SUPERIOR COURT OF THE DISTRICT OF COLUMBIA

Criminal Division

UNITED STATES OF AMERICA	:	
	:	**Case No. 2018 CF1 004356**
v.	:	
	:	**Hon. Robert Okun**
ROBERT GREEN	:	

ORDER

Pending before the Court are the following filings: 1) the Defendant's Motion to Exclude Expert Testimony in Firearms Identification ("the Defendant's Motion"), filed November 4, 2021; 2) the Government's Opposition to the Defendant's Motion, filed January 18, 2022; 3) the Defendant's Consolidated Reply to the Government's Opposition, filed September 29, 2022; 4) the Brief of *Amici Curiae* Criminal Law Scholars, Scientists, Statisticians and the Innocence Project, Inc., in Support of Defendant Robert Green, also filed September 29, 2022; 5) the Government's Sur-Reply to the Defendant's Reply, filed March 28, 2023; and 6) the Government's Notice of Proposed Language for Use by the Government's Firearm and Toolmark Expert ("the Government's Notice"), filed July 10, 2023.

For the reasons set forth below, the Defendant's Motion will be denied. More specifically, the Court rejects the Defendant's request that the Government's firearms expert either be precluded from testifying or be limited to discussing class characteristics of the ammunition at issue in this case. However, the Court also rejects the Government's request that its expert be allowed to state that his examination of the ammunition provides extremely strong support for the proposition that the ammunition at issue came from the Defendant's firearm. Instead, the Court will adopt an intermediate position that it believes is most closely aligned with both the current state of the science and the relevant case law and will allow the Government's firearms expert to testify in accordance with the following limitations:

1) Where the examiner has observed consistency in class characteristics and sufficient agreement in individual characteristics, he could testify that the relevant cartridge casings are consistent with having been fired from a particular firearm;
2) The examiner would make clear that his opinion is based on a determination of class characteristic agreement, and his subjective determination of sufficient agreement in individual characteristics or random imperfections;
3) The examiner would state that his conclusions are not based on a statistically derived or verified measure and that there currently is not a generally accepted statistical method for conveying the weight of an identification; and
4) The examiner would not state that his conclusions are to a 100% certainty, to a reasonable degree of scientific certainty, or based on a comparison to all other firearms or toolmarks.

RELEVANT PROCEDURAL HISTORY

On May 19, 2020, the Government filed its expert notice for firearms examiner Chris Monturo. In this notice, the Government stated that Mr. Monturo would testify as an expert in firearms and toolmark examination and identification and would also testify as a fact witness because he had conducted an examination of the firearms evidence in the case. More specifically, the notice indicated that Mr. Monturo would state that he had examined the gun that was recovered in this case (a Glock model 30S .45 Auto semiautomatic pistol), along with five cartridge cases and five bullets recovered in the case, and would opine, without qualification, that the gun fired all five of the cartridge cases and four of the bullets.[1]

On November 4, 2021, the Defendant filed his Motion seeking to limit expert testimony about firearms identification so that the Government's firearms expert would: 1) be precluded from testifying that, based on pattern-matching, casings and bullets were fired from a particular firearm; and 2) be limited to discussing class characteristics during his testimony about markings on ammunition. In support of this Motion, the Defendant asserted that there was a lack of scientific

[1] Mr. Monturo would state that the fifth bullet had similar class characteristics to the other four bullets but that no further conclusions could be reached because of damage done to this fifth bullet.

support for the proposition that a firearms examiner could reliably match the markings on expelled casings and bullets to a particular firearm.

On January 18, 2022, the Government filed its Opposition, arguing that its firearms expert's testimony should not be limited in the manner proposed by the Defendant and that the expert instead should be allowed to testify consistent with the limitations set forth in the Department of Justice's Uniform Language for Testimony and Reports for the Forensic Firearms/Toolmarks Discipline – Pattern Matching Examination ("ULTR"). In support of its Opposition, the Government stated that the Defendant had ignored recent scientific evidence showing that firearms and toolmark identification is reliable under Federal Rule of Evidence 702, because it is testable, subject to peer review, governed by standards, generally accepted in the field, and has a low error rate. The Government also claimed that the Defendant relied upon the findings of three outdated reports – 1) the 2008 *Ballistics Imaging* Report by the National Research Council ("NRC") ("2008 NRC Report"); 2) the 2009 *Strengthening Forensic Science in the United States: A Path Forward* Report by the NRC ("2009 NRC Report"); and 3) the 2016 *Forensic Science in Criminal Courts: Ensuring Validity of Feature-Comparison Methods* Report by the President's Council of Advisors on Science and Technology ("2016 PCAST Report") – and that these reports had been supplanted by several more recent studies that addressed the criticisms set forth in these three reports.

On September 29, 2022, the Defendant filed his Reply to the Government's Opposition. In his Reply, the Defendant questioned the validity of the studies cited by the Government in its Opposition and requested that the Court preclude any unqualified statements concerning the source of ammunition, including testimony that ammunition is "consistent with" being fired from a particular firearm.

Also on September 29, 2022, *Amici Curiae* filed its Brief in Support of the Defendant. In this brief, *amici* asserted that the scientific community did not believe that firearms and toolmark analysis is reliable or scientifically validated and requested that the Court preclude any testimony that implied a statistical basis for the expert's conclusions.

On March 28, 2023, the Government filed its Sur-Reply to the Defendant's Reply. In its Sur-Reply, the Government argued that the Defendant's Reply did not undermine the validity of the studies conducted subsequent to the NRC and PCAST reports described above and noted that an additional large-scale study had been performed subsequent to the filing of the Government's Opposition that confirmed the low error rates in the field of firearm and toolmark analysis.

On July 10, 2023 the Government filed its Notice setting forth the proposed language its firearms expert would offer at trial. More specifically, the Government proposed the following language, consistent with the ULTR and the Organization of Scientific Area Committees for Forensic Sciences ("OSAC") proposed Standard Scale of Source Conclusions and Criteria for Toolmark Examinations:

> *Source Identification* – The firearms and toolmark expert will testify that where he/she made a source conclusion that his/her opinion is that the observed class characteristics and corresponding incidental characteristics provide extremely strong support for the proposition that the two toolmarks originated from the same source and extremely weak support for the proposition that the two toolmarks originated from different sources.
>
> He/she may state it is his/her opinion, based on his examination of class and incidental characteristics, research supporting the ability to use incidental characteristics to tell one firearm from another, research indicating that firearms and toolmark examiners generally do so with low error rates, and verification of his/her conclusions in this case, that the most reasonable explanation is that the relevant items were fired from the same firearm and/or unknown/unidentified firearm.
>
> The examiner will inform the jury that his/her conclusion is not based on a scientifically derived or verified measurement. The examiner will also inform the jury that, at present, there is no generally accepted statistical way to measure/convey the weight of the evidence in firearms and toolmark analysis.

> The examiner will inform the jury that he/she cannot state that the two toolmarks originated from the same source to the exclusion of all other sources or with 100% certainty.
>
> The examiner may inform the jury that, to address the potential for human error, his/her conclusions were verified.

The Court conducted an evidentiary hearing in this case on eleven different days over the course of more than five months (July 12-14, September 12-14, November 1-3, November 16 and December 19, 2023) during which it heard from four expert witnesses for the Government, two expert witnesses for the Defendant, and one expert witness for the *Amici*.

On January 30, 2024, the Government filed an expert notice for one of the witnesses who testified at the hearing, firearms examiner Dr. Eric Warren. In this notice, the Government stated that Dr. Warren would testify at trial as an expert in firearms and toolmark examination and also would testify as a fact witness because he had conducted a re-examination of the firearms evidence in the case. More specifically, the notice indicated that Dr. Warren would state that he had examined the gun and ammunition that were recovered in this case would opine that the gun fired the cartridge cases but that he could not reach a conclusion as to whether the gun fired the bullets. The language that Dr. Warren would use in describing his conclusions would be consistent with the ULTR and the OSAC guidelines discussed above.

Finally, on March 8, 2024, the Government and the Defendant submitted their post-hearing briefs, setting forth their competing positions as to whether the evidence supported the admission of the firearm identification testimony proposed by the Government. In the Government's post-hearing brief, the Government argued that the evidence showed that: 1) toolmarks are sufficiently distinct that firearm examiners can distinguish one firearm from another, and the ammunition from different firearms, based on their toolmarks; 2) firearm examiners can discern the differences in toolmarks in an accurate manner because numerous studies have showed that firearms examiners

have a low error rate; 3) firearm examiners reach their results in a repeatable and reproducible manner; and 4) the Government's proposed language for the conclusions reached by firearms examiners is consistent with the scientific evidence.

By contrast, in the Defendant's post-hearing brief, the Defendant argued that firearms examiners should be limited to discussing class characteristics in their testimony because the error rate studies do not demonstrate the reliability of testimony based on an examination of individual characteristics. The Defendant further argued that the Government failed to establish by a preponderance of the evidence that there are valid estimates of error rates for firearms comparison, asserting that the Government's witnesses at the hearing in this case either lacked the knowledge to properly evaluate the error rate studies or applied flawed reasoning when doing so, and that his witnesses and the *Amici*'s witness provided much more probative testimony demonstrating the flaws in these studies.

FIREARMS IDENTIFICATION

A. Firearms Identification Generally

Firearms identification is a subset of toolmark identification and "rests on the notion that manufacturing processes leave behind 'toolmarks' when a hard object, the tool, comes into contact with the relatively softer manufactured object," *United States v. Tibbs,* 2019 D.C. Super. LEXIS 9, at *4 (Sept. 5, 2019) (citing 2009 NRC Report at 150), such as "the marks that result when the internal parts of a firearm make contact with the brass and lead that comprise ammunition." *Abruquah v. State*, 296 A.3d 961, 973 (Md. 2023) (internal quotation marks and citations omitted). Moreover, "firearms identification is based on the premise that no two firearms will make identical marks on a bullet or cartridge case . . . because the method of manufacturing firearms results in

the interior of each firearm being unique and, therefore, making unique imprints on ammunition components fired from it." *Id*.

The patterns and marks left on bullets and cartridge cases are classified into three categories—class characteristics, sub-class characteristics, and individual characteristics. *Id.* at 974; *Tibbs*, 2019 D.C. Super. LEXIS 9, at *6. Class characteristics are common to all bullets and cartridge cases fired from a particular "type of firearm, such as the caliber and the number of lands and grooves on a bullet." *Id.* Sub-class characteristics are common to a "group of guns within a certain make or model, such as those manufactured at a particular time and place." *Abruquah*, 296 A.3d at 974. Finally, individual characteristics are unique to an individual firearm and "include [r]andom imperfections produced during manufacture or caused by accidental damage." *Id.* (citation and internal quotation marks omitted); *Tibbs*, 2019 D.C. Super. LEXIS 9, at *6 ("individual characteristics are microscopic markings produced during manufacture by the random and constantly-changing imperfections of tool surfaces as well as by subsequent use or damage to the firearm.").

B. AFTE Methodology

The leading methodology used by firearms examiners is the Association of Firearm and Tool Mark Examiners ("AFTE") Theory of Identification. *Abruquah,* 296 A.3d at 974. Examiners using the AFTE Theory of Identification employ a two-step process. *Id.* At step one, the examiner evaluates the class characteristics of the known and unknown samples and if the class characteristics do not match, the firearm that produced the known sample is excluded as the source of the unknown sample. *Id.* at 975, citing AFTE, Summary of the Examination Method, available at https://afte.org./resources/swggun-ark/summary-of-the-examination-method. If the class characteristics match, the second step of analysis involves a "comparative evaluation . . . utilizing

a comparison microscope" during which the examiner attempts to determine whether "any marks present are subclass characteristics and/or individual characteristics and the level of correspondence of any individual characteristics." *Id.*

The AFTE has a range of conclusions that an examiner can reach after comparing individual characteristics. First, an examiner can make an "identification," which occurs when there is "agreement of a combination of individual characteristics and all discernible class characteristics where the extent of agreement exceeds that which can occur in the comparison of toolmarks made by different tools and is consistent with the agreement demonstrated by toolmarks known to have been produced by the same tool." AFTE Range of Conclusions, available at https://afte.org/about-us/what-is-afte/afte-range-of-conclusions. In addition, the agreement of individual characteristics must be of "a quantity and quality that the likelihood that another tool could have made the mark is so remote as to be considered a practical impossibility." AFTE Theory of Identification as it Relates to Toolmarks, available at https://afte.org/about-us/what-is-afte/afte-theory-of-identification.

Second, an examiner can make an "inconclusive" determination if there is full agreement of all discernible class characteristics, but not sufficient agreement of individual characteristics. AFTE Range of Conclusions. There are three varieties of inconclusive determinations: "(a) when there is some agreement of individual characteristics . . . but insufficient [agreement] for an identification [Inconclusive A]; (b) when there is neither agreement nor disagreement of individual characteristics [Inconclusive B]; and (c) when there is disagreement of individual characteristics, but insufficient [disagreement] for an elimination [Inconclusive C]." *Id.*

Finally, an examiner can make an "elimination" when there is "[s]ignificant disagreement of discernible class characteristics and/or individual characteristics." *Id.*

The AFTE acknowledges that firearms identification is "subjective in nature" and that there are no objective criteria to determine the "quantity or quality" of agreement demonstrating a "practical impossibility" of a different tool having made the same mark. AFTE Theory of Identification; *Abruquah*, 296 A.3d at 976.

C. Reports and Studies Examining Firearms Identification and AFTE Methodology

Beginning around 2008, a number of federally funded reports were issued that questioned the reliability of firearms identification generally and the AFTE methodology more specifically. *See Gardner v. United States*, 140 A.3d 1172, 1183 (D.C. 2016). For example, the NRC, an arm of the National Academy of Sciences, issued a report on ballistics imaging in 2008, in which it stated that "[t]he validity of the fundamental assumptions of uniqueness and reproducibility of firearms-related toolmarks has not yet been fully demonstrated." 2008 NRC Report at 3, cited in *Gardner*, 140 A.3d at 1183. The following year, the NRC issued a broader report about the forensic sciences, in which it again noted that firearms identification was "a subjective decision based on unarticulated standards and no statistical foundation for estimation of error rates." 2009 NRC Report, Def. Ex. 11A at 153-154. The Report also noted that the scientific community had not conducted sufficient studies to establish the reliability of firearms identification and recommended that additional studies be performed to address this shortcoming. *Id.* at 154.

Finally, in 2016, the President's Council of Advisors on Science and Technology ("PCAST") issued a report that also questioned the validity of firearms identification. More specifically, the Report noted that the AFTE Theory of Identification was circular and that it fell short of the criteria for foundational validity. *See* Govt. Ex. 39, 2016 PCAST Report at 104-106; *see also Williams v. United States*, 210 A.3d 734, 741 n.15 (D.C. 2019) (the PCAST Report "deemed inadequate the studies that have thus far been done to support the proposition that every

gun produces unique toolmarks, such that a gun can be matched to a fired bullet or vice versa.") (internal quotation marks and citation omitted).

The PCAST Report did note that one study had been conducted since the 2009 NRC Report that was the "first appropriately designed black-box study of firearms," a study that was commissioned and funded by the Defense Department's Forensic Science Center and conducted by the Ames Laboratory, a Department of Energy national laboratory affiliated with Iowa State University ("Ames I study"). Def. Ex. 14 at 11; *see also* Govt. Ex. 39 at 111 ("At present, there is only a single study that was appropriately designed to test foundational validity and estimate reliability (Ames Laboratory Study). . . . Because there has been only a single appropriately designed study, the current evidence falls short of the scientific criteria for foundational validity.").[2] The authors of the PCAST Report further noted that there was a need for "additional, appropriately designed black-box studies to provide estimates of reliability." *Id.* Further, in an Addendum to the Report, the authors stated that the scientific validity of firearms identification "require[d] at least two properly designed studies to ensure reproducibility." Def. Ex. 15 at 7. The Report did not establish an error rate that would demonstrate the reliability of firearms identification but it did note that, for a method to "be considered reliable, the [false positive rate] should certainly be less than 5 percent and it may be appropriate that it be considerably lower, depending on the intended application." Govt. Ex. 39 at 152.

[2] The PCAST Report 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 samples." *Id.* at 48. It should be noted that OSAC defines a "black-box study" as one that "assesses the accuracy of examiners' conclusions without considering how the conclusions were reached. The examiner is treated as a 'black-box' and the researcher measures how the output of the 'black-box,' i.e., the examiner's conclusion, varies depending on the input, which in this case is the test specimens presented for analysis." Govt. Ex. 36 at 12. OSAC was established in 2014 to "coordinate development of standards and guidelines . . . to improve quality and consistency of work in the forensic science community." Govt. Ex. 39, Def. Ex. 14 at 22.

In response to these federal reports questioning the validity of firearms identification and calling for more studies, a number of researchers conducted in-depth studies as to the reliability of firearms identification. For example, David Baldwin *et al.* conducted the Ames I study, entitled "A Study of False-Positive and False-Negative Error Rates in Cartridge Case Comparisons," in which 218 examiners evaluated fifteen sets of three known and one unknown cartridge cases fired from a collection of twenty-five new Ruger SR9 handguns. *See* Govt. Ex. 35, Def. Ex. 29, Ames I Study at 3. The study showed that, with respect to the 1090 same-source comparisons made by the examiners, there were four false eliminations, eleven inconclusive results, and 1075 correct identifications, a 0.367% false elimination rate. *Id.* at 15.[3] With respect to the 2178 different-source comparisons made by the examiners, there were twenty-two false identifications, 735 inconclusive results, and 1421 correct eliminations, a 1.01% false identification rate. *Id.* at 16.[4]

The Ames Laboratory, in conjunction with the Federal Bureau of Investigation ("FBI"), conducted another black-box validation study concerning firearms identification between 2016 and 2020, and issued a report in 2020 entitled "Report: Validation Study of the Accuracy, Repeatability, and Reproducibility of Firearm Comparisons" ("Ames II Study"). *See* Govt. Ex. 36, Def. Ex. 30, Ames II Study at 2.[5] In this study, 173 examiners evaluated fifteen sets of two known cartridge cases and one questioned cartridge case and another fifteen sets of two known bullets

[3] A "same-source" comparison involves the examiner comparing ammunition that was fired from the same firearm. The false elimination percentage was calculated by including the number of inconclusive results in the denominator (i.e., the number of false eliminations is divided by the number of correct identifications plus the number of inconclusive results). *Id.*

[4] A "different-source" comparison involves the examiner comparing ammunition that was fired by different firearms. The false identification percentage was calculated by including the number of inconclusive results in the denominator (i.e., the number of false identifications is divided by the number of correct eliminations plus the number of inconclusive results). *Id.*

[5] "Accuracy" is defined in the Ames II Study as the "ability of an examiner to correctly identify a known match or eliminate a known nonmatch." *Id.* at 10. "Repeatability" is defined in the Ames II Study as the "ability of an examiner, when confronted with the exact same comparison once again, to reach the same determination as when first examined." *Id.* Finally, "reproducibility" is defined in the Ames II Study as the "ability of a second examiner to evaluate a set previously viewed by a different examiner and reach the same conclusion." *Id.* at 11.

and one questioned bullet. *Id.* at 23. The study showed that, with respect to the 1405 same-source bullet comparisons made by the examiners, there were forty-one false eliminations, 127 Inconclusive A results, 125 Inconclusive B results, 36 Inconclusive C results, and 1076 correct identifications, a 2.92% false elimination rate. *Id.* at 34-35.[6] Regarding the 1420 same-source cartridge case comparisons made by the examiners, there were twenty-five false eliminations, 177 Inconclusive A results, 140 Inconclusive B results, 22 Inconclusive C results, and 1056 correct identifications, a 1.76% false elimination rate. *Id.*

The study also showed that, with respect to the 2842 different-source bullet comparisons made by the examiners, there were twenty false identifications, 268 Inconclusive A results, 848 Inconclusive B results, 745 Inconclusive C results, and 961 correct eliminations, a 0.70% false identification rate. *Id.*[7] Regarding the 2835 different-source cartridge case comparisons made by the examiners, there were twenty-six false identifications, 177 Inconclusive A results, 637 Inconclusive B results, 620 Inconclusive C results, and 1375 correct eliminations, a 0.92% false identification rate. *Id.*

In sum, therefore, the Ames II study demonstrated that the examiners had a 0.70% false identification rate for bullets and a 0.92% false identification rate for cartridge casings, while they had a 2.92% false elimination rate for bullets and a 1.76% false elimination rate for cartridge cases.

The parties have cited numerous other post-PCAST studies to show the reliability, or lack of reliability, of firearms identification, including the following eight studies: 1) "Validity of forensic cartridge-case comparisons," by Max Guyll *et al.*, published in the Proceedings of the National Academy of Sciences, 2023, Vol. 120, No. 20, *see* Govt. Ex. 10; 2) "Results of the 3D Virtual Comparison Microscopy Error Rate (VCMER) Study for Firearm Forensics," by Chad

[6] These percentages were calculated in a manner consistent with the Ames I study. *See* n.3, *supra*.
[7] These percentages were calculated in a manner consistent with the Ames I study. *See* n.4, *supra*.

Chapnick *et al.*, published in the Journal of Forensic Sciences, 2021:66:557-570, *see* Govt. Ex. 37; 3) "Development and Validation of a Virtual Examination Tool for Firearm Forensics," by Pierre Duez *et al.*, published in the Journal of Forensic Sciences, 2018, Vol. 63, No.4, *see* Govt. Ex. 38; 4) "Isolated Pairs Research Study," by Mark Keisler *et al.*, published in the AFTE Journal, Winter 2018, Vol. 50, No. 1, *see* Govt. Ex. 40; 5) "Blind testing in firearms: Preliminary results from a blind quality control program," by Maddisen Neuman *et al.*, published in the Journal of Forensic Sciences, 2022:67:964-974, *see* Govt. Ex. 41, Def. Ex. 26; 6) "Evaluating firearm examiner conclusion variability using cartridge case reproductions," by Eric Law *et al.*, published in the Journal of Forensic Sciences, 2021:00:1-17, *see* Def. Ex. 3; 7) "Beretta barrel fired bullet validation study," by Jaimie Smith, published in Journal of Forensic Sciences, 2021:66:547-556, *see* Def. Ex. 13; and 8) "An Assessment of the Foundational Validity of Firearms Identification Using Ten Consecutively Button-Rifled Barrels," by Brandon Best *et al.*, published in the AFTE Journal, Volume 54, Number 1, Spring 2022. *See* Def. Ex. 50.

While all of these studies are relevant to the Court's analysis, the Court will only discuss two of these studies in detail – the Guyll and Neuman studies.[8] In the Guyll study, 228 examiners evaluated four sets of cartridge cases that were fired by either a Beretta Model 92FS firearm or a HiPoint Model C9 firearm, with each set consisting of three known samples and one questioned sample. *See* Govt. Ex. 10 at 9. In this study, the authors calculated error rates in two different manners. First, they excluded the number of inconclusive results from both the numerator and

[8] The Court is focusing on these studies because of their scope and, with respect to the Neuman study, because it involved firearm examiners doing actual case work rather than examiners conducting examinations that they knew were being conducted only for study purposes. However, the Court notes that many of the other studies submitted by the parties had similar error rates. *See, e.g.*, Keisler study, Gov. Ex. 40 at 2 (0% false positive error rate and 0% false negative error rate where 126 examiners made 1508 of 1512 correct comparisons with four inconclusive results and made 805 of 1008 correct exclusions with 203 inconclusive results).

denominator and second, they included the number of inconclusive results in both the numerator and denominator. *Id.* at 4-5.

The study showed that, with respect to the 454 same-source Beretta comparisons made by the examiners, there was one false elimination, fifty-six inconclusive results, and 397 correct identifications, a 0.3% false elimination rate if inconclusive results were excluded and a 0.2% false elimination rate if inconclusive results were included. *Id.* at 4. With respect to the 457 same-source HiPoint comparisons made by the examiners, there were no false identifications, three inconclusive results, and 454 correct identifications, a 0% false elimination rate, and the error rate was the same regardless of whether inconclusive results were included or excluded. *Id.*

With respect to the 451 different-source Beretta comparisons, there was one false identification, 178 inconclusive results, and 272 correct eliminations, a 0.4% false identification rate if inconclusive results were excluded and a 0.2% false identification rate if inconclusive results were included. *Id.* With respect to the 449 different-source HiPoint comparisons, there were four false identifications, 145 inconclusive results, and 300 correct eliminations, a 1.3% false identification rate if inconclusive results were excluded and a 0.99% false identification rate if inconclusive results were included. *Id.*

In sum, therefore, the Guyll study showed that the examiners had an overall 0.1% false elimination rate, regardless of whether inconclusive results were included or excluded, and had an overall 0.9% false identification rate if inconclusive results were excluded and an overall 0.6% false identification rate if inconclusive results were included in the calculations. *Id.* A more detailed summary table is listed below.

Summary of Results of Guyll, et. al. Study					
	Identification	Elimination	Inconclusive	Total	Error Rate*
Beretta same-source	397	1	56	454	0.2/0.3%†
HiPoint same-source	454	0	3	457	0%
Beretta different-source	1	272	178	451	0.2/0.4%†
HiPoint different-source	4	300	145	449	0.9/1.3%†
*The Defendant argues that inconclusive results should count as errors when calculating error rate. The Court rejects this view, discussed *infra* p.33. †Error rate depends on whether inconclusives are included in or excluded from denominator, discussed *infra*.					

Finally, in the Neuman study, the Houston Forensic Science Center conducted a blind quality control program in firearms identification from December 2015 to June 2021. Govt. Ex. 41, Def. Ex. 26 at 964. In this study, eleven examiners evaluated 558 pieces of ammunition, including 272 bullet items, 265 cartridge cases, 18 fragments and 3 pellets. *Id.* at 968. In this study, there were no false identifications or false eliminations, but there were inconclusive results in 40.3% of the evaluations where identifications or eliminations could have been made. *Id.*

SUMMARY OF EVIDENCE AT HEARING

The evidence presented at the hearing was very extensive. The Court will only summarize the most relevant portions of the evidence here and will further address this testimony in the analysis section of this order.

A. The Government's Witnesses

The Government called the following witnesses at the hearing: Ronald Nichols, Dr. Max Guyll, Dr. Bruce Budowle, and Dr. Eric Warren.

Mr. Nichols was qualified as an expert in firearm and toolmark examinations and in the scientific literature concerning those examinations. Mr. Nichols testified generally about the

toolmarks left on casings and bullets, the class characteristics, sub-class characteristics and individual characteristics of casings and bullets, and the range of conclusions that could be formed about these characteristics under the AFTE standards, the OSAC Firearms Subcommittee Draft Range of Conclusions, and the ULTR standards. Mr. Nichols testified that: 1) the premise that different tools produce different toolmarks on firearms had been established by numerous machine-based studies, and 2) the premise that trained examiners could accurately discern these differences had been established by sixteen error rate studies that showed a low error rate for firearms examinations. However, on cross-examination, Mr. Nichols acknowledged that the machine-based studies he cited involved computers making comparisons, not people, that error rate studies were not necessarily indicative of case work reliability, and that some of these studies showed significant variability in both reproducibility (with different examiners reaching different results) and repeatability (with the same examiner reaching different results).

In reaching his conclusion that firearms examiners had a low error rate, Mr. Nichols opined that inconclusive results should not be considered erroneous results, but on cross-examination he acknowledged that an inconclusive result can be considered erroneous if there is sufficient data upon which to make a conclusion, and he also agreed that studies that included inconclusive results in the denominator may have understated the true error rate. Finally, Mr. Nichols testified that there was general acceptance of the validity of firearms examination in the forensic community at large but conceded that many members in the forensic community had a financial interest in toolmark identification.

Dr. Guyll was qualified as an expert in the design of psychological studies, calculations relating to those studies, and studies concerning human decision-making. Dr. Guyll's testimony focused on the study that he conducted as described above. Dr. Guyll testified that he agreed with

the Government's proposed language for firearm examiner testimony based on his review of the relevant scientific literature. Dr. Guyll also testified that inconclusive results should not be considered erroneous results, noting that the opposite approach would create a false equivalency between situations where an examiner had no false positive results and 50 inconclusive results and an examiner had 50 false positive results and no inconclusive findings. Dr. Guyll also opined as follows with respect to inconclusive determinations:

> A correct decision is one that asserts a ground-truth state that corresponds to the actual ground-truth state. And an incorrect decision is one that asserts a ground state that is contrary to the ground-truth state. And so the conclusive decisions of identification and elimination, they assert a ground-truth state and must be either correct or incorrect in all cases. However, inconclusive decisions do not assert a ground-truth and therefore can be neither correct nor incorrect.

See 7/14/23 Guyll Tr. at 93.

On cross-examination, Dr. Guyll acknowledged that some reported error rates may be misleading to the extent that they consider inconclusive results to be correct results. To address this problem, Dr. Guyll calculated error rates both with and without inconclusive results.

Dr. Budowle was qualified as an expert in forensic science management and research, study design, error rates and the creation of forensic standards. Dr. Budowle testified about the peer review process in the firearms examination field and about the concept of convergent validity, under which the validity of a field can be assessed by reviewing the data drawn from multiple studies that address different issues within the field. Dr. Budowle agreed with Dr. Guyll that inconclusive results should not be considered errors when conducting error rate studies, noting that in some cases an inconclusive determination may be the most appropriate finding because the sample is damaged or otherwise of inadequate quality to make a reliable comparison. However, Dr. Budowle expressed some hesitation in using the "extremely strong support" language proposed

by the Government, stating that this language, which first arose in the DNA context, might not fit as well in the firearms context.

Finally, Dr. Warren was qualified as an expert in firearms and toolmark examination and research design. Dr. Warren's testimony focused largely on his recent article, entitled "The inconclusive category, entropy and forensic firearm identification," published in 2023 in Forensic Science International. In this article, and in his testimony, Dr. Warren indicated that firearms examiners should not be measuring error rates by using the binary model of whether the result was correct or incorrect, and instead should consider the inconclusive result as a meaningful outcome of a forensic examination that can be measured more accurately by using alternative statistical paradigms, such as the cross-entropy model or the game theory model. However, Dr. Warren noted that, due to the complexity of these models, it was unlikely that they could be used in the courtroom in the near future and also acknowledged on cross-examination that the studies he conducted involved firearm examiners who were looking at class characteristics rather than examiners making comparisons of the individual characteristics of a firearm.

B. The Defendant's Witnesses

The Defendant called the following witnesses at the hearing: Dr. Nicholas Scurich and Dr. Michael Rosenblum.

Dr. Scurich was qualified as an expert in human subject experimental study design and the calculation of associated statistics. Dr. Scurich first testified that many of the firearms studies used a closed set design (otherwise known as a set-to-set design) that can artificially improve the examiners' performance in the study. Dr. Scurich also criticized those firearms studies that deemed inconclusive results to be correct responses, stating that this framework artificially lowered the actual error rates in these studies. Dr. Scurich further testified that, although certain inconclusive

results may be accurate because of a lack of sufficient characteristics on the ammunition being compared, other inconclusive results are inaccurate if there are sufficient characteristics on which to make a comparison. Dr. Scurich therefore offered an alternative way in which to consider inconclusive results when measuring error rates – namely, the creation of a "blue ribbon" panel of experts to examine the evidence and decide whether particular evidence should be deemed inconclusive or, in the alternative, to look at those items in a study where a high percentage of examiners reached an inconclusive result and to then determine whether the inconclusive result is correct for that particular piece of evidence.

However, on cross-examination, Dr. Scurich agreed that many of the studies the Government was relying on in this case, including both Ames I and II and the Guyll study, did not use closed set designs. Dr. Scurich also acknowledged the practical infeasibility of his "blue ribbon" panel proposal, which would require the panel members to review every piece of evidence before the start of a particular study, and he did not offer an opinion as to how many people should be on his proposed panel, and did not specify the qualifications of the people on this panel other than to note that they would be firearms examiners, like those who are participating in the study.[9]

Dr. Rosenblum was qualified as an expert in study design and statistics. Dr. Rosenblum reviewed twenty-five black-box firearms studies, including most of those described above, and concluded that there were major flaws in all of these studies and that these flaws were "so severe, numerous, and pervasive as to completely undermine the scientific foundation of these studies." Def. Ex. 54 at 3. More specifically, Dr. Rosenblum testified that these studies were deficient because they lacked sample size calculations, the study conditions were not representative of real

[9] Because of a serious illness in Dr. Scurich's family, the Government was not able to complete its cross-examination of Dr. Scurich. Rather than strike Dr. Scurich's testimony, the Court allowed the Government to submit declarations from Dr. Guyll and Todd Weller to set forth their criticisms of Dr. Scurich's testimony.

casework, error rates were incorrectly computed because inconclusive results were either considered to be correct responses or ignored, the confidence intervals contained in some studies were invalid, and missing data in some studies was not properly addressed. Dr. Rosenblum also cited FDA guidance that recommended that inconclusive results should not be ignored in error rate studies and instead could be reported in two different sets of performance measures, first as test positive results and second as test negative results.

On cross-examination, Dr. Rosenblum admitted that he had little first-hand experience or information about the field of firearms examination, acknowledging that he was not familiar with AFTE methodology, and had never conducted a firearm examination, never witnessed the firing of a firearm, never toured a firearm manufacturing facility, and never designed a study to measure the accuracy of firearms examinations. Dr. Rosenblum also conceded that the FDA guidance he cited did not actually apply to firearms examinations because that guidance does not apply to tests that "can (per the test instructions) produce a result which is anything other than positive or negative." Govt. Ex. 118 and Def. Ex. 55 at 18. Further, Dr. Rosenblum conceded that, even though he criticized the firearms studies for lack of sample size calculations and for not being representative of real case work, certain of his own studies similarly lacked sample size calculations and were not representative of real case work. Finally, although Dr. Rosenblum criticized the confidence intervals used in some firearms studies for the numbers they used, he admitted that the numbers he chose in his declaration when discussing a hypothetical firearms study were not based on any data and instead were arbitrarily assigned to illustrate his points.

C. Amici's Witness

Finally, *amici* called Dr. William Thompson as a witness at the hearing. Dr. Thompson was qualified as an expert in human judgment and decision-making, bias reduction procedures in

forensic laboratories, and validation and reporting concerning forensic science. Dr. Thompson agreed with some of Dr. Rosenblum's critiques of the firearms studies, including their treatment of inconclusive results, and testified about certain additional theories that he believed help explain some of the differing results in these studies. More specifically, Dr. Thompson testified that the results in studies can be affected by such variables as what examiners are told prior to conducting their examinations and the perceived consequences of providing a false result, which, Dr. Thompson believed, explained the high rate of inconclusive results in many firearms studies. Dr. Thompson also testified that error rates can be affected by examiners having different decision thresholds before they determine whether a result is an identification, inconclusive or an exclusion. Finally, Dr. Thompson testified about the perceived strength of forensic examiner's statements concerning their source conclusions, questioning whether there was a sufficient basis for firearms examiners to state that there was extremely strong support for their conclusions.

On cross-examination, Dr. Thompson acknowledged that not all inconclusive results should be considered false identifications. Moreover, in contrast to Dr. Rosenblum's highly critical assessment of firearms studies, Dr. Thompson testified that the probative value of firearms and toolmark studies provided moderate or moderately strong support for firearm and toolmark identification. In addition, because some of Dr. Thompson's testimony exceeded the scope of his expert notices, the Government was permitted, without objection, to submit a declaration from Dr. Budowle, along with Dr. Max Morris and Todd Weller, to address this portion of his testimony. In this declaration, the authors criticized the analysis used by Dr. Thompson in support of his signal detection model when applied to the Ames II data.[10]

[10] The signal detection model described in Dr. Thompson's testimony is used to predict how correct positive and negative identification rates would vary if the examiners changed their decision thresholds before reaching a conclusion that the result was an identification, inconclusive or an exclusion.

RELEVANT CASE LAW

The relevant standards for the admissibility of expert testimony were set forth by the Court of Appeals in *Motorola Inc. v. Murray*, 147 A.3d 751 (D.C. 2016) (*en banc*). In that case, the Court of Appeals abandoned the general acceptance test of admissibility set forth in *Frye v. United States*, 293 F.1013 (D.C. Cir. 1923), and *Dyas v. United States*, 376 A.2d 827 (D.C. 1977), and instead adopted the admissibility standards set forth in Rule 702 of the Federal Rules of Evidence. *Motorola*, 147 A.3d at 752. Pursuant to that rule, a witness "who is qualified as an expert by knowledge, skill, experience, training or education may testify in the form of an opinion or otherwise if: (a) the expert's scientific, technical or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue; (b) the testimony is based on sufficient facts or data; (c) the testimony is the product of reliable principles and methods; and (d) the expert has reliably applied the principles and methods to the facts of the case." *Id.* at 756. [11]

The principles set forth in Rule 702, in turn, were based on the Supreme Court's opinion in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 587 (1993), and its progeny. *See Motorola*, 147 A.3d at 756 (Rule 702 was amended in 2000 to reflect the Supreme Court's guidance in *Daubert*). In *Daubert*, the Supreme Court held that, when a party proffers expert scientific testimony, the trial court must make "a preliminary assessment of whether the reasoning or methodology is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue." *Daubert*, 509 U.S. at 592-93. Although the Court eschewed "a definitive checklist or test" in making the reliability assessment, the Court did suggest certain factors to be considered, including whether the theory or technique has been tested, whether it "has been subjected to peer review and publication," "the known or potential rate of error," and "the

[11] Rule 702(d) was amended effective December 1, 2023, and now states "the expert's opinion reflects a reliable application of the principles and methods to the facts of the case. This change does not alter the Court's analysis.

existence and maintenance of standards controlling the technique's operation." *Id*. at 593-94. Further, although "general acceptance" was no longer the linchpin of admissibility, it could "have a bearing on the inquiry." *Id.* at 594.

The Supreme Court subsequently noted that the test of reliability is a "flexible" one and that *Daubert*'s list of factors to be considered "neither necessarily nor exclusively applies to all experts or in every case. Rather, the law grants [the trial] court the same broad latitude when it decides *how* to determine reliability as it enjoys in respect to its ultimate reliability decision." *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 141-42 (1999). Furthermore, the Court of Appeals has noted that the trial court's gatekeeping role under *Daubert* was not meant to "displace the normal tools of the adversary system . . . [v]igorous 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." *Motorola*, 147 A.3d at 754.

This jurisdiction has allowed the admission of expert testimony concerning firearm identification for more than a hundred years. *See Laney v. United States*, 294 F. 412, 416 (D.C. Cir. 1923) (upholding admission of expert testimony "tending to establish that the bullet, extracted from the head of the deceased, was shot from the pistol found in the defendant's possession"). However, the Court of Appeals has limited the proper scope of expert testimony from a firearm or toolmark expert on a number of occasions over the past several years. For example, in *Gardner*, *supra*, the Court held that a firearm and toolmark expert "may not give an unqualified opinion, or testify with absolute or 100% certainty, that based on ballistics pattern comparison matching a fatal shot was fired from one firearm to the exclusion of all other firearms." 140 A.3d at 1177. The Court further noted that its holding was limited "in that it allows toolmark experts to offer an opinion that a bullet or shell casing was fired by a particular firearm, but it does not permit them

to do so with absolute or 100% certainty," and noted that it had doubts as to whether toolmark experts should be allowed to state their opinions "with a reasonable degree of certainty." *Id.* at 1184, n.19.

In *Williams, supra*, the Court of Appeals reiterated that it is error to allow a firearm and toolmark examiner to "provide unqualified opinion testimony that purports to identify a specific bullet as having been fired by a specific gun via toolmark pattern matching." 210 A.3d at 743. The Court of Appeals did not resolve the Government's argument that *Gardner* only prohibited certainty statements and otherwise continued to authorize opinion testimony identifying a specific bullet as having been fired by a specific gun, because the examiner in that case had given a certainty statement. *Id.* at 741-42. However, the Court noted that the Government's argument was "difficult to square" with the Court's holding in *Gardner* that the trial court had erred by admitting the examiner's unqualified opinion that a specific gun was the murder weapon. *Id.* at 739. The Court also noted that its opinion did "not limit firearms and toolmark examiners from making other observations about the ballistics evidence recovered in a particular case," because those observations were not at issue in the case. *Id.* at 743, n.19; *see also Bellinger v. United States*, 294 A.3d 1094, 1101 (D.C. 2023) ("we have not foreclosed 'all firearms and toolmark evidence,' but rather such evidence that 'unqualifiedly connects a specific bullet to a specific gun.'") (quoting *Williams*, 210 A.3d at 742-43).

In *Gordon v. United States*, 285 A.3d 199 (D.C. 2022), the Court of Appeals held that the trial court "did not err – let alone plainly err" where it did not strike the testimony of a firearms expert that six casings were "most likely" fired from a Glock semi-automatic pistol, that two bullets were "consistent" with a Glock, and that all six casings were fired from "the same gun," where the expert did not conclusively say that the casings and bullets were fired from a particular gun. *Id.* at

219-220. In reaching this conclusion, the Court noted that the expert was careful to qualify his opinion and "only opined on the fact that the six cartridges were most likely fired from a similar type of *unspecified* gun." *Id.* at 220 (emphasis in original).

Finally, in *Geter v. United States*, 306 A.3d 126 (D.C. 2023), the Court of Appeals held that this Court plainly erred when it admitted testimony that five cartridges were "fired from the same gun, and they all matched the[] test fires conducted on the gun recovered from the scene." *Id*. at 132. The Court of Appeals rejected the Government's argument that the testimony was not erroneously admitted because the expert did not provide an unqualified opinion. Rather, the Court stated that the expert's opinion contravened *Gardner* and *Williams* because the expert testified that the inside of the gun had "unique" markings that were transferred to the shell casing when the weapon was fired so that the import of his testimony was that the markings on the shell casings were unique to a specific gun. *Id.* The Court also stated that the testimony was erroneously admitted because "the research does not exist to say that a specific bullet can be matched to a specific gun based on pattern matching." *Id.* at 133. Further, the Court indicated that the expert's testimony was impermissibly "unqualified" because, after acknowledging "that pattern matching could be inconclusive in some cases . . . he testified, without reservation, that . . . these five cartridge casings came from this firearm." *Id.* Finally, the Court rejected the Government's argument that the testimony was permissible under footnote 19 in *Gardner* (which allowed toolmark experts to offer an opinion that a bullet or shell casing was fired by a particular firearm, but not to do so with absolute or 100% certainty), because the expert's testimony matched a

specific bullet to a specific gun "without *any* explicit statement of uncertainty or qualification." *Id.* at 135 (emphasis in original).[12]

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

By contrast, in *United States v. Harris*, 502 F.Supp.3d 28 (D.D.C. 2020), Judge Contreras disagreed with Judge Edelman's analysis, after conducting an evidentiary hearing, and held that the Government's firearms and toolmark expert could testify that casings were fired from the same firearm when all class characteristics were in agreement and "the quality and quantity of corresponding individual characteristics is such that the examiner would not expect to find that same combination of individual characteristics repeated in another source and has found insufficient disagreement of individual characteristics to conclude that they originated from different sources." 502 F.Supp.3d at 45. Judge Contreras also noted with approval that the Government had agreed that its expert would not use terms such as "match" or state his opinion

[12] The Court noted that the expert's opinion in *Gardner* was "unqualified" because the expert represented without any qualification that a specific bullet came from a specific firearm, citing the Merriam-Webster's Dictionary definition of "unqualified" as meaning "not modified or restricted by reservations." *Id.*at 133. The Court also pointed out that the expert's testimony was unqualified because, when asked if the "bullet was just consistent with" being fired from the purported murder weapon, the expert went beyond the "consistent with" language in the question and instead answered "[i]t was fired from the pistol" and then reasserted "'[i]t was identified as having been fired from [the pistol]." *Id.*at 134. By contrast, the Court of Appeals stated that its opinion in *Gordon* provided guidance on how explicitly a firearms and toolmark examiner's testimony must be qualified when providing testimony that purports to link specific shell casings to a specific gun, noting that the trial court in that case did not err by failing to strike a firearms and toolmark examiner's testimony where the examiner used phrases such as "most likely" and "consistent with." *Id.* at 134, n.7.

with any level of statistical or scientific certainty or to the "exclusion of all other firearms." *Id.* at 44.

Finally, many state and federal courts in other jurisdictions have also addressed this issue, with varying results. Of particular note, the Maryland Supreme Court recently held that the trial court committed reversible error where the firearms examiner offered an "unqualified opinion that the crime scene bullets were fired from [the defendant's] gun." *Abruquah*, 296 A.3d at 968. However, the Court also stated that the firearms methodology applied in that case could support "reliable conclusions that patterns and markings on bullets are consistent or inconsistent with those on bullets fired from a particular firearm." *Id.*; *see also id.* at 998 ("the circuit court did not abuse its discretion in ruling that [the firearms expert] could testify about firearm identification generally, his examination of the bullets and bullet fragments found at the crime scene, his comparison of that evidence to bullets known to have been fired from [the defendant's revolver], and whether the patterns and markings on the crime scene bullets are consistent or inconsistent with the patterns and markings on the known bullets.").

ANALYSIS

A. Preliminary Considerations

As a threshold matter, the Court notes that it found the numerous experts who testified in the hearing to be credible, well-educated, and articulate witnesses. However, the Court notes that none of the defense and *amici* witnesses actually had any first-hand experience with firearms and toolmark identification, unlike two of the Government's witnesses, Ronald Nichols and Dr. Eric Warren, who had decades of experience as firearms examiners. Moreover, none of the defense or *amici* witnesses had been involved in designing their own firearm identification error rate studies, while one of the Government's witnesses, Dr. Guyll, personally oversaw an extensive firearm

identification error rate study that has proved particularly helpful to the Court in its analysis. Thus, while all of the experts helped inform the Court's opinion, the Government's experts, on the whole, were particularly valuable to the Court's analysis.

B. Testability

Although the parties vigorously dispute the adequacy of the testing that has been conducted to measure the accuracy of firearms and toolmark identification, they do not dispute the fact that this technique is testable. Further, even those courts that have been imposed significant limits on the testimony of firearm examiners have agreed that the technique is testable. *See, e.g., Tibbs*, 2019 D.C. Super. LEXIS 9, at *24 (finding the technique to be testable after noting that "virtually every court that has evaluated the admissibility of firearms and toolmark identification has found the AFTE method to be testable . . . ."); *Abruquah*, 296 A.3d at 988 ("it is undisputed that firearms identification can be tested."). The Court agrees that the accuracy of firearms and toolmark identification can be tested and this factor weighs in favor of admissibility.

C. Peer Review and Publication

This factor considers whether a methodology has been submitted to the "scrutiny of the scientific community" because doing so "increases the likelihood that substantive flaws in methodology will be detected." *Daubert*, 509 U.S. at 593. For the reasons set forth below, the Court finds that this factor also weighs in favor of admissibility.

The Defendant argues that this factor should not weigh in favor of admissibility because most of the studies in the field have been published in the AFTE Journal, a publication produced by the Association of Firearm and Toolmark Examiners. Citing *Tibbs*, the Defendant argues that the articles published in this journal do not satisfy the *Daubert* standard because the publication is composed entirely of practitioners who have a vested self-interest in establishing the scientific

validity of firearm and toolmark identification. *See Tibbs*, 2019 D.C. Super. LEXIS 9, at **35-36 (finding that this factor did not weigh strongly in favor of admissibility because "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.").

The Court does not agree with the Defendant for two primary reasons. First, although the Court is mindful of the concerns raised by Judge Edelman about the weight to be accorded to studies published in this journal, the Court gives more weight to the results of some of these studies than Judge Edelman appeared to do because of changes that the AFTE Journal made after the issuance of the *Tibbs* opinion. More specifically, subsequent to the issuance of the *Tibbs* opinion, the AFTE Journal modified its review process and now uses a double-blind peer review process where the author and reviewer do not know each other's identity, instead of an open review process where the author and reviewer know each other's identity and can contact each other during the review process. *See* AFTE Peer Review Process – January 2020, cited at https://afte.org/afte-journal/afte-journal-peer-review-process; *see also United States v. Cloud*, 576 F.Supp.3d 827, 841 (E.D. Wa. 2021); *United States v. Rhodes*, 2023 U.S. Dist. LEXIS 7528, at *14 (D. Ore. 2023). As Judge Edelman noted in *Tibbs*, the double-blind peer review process "constitutes the standard among scientific publications and guards against personal and institutional biases by shielding both reviewer and author from the identity of the other." 2019 D.C. Super. LEXIS 9, at *30. Thus, the Court gives more weight to the firearms studies published in the AFTE Journal beginning in January 2020, when the Journal modified its peer review process. *See, e.g.,* Best, B.A. & Gardner,

E.A. (2022) "An Assessment of the Foundational Validity of Firearms Identification Using Ten Consecutively Button-Rifled Barrels," published in AFTE Journal, 54(1).[13]

Second, and importantly, the parties have produced a plethora of studies that were published in peer-reviewed journals that have no affiliation with the AFTE. Indeed, Todd Weller's declaration cites to more than forty scientific studies involving firearms and toolmark identification that were published in ten journals other than the AFTE Journal, *see* Govt. Opp., Appendix A, Weller Decl. at 52-55, and Dr. Rosenbloom's declaration demonstrates that he reviewed thirteen black-box studies published in such journals, including eight that were published in the Journal of Forensic Sciences, four that were published in Forensic Science International and one that was published in the Proceedings of the National Academy of Sciences. *See* Def. Ex. 54 at 17-19. As Judge Edelman noted in *Tibbs*, such journals provide "meaningful review" and are unaffiliated with "AFTE, any crime lab, or any individual with a financial or professional interest in the validation of the field of firearms and toolmark analysis." 2019 D.C. Super. LEXIS 9, at *27; *see also United States v. Chavez*, 2021 U.S. Dist. LEXIS 237830, at *8-9 (N.D. Cal. Dec. 13, 2021) ("even if the Court were to fully discount the studies published in the AFTE Journal, these are not the only examples of peer review research published in the field," citing studies published in the Journal of Forensic Sciences, and a non-exhaustive list of over fifty peer-reviewed studies published outside the AFTE Journal, and noting that these studies fulfilled the required publication and peer review requirement); *Harris*, 502 F.Supp.3d at 40 (same).

In addition, although the Ames I and II studies have not been published in a peer-reviewed journal, the PCAST Report stated that the Ames I study was "appropriately designed to test

[13] The Court also notes that many courts have not discounted the weight of the studies published in the AFTE Journal and instead have concluded "that publication in the AFTE Journal satisfies this prong of the *Daubert* admissibility analysis." *Harris*, 502 F.Supp.3d at 40 (citing cases).

foundational validity and estimate reliability" and the Ames II study was similarly designed. *See* PCAST Report, Govt. Ex. 39 at 111. Thus, although these two studies have not been published in a peer-reviewed journal, they have been subject to exacting scrutiny from the scientific community.

In sum, given the changes made in the review process at the AFTE Journal and, more importantly, the number and quality of studies published in journals unaffiliated with the AFTE, the Court finds that this factor weighs in favor of admissibility.

D. Error Rate

The most hotly disputed issue between the parties is the error rate for firearms and toolmark identification. The Government argues that numerous well-conducted, black-box studies have consistently showed that firearms examiners have a low error rate, both in terms of false identifications and false exclusions, when conducting their examinations. The Defendant, by contrast, argues that these studies are significantly flawed, and that the error rates are much higher than reported in these studies because these studies either consider inconclusive results to be correct responses or do not include these results when determining the error rate.

As a preliminary matter, the Court must address the threshold issue of how to assess inconclusive responses, because this assessment can affect the error rates reported in these studies. The Government argues that inconclusive results either should be considered to be correct responses or should not be considered as either correct or incorrect responses, while the Defendant argues that these results should be considered incorrect responses because they differ from the ground truth of whether the ammunition being examined came from the firearm being examined or a different firearm.

The Court agrees with the Government that inconclusive results should not necessarily be considered incorrect responses. First, with respect to the Defendant's ground truth argument, the Court notes that one of the Defendant's experts and the *Amici*'s expert acknowledged that some ammunition could lack sufficient characteristics for an examiner to make an accurate comparison, so that an inconclusive determination could not properly be considered an incorrect answer. *See, e.g.,* Scurich 9/14/23 Tr. at 96; 11/3/23 Thompson Tr. at 44. Indeed, even courts that have placed significant limits on the testimony that could be offered by firearms examiners have rejected the argument that inconclusive results should be considered incorrect answers. *See, e.g.*, *Tibbs*, 2019 Super. LEXIS 9, at *17 (rejecting Dr. Scurich's opinion that inconclusive results should be viewed as false positive errors, and stating that this opinion "fails to make logical sense: while under laboratory conditions such inconclusives are surely some type of error, it does not follow that inconclusives are functionally the same as a false conclusion by an examiner who attributes a cartridge casing to a gun that did not fire it."); *see also Rhodes*, 2023 U.S. Dist. LEXIS 7528, at *12 ("[w]ith respect to Dean Faigman and Dr. Scurich's opinions that the error rates are unreliable because they do not include inconclusive results [as errors], this Court finds those opinions unsupported.").

Moreover, many courts have noted that the primary reliability concern tied to firearm and toolmark examination is that an examiner will falsely identify ammunition as having come from a firearm that is linked to the defendant in that case – in other words, that the examiner will make a false identification – and this concern is not raised when an examiner makes an inconclusive determination. *See, e.g., United States v. Hunt*, 63 F.4th 1229, 1240 n.7 (10th Cir. 2023) (the Ames Study false-positive rate furnishes the relevant error rate because "a false positive identification . . . is the type of error that could lead to a conviction premised on faulty evidence," and further

noting "that there is no harm to the defendant if a toolmark examiner makes an inconclusive finding."); *see also Rhodes*, 2023 U.S. Dist. LEXIS 7528, at *12 ("while the relatively high rates of inconclusive results in error rate studies may be relevant to certain policy determinations – such as failing to eliminate a suspect from consideration – those concerns do not relate to the issue of paramount importance to trial courts; *i.e.*, the false positive rate which may result in a wrongful conviction."); *Harris*, 502 F.Supp.3d at 39 ("the critical inquiry under this factor is the rate of error in which an examiner makes a false positive identification . . . ."); *Abruquah*, 296 A.3d at 1013 ("we are concerned with the likelihood that *when* application of the AFTE Theory yields a conclusive determination – here a positive identification – the result is correct.") (Gould, J., dissenting).

This Court agrees that the critical inquiry with respect to the error rate factor is the rate of false identifications, which is the error of most concern in the context of a criminal proceeding.[14] The Court further finds that the most accurate way to address this concern is to determine the false identification rate by dividing the number of false identifications made by the total number of comparisons made, including the inconclusive results. Using this framework, the Court finds that the error rate, as measured by the false identification rate, is extremely low, ranging from 0% (Neuman study) to 1.01% (Ames I study), and this low error rate strongly favors the admission of the evidence.

However, the Court also agrees with the Defendant that inconclusive responses are not necessarily correct, as was acknowledged by some of the Government's expert witnesses. *See*, *e.g.*,

[14] Although the Court agrees that the rate of false identifications is the critical inquiry, the Court notes that false eliminations can be inculpatory as well if the false elimination involves a firearm allegedly possessed by a third party where the defendant is raising a third-party perpetrator defense under *Winfield v. United States*, 676 A.2d 1 (D.C. 1996).

7/12/23 Nichols Tr. at 168 ("when there's sufficient data upon which to base a more definitive conclusion, an inconclusive would be an incorrect answer."). The Court also is aware that some courts have found that inconclusive results should not be considered either correct or incorrect, and they have calculated the error rate by excluding inconclusive results entirely from their analysis. *See, e.g., United States v. Felix*, 2022 U.S. Dist. LEXIS 213513, at *49 (D.V.I. 2022) ("The Court agrees with Dr. Scurich that inconclusive responses are not properly classified as correct responses, because counting them as such would likely artificially *deflate* the error rate . . . [h]owever, counting inconclusive responses as errors is not a satisfactory solution because counting them as such would likely artificially *inflate* the error rate . . . ."); *Abruquah*, 296 A.3d at 1014 ("excluding inconclusives from error rate calculations when assessing the reliability of a positive identification is not only an acceptable approach, but the preferred one, at least according to PCAST.") (Gould, J., dissenting). Indeed, both the PCAST report and Dr. Guyll have endorsed or calculated error rates by using two alternative calculations – first, by including inconclusive results in the denominator and second by excluding inconclusive results entirely from the calculation. *See* PCAST Report, Govt. Ex. 39 at 110-111 (describing Ames as a study that could be used to establish the foundational validity of firearms and toolmark examination even though it had an inconclusive rate of 33.7% and considering error rates with and without inconclusive results); Guyll, Govt. Ex. 10 at 4-5 *see also* Arkes and Koehler, Govt. Ex. 56 at 164 ("our view is that inconclusives should not be scored as either correct or incorrect when computing forensic error rates").

Nonetheless, even if this Court were to exclude inconclusive determinations entirely from its analysis, the error rate would still be quite low and this factor would still favor admission of this evidence. More specifically, when excluding inconclusive results from the error rate

calculations, the most comprehensive studies all show error rates that are sufficiently low that they would be considered reliable pursuant to the PCAST Report, which indicated that false positive rates should be lower than five percent in order to be considered reliable. *See* PCAST Report at 152; *see also United States v. Brown*, 973 F.3d 667, 704 (7th Cir. 2020) (noting that error rate was low for *Daubert* purposes where the error rates in the reliability studies were "in the single digits"). For example, the Ames I study had a 1.524% false identification rate and a 0.371% false elimination rate if inconclusive results are omitted; the Ames II study had a 2.039% false identification rate and a 3.670% false elimination for the bullets examined in that study, and a 1.428% false identification rate and a 2.312% false elimination rate for the cartridges examined in that study, if inconclusive results are omitted; and Dr. Guyll's study had an overall 0.9% false identification rate and an overall 0.1% false elimination rate if inconclusive results are omitted.

The parties devote much of their error rate arguments to the related issue of whether the results from firearm examinations are repeatable (meaning the same examiner will reach the same result when he or she re-examines the same items of evidence) and reproducible (meaning different examiners will reach the same result when he or she examines the same items of evidence). Most courts have not focused on these metrics when evaluating the error rates for firearms examiners, *but see Abruquah*, 296 A.3d at 993-94, and it is not clear how much weight repeatability and reproducibility should be given when evaluating error rates. Indeed, as Judge Gould noted in *Abruquah*, "consistent results from separate examinations of the same sample, by either the same or a different examiner, are not necessarily desirable. Certainly, consistency is good if the initial determination is correct. But consistency is undesirable if the initial determination is wrong, in which case we would prefer disagreement. That is, we would prefer that the same examiner or another examiner get it right the second time rather than repeat the mistake. Disagreement with an

incorrect determination would increase our confidence that a peer review process would catch and correct mistakes, particularly false positives." 296 A.3d at 1016-17.

Moreover, this Court would reach the same result even if repeatability and reproducibility were relevant to the error rate analysis. More specifically, in the Ames II study, if the three types of inconclusive results are viewed as the same result – namely, inconclusive – the repeatability rates are 83.4% and 83.6% for matching and non-matching bullet comparison sets, and 80.9% and 78.9% for matching and non-matching cartridge sets, *see* Govt. Ex. 36 at 42, Table XII, while the reproducibility rates are 72.3% and 54.6% for matching and non-matching bullet sets, and 70.4% and 54.9% for matching and non-matching cartridge cases. *See* Govt. Ex. 36 at 49, Table XVII. The Court agrees with the Government that it is proper to aggregate inconclusive results for the purposes of determining repeatability and reproducibility, because, as Judge Gould noted in *Abruquah*, "if we return to the primary focus of our inquiry – the reliability of a conclusive determination – we should be less concerned with movement within the inconclusive categories." 296 A.3d at 1018. Further, it is important to note that the vast majority of disagreements occurred on the margins – from inconclusive to identification or exclusion - and were not reversed conclusions where an examiner went from exclusion to identification or vice versa. Indeed, in Ames II, reversed conclusions occurred in only 2.68% and 0.11% of bullet examinations of

matching and non-matching sets, respectively, and 2.14% and 0.32% of cartridge examinations in matching and non-matching sets, respectively. *See* Govt. Ex. 6 at Table 5.[15]

In addition, the Court rejects the Defendant's argument that the error rates from these studies are likely to be lower than the error rates in practice. To the contrary, this argument is directly contradicted by the results of the Neuman study, in which the Houston Forensic Science Center conducted a blind quality control program in firearms identification from December 2015 to June 2021, a period of five and a half years of actual case work and found no false identifications or false eliminations.[16] Indeed, as one court recently noted, because many studies use consecutively manufactured items with the least variation in toolmarks to be compared or steel bullets that do not mark as easily as brass, "these extra hurdles in the academic studies show their resulting error rate is likely higher than the error rate in practical field work." *United States v. James*, 2024 U.S. Dist. LEXIS 7615, at *9 (D.V.I. 2024).

For all these reasons, this Court joins the overwhelming majority of courts to have addressed this issue after an evidentiary hearing and finds that the error rate for firearm and toolmark identification is low and that this factor favors admission of this evidence. *See, e.g.*, *James*, 2024 U.S. Dist. LEXIS, at **15-16; *Rhodes*, 2023 U.S. Dist. LEXIS 728, at **11-13; *Felix*, 2022 U.S. Dist. LEXIS 213513, at **50-51; *Harris*, 502 F.Supp.3d at 39; *United States v. Romero-*

[15] It should be noted that the results of Mr. Monturo's examination were not identical to the results of Dr. Warren's examination. Although both experts opined that the five cartridge casings they examined were fired by the gun at issue, Mr. Monturo opined that four of the bullets were fired by the same gun and that he could not reach a conclusion as to the fifth bullet, while Dr. Warren opined that he could not reach a conclusion as to any of the five bullets. This disagreement is similar to the vast majority of disagreements in the Ames II study, because it does not involve a reversed result from identification to exclusion or vice versa. The Court also notes that both Mr. Monturo's examination and Dr. Warren's examinations were subsequently verified by another examiner. This second layer of review further diminishes the likelihood of a false identification being presented to the jury. *See, e.g., Chavez*, 2021 U.S. Dist. LEXIS 237830, at *12 (the use of a verifier provides a second layer of review, and the possible use of a defense firearms expert can provide a third level of review, and these additional layers of review "would drastically reduce the ultimate false positive rate at trial.").

[16] Some examiners in the Neuman study learned that they were being tested, but their error rates and inconclusive rates were very similar to those examiners who never learned they were being tested. *See* Govt. Ex. 14 at 24.

Lobato, 379 F.Supp.3d 1111, 1119-20 (D. Nev. 2019); *United States v. Johnson*, 2019 U.S. Dist. LEXIS 39590, at **55-57 (S.D.N.Y. 2019).

E. Existence of Standards

The firearms and toolmark examiner conducts examinations pursuant to a number of standards. First, the AFTE has published several standards to guide firearms examiners, including the AFTE Training Manual, the AFTE Technical Procedures Manual, the AFTE Glossary and the AFTE Theory of Identification. *See* Govt. Ex. 2 at 50. In addition, OSAC has published guidelines for examiners, *see id.*, and the Department of Justice has published guidelines for the language to be used in testimony and reports by firearms examiners. *See* United States Department of Justice Uniform Language for Testimony and Reports for the Forensic Firearms/Toolmarks Discipline – Pattern Matching Examination.

However, as noted above, the AFTE Theory of Identification, which is the primary standard governing the conclusions reached by firearm examiners, is a subjective standard, without any numerical benchmarks to be used in reaching any particular conclusion. Moreover, it appears to be entirely tautological, under which an examiner can find sufficient agreement between two items if the agreement in toolmarks is significant. Therefore, the Court agrees with the majority of courts, including those that have liberally allowed the use of such testimony, in finding that this factor weighs against admissibility. *See, e.g.*, *Harris* 520 F.Supp.3d at 41-42; *Chavez*, 2021 U.S. Dist. LEXIS 237830, at *14 ("[T]his is the *Daubert* factor on which firearm toolmark analysis scores the lowest.") (citation omitted).

F. General Acceptance

As noted above, general acceptance is not a linchpin for admissibility, but it "can be an important factor in ruling particular evidence admissible." *Motorola*, 147 A.3d at 758 (quoting

Daubert, 509 U.S. at 594). Further, as the Maryland Supreme Court noted in *Abruquah*, the question of whether the AFTE Theory of Firearms Identification is generally accepted by the relevant community is "largely dependent on what the relevant community is." 296 A.3d at 994.

If the Court were to consider the community of firearm examiners to be the relevant community, it seems clear that there is general acceptance of this theory. *Id.* On the other hand, if the Court were to consider the relevant community to consist of scientists, statisticians, and other academics who have studied the issue, the question of general acceptance is much less clear. Indeed, as this hearing has demonstrated, there is a heated debate between non-firearm examiners as to the validity of firearms identification generally and the AFTE Theory more specifically.

Although the majority of courts that have addressed this issue have found that the AFTE Theory of Firearms Identification is generally accepted in the firearms examination community, this Court agrees with the *Abruquah* Court that the relevant community for the purpose of general acceptance should not be limited to either firearm examiners or non-firearm examiners and instead "consists of both firearms examiners and the broader scientific community that has weighed in on the reliability of the methodology." *Id.* Likewise, and largely for the reasons set forth in *Abruquah*, the Court finds this factor to be neutral. *Id.*

G. Overall Evaluation of Relevant Factors

The Court finds that the relevant factors support the admission of expert testimony about both the individual and class characteristics of the casings at issue in this case, and of expert testimony that the casings recovered at the crime scene are consistent with the gun recovered from the crime scene. Although the standards used by firearm and toolmark examiners are subjective and lack statistical benchmarks (a factor weighing against the admission of this testimony), the remaining factors weigh heavily in favor of the admission of this testimony, particularly the low

error rate associated with firearms and toolmark analysis, a factor that both parties agree is the most important factor for the Court's consideration. Put another way, the Court finds that the Government has established, by a preponderance of the evidence, that the existing empirical data shows both that firearms produce unique toolmarks on ammunition and that firearms examiners can accurately discern and compare those toolmarks when deciding whether a particular piece of ammunition was fired by a particular firearm.

The more difficult question is whether the evidence also supports the Government's request that its firearms expert be allowed to testify that the observed class and individual characteristics provide "extremely strong support" for the proposition that the two toolmarks originated from the same source and "extremely weak support" for the proposition that the two toolmarks originated from different sources. Although Dr. Guyll testified in favor of this language, one of the Government's other experts, Dr. Budowle, who is a renowned expert in the field of DNA analysis, expressed hesitation with the "extremely strong support" language, indicating that this language, which arose in the DNA context, might not fit as well in the firearms context. *See* 9/13/23 Budowle Tr. at 192 ("And translating a verbal scale for human DNA or biological stuff to firearms may not be in itself a meaningful one because different populations, different assumptions, different structures.").

The Court shares Dr. Budowle's hesitation. There is a strong statistical basis for this language in the DNA context, where comparisons are made with statistical benchmarks, and where DNA analysts conclude that there is "very strong support" for inclusion only where it is at least one million times more likely that the DNA profile originated from the defendant than from an unknown, unrelated contributor. *See* Department of Justice Uniform Language for Testimony and Reports for Forensic Autosomal DNA Examinations Using Probabilistic Genotyping Systems at

4. There are no equivalent statistical benchmarks in the firearms contexts, and the likelihood ratios described by Dr. Guyll in this context did not compare to the likelihood ratios justifying such language in the DNA context.[17] Likewise, Dr. Thompson testified that the likelihood ratios only demonstrated moderate or moderately strong support for firearm and toolmark examination, using the verbal scale employed by analysts in the DNA context, not very strong support. *See Amici* Ex. 3 at 14, Table 1. For these reasons, the Court will deny the Government's request that its firearms expert be allowed to testify that there is "extremely strong support" for the proposition that two toolmarks originated from the same source and "extremely weak support" for the proposition that the two toolmarks originated from different sources.

H. Consistency with the Relevant Case Law

The Court believes its decision in this case is consistent with the Court of Appeals' case law for the following reasons. First, as noted above, the scientific evidence before the Court in this case is dramatically different from the factual records before the Court of Appeals in cases like *Gardner*, *Williams*, and *Geter*. In those cases, the only reports mentioned by the Court of Appeals were the 2008 and 2009 NRC reports in *Gardner*, 140 A.3d at 1183, the NRC reports and the PCAST Report in *Williams*, 210 A.3d at 739-41, and the PCAST Report in *Geter*. 306 A.3d at 134. By contrast, this Court has not only considered the NRC reports and the PCAST Report, but it also has had a chance to closely review ten additional studies, whose designs meet the PCAST recommendations, in reaching its determination that there is a sufficient scientific foundation for firearms examiners to offer conclusions consistent with this Court's Order. *See* Govt. Ex. 4-5, 335-36 (Ames I); 6, 19, 43 (Ames II); 10-11 (Guyll); 37 (Chapnick); 38 (Duez); 40 (Keisler); 41, 55

[17] The likelihood ratio in Dr. Guyll's study for identifications was 177.458, *see* Govt. Ex. 14 at 19, a number that is much lower than the one million likelihood ratio that is needed for a DNA analyst to testify that there is "very strong support" for inclusion in the DNA context.

(Neuman); Def. Ex. 3 (Law); Def. Ex. 13 (Smith); and Def. Ex. 50 (Best). The additional scientific evidence presented in this case is exactly what the Court of Appeals identified as the pre-requisite for opinion testimony that could go beyond the limits set by the Court of Appeals in cases like *Gardner*, *Williams*, and *Geter*. *See, e.g., Williams*, 210 A.3d at 743 ("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. Rather, we conclude only that we do not have such a foundation in this case.").[18]

Second, and just as importantly, this Court's order does not permit the admission of testimony from a firearms examiner that unqualifiedly connects a specific piece of ammunition to a specific firearm. Thus, this case is unlike *Gardner*, where the trial court allowed the examiner to testify "unequivocally" and without any qualification or limitation that the silver gun at issue in that case "fired the killing shot." 140 A.3d at 1182.[19] This case also is unlike *Williams*, where the trial court allowed the examiner to testify that the gun recovered from the defendant's apartment "matched" the bullets recovered from the decedent's SUV, that these bullets "were fired from this firearm," and that there was no doubt in his mind that the bullets recovered from the decedent's SUV were fired from the gun found in the defendant's bedroom. 210 A.3d at 737.[20] Finally, this case is not like *Geter*, where this Court allowed the examiner to testify "without reservation" that

[18] To be clear, this Court is not endorsing unqualified opinion testimony in this case. *See infra.* Rather, the Court is citing *Williams* for the common-sense proposition that the trial court must make its decision based on the evidence that was presented in the hearing at issue, and that it is not foreclosed from evaluating this evidence differently than the Court of Appeals previously evaluated this type of evidence when the factual record before the Court of Appeals was significantly different from the factual record before this Court.

[19] More specifically, the expert testified on direct that the bullet recovered from the decedent's head was fired from the silver gun, testified on cross-examination that the bullet was fired from the silver gun, and testified on re-direct, after being asked whether the bullet was fired from the silver gun or was "consistent with" being fired from the silver gun, that the bullet "was identified as having been fired from [the silver gun]." 140 A.3d at 1182. Thus, the expert not only testified without any qualification or limitation numerous times, but he also explicitly rejected the alternative wording that would have placed a qualification on his testimony.

[20] More specifically, when the examiner was asked on redirect whether there was "any doubt in [his] mind" that the bullets recovered from the decedent's SUV were fired from the gun found in the defendant's bedroom, the examiner responded "[n]o, sir." 210 A.3d at 737.

the cartridge casings found at the crime scene came from the gun found at the scene because the casings found at the scene "matched" the casings that were test fired from the gun found at the scene. *Geter*, 306 A.3d at 132-33.

By contrast, in this case, the Court will be placing several significant limitations on the examiner's testimony. First, the Court will not permit the examiner to state that his conclusions are to a 100% certainty, to a reasonable degree of scientific certainty, or based on a comparison to all other firearms or toolmarks. Second, the examiner will have to qualify his opinion by stating that his conclusions are not based on a statistically derived or verified measure and that there is not a generally accepted statistical method for conveying the weight of an identification. Third, the examiner will have to qualify his opinion by making clear that his opinion is based on his subjective determination of sufficient agreement in individual characteristics or random imperfections. And finally, the examiner will have to qualify his opinion by testifying that the relevant cartridge casings are "consistent with" having been fired from the firearm at issue, not that they were fired from the firearm at issue, and not even, as the Government has proposed, that there is extremely strong support for the proposition that the casings were fired from the firearm at issue.

These qualifications ensure that the Court's order is consistent with the Court of Appeals' holdings in *Gordon*, *Williams*, and *Geter*, and likewise is consistent with the Court of Appeals' description in *Geter* of the type of language that would constitute a sufficient qualification of the firearm examiner's opinion. *See Geter*, 306 A.3d at 134, n.7 (noting with approval the qualifying language used in *Gordon*, 285 A.3d at 219-20, where the examiner testified that the casings "most likely" were fired by a Glock semiautomatic and that the bullets were "consistent" with being fired

from a Glock.).[21] The Court also notes that its "consistent with" qualification is consonant with the qualifications on firearm examiner testimony that were recently endorsed by the Maryland Supreme Court in *Abruquah*, 296 A.3d at 967, 998 ("The reports, studies and testimony presented to the circuit court demonstrate that the firearms identification methodology employed in this case can support reliable conclusions that patterns and markings on bullets are consistent or inconsistent with those on bullets from a particular firearm," and the trial court did not abuse its discretion in allowing the examiner to testify "whether the patterns and markings on the crime scene bullets are consistent or inconsistent with the patterns and markings on the known bullets.").

In sum, given the abundant and convincing scientific evidence presented to the Court in this case, and given the significant qualifications that the Court will impose on the firearm examiner's testimony in this case, the Court believes that its Order fully complies with the relevant case law on this issue as explicated by the Court of Appeals in *Gardner* and its progeny.

CONCLUSION

The scientific landscape of firearm examination and identification testimony has changed markedly since the 2008 and 2009 NRC reports and the 2016 PCAST report, which were the basis of the Court of Appeals' conclusions in cases such as *Gardner*, *Williams*, and *Geter* that there was an insufficient empirical foundation for a firearms examiner to link a particular gun to a particular piece of ammunition. In contrast to the lack of any empirical data in those cases, this Court has had the opportunity to consider and evaluate an extensive body of scientific literature and expert testimony that convincingly shows that guns do produce unique toolmarks on ammunition and that

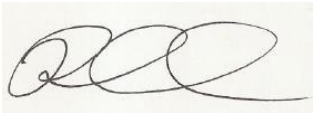
[21] The Court recognizes that the firearms examiner in *Gordon* did not link the bullets at issue to a specific gun, unlike the proposed testimony in this case. But that fact was not the relevant consideration for the Court of Appeals; rather, the relevant consideration for the Court of Appeals can be found in the language that the Court of Appeals quoted in showing that the examiner's testimony was sufficiently qualified – namely, that the examiner's opinions were modified by the adjectival phrases "most likely" and "consistent with." *Geter*, 306 A.3d at 134, n.7.

examiners can reliably evaluate those toolmarks and find connections between guns and ammunition based on those toolmarks. However, the Court is not persuaded that the science has advanced to the point where an examiner can testify that there is "extremely strong support" for the proposition that the toolmarks at issue originated from the same source. For these reasons, the Court will allow the examiner to testify as to both the class and individual characteristics he has observed in the ammunition at issue and to offer an opinion as to the source of the ammunition in the manner set forth on page 2 of this Order.

Accordingly, it is this 1st day of April, 2024, hereby

ORDERED that Defendant's Motion to Exclude Expert Testimony in Firearms Identification is **DENIED**; and it is further

ORDERED that the Government's firearms expert will be permitted to offer testimony in the manner set forth on page 2 of this Order.[22]



Judge Robert Okun
(Signed in Chambers)

Copies to:

Mani Golzari, Candace Mitchell, Jessica Willis
Defendant's Counsel

Sharon Donovan, Michael Spence, Monica Trigoso
United States Attorney's Office

Donald Salzman, Tania Brief, Marley Ann Brumme, James Moody
Counsel for Amici Curiae

[22] The Court expresses its appreciation for the high quality of the advocacy of both parties and *amici* in this case, which reflected a degree of preparation that was very impressive.