is a proper subject of expert testimony under Rule 702 and *Daubert* after considering the 2009 NAS Report); *State v. Romero*, 341 P.3d 498, 498 (Ariz. App. Div. 2 2014) (same); *United States v. Casey*, 928 F. Supp. 2d 397, 400 (D.P.R. 2013) (ballistics evidence admissible under Rule 702 and *Daubert*); *State v. Langlois*, 2 N.E.3d 936, 950 (Ohio Ct. App. 2013) (firearm and toolmark identification satisfies the test for reliability under Rule 702); *Otero*, 849 F. Supp. 2d at 438 (D.N.J. 2012) (essential foundations for the admission of expert testimony under Rule 702 established by the government); *United States v. Willock*, 696 F. Supp. 2d 536, 568 (D. Md. 2010); *Taylor*, 663 F.

---

[51] In *Tibbs*, Judge Edelman concluded that the 2008 NAS Report "directly addressed the sufficiency of the published studies purporting to show a low error rate in the field of firearm and toolmark identification." Judge Edelman's holding failed to consider the plain and unambiguous sworn statement by Dr. Rolph, refuting his finding, as well as judges who cited Dr. Rolph's statement in coming to the opposite conclusion. (*See* Govt. Ex. 25, Dr. Rolph's Statement.)

Supp. 2d at 1180 ("The evidence before the Court indicates that when a bullet is fired from a gun, the gun will impart to the bullet a set of markings that is, at least to some degree unique to that gun. The evidence further indicates that an experienced firearms examiner can make observations of those markings, using a method that has been peer-reviewed, that allow him, in some cases, to form an opinion that a particular bullet was or was not fired from a particular gun. The court therefore concludes that the firearms identification testimony is admissible under Rule 702 and *Daubert*."); *State v. Lee*, 2017 WL 1494012, *10 (4th Cir. Apr. 26, 2017) ("[E]ven after publication of the NAS Report, courts have addressed, in detail, the reliability of [firearms and toolmark identification] testimony and ruled it admissible, although to varying degrees of specificity."); *Spears v. Ryan*, 2016 WL 6699681, *5 (D. Ariz. Nov. 15, 2016) ("[T]he NAS Report would have had no effect on the admissibility of the toolmarks evidence in this case."); *Napier v. Commonwealth*, 2014 WL 3973113, *9 (Ky. Ct. App., Aug. 15, 2014) (It was not the purpose of the 2009 NAS Report to opine on the long-established admissibility of toolmark and firearms testimony in criminal prosecutions and there was no error in taking judicial notice of scientific reliability of ballistic analysis under *Daubert*); *United States v. Sebbern*, 2012 WL 5989813, at * 8 (E.D.N.Y. 2012) (no need for a *Daubert* hearing before admitting ballistics evidence); *United States v. Cerna*, 2010 WL 3448528, at *5 (N.D. Cal. Sept. 1, 2010) (the NAS report "does not necessitate exclusion of expert [ballistics] testimony").

For these reasons, Judge Contreras concluded that the 2008 Ballistic Imaging Report and 2009 NAS Report are "outdated by over a decade due to intervening scientific studies," and therefore "have been repeatedly rejected by courts as a proper basis to exclude firearm and toolmark testimony." *Harris*, 502 F.Supp.3d at 35.

Conclusion

For the foregoing reason, Respondent's Motion to Exclude or, in the Alternative, Limit Firearm and Toolmark Examiner Testimony should be denied.

Respectfully

KARL A. RACINE
Attorney General for the District of Columbia

ELIZABETH WIESER
Deputy Attorney General, Public Safety Division

JULLIAN BREVARD
Chief, Juvenile section

By: 

D. Andrew Zirpoli #462172
Assistant Attorney General
400 6th Street, N.W. – Suite 9100
Washington, D.C. 20001
(202) 727-6341
202 727-3745 *fax*
andrew.zirpoli@dc.gov

Stephanie Daigle
Assistant Attorney General

CERTIFICATE OF SERVICE

I hereby certify that on April 5, 2022, a copy of the Government's Opposition to Respondent's Motion to Exclude or, in the Alternative, Limit Firearm and Toolmark Examiner Testimony motion was electronically served via CasefileXpress and by email to:

Julie Swaney, Esquire,
Counsel for Respondent
jswaneylaw@gmail.com

_____

D. ANDREW ZIRPOLI

SUPERIOR COURT OF THE DISTRICT OF COLUMBIA
Family Court—Juvenile Branch

In the Matter of	)	[redacted]
	)	
[redacted]	)	[redacted]
	)	
Respondent	)	Judge Andrea Hertzfeld

ORDER

Upon consideration of the Respondent's Motion to Exclude or, in the Alternative, Limit Firearm and Toolmark Examiner Testimony, the Government's Opposition thereto and the record herein, it is this _____ day of March, 2022,

ORDERED: That the Motion for Bill of Particulars hereby is, **DENIED.**

SO ORDERED.

_____
ANDREA HERTZFELD
Associate Judge

Copies to:

D. Andrew Zirpoli
Assistant Attorney General
andrew.zirpoli@dc.gov

Stephanie M. Daigle
Assistant Attorney General
Stephanie.Daigle@dc.gov

Julie Swaney, Esquire,
Counsel for Respondent
jswaneylaw@gmail.com