

**SUPERIOR COURT OF THE DISTRICT OF COLUMBIA  
CRIMINAL DIVISION – FELONY BRANCH**

<b>UNITED STATES</b>	:	
	:	<b>Case No. 2016 CF1 19431</b>
<b>v.</b>	:	
	:	<b>Judge Todd E. Edelman</b>
<b>MARQUETTE TIBBS</b>	:	

**MEMORANDUM OPINION**

In this case, the defense raised and extensively litigated its objection to the government’s proffer of expert testimony regarding firearms and toolmark identification, a species of specialized opinion testimony that judges have routinely admitted in criminal trials. Specifically, the government sought to introduce the testimony of the firearms and toolmark examiner who used a high-powered microscope to compare a cartridge casing found on the scene of the charged homicide with casings test-fired from a firearm allegedly discarded by a fleeing suspect. According to the government’s proffer, this analysis permitted the examiner to identify the recovered firearm as the source of the cartridge casing collected from the scene. The defense argued that such a conclusion does not find support in reliable principles and methods, and thus must be excluded pursuant to the standard set by the District of Columbia Court of Appeals in *Motorola Inc. v. Murray*, 147 A.3d 751 (D.C. 2016) (en banc); by the United States Supreme Court in *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579 (1993); and by Federal Rule of Evidence 702.

Courts across the country have regularly admitted such source attribution statements from firearms and toolmark examiners, without restriction, for several decades. However, on the heels of several major reports emanating from outside of the judiciary calling into question the

foundations of the firearms and toolmark identification discipline, recent decisions of the District of Columbia Court of Appeals have imposed significant limitations on the conclusions that an expert in this field can render in court.

After conducting an extensive evidentiary hearing in this case—one that involved detailed testimony from a number of distinguished expert witnesses, review of all of the leading studies in the discipline, pre- and post-hearing briefing, and lengthy arguments by skilled and experienced counsel—this Court ruled on August 8, 2019 that application of the *Daubert* factors requires substantial restrictions on specialized opinion testimony in this area. Based largely on the inability of the published studies in the field to establish an error rate, the absence of an objective standard for identification, and the lack of acceptance of the discipline’s foundational validity outside of the community of firearms and toolmark examiners, the Court precluded the government from eliciting testimony identifying the recovered firearm as the source of the recovered cartridge casing. Instead, the Court ruled that the government’s expert witness 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. The Court issues this Memorandum Opinion to further elucidate the ruling it made in open court.

## **I. BACKGROUND**

### *A. Firearms and Toolmark Identification: The Basics*

Numerous reports and court decisions have described in detail the theory and methodology behind the forensic discipline of firearms and toolmark identification. *See, e.g., United States v. Johnson*, (S5) 16 Cr. 281 (PGG), 2019 U.S. Dist. LEXIS 39590, at \*16–21,

2019 WL 1130258, at \*5–7 (S.D.N.Y. Mar. 13, 2019); *United States v. Simmons*, Case No. 2:16cr130, 2018 U.S. Dist. LEXIS 18606, at \*5–11, 2018 WL 1882827, at \*2–3 (E.D. Va. Jan. 12, 2018); *United States v. Otero*, 849 F. Supp. 2d 425, 427–28 (D.N.J. 2012); *United States v. Monteiro*, 407 F. Supp. 2d 351, 359–61 (D. Mass. 2006); *United States v. Green*, 405 F. Supp. 2d 104, 110–12 (D. Mass. 2005); Nat’l Res. Council, Nat’l Academies, *Strengthening Forensic Science in the United States: A Path Forward* 150–51, 152–53 (2009) [hereinafter *2009 NRC Report*]. In short, this field endeavors to match the components of spent ammunition, i.e., bullets and cartridge casings, to a particular firearm. *See Monteiro*, 407 F. Supp. 2d at 359. Firearms and toolmark identification is a specialized area of forensic toolmark identification, a discipline concerned with matching toolmarks to the specific tools that made them. *Otero*, 849 F. Supp. 2d at 427. Forensic toolmark identification 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. *2009 NRC Report* at 150.

The discipline of firearms and toolmark identification derives from the theory that the tools used in the manufacture of firearms leave distinct markings on the internal components of a firearm, such as the barrel, breech face, and firing pin. *Otero*, 849 F. Supp. 2d at 427. These distinct markings, sometimes referred to as “individual characteristics,” are said to result from the cutting, drilling, grinding, and hand-filing involved in the firearm manufacturing process. *Monteiro*, 407 F. Supp. 2d at 359. Such markings are supposedly individualized to each particular firearm as a result of the changes undergone by the tool being used to manufacture the firearm each time it cuts and scrapes metal to produce a new weapon. *Otero*, 849 F. Supp. 2d at 427. According to the theory, no two firearms, even those consecutively produced on the same production line, should bear microscopically identical toolmarks. *See id.*

When a firearm discharges a round of ammunition, the components of that ammunition come into contact with the internal components of the firearm. *Monteiro*, 407 F. Supp. 2d at 359–60. According to the proponents of firearms and toolmark identification, the tool markings on the firearm then transfer to the ammunition’s components. *Id.* at 360. The theory underlying firearms and toolmark identification ultimately hypothesizes that “no two firearms should produce the same microscopic features on bullets and cartridge cases such that they could be falsely identified as having been fired from the same firearm.” *Id.* at 361 (citation omitted). Stated more simply, firearms and toolmark examiners believe they can trace the toolmarks left on spent ammunition back to a particular firearm and that firearm only. *See 2009 NRC Report* at 150.

Trained firearms examiners generally follow a particular methodology in attempting to reach conclusions as to the source of a bullet or cartridge casing. By using a comparison microscope to examine the markings on ammunition test fired from a particular firearm and those on spent ammunition recovered from a crime scene, trained firearms examiners attempt to determine whether the spent ammunition was fired from that particular firearm. *See Monteiro*, 407 F. Supp. 2d at 361. When making these comparisons, examiners observe three types of characteristics of the ammunition—class, subclass, and individual characteristics. *Otero*, 849 F. Supp. 2d at 428. “Class characteristics are gross features common to most if not all bullets and cartridge cases fired from a *type* of firearm,” such as caliber and the number of lands and grooves on a bullet. *Id.* (emphasis added). These characteristics are predetermined at manufacture, *Simmons*, 2018 U.S. Dist. LEXIS 18606, at \*8, 2018 WL 1882827, at \*2, and have been described as “family resemblances,” *Monteiro*, 407 F. Supp. 2d at 360. Subclass characteristics appear on a smaller subset of a particular make and model of firearm, such as a group of guns

produced together at a particular place and time. *Id.* They are produced incidental to manufacture, sometimes as the result of being manufactured by the same irregular tool. *Otero*, 849 F. Supp. 2d at 428. 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. *Id.* These are the markings purported to be unique to a particular firearm and that permit an individualized source determination—in other words, a conclusion that a particular firearm discharged a particular component of ammunition. *See United States v. Taylor*, 663 F. Supp. 2d 1170, 1174 (D.N.M. 2009).

The forensic examination begins with the identification of class characteristics. *2009 NRC Report* at 152. If the observable class characteristics differ between the recovered and test fired ammunition, the examiner can immediately eliminate the recovered firearm as the source of the recovered ammunition. President’s Council of Advisors on Sci. and Tech., Executive Off. of the President, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods* 104 (2016) [hereinafter PCAST Report]. If the class characteristics match, the examiner will use the comparison microscope to identify and compare the individual characteristics in both samples. *Id.* Under the theory of identification promulgated by the Association of Firearm and Tool Mark Examiners (“AFTE”) and discussed in detail *infra* at Section III(D), an examiner may declare the two samples to be of common origin (i.e., fired from the same gun) if she finds “sufficient agreement” between their individual characteristics. *See 2009 NRC Report* at 153. Dissimilarities in observed subclass and/or individual characteristics can allow an examiner to exclude or eliminate the firearm as the source of the questioned sample of ammunition. The examiner may also render an inconclusive determination when there is agreement between the two samples’ class characteristics but insufficient agreement or

disagreement between their individual characteristics to make an identification or exclusion determination. *See Johnson*, 2019 U.S. Dist. LEXIS 39590, at \*9, 2019 WL 1130258, at \*3.

*B. Proffered Firearms and Toolmark Evidence in this Case, and the Defendant's Motion to Exclude*

Mr. Tibbs is charged with one count of first degree murder while armed as well as other related offenses. According to the government, a .40 caliber Smith & Wesson cartridge casing from a semi-automatic weapon was recovered from the scene of the homicide on November 11, 2016. The government alleges that a police officer observed Mr. Tibbs discarding a .40 caliber Smith & Wesson semi-automatic pistol shortly after the homicide occurred. On December 21, 2016, District of Columbia Department of Forensic Sciences Examiner Christopher Coleman prepared a report of examination, which indicated the recovered cartridge casing “was microscopically examined and identified as having been fired in [the recovered pistol], based on breechface marks and firing pin aperture shear marks.” Christopher Coleman, D.C. Dep’t of Forensic Sci., *Report of Examination: Firearms Examination Unit Report 1* (Dec. 21, 2016), Def.’s Mot. Ex. A, at 3 (Dec. 18, 2018).

Through his counsel, Mr. Tibbs challenged the admissibility of Mr. Coleman’s opinion testimony with regard to firearms and toolmark identification. Specifically, the Defendant filed his Motion to Exclude the Testimony of Government’s Proposed Expert Witness in Firearms Examination (“Defendant’s Motion”) on December 18, 2018. The government filed its Opposition to Defendant’s Motion on January 24, 2019; the Defendant filed a Reply on March 23, 2019, to which the government filed a Surreply on April 15, 2019. The defense supplemented its pleadings with affidavits from Professor David Faigman and Dr. Nicholas

Scurich, while the government submitted a declaration from Todd J. Weller, a report by Dr. Nicholas Petraco, and an affidavit from Dr. Bruce Budowle.

The Court conducted an extensive hearing on Defendant's Motion during the week of May 13, 2019, hearing lengthy testimony from Dr. Petraco, Mr. Weller, Dr. Scurich, and Professor Faigman. The parties' arguments on these issues spanned several days and finally concluded on June 10, 2019. Subsequent to the conclusion of the hearing, the Court provided the parties with the opportunity to file supplemental pleadings on the effect of the District of Columbia Court of Appeals' June 27, 2019 decision in *Williams v. United States (Williams II)*, 210 A.3d 734 (D.C. 2019), on the Court's resolution of Defendant's Motion; the parties each filed such a brief on July 10, 2019.<sup>1</sup>

In his written pleadings, the Defendant asked the Court to exclude all testimony regarding firearms examination and identification in this case. In the alternative, he requested that the Court preclude Mr. Coleman from testifying that the recovered pistol fired the recovered cartridge casing, and limit his testimony to a conclusion that he could not exclude the recovered firearm as the source of the recovered cartridge casing. At the hearing, Mr. Tibbs proposed alternative restrictions on Mr. Coleman's proposed testimony but ultimately conceded that Mr. Coleman should at least be permitted to testify about his comparison of class characteristics between the recovered and test fired cartridge casings.

---

<sup>1</sup> On June 27, 2019, the government also filed a Motion to Correct Factual Inaccuracies in the Record. The Defendant filed his Reply on August 2, 2019.

## II. LEGAL STANDARD

### A. *Daubert and Rule 702: General Principles*

In 2016, the District of Columbia Court of Appeals, sitting *en banc*, abandoned this jurisdiction's previous standard for the admissibility of expert opinion testimony. *Motorola*, 147 A.3d at 756–57. That standard, commonly referred to as the *Frye/Dyas* test, was originally developed by the United States Court of Appeals for the District of Columbia, and held that a scientific technique or principle could serve as the subject of expert testimony to the extent it had been “general[ly] accept[ed]” within its field of origin. *See Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923). *See generally Dyas v. United States*, 376 A.2d 827, 831–32 (D.C. 1977). In *Motorola*, the Court of Appeals adopted the admissibility standard announced by the United States Supreme Court in *Daubert*—the same standard that has been applied in federal courts for over twenty years and that now appears in Federal Rule of Evidence 702. *See Motorola*, 147 A.3d at 756–57.

*Daubert* itself repudiated *Frye* by holding its standard had been “superseded by the adoption of the Federal Rules of Evidence” and, in particular, by Rule 702. *See* 509 U.S. at 587–89. The Supreme Court stated that trial judges considering the admissibility of proffered expert opinion testimony must conduct a “preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue.” *Id.* at 592–93. Thus, under *Daubert* and Rule 702, the admissibility of proffered expert opinion testimony does not exclusively rest on the acceptance of the opinion's underlying theory or methodology within a community of scientists or practioners. *See id.* at 594–95. Nor does it turn on the trial judge's view on the ultimate accuracy of the offered conclusion. *See id.* at 595. Instead, the admissibility inquiry

focuses on whether reliable principles and methods support the proposed testimony and on whether those principles and methods were reliably applied in the case at hand. *Id.* at 594–95; *see also Motorola*, 147 A.3d at 754. Rule 702 articulates the elements of the *Daubert* inquiry:

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.

In changing the standard for the admissibility of expert opinion testimony, *Daubert* also modified the judge’s role in making the admissibility determination. A judge must serve as a gatekeeper to “ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable.” *Daubert*, 509 U.S. at 589.<sup>2</sup> Indeed, *Daubert*, its progeny, and subsequent amendments to Rule 702 “gave to the courts a more significant gatekeeper role with respect to the admissibility of scientific and technical evidence than courts previously had played.” *United States v. Glynn*, 578 F. Supp. 2d 567, 569 (S.D.N.Y. 2008). *Daubert* noted that such an assessment would involve the examination of a diverse set of factors. *See* 509 U.S. at 593. Envisioning a flexible inquiry, the Supreme Court did “not presume to set out a definitive checklist or test.” *Id.* at 593–94. It did, however, enumerate five factors that would generally guide a trial court’s admissibility inquiry:

- (1) whether a theory or technique can be (and has been) tested;

---

<sup>2</sup> In *Kumho Tire. Co. v. Carmichael*, the United States Supreme Court held that the *Daubert* reliability standard applies not just to expert testimony based on “scientific” knowledge, but to testimony based on “technical” or “other specialized” knowledge as well. 526 U.S. 137, 149 (1999).

- (2) whether the theory or technique has been subjected to peer review and publication;
- (3) the theory's or technique's known or potential rate of error;
- (4) the existence and maintenance of standards controlling the technique's operation; and
- (5) whether the theory or technique is generally accepted within the relevant scientific community.

*Id.*; see also *Motorola*, 147 A.3d at 754.

The proponent of the expert testimony bears the burden of proving its reliability by a preponderance of the evidence. *Cf. Daubert*, 509 U.S. at 592 n.10. Our Court of Appeals has consistently held that admissibility determinations are within the discretion of the trial court. *See, e.g., Johnson v. United States*, 960 A.2d 281, 296 (D.C. 2008) (citing *Dockery v. United States*, 853 A.2d 687, 697 (D.C. 2004); *Smith v. United States*, 686 A. 2d 537, 542 (D.C. 1996))

*B. Daubert and Firearms and Toolmark Identification*

1. Mr. Tibbs's *Daubert* challenge

Mr. Tibbs raised a general challenge to the reliability of the principles and methods underlying firearms and toolmark identification. *See generally* Def.'s Mot. Accordingly, he at times moved to exclude all such evidence. At other points in his pleadings and arguments, however, he offered a series of concessions and alternative proposals as well. As described in the Court's August 8, 2019 oral ruling, the undersigned found it useful to conceptualize Mr. Tibbs's challenge in several different ways. The Court could have analyzed the issues raised in Defendant's Motion by first determining whether the discipline of firearms and toolmark identification generally employs reliable principles and methods—such that it is admissible under *Daubert*, *Motorola*, and Rule 702—and subsequently, whether *Daubert* requires any

limitations on the proffered testimony. Alternatively, the Court could have treated Mr. Tibbs's challenge as requiring two separate *Daubert* inquiries: (1) whether the Court could characterize the underlying theory of firearms and toolmark identification—the theory that manufacturing tools leave certain unique marks on firearms, and that firearms therefore leave unique and/or identifiable marks on bullets and cartridge casings—as reliable; and (2) whether the Court could conclude that a firearms examiner's opinion that she can compare bullets or cartridge casings and make an accurate source attribution statement (that is, a conclusion that a particular firearm fired a particular bullet or cartridge casing) finds support in reliable principles and methods. Regardless of the framework under which Mr. Tibbs's challenge was to be evaluated, Defendant's Motion ultimately required the Court to determine what type of opinion, if any, can be rendered with respect to firearms and toolmark evidence.

## 2. The limited persuasive value of existing case law

Judges across the United States have considered similar challenges to firearms and toolmark identification evidence. Of course, “for many decades ballistics testimony was accepted almost without question in most federal courts in the United States.” *Glynn*, 578 F. Supp. 2d at 569. Based on the pleadings in this case, as well as the Court's own research, there do not appear to be any reported cases in which this type of evidence has been excluded in its entirety. Earlier this year, the United States District Court for the District of Nevada also surveyed the relevant case law and concluded that no federal court had found the method of firearms and toolmark examination promoted by AFTE—the method generally used by American firearms examiners and employed by Mr. Coleman in this case—to be unreliable. *United States v. Romero-Lobato*, 379 F. Supp. 3d 1111, 1117 (D. Nev. 2019); *see also Simmons*,

2018 U.S. Dist. LEXIS 18606, at \*28, 2018 WL 1882827, at \*9 (“Defendants concede, as they must, that no court has ever *totally* rejected firearms and toolmark examination testimony.”); *State v. DeJesus*, 7 Wn. App. 2d 849, 864 (2019) (“[T]he judicial decisions uniformly conclude toolmark and firearms identification is generally accepted and admissible at trial.”).

In evaluating the persuasive weight of these decisions, however, the undersigned could not help but note that, despite the enhanced gatekeeping role demanded by *Daubert*, *see* 509 U.S. at 589, the overwhelming majority of the reported post-*Daubert* cases regarding this type of expert opinion testimony have not engaged in a particularly extensive or probing analysis of the evidence’s reliability. In 2009, the National Research Council (“NRC”) specifically criticized the judiciary’s treatment of issues relating to the admissibility of firearms and toolmark evidence and the judiciary’s failure to apply *Daubert* in a meaningful fashion. In the NRC’s view, “[t]here is little to indicate that courts review firearms evidence pursuant to *Daubert*’s standard of reliability.” *2009 NRC Report* at 107 n.82. The NRC observed that trial judges

. . . often affirm admissibility citing earlier decisions rather than facts established at a hearing. Much forensic evidence—including, for example, bite marks and firearm and toolmark identification—is introduced in criminal trials without any meaningful scientific validation, determination of error rates, or reliability testing to explain the limits of the discipline.

*Id.* at 107–08 (footnote and internal quotation marks omitted). Without disparaging the work of other courts, the NRC’s critique of our profession rings true, at least to the undersigned: many of the published post-*Daubert* opinions on firearms and toolmark identification involved no hearing on the admissibility of the evidence or only a cursory analysis of the relevant issues. Our Court of Appeals has noted that “[t]here is no ‘grandfathering’ provision in Rule 702.” *Motorola*, 147 A.3d at 758. Yet, the case law in this area follows a pattern in which holdings supported by limited analysis are nonetheless subsequently deferred to by one court after another. This pattern

creates the appearance of an avalanche of authority; on closer examination, however, these precedents ultimately stand on a fairly flimsy foundation. The NRC credited Professor David Faigman—one of the defense experts who testified at the *Daubert* hearing in this matter—with the observation that trial courts defer to expert witnesses; appellate courts then defer to the trial courts; and subsequent courts then defer to the earlier decisions. *See 2009 NRC Report* at 108 n.85.

It is difficult to avoid the conclusion that, despite the criticisms of the NRC and other bodies, the judicial branch has demonstrated an aversion to meaningful hearings on this issue. In 2005, Judge Nancy Gertner of the United States District Court for the District of Massachusetts commented, “every single court post-*Daubert* has admitted [firearms identification] testimony, sometimes without any searching review, much less a hearing.” *Green*, 405 F. Supp. 2d at 108 (emphasis omitted). Indeed, in 2012, the United States District Court for the Eastern District of New York could identify only four federal cases in which a judge had conducted a *Daubert* hearing on the admissibility of firearms and toolmark evidence. *United States v. Sebborn*, 10 Cr. 87 (SLT), 2012 U.S. Dist. LEXIS 170576, at \*17–18, 2012 WL 5989813, at \*6 (E.D.N.Y. Nov. 30, 2012). Since then, few other federal courts have held similar hearings.<sup>3</sup> *See Romero-Lobato*, 379 F. Supp. 3d at 1114; *Johnson*, 2019 U.S. Dist. LEXIS 39590, at \*4–5, 2019 WL 1130258, at \*2; *Simmons*, 2018 U.S. Dist. LEXIS 18606, at \*3, 2018 WL 1882827, at \*1; *United States v. Wrensford*, Criminal No. 2013-0003, 2014 U.S. Dist. LEXIS 102446, at \*2, 2014 WL 3715036, at \*1 (D. V.I. July 28, 2014). In most cases, courts resolved the objection to firearms and toolmark identification testimony without conducting any hearing at all. *See, e.g., United States v. Hylton*, Case No. 2:17-cr-00086-HDM-NJK, 2018 U.S. Dist. LEXIS 188817, at \*6, 2018 WL

---

<sup>3</sup> Because many decisions on evidentiary issues do not result in the issuance of a reported or written opinion, the weight of authority from other courts and jurisdictions cannot be precisely determined. *See 2009 NRC Report* at 97.

5795799, at \*3 (D. Nev. Nov. 5, 2018); *United States v. White*, 17 Cr. 611 (RWS), 2018 U.S. Dist. LEXIS 163258, at \*5, 2018 WL 4565140, at \*2 (S.D.N.Y. Sept. 24, 2018); *United States v. Johnson*, Case No. 14-cr-00412-TEH, 2015 U.S. Dist. LEXIS 111921, at \*11, 2015 WL 5012949, at \*4 (N.D. Cal. Aug. 24, 2015); *United States v. Ashburn*, 88 F. Supp. 3d 239, 244 (E.D.N.Y. 2015). Even in the few cases in which a *Daubert* hearing was conducted, it most often consisted only of the testimony of the examiner who worked on the case at issue, rather than of experts with a broader understanding of the foundational validity of the field.<sup>4</sup> *See Romero-Lobato*, 379 F. Supp. 3d at 1115; *Johnson*, 2019 U.S. Dist. LEXIS 39590, at \*3–5, 2019 WL 1130258, at \*1–2; *Simmons*, 2018 U.S. Dist. LEXIS 18606, at \*3, 2018 WL 1882827, at \*1. The Court does not suggest that these decisions represent an abuse of discretion by the judges who issued them. The seemingly perfunctory nature of many of these written decisions does, however, lessen the persuasive weight of what would have otherwise been afforded to a near unanimous set of judicial opinions.

### 3. Judicial restrictions on firearms and toolmark identification testimony

Although, as stated *supra*, no trial court has entirely excluded firearms and toolmark evidence in its entirety, some judges admitting firearms and toolmark evidence have recently restricted the conclusions examiners can render before a jury. *See Romero-Lobato*, 379 F. Supp. 3d at 1117; *DeJesus*, 7 Wn. App. 2d at 864 (“Courts have considered scholarly criticism of the methodology, and occasionally placed limitations on the opinions experts may offer based on the

---

<sup>4</sup> Some trial courts have conducted full evidentiary hearings on the admissibility of firearms and toolmark identification evidence. *See Wrensford*, 2014 U.S. Dist. LEXIS 102446, at \*2, 2014 WL 3715036, at \*1; *Monteiro*, 407 F. Supp. 2d at 355. Others have even considered the recent critiques of firearms and toolmark identification. *See Romero-Lobato*, 379 F. Supp. 3d at 1117–22. These three courts admitted testimony similar to that proffered in this case under the *Daubert* framework. *See Romero-Lobato*, 379 F. Supp. 3d at 1123; *Wrensford*, 2014 U.S. Dist. LEXIS 102446, at \* 58, 2014 WL 3715036, at \*18; *Monteiro*, 407 F. Supp. 2d at 372.

methodology.”). For example, at least one judge has precluded the sponsor of such evidence from referring to it as a “science.” *Glynn*, 578 F. Supp. 2d at 568–69. Other courts have prohibited examiners from stating their conclusions to an absolute or statistical certainty. *See, e.g., Monteiro*, 407 F. Supp. 2d at 372. Some of these judges have permitted examiners to state their opinions only to a “reasonable degree of ballistic certainty” or a “reasonable degree of certainty in the ballistics field,” *see Ashburn*, 88 F. Supp. 3d at 249; *Monteiro*, 407 F. Supp. 2d at 372; *Simmons*, 2018 U.S. Dist. LEXIS 18606, at \*30, 2018 WL 1882827, at \*10, while others have precluded any reference to the concept of “certainty,” regardless of what modifiers the examiner may attach, *see White*, 2018 U.S. Dist. LEXIS 163258, at \*7, 2018 WL 4565140, at \*3; *United States v. Willock*, 696 F. Supp. 2d 536, 549 (D. Md. 2010); *Glynn*, 578 F. Supp. 2d at 568–69. A number of courts have prevented examiners from stating that recovered ballistics evidence can be matched to a firearm to the exclusion of all other firearms. *See Taylor*, 663 F. Supp. 2d at 1180; *Green*, 405 F. Supp. 2d at 124.

Other judges have gone further in limiting expert opinion testimony regarding firearms and toolmark examination. In *Glynn*, a United States District Court Judge permitted a firearms examiner to state his conclusions of the match between the recovered ammunition and recovered firearm in terms of “more likely than not, but nothing more.” 578 F. Supp. 2d at 575 (internal quotation marks omitted). And in *State v. Terrell*, a state trial court judge referenced a case in which he had limited an examiner “to describing the similarities and dissimilarities between the known and unknown shell casings” and allowed her to conclude only that “the casings were consistent with having been fired from the subject hand gun.” CR170179563, 2019 Conn. Super. LEXIS 827, at \*19, 2019 WL 2093108, at \*5 (Mar. 21, 2019). Nonetheless, despite the handful of judges that have imposed these restrictions, “limitations on firearm and toolmark expert

testimony [have been] the exception rather than the rule.” *Romero-Lobato*, 379 F. Supp. 3d at 1117.

The District of Columbia Court of Appeals, in a series of cases, has similarly restricted the conclusions firearms examiners may offer in court. *See Williams II*, 210 A.3d at 738; *Gardner v. United States*, 140 A.3d 1172, 1184 (D.C. 2016); *Jones v. United States*, 27 A.3d 1130, 1139 (D.C. 2011). Although, as discussed in Section IV *infra*, some ambiguity exists as to the state of the law post-*Williams II*, there can be no dispute that these authorities preclude firearms examiners from stating their conclusions with absolute or 100% certainty. *See, e.g., Gardner*, 140 A.3d at 1177. Nor can these expert witnesses identify a particular firearm as the source of spent ammunition to the exclusion of all other firearms. *Id.* Furthermore, it is unlikely examiners are even able to state their conclusions “with a reasonable degree of certainty.” *See id.* at 1184 n.19 (“[W]e have doubts as to whether trial judges in this jurisdiction should permit toolmark experts to state their opinions with a reasonable degree of certainty.” (internal quotation marks omitted)). None of these precedents, however, entirely control the *Daubert* challenge posed by Defendant’s Motion. *Jones*, *Gardner*, and *Williams II* addressed the reliability of an examiner’s conclusion, but all three were decided prior to the Court of Appeals’ decision in *Motorola*—when the *Frye/Dyas* test still governed the admissibility of expert opinion testimony in the District of Columbia. None of them explicitly evaluated the admissibility of firearms and toolmark evidence under *Daubert* and Rule 702. And, while providing some examples of what firearms examiners *cannot* say in court, none of these cases provide definitive guidance as to what these witnesses *can* say.

#### 4. Conclusion

Granted, the precedents from other jurisdictions do provide at least some amount of guidance as to the challenge presented, and the Court of Appeals' recent opinions do have some bearing on the Court's present decision. However, particularly in light of the absence of any District of Columbia authority applying *Daubert* to firearms and toolmark identification testimony and the lack of any particularly persuasive authority from other jurisdictions, Defendant's Motion posed an issue of first impression. Accordingly, the Court undertook to determine the admissibility of the proffered testimony under *Daubert*, *Motorola*, and Rule 702. As explained by Judge Gertner, "*Daubert* plainly raised the standard for existing, established fields, inviting a reexamination even of generally accepted venerable, technical fields. Refusing to do so would be equivalent to grandfathering old irrationality." *Green*, 405 F. Supp. 2d at 118 (internal citations and quotation marks omitted).

### **III. APPLICATION OF THE *DAUBERT* FACTORS TO FIREARMS AND TOOLMARK ANALYSIS**

#### *A. Can and has the technique been tested?*

The first of the *Daubert* factors—whether the technique or process in question can and has been tested—represents a “key question” in determining whether expert testimony should be admitted. *Romero-Lobato*, 379 F. Supp. 3d at 1118. As described in the Advisory Committee Notes to Rule 702, the “testability” of a theory refers to “whether the expert’s theory can be challenged in some objective sense, or whether it is instead simply a subjective, conclusory approach that cannot be reasonably assessed for reliability.” As *Daubert* itself noted, “generating hypotheses and testing them to see if they can be falsified . . . is what distinguishes science from other fields of human inquiry.” *Daubert*, 509 U.S. at 593 (citation omitted).

“There appears to be little dispute that toolmark identification is testable as a general matter.” *Johnson*, 2019 U.S. Dist. LEXIS 39590, at \*44, 2019 WL 1130258, at \*15. Indeed, virtually every court that has evaluated the admissibility of firearms and toolmark identification has found the AFTE method to be testable and that the method has been repeatedly tested. *See, e.g., Romero-Lobato*, 379 F. Supp. 3d at 1118–19; *Simmons*, 2018 U.S. Dist. LEXIS 18606, \*18, 2018 WL 1882827, at \*6; *Ashburn*, 88 F. Supp. 3d at 245; *Otero*, 849 F. Supp. 2d at 433. Although the NRC and PCAST reports have levied significant criticism against firearms and toolmark analysis, courts have found that such reports do not affect the method’s testability. *See, e.g., Romero-Lobato*, 379 F. Supp. 3d at 1119; *see also Otero*, 849 F. Supp. 2d at 433 (“Though the methodology of comparison and the AFTE ‘sufficient agreement’ standard inherently involves the subjectivity of the examiner’s judgment as to matching toolmarks, the AFTE theory is testable on the basis of achieving consistent and accurate results.”). Additionally, some courts have cited annual proficiency testing undergone by firearms and toolmark examiners as further evidence of the method’s testability. *See Johnson*, 2019 U.S. Dist. LEXIS 39590, at \*45–46, 2019 WL 1130258, at \*15 (citing *United States v. Diaz*, No. CR 05-000167 WHA, 2007 U.S. Dist. LEXIS 13152, at \*15, 2007 WL 485967, at \*5 (N.D. Cal. Feb. 12, 2007)); *United States v. Johnson*, 2015 U.S. Dist. LEXIS 111921, at \*9, 2015 WL 5012949, at \* 3.

Here, the propositions advanced by the government in support of its proffer of the expert testimony at issue—namely, that firearms leave discernible toolmarks on bullets and cartridge casings fired from them, and that trained examiners can conduct comparisons to determine whether a particular gun has fired particular ammunition—can be, and have been, tested. The Defendant’s written pleadings and oral argument did not specifically contest this particular point, and the government met its burden with respect to testability.

B. *Has the theory or technique been subjected to peer review and publication?*

The second of the *Daubert* factors considers whether the theory or technique “has been subjected to peer review and publication.” *Motorola*, 147 A.3d at 754 (quoting *Daubert*, 509 U.S. at 593–94). As the Supreme Court emphasized in *Daubert*, “submission to the scrutiny of the scientific community is a component of ‘good science,’ in part because it increases the likelihood that substantive flaws in methodology will be detected.” 509 U.S. at 593. While the existence of peer reviewed literature can help determine a methodology’s reliability under *Daubert*, the “fact of publication (or lack thereof) in a peer reviewed journal” is not dispositive. *Id.*; see also *Romero-Lobato*, 379 F. Supp. 3d at 1119; *United States v. Mouzone*, 696 F. Supp. 2d 536, 571 (D. Md. 2009).

Evidence presented at the hearing demonstrated that studies assessing the foundational validity and reliability of the type of firearms pattern matching evidence proffered here—that is, studies that attempt to show whether trained firearms examiners can accurately attribute a particular firearm as the source of a particular cartridge casing or bullet—have been published and subjected to varying types of review. Two of the studies in this area, the 2019 study by James E. Hamby et al., *A Worldwide Study of Bullets Fired from 10 Consecutively Rifled 9MM RUGER Pistol Barrels—Analysis of Examiner Error Rate*, 64 J. Forensic Sci. 551 (2019) [hereinafter 2019 Hamby Study], and the 2016 study by Tasha P. Smith et al., *A Validation Study of Bullet and Cartridge Case Comparisons Using Samples Representative of Actual Casework*, 61 J. Forensic Sci. 692 (2016) [hereinafter 2016 Smith Study], were published in the *Journal of Forensic Sciences*, and thus have undergone meaningful peer review. The *Journal of Forensic Sciences* employs “double-blind” peer review, a type of review process used throughout many scientific disciplines and designed to limit various types of bias by requiring that neither the

study's authors nor the journal's reviewers know the identity of the other. Scurich Test. May 15, 2019, 37:3-7; Expert Report of Nicholas Scurich, PhD, 6 [hereinafter Scurich Report] (citing Author Guidelines, <https://onlinelibrary.wiley.com/page/journal/15564029/homepage/forauthors.html> (last visited August 28, 2019)). Further, this particular publication is an independent journal, 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.

However, most of the other studies in this field—including the vast majority of those relied upon by the government and the expert witnesses it presented at the *Daubert* hearing—have been published in the *AFTE Journal*, a publication produced by the Association of Firearm and Toolmark Examiners. The government's experts, Mr. Weller and Dr. Petraco, contended that the studies published in the *AFTE Journal* are subjected to both pre- and post-publication peer review. Prior to publication, articles submitted to the *AFTE Journal* are reviewed by AFTE members; the *AFTE Journal* utilizes an "open" pre-publication peer review process in which the author and the reviewers know each other's identity and may communicate directly during the review period. Scurich Report 7 (citing AFTE Peer Review Process – August 2009, <https://afte.org/afte-journal/afte-journal-peer-review-process> (last visited Aug. 28, 2019)). Both government experts primarily focused on post-publication peer review, and characterized letters to the editor in response to a published study as part of the *AFTE Journal*'s peer review process. Suppl. Decl. of Todd J. Weller 7–8 [hereinafter Weller Suppl. Decl.]; Report of Dr. Nicholas Petraco 1–2 [hereinafter Petraco Report]; Petraco Test. May 13, 2019, 20:7–18. Further, Dr. Petraco also discussed the publication of "counter studies" as part of the peer review process. Petraco Report at 2.

Other courts considering challenges to this discipline under *Daubert* have concluded that publication in the *AFTE Journal* satisfies this prong of the admissibility analysis. *See, e.g., Romero-Lobato*, 379 F. Supp. 3d at 1119 (citing *Ashburn*, 88 F. Supp. 3d at 245–46; *Otero*, 849 F. Supp. 2d at 433; *Taylor*, 663 F. Supp. 2d at 1176; *Monteiro*, 407 F. Supp. 2d at 366–67); *Mouzone*, 696 F. Supp. 2d at 571. It is striking, however, that these courts devote little attention to the sufficiency of this journal’s peer review process or to the issues stemming from a review process dominated by financially and professionally interested practitioners, and instead, mostly accept at face value the assertions regarding the adequacy of the journal’s peer review process. *See, e.g., Romero-Lobato*, 379 F. Supp. 3d at 1119; *Johnson*, 2019 U.S. Dist LEXIS 39590, at \*49–50, 2019 WL 1130258, at \*16–17; *Ashburn*, 88 F. Supp. 3d at 245–46; *Wrensford*, 2014 U.S. Dist. LEXIS 102446, at \*43–44, 2014 WL 3715036, at \*13; *Otero*, 849 F. Supp. 2d at 433; *Monteiro*, 407 F. Supp. 2d at 366–67.<sup>5</sup>

In the undersigned’s view, three aspects of publication in the *AFTE Journal* make this journal’s review process far less meaningful (and its published articles that much less reliable) than *Daubert* contemplates. First, as noted *supra*, the *AFTE Journal* peer review process itself is “open,” meaning that both the author and reviewer know the other’s identity and may contact each other during the review process. Scurich Report 7 (citing *AFTE Peer Review Process – August 2009*, <https://afte.org/afte-journal/afte-journal-peer-review-process> (last visited Aug. 28, 2019)). This open process seems highly unusual for the publication of empirical scientific research, as Dr. Scurich testified and as Dr. Petraco admitted in his written report. Scurich Test.

---

<sup>5</sup> Indeed, one court has recently found that the PCAST and NRC Reports themselves—despite their negative treatment of the established validity of firearms and toolmark evidence—constitute relevant peer review of the articles published in the *AFTE Journal*. *See Romero-Lobato*, 379 F. Supp. 3d at 1119. If negative post-publication commentary from an external reviewing body can satisfy this prong of the *Daubert* analysis, then the peer reviewed publication component would be more or less read out of *Daubert*, leaving behind only the requirement of *some type* of publication.

May 15, 2019, 28:17–18; Petraco Report at 2. The practice of double-blind peer review, by contrast, 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. Mr. Weller, even while defending the *AFTE Journal*'s open process, acknowledged that the publication is now moving toward a blind peer review process. Weller Test. May 14, 2019 (1), 23:18; Weller Suppl. Decl. 8. While neither *Daubert*, *Motorola*, nor Rule 702 mandate any specific type of peer review process, the *AFTE Journal*'s use of a so-called “open” process diminishes the extent to which proponents of firearms and toolmark identification evidence can claim that its articles have been subjected to meaningful, stringent peer review.

Second, AFTE does not make this publication generally available to the public or to the world of possible reviewers and commentators outside of the organization's membership. Of course, an interested party can receive the publication by joining AFTE, if such a person meets the organization's membership requirements, or can pay to access specific articles. Weller Test., May 14, 2019 (1), 18:16–21. But unlike other scientific journals, the *AFTE Journal* is not more broadly available and cannot even be obtained in university libraries. *Id.* 18:11–13. Such restricted access effectively forecloses the type of review of the journal's publications by a wider community of scientists, academics, and other interested parties that could serve as an important mechanism for quality assurance. Indeed, a National Commission on Forensic Science's (NCFS) publication listed among the criteria for “foundational, scientific literature supportive of forensic practice” that the articles be “published in a journal that is searchable using free, publicly available search engines (e.g. Pub Med, Google Scholar, National Criminal Justice Reference Service) that search major databases of scientific literature (e.g. Medline, National Criminal Justice Reference Service Abstracts Database, and Xplore)” and “published in a journal that is

indexed in databases that are available through academic libraries and other services (e.g. JSTOR, Web of Science, Academic Search Complete, and SciFinder Scholar).” Nat’l Comm’n on Forensic Sci., *Scientific Literature in Support of Forensic Science and Practice*, 3 (2015), [justice.gov/archives/ncfs/file/786591 /download](https://www.justice.gov/archives/ncfs/file/786591/download) [hereinafter NCFS Report].<sup>6</sup> The *AFTE Journal*, by generally limiting the review of its publications and making them available only to its members or others who pay, avoids the scrutiny of scientists and academics outside the field of firearms and toolmark analysis. These limitations significantly diminish the stringency of the review that a study published in the *AFTE Journal* can be said to have undergone, even after its publication.

Third, the very nature of AFTE impacts the meaningfulness of its review process. The *AFTE Journal* is published by the largest organization of practicing firearms and toolmark examiners, and its articles are reviewed by members of an editorial board composed entirely of members of AFTE. Scurich Report 7 (citing AFTE Peer Review Process – August 2009, <https://afte.org/afte-journal/afte-journal-peer-review-process> (last visited Aug. 28, 2019)). This oversight structure may create a threshold issue in terms of quality of peer review: as Dr. Scurich pointed out, those who review the *AFTE Journal*’s articles may be trained and experienced in the field of firearms and toolmark examination, but do not necessarily have any specialized or even relevant training in research design and methodology. Scurich Report 7–8. Perhaps more importantly, members of the *Journal*’s editorial board—those who review its articles prior to publication—have a vested, career-based interest in publishing studies that validate their own field and methodologies. In contrast with this particular publication’s editorial structure, the National Commission on Forensic Science has specifically stated that foundational

---

<sup>6</sup> Although surely not what the NFCS’s recommendations contemplate, AFTE’s website indicates that the public may search its articles’ abstracts and keywords in its own index available on the AFTE website. See What is the Journal?, <https://afte.org/afte-journal/what-is-the-journal> (last visited Aug. 28, 2019).

scientific literature should be “published in a journal that utilizes rigorous peer review with independent *external* reviewers to validate the accuracy in its publications and their overall consistency with scientific norms of practice.” NCFCS Report at 3 (emphasis added). The *AFTE Journal* is thus, in a sense, “comparable to talk within congregations of true believers” rather than an example of “the desired scientific practice of critical review and debate mentioned in *Daubert*.” David H. Kaye, *How Daubert and its Progeny Have Failed Criminalistics Evidence and a Few Things the Judiciary Could Do About It*, 86 Fordham L. Rev. 1639, 1645 (2018).

While the Court does not doubt the good faith of AFTE or those who serve on the editorial board of the *AFTE Journal*, neither can it ignore this intrinsic bias and lack of independence when analyzing the nature of peer review this journal utilizes.<sup>7</sup> Discussing a similar journal within the field of handwriting analysis, Judge Jed. S. Rakoff of the United States District Court for the Southern District of New York highlighted the issue central to the question of whether publication in the *AFTE Journal* should qualify as peer reviewed publication under *Daubert*: the very meaning of the term “peer.” As Judge Rakoff reasoned:

Of course, the key question here is what constitutes a ‘peer,’ because just as astrologers will attest to the reliability of astrology, defining ‘peer’ in terms of those who make their living through handwriting analysis would render this *Daubert* factor a charade. While some journals exist to serve the community of those who make their living through forensic document examination, numerous courts have found that ‘[t]he field of handwriting comparison . . . suffers from a lack of meaningful peer review’ by anyone remotely disinterested.

*Almeciga v. Ctr. for Investigative Reporting, Inc.*, 185 F. Supp. 3d 401, 420 (S.D.N.Y. 2016)

(citation omitted). So, too, with the field of firearms and toolmark analysis: although studies analyzing error rates among practicing firearms and toolmark examiners have, on two occasions, been published in other journals utilizing double-blind peer review presumably performed by

---

<sup>7</sup> At least one other court has made similar observations regarding the *AFTE Journal*’s lack of independence. See *Green*, 405 F. Supp. 2d at 109 n.7.

disinterested referees, the vast majority of published articles in the field have not undergone peer review by a “competitive, unbiased community of practitioners and academics, as would be expected in the case of a scientific field.” *Id.* (internal quotation marks omitted); *see also United States v. Starzepyzel*, 880 F. Supp. 1027, 1037–38 (S.D.N.Y. 1995).

Overall, 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. Ultimately, the Court has seen only two meaningfully peer reviewed journal articles regarding the foundational validity of the field, as the vast majority of the studies are published in a journal that uses a flawed and suspect review process. While the implications of these conclusions arise again with respect to the third *Daubert* factor regarding the demonstrated rate of error, this factor on its own does not, despite the sheer number of studies conducted and published, work strongly in favor of admission of firearms and toolmark identification testimony.

*C. Does the methodology have a known or potential rate of error?*

The parties focused most of their attention on the third *Daubert* factor—“the known or potential rate of error.” And with good reason: determining the error rate for a particular methodology appears essential to determining its ultimate reliability. On this question, the undersigned agrees with one of the essential premises of the 2016 PCAST Report:

Scientific validity and reliability require that a method has been subjected to empirical testing, under conditions appropriate to its intended use, that provides valid estimates of how often the method reaches an incorrect conclusion. For subjective feature-comparison methods, appropriately designed black-box studies are required, in which many examiners render decisions about many independent tests (typically, involving “questioned” samples and one or more “known”

samples) and the error rates are determined. Without appropriate estimates of accuracy, an examiner's statement that two samples are similar – or even indistinguishable – is scientifically meaningless: it has no probative value, and considerable potential for prejudicial impact. Nothing – not training, personal experience nor professional practices – can substitute for adequate empirical demonstration of accuracy.

PCAST Report at 46. Likewise, an expert witness's ability to explain the methodology's error rate—in other words, to describe the limitations of her conclusion—is essential to the jury's ability to appropriately weigh the probative value of such testimony. As Judge Rakoff stated in *United States v. Glynn*: “The problem is how to admit [ballistics comparison evidence] into evidence without giving the jury the impression – always a risk where forensic evidence is concerned – that it has greater reliability than its imperfect methodology permits.” 578 F. Supp. 2d at 574.

Courts considering this issue have rather uniformly weighed this third *Daubert* factor in favor of admissibility. A few courts have characterized the calculation of an error rate for firearms and toolmark pattern matching evidence as an impossible or exceedingly difficult task and acknowledged that an error rate is “presently unknown.” *Johnson*, 2019 U.S. Dist. LEXIS 39590, at \*55, 2019 WL 1130258, at \*18 (citing *Ashburn*, 88 F. Supp. 3d at 246; *Diaz*, 2007 U.S. Dist. LEXIS 13152, at \*27, 2007 WL 485967, at \*9); *Romero-Lobato*, 379 F. Supp. 3d at 1119 (quoting *Monteiro*, 407 F. Supp. 2d at 367); *Ashburn*, 88 F. Supp. 3d at 246. The vast majority of courts have nonetheless accepted the notion that existing studies support the conclusion that the discipline's error rate is quite low—between one and two percent. *Romero-Lobato*, 379 F. Supp. 3d at 1119–20; *Johnson*, 2019 U.S. Dist. LEXIS 39590, at \*56–57, 2019 WL 1130258, at \*18–19; *Johnson*, 2015 U.S. Dist. LEXIS 111921, at \*10, 2015 WL 5012949, at \*4 (citing *Otero*, 849 F. Supp. 2d at 433–34); *Ashburn*, 88 F. Supp. 3d at 246. Indeed, one court

ratified the assertion that the error rate for this discipline is “almost zero.” *Wrensford*, 2014 U.S. Dist. LEXIS 102446, at \*56–57, 2014 WL 3715036, at \*17.

In spite of the court system’s widespread acceptance of the discipline’s assertion that it enjoys low error rates, several extensive reports originating from institutions independent of the judiciary have recently taken a different view of the sufficiency of the existing studies in establishing an error rate and in validating the discipline in general. Two National Research Council reports have directly addressed the sufficiency of the published studies purporting to show a low error rate in the field of firearms and toolmark identification. In the first report, the NRC commented:

The validity of the fundamental assumptions of uniqueness and reproducibility of firearms-related toolmarks has not yet been fully demonstrated. . . . A significant amount of research would be needed to scientifically determine the degree to which firearms-related toolmarks are unique or even to quantitatively characterize the probability of uniqueness.

Nat’l Research Council, *Ballistics Imaging 3* (2008) [hereinafter *2008 NRC Report*]. Similarly, the NRC’s second report noted, “[s]ufficient studies have not been done to understand the reliability and repeatability of the methods.” *2009 NRC Report* at 154. Finally, and most recently, PCAST concluded that most of the studies

involved designs that are not appropriate for assessing the scientific validity or estimating the reliability of the method as practiced. Indeed, comparison of the studies suggests that, because of their design, many frequently cited studies seriously underestimate the false positive rate. . . . The scientific criteria for foundational validity require appropriately designed studies by *more than one group* to ensure reproducibility. Because there has been only a single appropriately designed study [the Baldwin/Ames Laboratory study], the current evidence falls short of the scientific criteria for foundational validity. There is thus a need for additional, appropriately designed black-box studies to provide estimates of reliability.

PCAST Report at 111. Together, these reports raise significant questions as to the extent to which courts should rely on certain studies and the low error rates they claim when evaluating this evidence under *Daubert*.

As a general matter, those courts that have found low error rates for this discipline appear to have done so by simply accepting the conclusions of the studies as presented and without any analysis of the methodological or other issues presented in them. *See, e.g., Otero*, 849 F. Supp. 2d at 434; *Romero-Lobato*, 379 F. Supp. 3d at 1119–20; *Johnson*, 2019 U.S. Dist LEXIS 39590, at \*56–57, 2019 WL 1130258, at \*18–19; *Johnson*, 2015 U.S. Dist LEXIS 111921, at \*10, 2015 WL 5012949, at \*4; *Ashburn*, 88 F. Supp. 3d at 246.<sup>8</sup> However, after extensive review of the testimony of the expert witnesses and of the studies about which those experts testified, the undersigned finds it difficult to conclude that the existing studies provide a sufficient basis to accept the low error rates for the discipline that these studies purport to establish. Although the Defendant and the government provided expert testimony and argument on a range of issues presented by these studies, three main problems with the design and interpretation of these studies provide the greatest cause for concern. First, most of the studies suffer from basic, threshold design flaws that undermine the value of their stated results. Second, the reliance of most of these studies on “closed” and/or “set-based” design structures substantially limit the reliability of the error rates claimed in these studies. Third, and perhaps most significantly, the

---

<sup>8</sup> To be sure, a few judges who have admitted firearms and toolmark identification testimony have addressed, at least in some fashion, various criticisms of the discipline related to the methodology’s error rate and its calculation. *See Romero-Lobato*, 379 F. Supp. 3d at 1120; *Ashburn*, 88 F. Supp. 3d at 246; *Otero*, 849 F. Supp. 2d at 434; *Taylor*, 663 F. Supp. 2d at 1177. In response to the PCAST Report’s criticism regarding the general lack of adequately designed studies for firearms and toolmark validation, the United States District Court for the District of Nevada explained that it would not “adopt such a strict requirement for which studies are proper and which are not.” *Romero-Lobato*, 379 F. Supp. 3d at 1120. The court went on to find that “*Daubert* does not mandate such a prerequisite for a technique to satisfy its error rate element.” *Id.* The United States District Court for the Eastern District of New York rejected a separate criticism levied by the 2009 *NRC Report*—that “the lack of objective standards prevents a ‘statistical foundation for estimation of error rates’”—and argued that the “information derived from [] proficiency testing is indicative of a low error rate[.]” *Ashburn*, 88 F. Supp. 3d at 246 (first quoting 2009 *NRC Report* at 154; then quoting *Otero*, 849 F. Supp. 2d at 434).

studies permit participants to label toolmark comparisons as “inconclusive” without adequately assessing the impact of such inconclusive determinations on the results of the study as a whole.

1. Most of the studies in the field of firearms and toolmark analysis suffer from basic, threshold design flaws.

Generally, studies published within the area of firearms and toolmark analysis are designed exclusively by toolmark examination professionals who have no experience or training in research methods or decision science. Though these professionals have varying levels of experience within the field of firearms and toolmark analysis, there is no indication that they have experience or training in human subjects research that would facilitate the design of studies that, for example, account for test-taking biases and achieve consistent results by providing specific and uniform procedures for test takers to follow. *See* Scurich Test., May 14, 2019 (2), 79:20–22, 80:3–10.

Concerns with test-taking biases arise from the notion that a person being tested on her ability to perform a task will, consciously or not, perform differently while being monitored, either guessing the purpose of the test and responding accordingly, Faigman Test., May 16, 2019, 84:23–85:6, or being influenced by a test designer’s cues toward one response over another, Angela Stroman, *Empirically Determined Frequency of Error in Cartridge Case Examinations Using a Declared Double-Blind Format*, 46 AFTE J. 157, 157 (2014) [hereinafter 2014 Stroman Study]; *see also* 2009 NRC Report at 122–24. A test-taker may, consciously or not, try harder or behave more conservatively to avoid being wrong and thus appear to be performing the task better than she would under other circumstances. *See* 2016 Smith Study at 693 (noting possible “fear of answering incorrectly” when taking a test lacking anonymity). Mr.

Weller, having personally participated in research studies in this field, testified that questions regarding test-taking bias need not concern the courts:

I think if you ask a human factor person that is always a concern; the concept of test taking bias; that decisions, there may be a subconscious thing that is going on. So, the test may not be completely reflective of true casework decisions. From my own perspective, I treated the case samples in the same way I would treat casework and I used the same methods and comparison techniques and my own criteria to reach those conclusions. So, I appreciate the concern. I don't know how tangible that concern is and how you rectify that potential problem.

Weller Test., May 14, 2019 (1), 30:20–31:7.<sup>9</sup> The Court simply cannot accept the conclusion that a recognized bias-related concern should not be a concern at all because a person participating in a study did not himself perceive any impact of that bias. This is, of course, precisely the problem with biases, which have their greatest impact whenever and wherever they operate completely unacknowledged. *See 2009 NRC Report* at 124. Based on the evidence adduced at the hearing, it appears that the studies relied upon by the government do not address the potential impact of such biases.

A more concrete study design concern stems from the lack of clarity in these studies as to how the test-takers were expected to perform the work, and the resulting lack of information about what practices and procedures the test-takers actually followed when participating in a study. Many of the studies failed to instruct their participants clearly on whether to follow the testing policies and protocols of their individual laboratories, or to conduct the comparisons in a particular manner in order to ensure uniformity. *See, e.g., 2014 Stroman Study* at 169 (instructing examiners to follow their “normal” procedures); Mark A. Keisler et al., *Isolated Pairs Research Study*, 50 AFTE J. 56, 58 (2018) [hereinafter 2018 Keisler Study] (instructing examiners to complete the research study like they would casework, but noting it was “unclear if

---

<sup>9</sup> Mr. Weller’s training and experience, which involves a Master of Science degree in Forensic Science as well as over ten years of training and casework experience in firearms and toolmark analysis, *see* Decl. of Todd J. Weller 1, does not include any training or experience in decision science.

participants . . . deviated from laboratory policy”); 2016 Smith Study at 698 (failing to instruct examiners but noting factors “such as a laboratory’s quality assurance program (which includes verifications and peer review), would influence error rates in casework”). This inconsistency poses a significant interpretive problem because different labs have different policies for how to conduct toolmark examinations. Scurich Test., May 15, 2019, 53:12–19; Faigman Test., May 16, 2019, 85:24–86:6. For example, some lab policies require a second examiner to verify a first examiner’s work while others do not; similarly, some labs have policies that prohibit rendering a conclusion of “exclusion” when class characteristics are all in common, while others do not have such a policy. *See, e.g.*, 2018 Keisler Study at 58. In other words, in many of the studies that the government and its experts rely on, it is unknown whether one or more of the test participants had a colleague verify his or her work, and whether reported “inconclusives” were only deemed inconclusive due to adherence with a policy demanding such a result rather than on an actual analysis of the patterns on a particular bullet or casing.<sup>10</sup> These design issues prevent the Court from evaluating whether the test-takers in these studies *were even taking the same test*—as it cannot be determined what instructions each examiner followed in completing the comparisons—and thus reduce the ability of these studies to support the foundational validity of the field.

Yet another study design issue relates to the manner in which the test administrators selected practicing examiners to participate in the studies. Scurich Test., May 14, 2019 (2), 93:9–20, 93:22–94:1. Some studies provided no information regarding how their participants were selected and recruited, *see, e.g.*, 2018 Keisler Study, but those studies that did indicated that

---

<sup>10</sup> In one frequently-cited study, the test designers simply did not make clear whether their participants were to follow their specific lab’s policies. 2018 Keisler Study at 58; Faigman Test., May 16, 2019, 85:24–86:6. The same study recognized this concern and specifically asked participants what their labs’ policies were with respect to not excluding samples with matching class characteristics. 2018 Keisler Study at 58. However, when analyzing its data, that study made no attempt to disaggregate that data by the different policies used. *Id.* at 57–58.

they had solicited volunteer participation from AFTE membership lists or from groups of employees in specific crime laboratories: one study, for example, used only examiners employed by a Federal Bureau of Investigation laboratory, Charles DeFrance and Michael D. Van Arsdale, *Validation Study of Electrochemical Rifling*, 35 AFTE J. 35, 36 (2003) [hereinafter 2003 DeFrance Study]; another engaged a third party to solicit volunteers from laboratories, 2016 Smith Study at 693; and two others recruited volunteers via email, using a list of AFTE members, Thomas G. Fadul, Jr., et al., *An Empirical Study to Improve the Scientific Foundation of Forensic Firearm and Tool Mark Identification Utilizing 10 Consecutively Manufactured Slides*, 45 AFTE J. 376, 379 (2013) [hereinafter 2013 Fadul Study]; Thomas G. Fadul, Jr., et al., *An Empirical Study to Improve the Scientific Foundation of Forensic Firearm and Tool Mark Identification Utilizing Consecutively Manufactured Glock EBIS Barrels with the Same EBIS Pattern*, Final Report on Award Number 2010-DN-BX-K269, 16 (2013) [hereinafter Miami-Dade Study]. Other studies simply report that they used volunteers from laboratories or AFTE membership lists without clarifying further as to how the participants were recruited. David P. Baldwin et al., *A Study of False-Positive and False-Negative Error Rates in Cartridge Case Comparisons*, 7 (2014), <https://www.ncjrs.gov/pdffiles1/nij/249874.pdf> [hereinafter Ames Laboratory Study]; David J. Brundage, *The Identification of Consecutively Rifled Gun Barrels*, 30 AFTE J. 438, 440, 442 (1998) [hereinafter 1998 Brundage Study]; 2014 Stroman Study at 168. Still, others do not specifically describe their pool of participants, let alone how those participants were solicited to take part in the study. See 2019 Hamby Study; 2018 Keisler Study; Dennis J. Lyons, *The Identification of Consecutively Manufactured Extractors*, 41 AFTE J. 246 (2009). In spite of this vagueness in some of these articles, these studies generally appear to use a self-selected set of volunteers. While simply soliciting volunteers is obviously the easiest way

to perform these experiments, use of volunteers for what amounts to a proficiency examination does not provide the clearest indication of the accuracy of the conclusions that would be reached by average toolmark examiners. *Scurich Test.*, May 14, 2019 (2), 93:19–20.

These design issues do not necessarily invalidate the results of these studies, and *Daubert* does not necessarily require the proponent of a theory or methodology to present only studies with the best possible design. Undoubtedly, experts with extensive training in research methods could likely find fault with the methodology of any study. But these threshold design issues—perhaps the result of their designers not securing the assistance of individuals with design science expertise—surely impact the validity of these studies’ conclusions and limit their utility to some extent.

2. Because of their reliance on “closed” and “set-based” designs, the studies in the field of firearms and toolmark analysis do not provide reliable data regarding the ability of an examiner to match unknown and known samples.

In general, the firearms and toolmark identification field has produced two types of comparison studies—those that are referred to as “open” and “independent comparison” studies (also called “pairwise comparison” studies), and those that are referred to as “closed” and “set-based” studies. *See* PCAST Report at 106–10. In the “open” and “independent comparison” studies, participants are given an unknown sample and asked to determine whether it matches another specific sample. *Id.* at 110. Such a study may involve a series of separate comparisons, but each comparison presents as a separate problem. *See id.* Most importantly, not all of the unknown samples will have a matching known sample, so the participant will not have reason to know whether the correct match is present. *See id.* Based on the testimony at the hearing and the materials submitted by the parties, it appears that only two studies have been conducted using

this approach: the 2014 Ames Laboratory study and the 2018 Keisler study. In the Ames Laboratory study, participants were given a test kit consisting of fifteen separate problem sets for comparison. Ames Laboratory Study at 10. Each set contained three cartridge casings designated as being from the same “known” firearm and one cartridge casing designated as the “unknown” or “questioned” sample; unknown to the participants, each test kit contained five same-source pairs and ten different-source pairs. *Id.* Participants were asked to approach each of the fifteen problems separately and to render a conclusion, and they were not told whether any of the questioned samples would match the known samples. *Id.* Similarly, the Keisler study provided participants with a test kit made up of twenty sets of two cartridge casings each, and unknown to the participants, each test kit contained twelve same-source pairs and eight different-source pairs. 2018 Keisler Study at 56. Participants were asked to examine each pair separately from any other pair and to render a conclusion as to each pair. *Id.*

By contrast, virtually all studies published in this field utilize a “closed” universe, where a match is always present for each unknown sample, and a “set-based” design, where comparisons are made within a set of samples. *See* PCAST Report at 106. This methodology differs from the “open” and “independent comparison” studies because the comparisons are not divided up into individual problems for the participant to consider one at a time; instead, participants are either given a group of samples and asked to compare all of those samples to each other and to find matches, or participants are given a group of known samples and a group of unknown samples and asked to make comparisons between the two groups to find matches. *See id.* at 106–08. For example, the 2019 Hamby Study, using the same design and test kits as the 1998 Brundage Study and published incorporating all data from several iterations of Brundage’s original study over the last twenty-one years, provided participants with fifteen

questioned samples and ten pairs of known samples and asked the participants to make comparisons. 2019 Hamby Study at 556; 1998 Brundage Study at 440. Similarly, the two Fadul studies gave participants a quantity of questioned samples and a number of known samples and asked them to make comparisons between the two groups. 2013 Fadul Study at 380; Miami-Dade Study at 19. These studies, and others like them, often involved the use of an answer sheet to allow the participant to indicate the known sample to which an unknown sample could be matched. *See, e.g.,* Miami-Dade Study at 19.

During the hearing, counsel and witnesses debated the question of whether one of the study types better mimics casework. The PCAST report concluded that the “closed” and “set-based” studies did not replicate casework. PCAST Report at 106. The government expert witnesses, Mr. Weller and Dr. Petraco, disagreed with this contention. Weller Test., May 13, 2019, 126:21–127:19; Petraco Test., May 13, 2019, 71:15–21, 71:24–72:5. While the Court presently lacks sufficient information to resolve this empirical question, its answer would not provide much guidance for the *Daubert* question at issue here. As Dr. Scurich stated, the question of whether a study mimics real-world casework differs from the question of whether a study accurately measures the ability of examiners to make source determinations based on pattern matching. *See* Scurich Test., May 15, 2019, 77:20–24.

Having reviewed the studies and considered both parties’ arguments on the different study designs, the undersigned finds that the independent comparison studies, or “pairwise” studies, best test the validity of the assumptions underlying the firearms and toolmark analysis field and that the closed, set-based studies have inherent limitations that preclude them from providing substantial validation. This conclusion mirrors that of PCAST, which explained:

Specifically, many of the studies employ ‘set-based’ analyses, in which examiners are asked to perform all pairwise comparisons within or between small samples

sets. . . . The study design has a serious flaw, however: the comparisons are not *independent* of one another. Rather, they entail internal dependencies that (1) constrain and thereby inform examiners' answers and (2) in some cases, allow examiners to make inferences about the study design. . . . Because of the complex dependancies among the answers, set-based studies are not appropriately-designed black-box studies from which one can obtain proper estimates of accuracy. Moreover, analysis of the empirical results from at least some set-based studies ('closed-set' designs) suggest that they may substantially underestimate the false positive rate.

PCAST Report at 106. Of course, the PCAST report is hardly beyond critique, and the government's experts stated many valid criticisms of it throughout the hearing: the Council did not include anyone from the firearms and toolmark examination community, *id.* at v-ix; it criticized studies for lack of peer review but was not itself peer reviewed, Petraco Test., May 13, 2019, 34:20–24; and the report apparently miscounted or omitted data from several studies, Weller Test., May 13, 2019, 108:10–109:8. Despite these shortcomings, the Court finds the conclusions of PCAST (as echoed by Dr. Scurich at hearing) about the very limited utility of closed-set studies to have been essentially correct.

Closed, set-based studies have two significant problems that make them difficult to rely upon as evidence of the reliability of conclusions regarding toolmark evidence. First, a set-based study involves an unknown number of total comparisons that a participant makes in the process of matching samples to each other, which means that such a study cannot calculate a true error rate based on the total comparisons made. In other words, the total number of comparisons made remains unknown at the conclusion of the study because it is not known whether a participating examiner compared a particular unknown sample to only one other sample, or to a few of the other samples, or to all of the other samples before making a conclusion regarding that sample. One of the government's expert witnesses acknowledged this issue in his testimony and agreed that

in closed, set-based studies, it is not possible to know the total number of true different source comparisons performed and that a false positive error rate thus cannot be calculated. Weller Test., May 14, 2019 (2), 22:17–23.

Second, and perhaps more importantly, the participants in a closed, set-based study can see all of the questioned samples and all of the known samples at once and can thus employ inferences gained from looking at one of the individual problems in order to solve other individual problems. In independent comparison studies, the examiner simply makes a one-to-one comparison, an exercise well-suited to gauge her ability to look at two items and, based only on the features of those two items, make a determination of match. PCAST likened closed, set-based studies, by contrast, to a Sudoku puzzle, “where initial answers can be used to help fill in subsequent answers.” PCAST Report at 106. This puzzle analogy, which Dr. Scurich also employed to explain this pitfall of closed, set-based studies, identifies a substantial problem with the closed and set-based study design. Such a design allows participants to rely on their own decisions and inferences about some of the samples to make decisions regarding the remaining samples, which the defense aptly characterized as the “interdependency problem.” Tr. June 10, 2019, 20:20. In other words, the participant can rely on other, unrelated parts of the puzzle—or even the puzzle as a whole—to solve an individual part of the puzzle, and thus a match determination for each of the individual problems evaluated would depend not simply on one-to-one comparisons but also on information and inferences gleaned from other individual problems (or from the set as a whole). Such a study design does not provide a reliable measure of the ability of firearms and toolmark

examiners to make comparisons between known and unknown samples where such inferences are not available to be drawn.

Because of these significant limitations of the closed and set-based studies, the vast majority of studies that the field relies upon to establish its foundational validity simply do not provide an adequate basis to do so. Unfortunately, the only studies with the more appropriate design for assessing reliability—the Ames Laboratory study and the Keisler study—have not, as described *supra*, undergone meaningful, independent peer review prior to publication.<sup>11</sup>

3. The large number of “inconclusive” results, and the studies’ failure to address them, undermines the reliability of the studies’ claimed error rates.

The final, and perhaps most substantial, issue related to the studies proffered to support the reliability of firearms and toolmark analysis relates to how the studies address—or fail to address—the “inconclusive” answers (hereinafter “inconclusives”) frequently given by the examiners participating in these studies, and how such answers affect the error rate. In field work, examiners analyzing bullets and cartridge casings recovered from a crime scene and comparing them to test fired samples from a recovered firearm can reach three possible conclusions: they can conclude that the samples match, and thus make an “identification”; they can conclude the samples do not match, and thus make an “elimination”; or they can characterize the comparison as “inconclusive.” Inconclusive appears to be a reasonable and acceptable conclusion in casework, possibly because the firearm may not have left sufficient marks for comparison, *see* Weller Test., May 13, 2019, 117:15–19, or because environmental factors may change or distort the soft metal of a cartridge casing or bullet. As Judge Rakoff described, “[t]he

---

<sup>11</sup> The 2014 Ames Laboratory Study was made available on the internet without having undergone any clear peer review process, while the 2018 Keisler Study was published in the *AFTE Journal*.

bullets and/or shell casings recovered from the crime scene may be damaged, fragmented, crushed or otherwise distorted in ways that create new markings or distort existing ones.” *Glynn*, 578 F. Supp. 2d at 573.

Nevertheless, the methods used in the proffered laboratory studies make a compelling case that inconclusive should not be accepted as a correct answer in these studies. First and foremost, the study designers make efforts to control the effects of the environment on the samples. Rather than being fired such that the casings or bullets could roll, hit walls or cars, or be stepped on or exposed to the weather, these studies use samples collected under test fire conditions. In the Ames Laboratory study, for example, all of the test fired casings were collected in a brass catcher, and any that fell out of the catcher and hit the floor were discarded. Ames Laboratory Study at 12.

Additionally, most of the studies involved some quality assurance mechanism to ensure that the samples to be examined by the participants had sufficient markings for comparison purposes before the test kits were supplied to the examiners. For example, one study involved several test fires to account for a so-called “break-in period” to ensure that the newly-manufactured firearms were producing consistent markings, and the study designers checked the samples to ensure that the markings were then consistent. 2003 DeFrance Study at 35.<sup>12</sup> In the two Fadul studies, study designers personally inspected every tenth test set to ensure that the samples had sufficient markings for comparison purposes. 2013 Fadul Study at 382; Miami-Dade Study at 19. Another study involved a “pre-test” that was conducted to review the test sets before they were delivered to participants. 2009 Lyons Study at 250–51. The 2018 Keisler

---

<sup>12</sup> The notion of a “break-in period,” during which time a firearm does not make consistent markings, would seem to undercut the general premise underlying the entire field of firearms and toolmark analysis—that is, that firearms reliably leave unique markings on casings and bullets fired based on marks left during the manufacturing process.

Study, at 57, and the 2016 Smith Study, at 694, also noted that the samples used for comparison had been deemed determinable.

The government and its expert witnesses view the number of inconclusives given by examiners in these published studies as irrelevant to the ultimate issue before the Court. Based on the premise that declaring a comparison inconclusive has no probative value, the government argues that such an opinion would not be given in court, and thus need not be a factor in assessing the reliability of pattern matching within the field of firearms and toolmark analysis. In other words, the government and its experts contend that only identifications—i.e., “match conclusions”—and a false positive error rate calculated based upon identifications combine to establish reliability. PCAST addressed inconclusives in this manner—by removing them entirely from analysis of the studies and their data, PCAST Report at 153—as did the United States District Court for the District of New Jersey in the only published opinion addressing this aspect of the studies, *see Otero*, 849 F. Supp. 2d at 434.<sup>13</sup>

However, in laboratory testing situations, in which samples were collected using procedures to minimize environmental alterations and in which samples were checked by test administrators to ensure they contained sufficient marks suitable for comparison purposes, a conclusion by an examiner characterizing the comparison as inconclusive should not qualify as a correct answer. Dr. Scurich opines, based on principles of mathematics and statistics in particular, that such responses should be viewed as false positive errors (i.e., included among false identifications),<sup>14</sup> but such a characterization fails to make logical sense: while under

---

<sup>13</sup> The studies themselves have treated inconclusives differently. For example, the Ames Laboratory study included the inconclusives in the denominator of the error rate calculation, such that inconclusives counted toward the total number of comparisons made, likely underestimating the overall error rate. Ames Laboratory Study at 15. The Lyons study, by contrast, treated an inconclusive response as a correct response in its calculations. Lyons Study at 254–55; Scurich Test., May 14, 2019 (2), 100:13–17, 19–21.

<sup>14</sup> The Court understands Dr. Scurich to reason as follows: (1) The only correct answers in laboratory studies are “identification” or “exclusion” because the samples are such that they can be identified, Scurich Test., May 14, 2019

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. While the Court does not accept Dr. Scurich's inclusion of inconclusives in the false positive error rate, it agrees with his essential premise that such responses should represent an error by the examiner. Under these controlled circumstances, an examiner who looks at a casing collected in a laboratory test fire and that has been examined by a test administrator to make sure it has markings suitable for comparison, and who nonetheless describes her comparison as yielding inconclusive results, *is making an error of some kind*. In these published studies, at the very least, the test taker giving an answer of inconclusive may simply be avoiding the most difficult problem on the test. Or it may be that the examiner's failure to identify or exclude the sample constitutes a mistake in her analysis. Alternatively, there may be some ambiguity, discussed at length in the Ames Laboratory study, regarding why some examiners make a determination of inconclusive, and whether some of those determinations are the result of laboratory policies against declaring exclusions when class characteristics are the same. Ames Laboratory Study at 18–19.

Based on the studies and the testimony of the government's expert witnesses, no adequate explanation has been offered regarding the reason for examiners returning inconclusives in these controlled circumstances. The government's experts insist that inconclusives should not be treated as any kind of error because inconclusive is not a conclusion at all. *See* Petraco Report at

3. Nevertheless, and again under these controlled circumstances, an inconclusive response *is a*

---

(2) 102:13–24; (2) In such a scenario involving a binary question, the basic principles of mathematics mean that the rate of true exclusions (called “specificity”) and the rate of false identifications or false positives (called “1 minus specificity”) must sum to 100% (i.e., of all the bullets that are known *not to match*, the percent declared “excluded,” and the percent declared “match” must sum to 100%), Scurich Test., May 14, 2019 (2), 86:6–19, 87:11–16, 87:21–88:2; and (3) Therefore, the false positive rate must equal 100% minus the percentage of correct exclusions, Scurich Test., May 14, 2019 (2), 87:12–16. For an example, out of all of the possible correct exclusions, if examiners correctly concluded “exclusion” 80% of the time, then it must be true that they reached incorrect conclusions the remaining 20% of the time. *See* Scurich Test., May 14, 2019 (2), 99:10–14.

conclusion, even if it is only a conclusion against making any other conclusion. In a recent article, Itiel Dror asserts that inconclusive determinations may be the result of “over-reliance” by forensic examiners on the option of “decid[ing] not to decide.” Itiel E. Dror & Glenn Langenburg, *“Cannot Decide”: The Fine Line Between Appropriate Inconclusive Determinations Versus Unjustifiably Deciding Not to Decide*, 64 J. Forensic Sci. 10, 11 (2019). Where there is sufficient information for concluding “identification” or “exclusion,” “[a]n inconclusive determination is an erroneous decision because the evidence does not support that decision.” *Id.* at 13. In the end, all that is known is that some examiners in these studies, taking these tests involving samples collected under carefully controlled circumstances, responded that the comparison was “inconclusive.”<sup>15</sup>

Viewing these inconclusives as an error of some type greatly affects the overall error rates produced by the studies. Focusing on the only two “open” studies, the Ames Laboratory study calculated a false positive error rate of 1.01%, while the Keisler study reported a false positive error rate of 0%. If the inconclusives are considered as errors, however, the Ames Laboratory study’s error rate among different source comparisons soars to 34.76% while the Keisler study’s error rate rises to 20.14%. Again, Dr. Scurich’s approach of treating inconclusives as false positives does not appropriately address the issue presented by inconclusives, but the large number of the inconclusives reported in the studies greatly reduces their persuasive force in establishing the ability of a firearms and toolmark examiner to make accurate source determinations. Indeed, even Dr. Petraco acknowledged that the number of inconclusives increased uncertainty about calculations of the error rate, Petraco Test., May 13,

---

<sup>15</sup> Additionally, it is important to note that inconclusives appear more frequently in open studies compared to closed and set-based studies, *see* PCAST Report at 109, and more frequently when the compared samples are true exclusions. For example, the Ames Laboratory Study, at 16–17, reported 735 inconclusives for 2,178 true different-source comparisons compared to only eleven for 1,090 possible true same-source comparisons. The evidence and testimony presented in the hearing did not adequately account for these disparities.

2019, 26:10–12, while Mr. Weller testified that the questions surrounding inconclusives and the error rate calculation were, in his words, “not well studied,” Weller Test., May 14, 2019 (1), 53:25–54:1.

#### 4. Conclusion

Based on the basic design of the studies, the prevalence of closed-set studies, and the uncertain relationship in the open studies between inconclusives and the ultimate error rates, the undersigned was unable to conclude that the field has established a known or potential error rate with regard to the ability of a firearms and toolmark examiner to make a source determination. Dr. Petraco testified, and the government repeated several times in argument, that no studies refute the proposition that “firearms examiners can identify bullets or fired cartridge casings to particular guns with a high degree of accuracy.” Petraco Test., May 13, 2019, 12:24–13:4. This formulation of the issue turns both the scientific method and the *Daubert* burden of proof on their heads: instead, the question before the Court turns on whether the government can establish the foundational validity of the discipline, not whether the opposing party can prove a negative.

With regard to the proffered discipline, most of the studies on which the government relies involved closed-set designs that cannot provide an accurate accounting of the error rate. While the two studies that employ an open, independent comparison design could yield an accurate error rate measurement, neither was subjected to meaningful peer review, and both were plagued by a large number of “inconclusive” responses. Under such circumstances, the Court cannot conclude that the government has established that this forensic discipline has established a “known or potential rate of error.” *See Motorola*, 147 A.3d at 754 (citing *Daubert*, 509 U.S. at 593–94). While other studies being conducted now or in the future may change this conclusion,

the Court finds that this factor currently weighs against the admission of source attribution statements made by a firearms and toolmark examiner.

*D. Is there a standard controlling the technique's operation?*

The fourth *Daubert* factor requires an inquiry into “the existence and maintenance of standards controlling the technique's operation.” *Motorola*, 147 A.3d at 754 (quoting *Daubert*, 509 U.S. at 593–94). As described *supra*, the operative standard for firearms and toolmark identification is known as the “AFTE theory of identification,” which states that the examiner can make a conclusion of common origin when microscopic surface contours of the toolmarks are in “sufficient agreement.” PCAST Report at 59–60 (citing Ass’n of Firearm and Tool Mark Examiners, *Theory of Identification as it Relates to Tool Marks: Revised* 43 AFTE J. 287 (2011)). Stated in full, the AFTE Theory of Identification reads as follows:

1. The theory of identification as it pertains to the comparison of toolmarks enables opinions of common origin to be made when the unique surface of two toolmarks are in “sufficient agreement.”
2. This “sufficient agreement” is related to the significant duplication of random toolmarks as evidenced by the correspondence of a pattern or combination of patterns of surface contours. Significance is determined by the comparative examination of two or more sets of surface contour patterns comprised of individual peaks, ridges and furrows. Specifically, the relative height or depth width, curvature and spatial relationship of the individual peaks, ridges and furrows within one set of surface contours are defined and compared to the corresponding features in the second set of surface contours. Agreement is significant when the agreement in individual characteristics exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools and is consistent with agreement demonstrated by toolmarks known to have been produced by the same tool. The statement that “sufficient agreement” exists between two toolmarks means the agreement of individual characteristics is of a quantity and quality that the likelihood another tool could have made the mark is so remote as to be considered a practical impossibility.
3. Currently the interpretation of individualization/identification is subjective in nature, founded on scientific principles and based on the examiner’s training and experience.

*Id.*

As other courts have noted, and as the Defendant argues here, one of the primary challenges to firearms and toolmark identification stems from the methodology's lack of objective criteria for examiners to use in determining a "match." *See, e.g., Romero-Lobato*, 379 F. Supp. 3d at 1120. Courts that have admitted firearms and toolmark identification testimony in the face of a *Daubert* challenge have found the standard articulated in the AFTE theory of identification sufficient. *See Johnson*, 2019 U.S. Dist. LEXIS 39590, at \*51, 2019 WL 1130258, at \*17; *Johnson*, 2015 U.S. Dist. LEXIS 111921, at \*10–11, 2015 WL 5012949, at \*4; *Ashburn*, 88 F. Supp. 3d at 247; *Wrensford*, 2014 U.S. Dist. LEXIS 102446, at \*54–55, 2014 WL 3715036, at \*16. However, the AFTE theory of identification has been sharply criticized by a number of other courts as "inherently vague," *Glynn*, 578 F. Supp. 2d at 572; "inherently subjective," *Romero-Lobato*, 379 F. Supp. 3d at 1121; and "either tautological or wholly subjective," *Green*, 405 F. Supp. 2d at 114. As one United States District Court Judge noted, "the AFTE Theory appears to be more of a description of the process of firearm identification rather than a strictly followed charter for the field." *Monteiro*, 407 F. Supp. 2d at 371.

Both the NRC and PCAST lodged similar criticisms. The NRC focused its critique on this lack of an objective comparison standard:

AFTE has adopted a theory of identification, but it does not provide a specific protocol. . . . The meaning of "exceeds the best agreement" and "consistent with" are not specified, and the examiner is expected to draw on his or her own experience. This AFTE document, which is the best guidance available for the field of toolmark identification, does not even consider, let alone address, questions regarding variability, reliability, repeatability, or the number of correlations needed to achieve a given degree of confidence.

*2009 NRC Report* at 155. Calling this a "fundamental problem with toolmark and firearm analysis," *id.*, the NRC further stated, "even with more training and experience using newer techniques, the decision of the toolmark examiner remains a subjective decision based on

unarticulated standards and no statistical foundation for estimation of error rates.” *Id.* at 153–54. And, more recently, PCAST criticized the AFTE standard as “circular.” PCAST Report at 60.

In this case, the evidence supports—and the undersigned agrees with—all of these assessments of the AFTE theory of identification. By its own terms, it is a fundamentally subjective standard that can only be characterized as entirely tautological: an opinion of common origin can be rendered when the surfaces of the two examined items are in “sufficient agreement,” which exists not when some objective measure is satisfied, but when the examiner determines, based on her training and experience, that it would be a “practical impossibility” for the two items not to share a common origin. In other words, this protocol permits the ultimate finding of “sufficient agreement” whenever an individual examiner concludes that she would be hard pressed (for reasons not specified in the governing standard) to find such similar markings on casings or bullets fired by different firearms. Although AFTE has attempted to use terms like “sufficient agreement” to resemble terminology that one would find in an objective or scientific standard, in the end it simply leaves the determination of common origin to the standardless, undefined judgment of an individual examiner. Therefore, under this so-called standard, the process for determining what constitutes a “match” lacks defined criteria; it is merely unconstrained subjectivity masquerading as objectivity.

Courts that have admitted this type of expert opinion testimony have responded to such criticisms about the standard’s subjective nature by correctly noting that “[t]he mere fact that an expert’s opinion is derived from subjective methodology does not render it unreliable.” *Romero-Lobato*, 379 F. Supp. 3d at 1120 (citing *Ashburn*, 88 F. Supp. 3d at 246–47; *Cohen v. Trump*, Case No.: 3:13-cv-2519-GPC-WVG, 2016 U.S. Dist. LEXIS 117059, at \*35, 2016 WL 4543481,

at \*11 (S.D. Cal. Aug. 29, 2016)). Even the *Romero-Lobato* court, which found the lack of objective criteria to weigh against admissibility, explained:

[Rule 702] does not impose a requirement that the expert must reach a conclusion via an objective set of criteria or that he be able to quantify his opinion with a statistical probability. Such requirements would, in most circumstances, exclude psychologists, physicians, and lawyers from testifying as expert witnesses. Of course, a litigant would be hard pressed to make a good faith argument that the methods used by mainstream medical and legal experts are unreliable under *Daubert*.

379 F. Supp. 3d at 1120.

Of course, expert witnesses in many fields testify to subjective opinions. For example, an assessor testifying about home values would provide a subjective opinion about the value of a particular piece of property, but that assessor would be able to describe the basis of her opinion in objective terms, premised on a comparison with other properties that are similar in certain defined ways (such as the number of bedrooms, total square footage, or specific location), or on a general change in home values in a particular neighborhood since the last time the house was sold. Such an opinion would ultimately be subjective in nature, but it would be grounded in objective criteria, the applicability of which can be analyzed, debated, and critiqued, and not simply on the assessor's judgment, based on her experience, as to what the property is worth. Similarly, an expert in a medical malpractice case testifying about whether a doctor satisfied a particular standard of care would base her subjective opinion on objective criteria in the form of promulgated and practiced nationwide standards of care within that medical specialty and not in her personal opinion, based on her own training and experience, as to what that standard should be.

The AFTE theory of identification is more subjective than such other examples of subjective opinions. “[B]allistics comparison lacks defining standards to a degree that exceeds most other kinds of forensic expertise.” *Glynn*, 578 F. Supp. 2d at 574. Unlike the standards

underlying opinions in other fields, the AFTE theory provides no objective yardstick to support or explicate the expert's opinion; instead, the expert is left to rely on her own thoughts and conclusions based only on the vagaries of her own training and experience. An opinion that “the agreement in individual characteristics exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools” and “the agreement of individual characteristics is of a quantity and quality that the likelihood another tool could have made the mark is so remote as to be considered a practical impossibility” relies entirely on subjective judgment, without any underlying objective criteria that the examiner must reference or apply. For all of these reasons, this fourth *Daubert* factor strongly militates against the admission of expert witness testimony in the field of firearms and toolmark analysis.

*E. To what degree is the technique accepted within the scientific community?*

The final enumerated *Daubert* factor—the “degree of acceptance within [a relevant] scientific community”—incorporates, at least to some extent, the *Frye/Dyas* principles that the general acceptance of theories speaks to their validity. *See Daubert*, 509 U.S. at 594; *see also Motorola*, 147 A.3d at 754. As stated in *Daubert*, “[w]idespread acceptance can be an important factor in ruling particular evidence admissible, and a known technique which has been able to attract only minimal support within the community may properly be viewed with skepticism.” 509 U.S. at 594 (citation and internal quotation marks omitted). Every published opinion evaluating the admissibility of firearms and toolmark evidence has found that the AFTE method enjoys general acceptance in the relevant community and that such acceptance weighs in favor of admissibility. *See, e.g., Romero-Lobato*, 379 F. Supp. 3d at 1122; *Johnson*, 2019 U.S. Dist. LEXIS 39590, at \*58, 2019 WL 1130258, at \* 19; *Johnson*, 2015 U.S. Dist. LEXIS 111921, at \*11, 2015 WL 5012949, at \*4; *Ashburn*, 88 F. Supp. 3d at 247; *Wrensford*, 2014 U.S. Dist.

LEXIS 102446, at \*45–46, 2014 WL 3715036, at \*14; *Taylor*, 663 F. Supp. 2d at 1178.

However, these precedents have generally limited the scope of the so-called “relevant community” to the specific community of firearms and toolmark examiners, or to those generally operating within the field of criminal forensics. *See Romero-Lobato*, 379 F. Supp. 3d at 1122; *Johnson*, 2019 U.S. Dist. LEXIS 39590, at \*58, 2019 WL 1130258, at \*19; *Johnson*, 2015 U.S. Dist. LEXIS 111921, at \*11, 2015 WL 5012949, at \*4; *Ashburn*, 88 F. Supp. 3d at 247; *Otero*, 849 F. Supp. 2d at 435.

In the undersigned’s view, if *Daubert*, *Motorola*, and Rule 702 are to have any meaning at all, courts must not confine the relevant scientific community to the specific group of practitioners dedicated to the validity of the theory—in other words, to those whose professional standing and financial livelihoods depend on the challenged discipline. As Judge Jon M. Alander of the Superior Court of Connecticut aptly stated, “[i]t is self evident that practitioners accept the validity of the method as they are the ones using it. Were the relevant scientific community limited to practitioners, every scientific methodology would be deemed to have gained general acceptance.” *Terrell*, 2019 Conn. Super. LEXIS 827, at \*14, 2019 WL 2093108, at \*4. Indeed, in other forensic science fields, techniques and methods that had gained “general acceptance” among practitioners have been deemed unreliable and have been excluded as a result of *Daubert* challenges. *See, e.g., United States v. Saelee*, 162 F. Supp. 2d 1097, 1101–05 (D. Alaska 2001) (forensic handwriting analysis).

Here, the government failed to show general acceptance outside of the field of firearms and toolmark practitioners of the theory that an examiner can microscopically analyze individual toolmarks on a cartridge casing or bullet and reach a reliable conclusion that a particular firearm fired that particular cartridge casing or bullet. The conclusions of the NRC and PCAST reports

indicate that the wider academic and scientific community does not necessarily generally accept this theory. With the majority of studies published by and for the review of professional firearms and toolmark examiners, there is currently insufficient evidence that this methodology is generally accepted as proven, established, or validated—a factor that weighs against admissibility.

*F. A balancing of these factors requires that the expert be constrained to testify only that the recovered firearm cannot be excluded as the source of the recovered casing.*

In weighing and applying these factors pursuant to *Daubert*, *Motorola*, and Rule 702, the Court found that—particularly in light of the inability of the published studies to establish an error rate, the absence of an objective standard for identification, and the lack of general acceptance of the foundational validity of the field outside of the community of practitioners within the field—reliable principles and methods do not adequately support the theory that a firearms examiner can identify a particular firearm as having fired a particular bullet or cartridge casing. Accordingly, the Court will not permit Mr. Coleman, the firearms examiner who conducted the comparison in the above-captioned case, to testify in the form of such a source attribution statement. Again, in light of the state of the evidence presented here, a conclusion that a particular firearm was the source of a particular bullet or cartridge casing does not yet find support in sufficiently reliable principles and methods.

Such a conclusion, however, does not require the exclusion of all specialized opinion testimony in the area of firearms and toolmark examination, nor does it equate to a finding that the entire discipline lacks foundational reliability. As such, the Court denied Defendant's request to exclude Mr. Coleman's testimony in its entirety. The defense has not challenged the general theory that tools used to create firearms leave accidental or incidental toolmarks on the

firearms, and that those toolmarks leave impressions that can be discerned on the contours of the bullets and cartridge casings discharged through the firearm; based on the evidence before it, the Court found that reliable principles support this theory, at least at that stated level of generality. Nor did the defense challenge the reliability of the basic method used by Mr. Coleman and other firearms examiners, i.e., the use of a comparison microscope to observe these marks on bullets and cartridge casings. In addition, reliable principles permit a conclusion that a firearm cannot be excluded as the source of a recovered casing or bullet; indeed, this limited conclusion is supported by the reliable principle that firearms leave toolmark impressions on discharged cartridge casings and the reliable method of viewing those impressions under a comparison microscope. As the defense acknowledges, such a conclusion does not imply a particular statistical weight, and furthermore, it does not stray into territory unsupported by reliable principles and methods, such as a conclusion that a firearm “matches” or was the source of a particular casing.

Accordingly, the Court ruled that the government’s proffered expert, Mr. Coleman, may testify and give general specialized opinion testimony in this case. Mr. Coleman may describe the work he performed and the comparisons he made; he may describe the basis of his conclusion regarding the physical consistency of the toolmarks that he observed; and he may make, as the Defendant concedes, a comparison of the samples based on class characteristics. In sum, Mr. Coleman may conclude that based on his examination and the consistency of the class characteristics and microscopic toolmarks, the recovered firearm cannot be excluded as the source of the cartridge casing found on the scene of the alleged shooting—in other words, that the firearm *may* have fired the recovered casing. Mr. Coleman may not state an ultimate conclusion in stronger terms. Similarly, Mr. Coleman will be precluded at any point in his

testimony from stating that individual marks are unique to a particular firearm or that observed individual characteristics can be used to “match” a firearm to a piece of ballistics evidence.

In fashioning this ruling, the Court found that the government’s alternative proposals for expressing Mr. Coleman’s opinion did not adequately address the concerns raised by the *Daubert* factors. The government’s proffer that Mr. Coleman could testify that, based on his training and experience, he believes that the recovered cartridge casing was fired from the recovered gun, represents no improvement over a simply-stated opinion that a recovered casing was fired from a particular gun, even if Mr. Coleman also expressed his opinion with the limitations on certainty statements imposed by the Court of Appeals. In this alternative, the expert would be characterizing his opinion as his own personal opinion—as any expert must—but would still be making a source attribution statement not sufficiently supported by reliable principles and methods.

Similarly, the Court strongly disagrees with the government that cross-examination could cure any reliability issues created by a source attribution statement. Of course, the *Daubert* decision recognized, and other courts have noted, that “[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.” 509 U.S. at 596; *see also Motorola Inc.*, 147 A.3d at 754.<sup>16</sup> While cross-examination may often play such a role, this discipline and the disputes surrounding it seem far too complex for a series of questions on cross-examination to allow a full understanding of the limitations of the field. Indeed, a full

---

<sup>16</sup> Some cases have premised findings of the ability of cross examination to illuminate questions regarding the foundational validity of this discipline on the supposed simplicity of the issues involved. “These weaknesses [in the methodology of toolmark identification] are also not particularly complicated or difficult to grasp, and thus are likely to be understood by jurors if addressed on cross-examination.” *Johnson*, 2019 U.S. Dist LEXIS 39590, at \*58, 2019 WL 1130258, at \*19; *see also Johnson*, 2015 U.S. Dist. LEXIS 111921, at \*8, 2015 WL 5012949, at \*3.

exploration of the issues surrounding the reliability of this evidence in the present case required several days of testimony from multiple expert witnesses, close evaluation of numerous applied-science studies, exploration into the studies' design and methodology and the problems arising therefrom, and advocacy by counsel on each side specially tasked with litigating forensic science issues. It would be fanciful to conclude that the normal adversarial process would enable a lay jury to adequately understand these issues, and it is similarly unrealistic to conclude that the average attorney in the average trial would be able to raise these issues in front of the jury in this fashion, particularly when this issue would be one among many issues to be presented to the jury in a trial. Ultimately, Judge Rakoff's characterization in *Glynn* captures the essence of this issue:

[O]nce expert testimony is admitted into evidence, juries are required to evaluate the expert's testimony and decide what weight to accord it, but are necessarily handicapped in doing so by their own lack of expertise. There is therefore is [sic] a special need in such circumstances for the Court, if it admits such testimony at all, to limit the degree of confidence which the expert is reasonably permitted to espouse.

578 F. Supp. 2d at 571.

For all of these reasons, the government's expert may testify that based on his examination, the recovered firearm cannot be excluded as the source of the cartridge casing found on the scene of the alleged shooting. This formulation of the expert's opinion is limited to the principles and methodologies which the evidence supports as sufficiently reliable. Any statements by the expert involving more certainty regarding the relationship between a casing and a firearm would stray into territory not presently supported by reliable principles and methodology.

#### IV. COHERENCE WITH RECENT DISTRICT OF COLUMBIA COURT OF APPEALS PRECEDENTS

The Court of Appeals issued its opinion in *Williams II* after the *Daubert* hearing was held in this case. Upon request of the Court, both parties filed additional pleadings to address what, if any, effect *Williams II* should have on the Court's present determination.

After his conviction for first-degree felony murder while armed and other related offenses, Marlon Williams appealed his convictions—arguing, *inter alia*, that the trial court should not have permitted the government's firearms examiner to testify, based on pattern-matching, that the gun recovered from Mr. Williams's apartment was the murder weapon. *Williams II*, 210 A.3d at 737 (citing *Williams v. United States (Williams I)*, 130 A.3d 343, 345, 347 (D.C. 2016)). At trial, the examiner testified that he microscopically examined the markings on three bullets recovered from the decedent's vehicle and that they matched the markings on the bullets test fired from the gun recovered from Mr. Williams's apartment. *Williams II*, 210 A.3d at 738. The expert further opined, "these three bullets were fired from [the recovered] firearm." *Id.* On re-direct, the examiner also testified that he had no "doubt in [his] mind" that the recovered bullets were fired from the recovered gun. *Id.* The Court of Appeals initially affirmed Mr. Williams's convictions, holding there had yet to be any precedent in the District of Columbia "limit[ing] a toolmark and firearms examiner's testimony about the certainty of his pattern-matching conclusions." *Williams I*, 130 A.3d at 347–48. On re-hearing, and relying on its intervening decision in *Gardner*, the Court of Appeals subsequently held it was error to allow the 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." *Williams II*, 210 A.3d at 742–43. In *Gardner*, the Court of Appeals had held that "a firearms 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 of Appeals did note, however, its decision allowed examiners to “offer an opinion that a bullet or shell casing was fired by a particular firearm,” just not with “absolute or 100% certainty.” *Id.* at 1184 n.19.

*Williams II* appears to extend, or at least clarify, the Court of Appeals’ holding in *Gardner*, even if not resolving the apparent contradiction between the language that appears in the text and in footnote 19 of the earlier case. *See Williams II*, 210 A.3d at 740–43. Not only does *Williams II* prohibit source attribution statements made with certainty, but it also prohibits any statement that conveys a “match” without sufficient qualification. *See id.* at 742–43. In *Gardner*, the “unqualified opinion” admitted in error was simply that the bullet recovered from the decedent’s body and cartridge casing recovered from the crime scene were fired from the recovered firearm. 140 A.3d at 1182. The testimony was, “[i]n essence,” that “[the recovered gun] was the murder weapon.” *Id.* On re-direct, the examiner reiterated his opinion by stating the recovered bullet “was fired from the pistol.” *Id.* Similarly, in *Williams II*, the examiner concluded, “these three bullets were fired from this firearm.” 210 A.3d at 738.<sup>17</sup> The Court of Appeals disparaged the government’s argument—repeated as one of the bases of the post-hearing briefs filed in this case—that *Gardner*’s limitation on firearms and toolmark testimony only applies to certainty statements. *See Williams II*, 210 A.3d at 740. In sum, *Williams II* barred “unqualified” statements of “match” and source attribution. *Id.* at 742–43. The Court of Appeals failed, and thought it unnecessary, to address what type of qualification could make such a statement admissible. *Id.* at 741–42 (“We ultimately conclude that we need not resolve

---

<sup>17</sup> On re-direct, the examiner said more about the uniqueness of the markings of the recovered firearm, *Williams II*, 210 A.3d at 738, but the Court of Appeals’ ruling did not turn on the examiner’s additional statements, *cf. id.* (“[W]e conclude that it was error to admit the examiner’s opinion testimony, based on pattern matching, that the gun recovered from Mr. Williams’s apartment was the murder weapon.”).

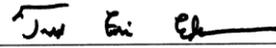
the ambiguity of *Gardner*'s footnote 19 in this case where the firearms and toolmark examiner not only testified, like the examiner in *Gardner*, that a specific bullet could be matched to a specific gun, but also that he did not have "any doubt" about his conclusion." Judge Catharine Easterly indicated in her concurrence, however, that an examiner might be able to testify that a specific bullet was fired by a specific gun if he could "reliably qualify" his opinion with a "verifiable error rate." *Id.* at 746 (Easterly, J., concurring).

The extent to which these cases should affect the Court's decision seems a bit unclear. *Williams II*, like *Gardner* before it, reviewed trials that occurred in the pre-*Motorola* era, but nonetheless invoked the language of reliability in a manner more consistent with *Daubert* and Rule 702 than *Frye* and *Dyas*. See *Williams II*, 210 A.3d at 742. Although the Court's present decision has been made pursuant to *Daubert* and Rule 702, it restricts the firearms examiner's testimony such that he may not make a source attribution statement connecting the firearm and cartridge casing. This ruling fully comports with, and may even be compelled by, the strictures imposed by *Williams II* and other relevant precedent.<sup>18</sup>

For these reasons, as well as any others stated on the record in open court on August 8, 2019, Defendant's Motion has been GRANTED IN PART and DENIED IN PART.

---

<sup>18</sup> Although not addressed by this Memorandum Opinion, Mr. Tibbs also challenges whether Mr. Coleman reliably applied the AFTE method in this case. Based on the Court's present understanding of this aspect of Defendant's argument, this challenge would only be appropriate if Mr. Coleman were permitted to testify to a "match" (*i.e.*, that the recovered cartridge casing was fired from the recovered firearm). That, of course, is not the case; Mr. Coleman is restricted to testifying to his work, his observations, and the ultimate conclusion that the recovered firearm cannot be excluded as the source of the cartridge casing. It is not evident to the Court that the Defendant's argument applies to Mr. Coleman's application of the methodology given the restriction on any ultimate conclusion he would render. Accordingly, Defendant's Motion as it relates to Mr. Tibbs's as-applied challenge is denied as moot.



---

Todd E. Edelman  
Associate Judge  
(Signed in Chambers)

Date: September 5, 2019

***Copies to:***

Jessica Willis, Esq.  
*Attorney for Defendant*  
JWillis@pdsdc.org

Prescott Loveland, Esq.  
*Attorney for Defendant*  
PLoveland@pdsdc.org

Michael Ambrosino, Esq.  
*Assistant United States Attorney*  
Michael.Ambrosino@usdoj.gov

Charles Willoughby, Jr., Esq.  
*Assistant United States Attorney*  
Charles.Willoughby2@usdoj.gov

Lindsey Merikas, Esq.  
*Assistant United States Attorney*  
Lindsey.Merikas@usdoj.gov