E-FILED; Wicomico Circuit Court
Docket: 6/14/2023 9:39 AM; Submission: 6/14/2023 9:39 AM
Envelope: 13047104

STATE OF MARYLAND	*	**IN THE CIRCUIT COURT**
vs.	*	**FOR WICOMICO COUNTY**
DELONTE BRYANT	*	**STATE OF MARYLAND**
Defendant		
	*	**CASE NO.: 22-K-15-000572**

* * * * * * * * * * * * *

OPINION AND ORDER OF THE COURT

On June 6, 2022, Defendant Delonte Bryant (hereinafter "the Defendant") filed a motion *in limine* to exclude the testimony of firearm toolmark examiner Susan Kim (hereinafter "Kim"),[1] challenging both the reliability of the employed methodology and her application of that methodology. *Motion In Limine to Exclude Firearms Identification Evidence as Unreliable Under Rochkind v. Stevenson, 471 Md. 1 (2020)* (hereinafter "Motion In Limine"), pp. 6-11, 15-22. On October 25, 2022, the State filed a response to the Motion In Limine, arguing that the underlying methodology was reliable and properly applied by Kim.

On April 24 and 25, 2023, the undersigned conducted a two-day evidentiary hearing during which the State called as witnesses Forensic Firearm and Toolmark Examiner Susan Kim (hereinafter "Kim") and Dr. Eric Warren (hereinafter "Warren") and the Defendant called as a witness Dean Faigman (hereinafter "Faigman"). The Court heard extensive testimony from these

[1] Although the motion did not refer to the examiner by name, instead generally referring to the examiner as "her," subsequent filings and the oral arguments make clear that Susan Kim is the "she" so referenced.

witnesses on firearm toolmark identification and study design methods. During the hearing, the State agreed to limit Kim's testimony in a manner identical to those agreed upon by the Government in *United States v. Harris,* 502 F.Supp.3d 28 (D. D.C. Nov. 2020). On May 1, 2023, the Defendant filed a post-hearing memorandum and on May 2, 2023, the State filed a response thereto. For the reasons set forth below, the Defendant's Motion *In Limine* is **DENIED** and Kim will be permitted to testify without restriction so long as such testimony is in accord with the Department of Justice Uniform Language for Testimony and Reports for the Forensic Firearms/Toolmarks Discipline—Pattern Matching Examination (hereinafter "DOJ ULTR").

BACKGROUND

The Defendant is charged by criminal indictment with first degree murder, second degree murder, first degree assault, second degree assault, use of a firearm in the commission of a felony or violent crime, possession of a firearm with a felony conviction, reckless endangerment, and wearing, carrying, and transporting a firearm on their person. On August 3, 2015, officers of the Salisbury City Police Department were dispatched to the area of Anne Street and Baker Street in Salisbury, Wicomico County, Maryland to investigate reports of a shooting. Upon arriving, investigating officers discovered the body of Dommier Deshields—who was pronounced deceased on arrival at the Peninsula Regional Medical Center—and later that day recovered shell casings from the scene. Over the course of the investigation, officers developed the Defendant as a suspect. They subsequently linked the Defendant to another shooting that had occurred on July 4, 2015, roughly one month prior, at which investigating officers had recovered shell casings as well as bullets. No firearm was recovered from either crime scene.

The recovered shell casings and bullets were sent for analysis and microscopically compared by Kim, who concluded that both sets of shell casings and the bullets were fired from the same firearm. The State intends to introduce expert testimony from Kim at trial, who will opine that she performed a microscopic comparison of the cartridge cases and bullets and conclude that they were fired from the same firearm. This proposed testimony is the subject of this Opinion and Order.

ANALYSIS

A. Legal Standards

Md. Rules, Rule 5-702 (hereinafter "Rule 5-702") provides that "[e]xpert testimony may be admitted, in the form of an opinion or otherwise, if the court determines that the testimony will assist the trier of fact to understand the evidence or to determine a fact in issue." Courts in making this determination "shall determine[:] (1) whether the witness is qualified as an expert by knowledge, skill, experience training or education[;] (2) the appropriateness of the expert testimony based on a particular subject[;] and (3) whether a sufficient factual basis exists to support the expert testimony." *Ibid.* The Court of Appeals adopted Rule 5-702 in 1994 after the United States Supreme Court issued its seminal opinion in *Daubert v. Merrell Dow Pharmaceuticals, Inc.,* 509 U.S. 579 (1993) and has described Rule 5-702 as the "counterpart to F[ed.] R[.] E[vid.] 702." *Rochkind v. Stevenson,* 471 Md. 1, 21 (2020). Although Rule 5-702 was adopted shortly after the issuance of *Daubert,* in which the United States Supreme Court held that Fed. R. Evid. 702 supplanted the existing expert testimony framework under *Frye v. U.S.,* 54 App.D.C. 46, 293 F. 1013 (D.C. 1923), Maryland courts continued to adhere to the *Frye*

framework as established in Maryland by *Reed v. State,* 283 Md. 374 (1978) for almost two decades thereafter.[2]

In *Rochkind,* the Court of Appeals joined a supermajority of jurisdictions when it abrogated *Reed* and "implement[ed] a single standard by which courts evaluate all expert testimony [under Rule 5-702]: *Daubert*." 471 Md. at 26. In so doing, the Court of Appeals established an expanded *Daubert* factor test and provided "thoughts for judges who shoulder the implementation of this new-to-Maryland evidentiary standard." *Id.* at 38.

The factors to be considered by a circuit court in evaluating expert testimony "include, but are not limited to:

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

(2) whether a theory or technique has been subjected to peer review and publication;

(3) whether a particular scientific technique has a known or potential rate of error;

(4) the existence and maintenance of standards and controls; []

(5) whether a theory or technique is generally accepted[;]

…

(6) whether experts are proposing to testify about matters growing naturally and directly out of research they have conducted independent of the litigation, or whether they have developed their opinions expressly for purposes of testifying;

(7) whether the expert has unjustifiably extrapolated from an accepted premise to an unfounded conclusion;

(8) whether the expert has adequately accounted for obvious alternative explanations;

[2] *See Rochkind*, 471 Md. at 12-26 for a history of the "jurisprudential drift" surrounding construction of Rule 5-702 under *Frye* and *Reed*.

(9) whether the expert is being as careful as he [or she] would be in his [or her] regular professional work outside his [or her] paid litigation consulting; and

(10) whether the field of expertise claimed by the expert is known to reach reliable results for the type of opinion the expert would give."[3]

Rochkind, 471 Md. at 35-36. The *Rochkind* court provided guidance, derived from *Daubert* and its progeny, for trial courts applying these factors.[4] First, "[t]he focus … must be solely on principles and methodology, not on the conclusions that they generate[,]" although "conclusion and methodology are not entirely distinct from one another." *Ibid* (citing *Joiner,* 522 U.S. at 146). Next, a court is not required "to admit opinion evidence that is connected to existing data only by the *ipse dixit* of the expert" and "may conclude that there is simply too great an analytical gap between the data and the opinion proffered." *Ibid.* Finally, "all of the *Daubert* factors are relevant to determining the reliability of expert testimony, yet no single factor is dispositive in the analysis[;] a trial court may apply some, all, or none of the factors depending on the particular expert testimony at issue." *Id.* at 37 (citing *Savage v. State,* 455 Md. 138, 184 (2017)).

As to the implementational thoughts provided to circuit courts, the *Rochkind* court stressed two salient concepts. First, the court emphasized the role of a trial court as gatekeeper and not ultimate arbiter of fact or truth—"[v]igorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means

[3] The Court of Appeals sourced the first five (5) rules from the *Daubert* decision itself and the latter five (5) rules from subsequent case law developments as expressed in the Fed. R. Evid. 702 Advisory Committee Note. *Rochkind,* 471 Md. at 36.

[4] The referenced progeny are the oft-cited cases of *General Elec. Co. v. Joiner,* 522 U.S. 136 (1997) and *Kumho Tire Co., Ltd. v. Carmichael,* 526 U.S. 137 (1999), collectively referred to alongside *Daubert* as the "*Daubert* trilogy."

of attacking shaky but admissible evidence." *Id.* at 38 (citing *Daubert,* 509 U.S. at 596); *see also State v. Matthews,* 479 Md. 278, 322-23 (2022) ("Our conclusion respects the trial court's role as a gatekeeper, and does not transform the trial court into an armed guard. As long as an expert's scientific testimony rests upon good grounds, based on what is known, it should be tested by the adversary process.") (internal quotations and citations omitted). Second, the adoption of *Daubert* "may mean, in a very real sense, that 'everything old is new again' with respect to some scientific and technical evidentiary matters long considered settled," meaning that trial courts must re-examine subject matter that was previously deemed reliable and admissible under *Reed*. *Ibid* (citing *U.S. v. Horn,* 185 F.Supp.2d 530, 554-55 (D. Md. 2002)). That said, unlike the situation faced by federal courts upon the landmark adoption of *Daubert*, everything old is *not* new again for Maryland trial courts in that they can and should "draw from and contribute to the broad base of case law grappling with scientific testimony" accreted over the three (3) decades following the adoption of *Daubert*. *Id.* at 34-35.

Finally, "the proponent of the expert testimony bears the burden of establishing admissibility, or coming forward with evidence from which the trial court could determine that the evidence is admissible under *Daubert.*" *U.S. Wind Inc. v. InterMoor, Inc.,* — F.Supp.3d —, 2022 WL 16923872 (D. Md. 2022) (citing *Daubert,* 509 U.S. at 592 n.10).

B. AFTE Theory, Critical Reports, and Persuasive Authority

i. Toolmark Identification and the AFTE Theory of Identification

The Defendant challenges the reliability of toolmark identification, the technique employed by Kim. Toolmark identification "is the practice of investigating whether a bullet, cartridge case, or

other ammunition component or fragment can be traced to a particular suspect weapon." *State v. Fleming,* 194 Md.App. 76, 100-01 (2010).

The firearm manufacturing process involves the "cutting, drilling, grinding, hand-filing, and, very occasionally, hand polishing" of firearm components. *U.S. v. Monteiro,* 407 F.Supp.2d 351, 359 (D. Mass. Jan. 2006). Proponents of toolmark identification posit that this manufacturing process, which involves the scraping of harder metal manufacturing tools against the softer metal of the firearm itself, creates distinguishable marks on the firearm known as toolmarks which correspond with imperfections on the manufacturing tools. Furthermore, over time, repeated use of the firearm may change these toolmarks further as the discharged rounds[5] scrape against and further warp the firearm components. A round fired from a firearm will necessarily contact components of the firearm—such as the breech face, firing pin, or barrel—which will in turn impart their particular marks and scrapes onto the even softer metal of the bullet or cartridge case. *Id.* at 360.

Toolmark examiners group and classify these toolmarks in three categories:

(1) *Class Characteristics*, which are present in all firearms of the same make and model, such as the direction and twist of barrel rifling;

(2) *Subclass Characteristics*, which are incidental to firearm manufacture and may only be present in a certain make or model depending on the time and place of manufacture, such as the use of a particularly dull tool to cut rifling; and

[5] A round consists of a bullet projectile and cartridge case, which itself contains primer and propellant responsible for creating the chemical reaction/explosion that propels the bullet.

(3) *Individual Characteristics,* which are "random imperfections produced during manufacture or caused by accidental damage" and are posited by proponents of the theory to be idiosyncratic to a particular firearm.

Id. at 359. The predominant theory of toolmark identification is the Association of Firearm and Tool Mark Examiner theory (hereinafter "AFTE theory" or "AFTE methodology"). The AFTE theory states that:

> "opinions of common origin [can] be made when the unique surface contours of two toolmarks are in 'sufficient agreement.' Significance is determined by the comparative examination of two or more sets of surface contour patterns comprised of individual peaks, ridges, and furrows within one set of surface 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 the same tool. The statement that 'sufficient agreement' exists between two toolmarks means that the agreement of individual characteristics is of a quantity and quality that the likelihood another tool could have made the mark is so remote as to be considered a practical impossibility."

AFTE Theory of Identification as it Relates to Toolmarks, State Ex. 2. Put more simply, the AFTE theory stands for the proposition that an examiner can conclude that two sets of bullets or cartridges were discharged from the same firearm when the observed similarities between the two sets of bullets or cartridge cases are better than the greatest level of observed similarities between any two sets of bullets or cartridge cases known to have been fired from different firearms.

It is undeniable that the theory as described is subjective at its core, reliant upon an opaque determination of sufficient agreement—even the AFTE itself describes the theory as "subjective in nature." *Ibid*. The AFTE theory does not quantify or specify how many of or to what extent the individual characteristics must be in agreement and relies on an otherwise undefined comparison against the "best agreement" between toolmarks produced by two different firearms. Researchers have attempted to metrologically[6] validate and perform toolmark analysis through the use of 3D imaging, machine learning, and confocal microscopes, but such efforts have yet to gain traction in the larger field due to their relative costliness.

Although the AFTE theory only defines the parameters of a matching set of toolmarks, toolmark examiners can nonetheless arrive at four (4) distinct conclusions upon comparing toolmarks:

(1) The two bullets/cartridges came from the same firearm, based on sufficient agreement between the individual characteristics as described above ("match" or "identification");
(2) The two bullets/cartridges came from a different firearm, usually ascertained by observed variance in class characteristics ("elimination");
(3) There is not enough evidence to arrive at either of the previous conclusions ("inconclusive"); and
(4) There are no discernible class or individual characteristics ("unsuitable").

United States v. Shipp, 422 F.Supp.3d 762, 771 (E.D. N.Y. Nov. 2019). A firearm examiner starts their examination by comparing class characteristics; if the class characteristics differ, the examiner can immediately conclude that the compared bullets or cartridges were not fired from the same firearm i.e., elimination. Committee on Identifying the Needs of the Forensic Sciences

[6] Metrology is defined as the scientific study of measurement.

Community, *Strengthening Forensic Science in the United States: A Path Forward,* National Academic Press (2009), p. 152 (hereinafter "NAS report"). If the class characteristics match, the examiner then uses a split-screen microscope to simultaneously compare the sets of toolmarks, ascertain whether sufficient agreement exists, and then arrive at one of the above-enumerated conclusions. *State v. Fleming,* 194 Md.App. at 104.

ii. NRC Report, NAS Report, and PCAST Report

Defendants challenging the reliability of toolmark identification and the AFTE theory, including the Defendant here, universally cite and rely upon three government reports critical of the toolmark identification field: (1) the National Research Council Report on Ballistic Imaging (hereinafter "NRC report");[7] (2) the NAS report; and (3) "Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods" (hereinafter "PCAST report").[8] As such, a brief exploration of each report is due.

a. NRC Report

The NRC report was authored in 2008 by a committee tasked with "assess[ing] the feasibility, accuracy and reliability, and technical capability of developing and using a national ballistics database as an aid to criminal investigations" on behalf of the federal government. Defense Ex. 5, p. 23. The committee was not charged with and did "not provide an overall assessment of firearms identification as a discipline nor d[id] it advise on the admissibility of firearms-related toolmark evidence in legal proceedings … the committee's charge [was] to determine the extent to which the toolmarks left on bullets and cartridge casings after firing a weapon can be captured

[7] National Research Council, et al., *Ballistic Imaging,* National Academic Press (2008).
[8] President's Council of Advisors on Science and Technology, *Reports to the President – Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods,* Executive Office of the President (2016).

by imaging technology … [and to] assess whether a ballistic imaging database … would be feasible and operationally useful." *Id.* at 24-25. That said, the NRC Report did find that the notion that individual characteristics are unique—a supposition underlying toolmark identification and the AFTE theory—"has not been fully demonstrated." *Ibid.*

Based on its nature and scope, it is clear that the NRC Report is of minimal probative value in assessing the validity of toolmark identification and the AFTE theory. Also, as noted by the State, the report evaluated a database system that was out-of-date even in 2008. State Ex. 7, p. 8 (hereinafter "Warren Dec'l"). As such, this Court will place little, if any, weight on the NRC report.

b. NAS Report

The NAS report was authored in 2009 by an NRC committee tasked with examining and assessing various forensic sciences, including toolmark identification. Joint Ex. 2, pp. 1-2. The report ultimately concluded that toolmark identification required additional studies to confirm the underlying supposition that toolmarks have sufficient variance to distinguish one source from another. The report noted that examiners or researchers generally cannot specify how many points of similarity are necessary for any confidence level because of a lack of knowledge as to whether and to what degree individual toolmarks are distinct, which would be required to make such a specification. *Id.* at 154. The report did conclude that "[i]ndividual patterns from manufacture or wear might, in some cases, be distinct enough to suggest one particular source, but additional studies should be performed to make the process of individualization more precise and repeatable." *Ibid.*

The NAS report persuasively demonstrates that the AFTE theory at that time required additional studies to corroborate its underlying supposition that toolmarks are individually distinguishable and unique, and that the AFTE theory lacks a precisely defined process necessary for achieving a degree of confidence. That said, its criticisms of the AFTE theory are less salient given that such studies have been conducted during the nearly fifteen (15) years following the release of the report, as discussed below. As such, the probative value of the report is limited to issues not subsequently addressed by developments in the field and literature.

c. PCAST Report

The PCAST report, released in 2016, was a follow up to the NAS report. Joint Ex. 3, p. 1. The report found that the foundational validity of toolmark identification had not been established by properly designed studies and literature. *Id.* at 106. The report highlighted perceived methodological flaws in then-extant studies: they employed either a "within-set"[9] or "closed-set"[10] study design rather than a "black box" design.[11] *Id.* at 106-08. The PCAST report noted that only one such black box study existed, the "Ames Laboratory study" (hereinafter "Ames I"), and that the results of Ames I varied significantly from those studies not employing a black box design. The report ultimately concluded that there should "be more than one such study[] to demonstrate reproducibility[] and that studies should ideally be published in the peer-reviewed

[9] "Within-set" involves studies in which examiners are presented with a collection of samples, from a handful of sources, and asked to determine which samples were shot from the same firearm. The PCAST report criticized existing studies employing this design because they resulted either in examiners declaring a staggering 97% of the samples as inconclusive or because the study authors did not adequately describe their results. *Id.* at 107.
[10] "Closed-set" involves studies in which the source gun is always present, making any examination "simpler than the problem encountered in casework, because the correct answer is always present in the collection." *Id.* at 108.
[11] A "black box" study is described as one "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." *Id.* at 6. Black box studies aim to control variables that would otherwise influence the quality of any resulting data; for example, a within-set study would allow participants to use process of elimination to deduce conclusions—or confirm their own conclusions—which would in turn artificially inflate the number of correct answers and accordingly skew and devalue any data.

scientific literature." *Ibid.* As such, according to the report, the field had not demonstrated its foundational validity[12] by failing to provide an appropriate number of methodologically valid studies.

Similar to the NAS report, the PCAST report identifies the core issue of toolmark identification and the AFTE theory: its underlying supposition of the existence of unique and distinct toolmarks and the lack of properly designed studies validating this supposition. These criticisms are more relevant given that the PCAST report was published in 2016, although they are also outdated in that the PCAST report is now around seven (7) years old and does not reflect subsequent changes in the toolmark identification literature.

iii. Survey of Persuasive Authority

The publication of the NAS and PCAST reports spurred a new wave of challenges to the admissibility of toolmark identification testimony, and some courts began limiting the scope of toolmark examiner testimony. Such cases are often cited by defendants, including the Defendant here, challenging the reliability of toolmark identification and the AFTE theory. Because there exists no post-*Rochkind* appellate authority in Maryland on the admissibility of toolmark identification testimony,[13] and because of the *Rochkind* court's recommendation to "draw from" existing case law, a brief exploration of these cases and other persuasive authority on point is warranted.

---

[12] The PCAST report defines foundational validity, at least in the context of forensic science methods, as requiring "that it be shown, based on empirical studies, to be *repeatable, reproducible, and accurate,* at levels that have been measured and are appropriate to the intended application. Foundational validity, then, means that a method can, *in principle,* be reliable. It is the *scientific* concept we mean to correspond to the *legal* requirement, in [F.R.E] 702(c), of reliable principles and methods." *Id* at 4-5 (emphasis in original).

[13] The defendant in *Abruquah v. State,* 471 Md. 249 *cert. granted,* 2020 WL 261722, levied a challenge to toolmark identification nearly identical to that here. However, this case is still pending decision by the Supreme Court of Maryland and therefore does not provide any guidance for trial courts.

U.S. v. Tibbs

In *Tibbs*, the Superior Court of the District of Columbia precluded a toolmark examiner from opining that a recovered firearm was the source of recovered cartridge casings and instead allowed the examiner only to conclude that "the [recovered] firearm cannot be excluded as the source of the casing." *U.S. v. Tibbs*, No. 2016-CF1-19431, 2019 WL 4359486 at *1 (D.C. Super. Ct. Sept. 5, 2019). Echoing criticisms of the field lodged by the NAS and PCAST reports, the *Tibbs* court took particular issue with the paucity of methodologically sound studies establishing foundational validity, lack of objective standards for establishing a match, and lack of general acceptance outside the field itself. *Id.* at 22. It also questioned the efficacy of cross-examination at eliciting for a jury the field's shortcomings, concluding that "this discipline and the disputes surrounding it seem far too complex for a series of questions on cross-examination" to adequately address, and emphasizing that courts admitting such testimony should "limit the degree of confidence which the expert is reasonably permitted to espouse." *Id.* at *23.

U.S. v. Davis

In *Davis,* the District Court for the Western District of Virginia disallowed a toolmark examiner from testifying that recovered cartridges were fired by the same gun, concluding that the two sets were a match/identification, opining that toolmarks bear signature patterns allowing them to be traced to a single firearm, or expressing a level of confidence in their opinion "to any specific level of certainty." *United States v. Davis,* 2019 WL 4306971, *8 (W.D.Va. Sept. 2019). The *Davis* court justified its holding by pointing to "the subjectivity of the field and the lack of any established methodology, error rate, or statistical foundation" evidencing the reliability of toolmark identification. *Id.* at *6.

U.S. v. Shipp

In *Shipp,* the U.S. District Court for the Eastern District of New York limited a toolmark examiner to testify "only that the toolmarks on the recovered bullet fragment and shell casings are consistent with having been fired from the same gun," prohibiting any testimony concluding that the "recovered firearm is, in fact, the source of the recovered fragment and shell casing." *United States v. Shipp,* 422 F.Supp.3d 762, 765-66 (E.D. N.Y. Nov. 2019). The *Shipp* court similarly noted concerns expressed by the NAS and PCAST reports in arriving at this conclusion, specifically highlighting a potential error rate of one false positive per 46 examinations, the circular nature of the AFTE theory, and a lack of general acceptance in the overall scientific community. *Id.* at 783.

U.S. v. Adams

In *Adams,* the U.S. District Court for the District of Oregon limited a toolmark examiner to testifying only to observed class characteristics without describing the applied methodology or any conclusions thereto. *United States v. Adams,* 444 F.Supp.3d 1248, 1267 (Mar. 2020). *Adams* varies from the above cases in that the Government "explicitly stated on the record that it did not intend to offer [the expert] as a technical witness" and instead endeavored to present them "as an expert scientific witness to combat the 'CSI effect.'" *Id.* at 1257. This distinction appears to have been a motivating factor behind the *Adams* court's decision to sharply limit the scope of examiner testimony—as the court noted, "[e]ven at its worst, comparison analysis has a very low rate of error and yields results that cannot be random. But it is not clear that those results are the product of a *scientific* inquiry." *Id.* at 1266-67 (emphasis in original).

U.S. v. Harris

In *Harris,* the U.S. District Court for the District of Columbia deemed toolmark identification and the AFTE theory "reliable and admissible" because "the only factor that does not favor admissibility is the lack of objective criteria under the fourth *Daubert* factor, but … the subjectivity of a methodology is not fatal under [Fed. R. Evid.] 702 and *Daubert*." *United States v. Harris,* 502 F.Supp.3d 28, 43 (D. D.C. Nov. 2020). The *Harris* court did enforce limitations on expert testimony to which the Government had already agreed, namely that the expert "will not use terms such as match, [] will not state his expert opinion with any level of statistical certainty, and [] will not use the phrases when giving his opinion of 'to the exclusion of all other firearms' or 'to a reasonable degree of scientific certainty.'" *Id.* at 44-45. The *Harris* court favorably cited the DOJ ULTR, observed that the Government's agreed limitations were in conformity, and clarified that the examiner's testimony should be in conformity with the DOJ ULTR. *Ibid;* State. Ex. 5.

Harris is of outsize importance because, as noted above, the State in the instant matter agreed to testimony limitations identical to those agreed to by the Government in *Harris*. As elaborated below, this Court agrees with the reasonableness of these limitations and will not limit Kim's testimony beyond those limitations agreed to by the State and imposed by the DOJ ULTR.

U.S v. Chavez

In *Chavez,* the United States District Court for the Northern District of California found that "the AFTE methodology is generally relevant and reliable" and declined to impose any limiting instructions on toolmark examiner testimony. *United States v. Chavez,* 2021 WL 5882466, *6 (Dec. 2021). Such a declination was premised on the fact that the Government stated that the expert would "not express identification conclusions with absolute certainty, or even that these

conclusions are held to a reasonable degree of scientific certainty, and [would] not testify that the identification is being made to the exclusion of all other firearms in the world." *Id.* at *2 (internal quotations omitted).

U.S. v. Cloud

In *Cloud,* the United States District Court for the Eastern District of Washington found the AFTE theory reliable under *Daubert*, with this finding premised on a conclusion that "subsequent validation studies have addressed the most significant of the PCAST Report's concerns." *United States v. Cloud,* 576 F.Supp.3d 827, 844-45 (E.D. Wash. Dec. 2021). While the *Cloud* court did not expressly impose any limitations on the expert testimony, it did state that, if the examiner "intends to go beyond testimony that merely notes the recovered cartridge casings could not be excluded as having been fired from the recovered," it would inform the jury of two deficiencies in efforts to validate toolmark identification.[14] *Id.* at 845.

U.S. v. Lee

In *Lee,* the United States District Court for the Northern District of Illinois found that toolmark identification "meets the *Daubert* standard" and permitted the toolmark examiner to testify without restriction. *United States v. Lee,* 2022 WL 3586164, *2 (N.D. Ill. Aug. 2022). The Government there represented that the toolmark examiner would not "testify as to a reasonable degree of scientific certainty regarding their findings." *Id.* at 3. Notably, the *Lee* court denied the defendant's request to hold a hearing on the matter, believing that "[a] *Daubert* hearing is not

[14] These deficiencies are that: "[(1)] only three studies that meet the minimum design standard have attempted to measure the accuracy of firearm/toolmark comparison and [(2)] these studies found false positive rates that could be as high as 1 in 46 in one study, 1 in 200 in the second study, and 1 in 67 in the third study, though this study has yet to be published and subjected to peer review." *Ibid.*

necessary" because the written filings and other judicial decisions "provide[d] the [c]ourt with a thorough and well-documented record from which to rule." *Ibid.*

U.S. v. Gist-Holden

In *Gist-Holden,* the United States District Court for the Northern District of Indiana "cho[]se to follow other courts' widespread acceptance of the AFTE methodology for toolmark identification" and allowed the toolmark examiner to testify without limitation. *U.S. v. Gist-Holden,* — F.Supp.3d —, *4 (N.D. Ind. Sept. 2022). In so concluding, the *Gist-Holden* court found that toolmark identification and the AFTE methodology "has been repeatedly tested," that "the rate of error is relatively low," that "the AFTE method has been subjected to peer review," and that the AFTE methodology is "a reliable methodology." *Id.* at *4-5.

U.S. v. Rhodes

In *Rhodes,* the United States District Court for the District of Oregon found that "[a]n analysis of the *Daubert* factors … compels a finding that the toolmark comparison evidence is admissible." *U.S. v. Rhodes,* 2023 WL 196174, p. *6 (D. Ore. Jan. 2023). The Government there did not agree to any limitations on testimony and the *Rhodes* court declined to impose any. Referring to the above-cited *Adams,* 444 F.Supp.3d at 1248, the *Rhodes* court acknowledged that "at least one judge in this District has found that toolmark analysis was not testable for *Daubert* purposes" but distinguished the case largely based on the variance in testimony quality between the two toolmark examiners. *Id.* at *2-3. The *Rhodes* court found toolmark testimony admissible because "the method is testable, has a low error rate, is based on identifiable standards, and is accepted in the relevant scientific community," while observing that "the AFTE Journal's peer review process has undergone substantial improvements in recent years…." *Ibid.* Also of note is that the

Rhodes court so concluded with the benefit of testimony from Faigman, who similarly testified in the instant matter. *Id.* at *4 ("With respect to Dean Faigman and Dr. Scurich's opinions…").

U.S. v. Blackman

In *Blackman,* the United States District Court for the Northern District of Illinois found the toolmark examiner testimony "sufficiently reliable under [Fed. R. Evid.] 702" and deemed "its probative value significant." *United States v. Blackman,* 2023 WL 3440384, p. *10 (D. Ill. May 2023). The court addressed the critiques of toolmark identification forwarded in the NAS and PCAST reports, noting that "numerous studies, including those conducted in the time since the [NAS and PCAST reports] were released, have continued to bolster the underlying premises of the field." *Id.* at *5. Like the *Harris* court, the *Blackman* court limited testimony to that "which remains consistent with DOJ [ULTR] guidelines." *Id.* at *10.

Jurisprudential Trend

The above authority appears to demonstrate a jurisprudential trend in how federal courts evaluate toolmark identification under Fed. R. Evid. 702. In 2019 and 2020, several courts were highly skeptical of the theory and placed significant limitations on permissible testimony in the wake of the relatively contemporary findings of the 2016 PCAST report. However, in the subsequent years leading up to the present, federal courts took note of developments in the field and its literature in response to the PCAST report and pared back their limitations—either limiting the examiner to testify in accordance with DOJ ULTR or imposing no limitation at all. These cases, and this observed trend, will inform this Court's decision-making process as it applies the *Daubert* factors under Rule 5-702 and *Rochkind*.

C. Expert Testimony is Admissible Under Rule 5-702

Having read the parties' briefing, reviewed the cited literature, and considered the testimony and argument of the *Rochkind* hearing, this Court finds that the testimony of Kim is admissible subject to the agreed upon limitations and its compliance with the testimony strictures of the DOJ ULTR.

1. *The Expert is Qualified as an Expert in Toolmark Identification*

The Court finds that Kim is qualified by her expertise, training, and education as an expert in the field of comparative toolmark identification. By the time she performed the examination in question, Kim had accrued around eighteen (18) years of experience as a toolmark examiner and averaged between ten to fifteen examinations per month, and she is now a supervisor at her laboratory. Kim holds a master's degree in forensic sciences, psychology, and criminology. Kim testified that she stays abreast of developments in the field by reading relevant literature and attending regional training seminars. Kim has received a competency certification from her lab and undergoes recurring proficiency testing, all of which she passed, and has been a regular member of the AFTE since 2002. Kim has been qualified as an expert 69 times in various state and federal courts, six (6) of which were in this Court, and has invariably been qualified as an expert.

2. Toolmark Identification and the AFTE Methodology Are Reliable

i. Testability

The first *Rochkind/Daubert* factor asks "whether a theory or technique can be (and has been) tested." *Rochkind,* 471 Md. at 35. Nearly every court, even those highly critical of toolmark identification evidence noted above, has found this to be the case. *See, e.g., Blackman,* 2023 WL 3440384 at *5 ("AFTE theory can be (and has been) tested."); *Rhodes,* 2023 WL 196174 at *3

("This Court finds that the AFTE methodology can be, and has been, tested."); *Chavez,* 2021 WL 5882466 at *2 ("The Court agrees with these cases and finds this [testability] factor weighs in favor of admissibility."); *Harris,* 502 F.Supp.3d at 38 ("the Court concludes that the testability factor supports admissibility…"); *Shipp*, 422 F.Supp.3d at 776 ("The Court finds that the AFTE Theory can be and has been tested and this factor therefore weighs in favor of admissibility."); *Tibbs,* 2019 WL 4359486 at *7 (noting that "[t]here appears to be little dispute that toolmark identification is testable as a general matter" and finding that it "can be, and ha[s] been, tested.") (internal quotations omitted). This Court agrees with the weight of authority that the AFTE theory can be and has been tested.

The Defendant argues that, although "[f]irearms techniques are testable … the lack of black box studies means that they had yet to be tested appropriately or adequately" and discounts subsequent black-box studies published since the release of the PCAST report on grounds that "they suffer significant methodological flaws…." *Faigman Dec'l.* at 23-24. This Court disagrees that the studies should be discounted.

The State has cited several peer-reviewed, black box studies, many of which support a finding of testability. *Warren Dec'l* at 8. One such study, a "black-box validation study designed to determine the accuracy, repeatability,[15] and reproducibility"[16] of toolmark identification found that "examiners score high in repeatability, i.e. their observed performance generally exceeds the statistically expected agreement by a fairly wide margin" and that "in general the determinations made by different examiners are reproducible … the general trend toward better observed

[15] Repeatability is defined as "[t]he ability of an examiner, when confronted with the exact same comparison once again, to reach the same determination as when first examined." Ames II. at 10.
[16] Reproducibility is defined as "[t]he ability of a second examiner to evaluate a set previously viewed by a different examiner and reach the same conclusion." *Id.* at 11.

agreement than expected agreement documents commonality in how the examination process is performed within the profession." S. Bajic, et al., *Report: Validation Study of the Accuracy, Repeatability, and Reproducibility of Firearm Comparisons*, Ames Laboratory-USDOE Technical Report # ISTR-5220, 2020, pp. 2, 73-74 (hereinafter "Ames II"). The Ames II study satisfies many of the criteria required of a reliable study as enumerated by PCAST and the Defendant: the study employed a black-box design, was peer-reviewed, and was conducted by the Ames Laboratory under the U.S. Department of Energy—a neutral source without any vested interest in validating the field of toolmark identification. While such studies may suffer from some methodological flaws, perhaps not adequately controlling for sources of human bias, these flaws are relatively minor and, in any event, were not emphasized by the PCAST Report as major methodological issues plaguing existing studies. The findings in Ames II demonstrate that the AFTE theory can be (and has been) tested.

Furthermore, as testified to by Kim and Warren, in practice a toolmark examiner is required to document their findings through a written report and have the findings validated by another qualified examiner. Such requirements "ensure sufficient testability and reproducibility to ensure that the results of the technique are reliable." *Harris,* 502 F.Supp.3d at 38 (internal citations omitted). For these reasons, the Court finds that the testability factor weighs in favor of admissibility.

ii. Peer Review and Publication

The second *Rochkind/Daubert* factor inquires "whether a theory or technique has been subjected to peer review and publication." *Rochkind,* 471 Md. at 35. This factor is of import because "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." *Daubert,* 509 U.S. at 593-94. Many courts, especially those more recently addressing this factor, have found that toolmark identification has been subjected to peer review and publication. *See, e.g., Blackman,* 2023 WL 3440384 at *7 ("The Court thus finds the peer review and publication standard factor satisfied."); *Rhodes,* 2023 WL 196174 at *3 (concluding that "toolmark analysis has been subjected to peer review and publication" while ultimately finding that "this factor is neutral."); *Chavez,* 2021 WL 5882466 at *2 ("The Court agrees with the weight of authority and finds [the peer review and publication] factor weighs in favor of admissibility."); *Harris,* 502 F.Supp.3d at 38 ("Because the toolmark identification methodology used by [the toolmark examiner] has been subject to peer review and publication, the Court finds this *Daubert* factor to also weigh in favor of admission."); *Shipp*, 422 F.Supp.3d at 776-77 ("even assigning limited weight to the substantial fraction of the literature that is published in the AFTE Journal, this factor still weighs in favor of admissibility … [i]ndeed, the scrutiny of PCAST and the flaws it perceived in the AFTE Theory inform much of the discussion" on critiques of the AFTE theory).

Courts arriving at different conclusions largely did so because, at that point, most if not all toolmark identification studies were published in the AFTE Journal which then suffered from flaws in its peer review process. *See Tibbs,* 2019 WL 4359486 at *9 (noting that other courts finding peer review and publication "devote little attention to the sufficiency of this [AFTE] journal's peer review process or to the issues stemming from a review process dominated by financially and professionally interested practitioners" and criticizing the then-contemporary practices of the AFTE Journal including the lack of a double-blind peer review process).

Such criticisms, while perhaps valid at the time, have since been addressed by the field. The AFTE itself adjusted its journal review process in January of 2020 in response to these criticisms. *See AFTE Peer Review Process – January 2020*, AFTE, https://afte.org/afte-journal/afte-journal-peer-review-process (last visited May 26, 2023). The AFTE journal peer review process now entails a "double-blind peer review process" and provides that "the article [in review] may be forwarded to a subject matter expert outside of the Editorial Review Panel." *Ibid.*

To any extent that criticisms of the AFTE journal peer review process are still valid, in that the journal is ultimately reviewed by peers within the AFTE community and thus is insulated from outside critique, such criticisms are inapplicable to additional studies and reports published by other entities. The State cited several studies published in the Journal of Forensic Sciences, which is published by an entity comprised of professionals from various fields including pathology, dentistry, toxicology, anthropology, and chemistry. *See About the AAFS*, AAFS, https://www.aafs.org/about-us (last visited May 26, 2023). Additionally, other peer reviewed reports cited by the State were issued by neutral parties outside the forensic science community. *See Warren Dec'l* at 8.

Furthermore, as noted above, some courts assessing this factor have found that the PCAST report alone satisfied the factor as articulated by the *Daubert* court. *See Shipp* at 777 ("[i]ndeed, the scrutiny of PCAST and the flaws it perceived in the AFTE Theory inform much of the discussion" of the AFTE theory). The *Daubert* court emphasized that this factor was important because "submission to the scrutiny of the scientific community … increases the likelihood that substantive flaws in methodology will be detected," and, according to the Defendant, the PCAST

Report indeed detected and laid bare these substantive flaws. *Daubert,* 509 U.S. at 593. For these reasons, the Court finds that the peer review and publication factor weighs in favor of admissibility.

iii. Known or Potential Error Rate

The third *Rochkind/Daubert* factor queries "whether a particular scientific technique has a known or potential rate of error." *Rochkind,* 471 Md. at 35. The parties here vociferously dispute how to correctly calculate error rates and interpret study data, with this dispute centering on how studies treat and categorize a finding of "inconclusive," as opposed to a finding of match/identification or elimination.

Ames II is useful in explicating this disagreement. In Ames II, study participants were provided test-packets containing fifteen (15) bullet comparison sets and fifteen (15) cartridge case comparison sets, and these sets had known ground truths[17] as established by the study. *Ames II,* p. 32. Examiners were tasked with examining and classifying these sets as either a match/identification, inconclusive, elimination, or unsuitable, with the inconclusives divided into three categories based on how close they were to a match or elimination. *Ibid.* The following table regarding examiner accuracy demonstrates this paradigm and shows some useful data.

[17] Ground truth means that there is an established, correct answer. In this context, it means that the sets of bullets or cartridge casings were either fired from the same firearm or fired from different firearms which corresponds to a "correct" finding of match/identification or elimination respectively.

Table V: First-Round Bullet and Cartridge case Summary Counts.

Bullet Evaluations by Set Type						
	ID	Inconclusive-A	Inconclusive-B	Inconclusive-C	Elimination	Other
Matching	1076	127	125	36	**41**	24
Nonmatching	**20**	268	848	745	961	49
Cartridge case Evaluations by Set Type						
	ID	Inconclusive-A	Inconclusive-B	Inconclusive-C	Elimination	Other
Matching	1056	177	140	22	**25**	25
Nonmatching	**26**	177	637	620	1375	40

Id. at 34. This table reflects the responses of submitting examiners who were asked to examine sets of bullets (the first group of rows) and sets of cartridge casings (the second group of rows). The study designers had ground truths established for the bullets and cartridge casing sets as either a match or elimination, with the leftmost column reflecting which rows correspond with ground truths of a match ("matching") or an elimination ("nonmatching"). *Ibid.* The subsequent columns reflect examiner responses and display the number of examiners who reached each of the four (4) aforementioned conclusions.[18]

There are several ways in which one can interpret these data and accordingly calculate an error rate, and this manner of interpretation is where the State and Defendant disagree. The State interprets these data in a manner identical to that of the study designers, calculating error rates based on the number of false positives (i.e. an examiner finds a match when the ground truth is an elimination) or false negatives (i.e. an examiner finds an elimination when the ground truth is a match); false positives and negatives are collectively referred to by the study as "hard errors."

[18] The table displays inconclusive determinations as either an Inconclusive-A, B, or C. These three categories of inconclusive represent how close an examiner was to arriving at a non-inconclusive conclusion. Inconclusive-A corresponds to inconclusive findings that border on a match, Inconclusive-C corresponds to inconclusive findings that border on an elimination, and Inconclusive-B corresponds to inconclusive findings that do not border on either a match or an elimination. *Ibid.*

Ibid; *see Warren Dec'l* at 7-8. These hard errors are bolded on the table. To calculate the false positive error rate and false negative error rate in bullet evaluations, the study designers divided the corresponding number of false positives or false negatives by the total number of results and arrived at a false positive error rate of 0.704% and a false negative error rate of 2.92% respectively.[19] Therefore, as construed by the study designers and the State, the numerator in each of these equations only consists of false positives or negatives and does not deem any inconclusive finding an error.

According to the Defendant, this method of calculation suffers from statistical and methodological shortcomings. Because the Ames II study had ground truths established, the Defendant reasons that any answer other than the ground truth itself constitutes an error. *Faigman Dec'l* at 18. This interpretation results in error rates vastly greater than under the previous interpretation: bullet and cartridge case rates would be 53% and 44% respectively. *Ibid*. The Defendant also argues that toolmark examiners, cognizant of the fact that inconclusives will not count as an incorrect answer, tend to provide more inconclusive answers than they would in casework. *Ibid; see also* ltiel E. Dror & Nicholas Scurich, *(Mis)Use of Scientific Measurements in Forensic Science,* 2 Forensic Science Int'l: *Synergy* 333, 336 (2020). This conundrum of error rate calculation is present in toolmark scholarship and is a hotly contested issue. *Compare. id., with* Alex Biedermann & Kyriakos N. Kotsoglou, *Forensic Science and the Principle of Excluded Middle.' "Inconclusive" Decisions and the Structure of Error Rate Studies,* 3 Forensic Science Int'l: Synergy 100147 (2021); *and* Max Morris, *Comments on: A Re-Analysis of*

[19] Described formulaically, for bullet evaluations these would be: **False Positive** = ((ID) / (False ID + Inconclusive A + Inconclusive B + Inconclusive C +Eliminations)) x 100%, or ((20)/(20+268+848+745+961)) x 100% and **False Negative** = ((Eliminations) / (ID + Inconclusive A + Inconclusive B + Inconclusive C + False Eliminations)) x 100%, or ((41)/(1076+127+125+36+41)) x 100%. *See ibid.*

Repeatability and Reproducibility in the Ames-USDOE-FBI Study, by Dorfman and Valliant, 10 Statistics and Public Policy No. 1 (2023).

That said, this Court need not weigh in on this academic debate. Recently, federal courts have convincingly reasoned that trial courts should only consider false positive identifications when evaluating this *Rochkind/Daubert* factor. *See, e.g., Blackman,* 2023 WL 3440384 at *6 (noting that "the critical inquiry under the [error rate] factor is the rate of error in which an examiner makes a false positive identification"); *Rhodes,* 2023 WL 196174 at *4 (considering only false positive identifications); *Harris,* 502 F.Supp.3d at 39 (agreeing with the Government that only false positive identifications are of note because only those "could lead to a conviction premised on faulty evidence."). In arriving at this decision, this Court finds the reasoning in *Rhodes* particularly convincing and relevant:

> "With respect to Dean Faigman['s] … opinion that the error rates are unreliable because they do not include inconclusive results, this Court finds those opinions unsupported. Or, perhaps more accurately, the Court finds that while the relatively high rate of inconclusive results may be relevant to certain policy determinations—such as failing to eliminate a suspect from consideration—those concerns do not relate to the issue of paramount importance to trial courts; i.e., the false positive rate which may result in a wrongful conviction … [p]ut another way, while an inconclusive result is an error insofar as it means the methodology did not produce an answer, it is not an error in the sense that it falsely attributes a cartridge or casing to the wrong firearm."

Rhodes, 2023 WL 196174 at *4. Furthermore, as mentioned above, the AFTE theory itself does not define the parameters of an inconclusive or elimination and only specifically describes the process for arriving at a conclusion of match/identification.

As such, while it may or may not be improper to discount inconclusives when calculating error rates from a study design perspective, such an exclusion is proper from the perspective of evidentiary admissibility at a criminal trial. The State has provided numerous studies with false positive rates averaging at 0.47%, with the highest at 1.01%. *Warren Dec'l* at 8.

That said, this Court is convinced by the Defendant's contention that examiners tend to find more inconclusives in studies than in field work, a concomitant of which being that examiners likely have higher false positive rates in practice than is reflected in the cited studies. However, despite this valid concern about the rate of inconclusive findings in the field, the Court nonetheless finds that the error rate factor weighs in favor of admissibility.

iv. Existence and Maintenance of Standards

The fourth *Rochkind/Daubert* factor asks a trial court to consider "the existence of maintenance of standards and controls." *Rochkind,* 471 Md. at 35. Unlike under the previous factors, reviewing courts have varied significantly in how they frame and weigh this factor. *Compare Blackman*, 2023 WL 3440384 at *7 ("Given the rigor of the underlying analysis and methodology … the Court finds that this factor does not help Defendants either."); *Rhodes,* 2023 WL 196174 at *5-6 ("the subjectivity present in the best-known non-match and sufficient agreement standards is subject to industry standards that safeguard the process … [t]his factor

favors admissibility."); *and Adams*, 444 F.Supp.3d at 1266 ("[The examiner] did identify some quality control mechanisms … he takes an annual proficiency test … every forensic toolmark test is reviewed by a second examiner … [the examiner] receive[s] training and procedures manuals from AFTE … [t]his amounts to quality control … [t]his *Daubert* factor therefore favors the Government…."), *with Chavez,* 2021 WL 5882466 at *4-5 (noting that "both courts and the scientific community have voiced serious concerns about the sufficient agreement standard" and finding that "this factor weighs against a finding of reliability…") (internal quotations and citations omitted); *Harris,* 502 F.Supp.3d at 41-42 ("The entire process of reaching a conclusion regarding the sufficient agreement in individual characteristics is one that relies wholly on the examiner's judgment, without any underlying numerical standards or guideposts to direct an examiner's conclusion … without more the Court cannot conclude this *Daubert* factor is met … this factor weighs against the admission of [examiner] testimony."); *Shipp,* 422 F.Supp.3d at 782 ("the court finds that the subjective and circular nature of the AFTE Theory weighs against finding that a firearms examiner can reliably identify" a match); *and Tibbs,* 2019 WL 4359486 at *21 ("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 … this fourth *Daubert* factor strongly militates against the admission of expert witness testimony…").

Judicial analysis of this factor varies from the previous factors in two ways. First, some of the above-cited courts made findings on this factor in disagreement with their overall conclusion of admissibility. *See, e.g., Adams,* 444 F.Supp.3d at 124 (significantly limiting examiner testimony despite finding that the factor favors the Government); *Harris,* 502 F.Supp.3d at 42-43 (declining to impose any limitations on testimony other than those consented to by Government despite

finding that the factor weighs against the admission of examiner testimony). Second, this factor has not been as heavily influenced by subsequent developments in scholarship of the field in response to the PCAST report and other criticisms, as the AFTE methodology itself remains largely unchanged.

This Court agrees with the reasoning employed by courts critical of the AFTE methodology under this factor and finds that it weighs against admissibility. Of course, "a partially subjective methodology is not inherently unreliable, or an immediate bar to admissibility;" such a requirement would bar swaths of reliable yet subjective testimony, such as a psychiatrist diagnosing a patient with a mental disorder, a histocytologist examining and identifying pollen grains under a microscope, or a radiologist interpreting X-rays or CT scans. *Harris,* 502 F.Supp.3d at 42.

However, unlike other subjective methodologies, the AFTE methodology does not attempt to define or enumerate the standards for what supports a particular finding and does not provide objective standards as quality control measures. The AFTE methodology also varies in that it ultimately *can* be subject to metrological validation, unlike other subjective determinations categorically incapable of quantification such as the testimony of a harbormaster opining on regional boating conditions. Because the AFTE methodology does not even attempt to objectively quantify or otherwise explicate the level of agreement required to distinguish a match from a non-match, and because it relies on an individual examiner's experience and memory of best-known non-matches, it falls short of the methodological standards imposed by other comparable subjective fields such as fingerprint analysis. *See Harris,* 502 F.Supp.3d at 41-42

(explaining the objective standards and quality control measures, and agreed-upon matching point standards, incorporated in the field of fingerprint analysis).

This is not to say that the AFTE methodology as applied by examiners in the field is entirely unmoored from any standards or controls such that examiners are "left to rely on [their] own thoughts and conclusions based only on the vagaries of [their] own training and experience." *Tibbs,* 2019 WL 4359486 at *21. To achieve accreditation, crime laboratories are required to develop and maintain standard operating procedures to be employed by their examiners whose validation by an accrediting body necessarily imposes a degree of uniformity of standards and procedures in the multitude of existing crime laboratories. *Warren Dec'l* at 7; *see also OSAC and AFTE Firearms Process Map,* National Institute of Standards and Technology, https://www.nist.gov/document/osac-firearms-process-mapjanzoz1. For example, Kim testified that, in performing her examination of the cartridge casings and bullets in question, she employed the standard operating procedures imposed by her laboratory at the time, which were also in conformity with AFTE guidelines. Toolmark examiners also require validation of their findings by a second examiner and are required to document their findings.

That said, while toolmark examiners impose various standards and processes that standardize and homogenize the application of the AFTE methodology across laboratories and individual examiners, the methodology itself is ultimately circular and tautological—reliant upon an opaque "sufficient agreement" finding that employs no objective guideposts or criteria to inform an examiner's conclusion. Accordingly, this factor weighs against admissibility.

v. General Acceptance

The last *Daubert* factor probes "whether a theory or technique is generally accepted." *Rochkind,* 471 Md. at 36. Although the *Daubert* court describes general acceptance as acceptance by "a relevant scientific community," it does not explain what a relevant scientific community is or how courts are to define its contours. *Daubert,* 509 U.S. at 595. Courts assessing toolmark identification under this factor appear to take three (3) approaches to defining a relevant scientific community.

First, some courts have defined the community as exclusively the field of toolmark identification itself and have universally found this factor to weigh in favor of admissibility. *See, e.g., Blackman,* 2023 WL 3440384 at *8 ("The standard is general—not universal—acceptance … AFTE remains the dominant approach in the field, and thus this Court easily concludes that the general acceptance factor has been satisfied and supports admissibility."); *Rhodes,* 2023 WL 196174 at *6 ("the weight of authority suggests that the AFTE method does enjoy general acceptance in the relevant scientific community—forensic ballistic examiners … this factor weighs in favor of admissibility."); *Harris,* 502 F.Supp.3d at 42-43 (noting that the AFTE methodology "enjoy[s] general acceptance as a reliable methodology in the relevant scientific community of examiners" and finding that "this factor weighs in favor of admitting" examiner testimony).

Other courts have defined this as the scientific community generally and found the factor to weigh against admissibility. *See, e.g., Adams,* 444 F.Supp.3d at 1266 ("The AFTE method … has been widely accepted within [the examiner's] own community of technical experts. But it has been heavily criticized by other members of the broader scientific community … I cannot find that the AFTE method enjoys general acceptance in the scientific community.") (internal

quotations omitted); *Shipp,* 422 F.Supp.3d at 782-83 ("Most courts have, in cursory fashion, identified toolmark examiners as the relevant community … [t]he court believes a broader definition is more appropriate … the court finds it appropriate to consider the opinions of the authors of the [NAS] report and the PCAST report who, while admittedly not members of the forensic ballistic community, are preeminent scientists and scholars … the AFTE Theory has not achieved general acceptance…."); *Tibbs,* 2019 WL 4359486 at *21 ("courts must not confine the relevant scientific community to the specific group of practitioners dedicated to the validity of the theory … [h]ere, the [G]overnment failed to show general acceptance outside the field of firearms and toolmark practitioners … a factor that weighs against admissibility."). One court elected to remain agnostic on the issue, instead relying on the *Daubert* court's phrasing of the factor in arriving at its conclusion. *Chavez,* 2021 WL 5882466 at *5 ("Setting aside whether this [the scientific community at-large] is a correct definition of the relevant scientific community … defendant cannot show that the technique in question has 'only minimal support' to warrant skepticism … the Court finds this factor weighs in favor of admissibility.") (quoting *Daubert,* 509 U.S. at 594).

This Court agrees with the analytical framework employed by the *Chavez* court and finds that this factor weighs in favor of admissibility. The *Daubert* court explicitly states that "a known technique which has been able to attract only minimal support within the community may properly be viewed with skepticism," meaning that such skepticism is not as warranted when the technique has more than minimal support within the community. *Daubert,* 509 U.S. at 594. Obviously, when the community is defined as the field of toolmark examiners, in which the AFTE methodology is predominant, it undoubtedly has more than minimal support. However,

even when the community is defined as the scientific community more generally, it nonetheless appears that the AFTE methodology has more than minimal support.

The State cited several studies and articles, performed or written by researchers outside the field of toolmark identification or forensic science generally, that validate toolmark identification or the AFTE methodology. *See, e.g.,* Max G. Morris, *Comments on: A Re-Analysis of Repeatability and Reproducibility in the Ames-USDOE-FBI Study, by Dorfman and Valliant,* 10 Statistics and Public Policy No. 1 (2023) (authored by a professor with the Department of Statistics and Industrial and Manufacturing Systems Engineering of the Iowa State University); Eric F. Law and Keith B. Morris, *Evaluating Firearm Examiner Conclusion Variability Using Cartridge Case Reproductions*, J Forensic Sci., 66 (5) (2021) (authored by members of the Department of Forensic and Investigative Science, West Virginia University). To be sure, toolmark identification and the AFTE methodology have detractors in the broader scientific community that forward cogent and well-received critiques. That said, it appears that—while certainly not accepted by the entire at-large scientific community—toolmark identification and the AFTE methodology enjoy more than "minimal support" within the scientific community regardless of how that community is defined. Accordingly, this factor weighs in favor of admissibility.

vi. Purpose of Developing Opinion

The sixth *Rochkind* factor, and the first not explicitly set forth as a factor in *Daubert*, requires a court to consider "whether experts are proposing to testify about matters growing naturally and directly out of research they have conducted independent of the litigation, or whether they have developed their opinions expressly for purposes of testifying." *Rochkind,* 471 Md. at 35. As noted above, the *Rochkind* court emphasized that a trial court's inquiry into reliability under Rule

5-702 is flexible and that "the law grants a trial court the same broad latitude when it decides *how* to determine reliability as it enjoys in respect to its ultimate reliability determination." *Id.* at 37 (emphasis in original). This Court finds that this factor is minimally relevant here and therefore this factor will not be considered in deciding reliability.

vii. Unjustifiable Extrapolation

The seventh *Rochkind* factor pertains to "whether the expert has unjustifiably extrapolated from an accepted premise to an unfounded conclusion." *Rochkind,* 471 Md. at 35. Here, it does not appear that Kim engaged in any unjustifiable extrapolation in arriving at her conclusion—she determined that the sets of casings and bullets shared class characteristics, observed sufficient agreement between breech face marks, ruled out any subclass characteristics that might mimic individual characteristics, and accordingly concluded that the compared sets were fired from the same firearm. The Court finds that this factor weighs in favor of admissibility.

viii. Adequate Accounting for Alternative Explanations

The eighth *Rochkind* factor addresses "whether the expert has adequately accounted for obvious alternative explanations." *Ibid.* Here, it appears that Kim has adequately accounted for alternative explanations. The most obvious alternative explanation as to why two sets of seemingly identical toolmarks were left on two distinct sets of ballistic evidence is that they were produced by two different guns manufactured in the same manufacturing run[20] and would therefore bear identical subclass characteristics that might mimic individual characteristics.

[20] Another way of phrasing this is that there were two separate guns both produced in the same facility in the same production line such that the same manufacturing tool(s) were used to produce both guns at roughly the same time, thus suggesting that identical toolmarks would be imparted by the manufacturing tool(s).

Here, Kim adequately accounted for subclass characteristics when performing her examination. She testified that cartridge cases express subclass characteristics through course, parallel marks running across the cartridge and that bullets usually express subclass characteristics through course marks that run across the entire length of the bullet. For the cartridge cases, Kim noted that she looked for finer agreement between the cartridges and noted marks inconsistent with subclass characteristic expressions—namely that the cartridges bore granular marks that ran in different directions. For the bullets, Kim explained that the marks she used for comparison were very short and did not run along the entire length of the bullet surface, thus accounting for any potential influence of subclass characteristics in arriving at her conclusion. Kim also stated that she did not make any identification based on marks in bullet groove impressions[21] because of their high potential for misidentification. As such, this factor weighs in favor of admissibility.

ix. Comparable Level of Care

The ninth *Rochkind* factor assesses "whether the expert is being as careful as he [or she] would be in his [or her] regular professional work outside his [or her] paid litigation consulting." *Rochkind*, 471 Md. at 36. As noted above, trial courts have wide discretion in which factors they consider when deciding reliability. Kim is not engaged in any "paid litigation consulting," being employed by the Maryland State Police laboratory, and therefore this factor is inapposite and will not be considered by the Court.

x. Claimed Expertise Reaches Reliable Results

The tenth and final *Rochkind* factor asks "whether the field of expertise claimed by the expert is known to reach reliable results for the type of opinion the expert would give." *Ibid.* As noted

[21] Kim explained that groove impressions are the impressions left on bullets by grooves on the rifling of a firearm. The raised areas of rifling are known as lands and the recessed areas are known as grooves.

under the first and third *Daubert* factors—testability and error rate—it appears that toolmark identification is known to reach somewhat reliable results. Accordingly, this factor weighs in favor of admissibility.

Balance of Factors Weighs in Favor of Admissibility

Balancing the ten (10) *Rochkind* factors, the Court finds that the proposed testimony of Kim is sufficiently reliable to warrant admission without the Defendant's proposed testimony restrictions. Other than the existence and maintenance of standards and controls, all relevant factors weigh in favor of admissibility. If this Court were deciding the instant matter between 2016 and 2020, it is quite possible that the testimony would be rendered inadmissible, or at least severely limited, by the significant flaws set forth in the NAS and PCAST reports. That said, in 2023, such criticisms are less valid. There now exists several black-box, peer reviewed studies supporting the underlying validity of the AFTE method—studies which have been published in journals other than the AFTE Journal by disinterested entities unassociated with the field. While it is undeniably true that the AFTE theory is subjective, "the subjectivity of a methodology is not fatal under Rule [5-]702 and *Daubert*." *Harris,* 502 F.Supp. at 43. It is also true that, aside from the subjectivity issue, the opacity of the "sufficient agreement" standard concomitantly lowers the degree of confidence with which examiners should express their findings. However, this issue is largely addressed by the testimony limitations as agreed to by the State, which will be further explored below.

3. Sufficient Factual Basis Exists Supporting Expert Testimony

As noted above, Rule 5-702 requires that a trial court determine "whether a sufficient factual basis exists to support the expert testimony." The Defendant asserts that the expert testimony

does not satisfy this requirement on three (3) grounds. First, the Defendant argues that there is insufficient data because Kim failed to provide documentation of the individual characteristics on which she relied during her examination. Motion In Limine, pp. 15-17. Next, the Defendant avers that Kim insufficiently accounted for subclass characteristics. *Id.* at 17-20. The Defendant supplements this argument by also noting that Kim did not have any recovered firearm at her disposal, stating that this compromised her ability to identify and consider subclass characteristics. *Id.* at 20-21. Finally, the Defendant alleges that Kim failed to adequately protect against unconscious bias, primarily because she was the verifying examiner when the evidence in question was first tested in 2015 and the primary examiner when it was retested in 2018. *Id.* at 21-23.

The factual basis portion of Rule 5-702 requires that "sufficient facts must underlie the expert's opinion that indicate the use of reliable principles and methodology in support of the expert's conclusions so that the opinion constitutes more than mere speculation or conjecture." *Parkway Neuroscience and Spine Institute, LLC v. Katz, Abosch, Windesheim, Gershman, & Freedman, P.A., et al.,* 255 Md.App. 596, 629-30 (2022) (citing *Exxon Mobil Corp v. Ford,* 433 Md. 426, 467-81 (2013)) (internal quotations omitted). Here, Kim noted and accounted for sufficient facts and used (somewhat) reliable principles and methodology in arriving at her conclusion, meaning that her opinion constitutes "more than mere speculation or conjecture," or *ipse dixit*. *Ibid.*

As to the first ground, the Defendant falsely equates a failure to document certain observations with a failure to make those observations. Although Kim did not document her assessment of subclass characteristics and did not fully document her consideration and assessment of individual characteristics, she testified as to her process and how she arrived at her conclusion—

and can similarly testify to such processes and conclusions thereto subject to cross-examination at trial. Furthermore, Warren—who is an expert in toolmark identification and examination, has been qualified as such in excess of 70 times, and runs a consulting firm that consults for both prosecutors and defendants—reviewed Kim's report, finding that there were no "red flags" and observing that the documentation was sufficient for him to evaluate Kim's process and accompanying conclusions. To any extent that contemporary protocol and standards require documentation of subclass characteristics, such a requirement was not imposed when Kim performed her 2018 examination and her examination was in accord with then-existing Maryland State Police laboratory standards and procedures.

This line of reasoning similarly applies to the second ground; while the documentation itself might not demonstrate that Kim sufficiently accounted for subclass characteristics when performing her examination, her testimony does. While the Defendant does forward compelling points about the difficulties facing toolmark examiners when comparing toolmarks without the benefit of a recovered firearm and finding a match based on breech face mark comparisons, which is the case here, such critiques do not make her opinion "mere speculation or conjecture." *See Gene C. Rivera et al.,* Subclass Characteristics in Smith & Wesson SW40VE Sigma Pistols, 3 AFTE Journal Vol. 39, 247 (Summer 2007). Rather, her opinion is proper fodder for cross-examination rather than exclusion—such critiques will be comprehensible for jurors and ultimately go to the weight of her testimony rather than its admissibility. It bears mentioning that Kim herself distinguished the articles cited by the Defendant from her observation of and reliance upon breech face marks, noting that the issues raised in these articles uniquely applied to Smith & Wesson pistols rather than all firearms.

As to the third ground, the fact that Kim was the verifying examiner in 2015 and the primary examiner in 2018 does not taint her conclusions with such a degree of confirmation bias to warrant exclusion. Kim was shielded from the initial results when she verified the primary examiner's conclusions in 2015 and, as Warren testified, casework as a primary examiner versus a verifying examiner requires different mindsets, different levels of documentation, and are essentially two different processes. While it is indeed true that Kim herself brought the fact that she was the verifying examiner to her supervisor's attention and was nonetheless assigned to the case for purposes of economy, nothing suggests that Kim had any recollection of her 2015 findings that would subsequently infect her 2018 finding with unacceptable levels of confirmation bias. It strains credulity to believe that Kim, who averaged between ten (10) and fifteen (15) examinations a month, would recall details of her verification of a case from which she was shielded when performing her primary examination around (3) years—and hundreds of examinations—later.

To be sure, some of the critiques forwarded by the Defendant spark some underlying concern in the validity of Kim's conclusion. However, shortcomings in an expert's conclusion, so long as they are supported by sufficiently reliable methodological underpinnings, "go[] to the weight, not the admissibility of the testimony." *Parkway Neuroscience,* 255 Md.App. at 630 (internal citation omitted). "Vigorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence." *Rochkind,* 471 at 38 (citing *Daubert,* 509 U.S. at 596). This description—shaky but admissible evidence—accurately describes the nature of Kim's testimony. Additionally, the testimony limitations to which the State agreed, discussed below, will attenuate or eliminate any veneer of scientific certainty that might unduly influence a jury's

understanding of the testimony and its probative value. As such, the Court finds that a sufficient factual basis exists supporting Kim's testimony such that it satisfies Rule 5-702.

Testimony Limitations

The testimony limitations which the State has imposed upon itself, as articulated in *Harris,* 502 F.Supp.3d at 44, sufficiently restrict Kim's testimony such that no additional limitations are required—so long as the testimony is in accord with the DOJ ULTR. As noted above, the *Harris* court articulated the limitations consented to by the Government but also instructed the expert "to abide by the expert testimony limitations detailed in the DOJ ULTR." *Id.* at 45. It is the State's intention to do the same.[22] Testimony subject to the agreed upon limitations and in conformity with the DOJ ULTR will accurately frame for the jury the reliability of the AFTE theory and toolmark identification generally and adequately account for the valid critiques forwarded by the Defendant.

CONCLUSION

For the aforementioned reasons, the Defendant's Motion *In Limine* to Exclude Firearms Identification Evidence as Unreliable Under Rochkind v. Stevenson, 471 Md. 1 (2020) is **DENIED.**

06/14/2023 9:08:43 AM

Judge Kathleen L. Beckstead

[22] For example, Kim "will not use terms such as match, [] will not state [her] expert opinion with any level of statistical certainty, and [] will not use phrases when giving [her] opinion of 'to the exclusion of all other firearms' or 'to a reasonable degree of scientific [or ballistic] certainty.'" *Id.* at 44-45 (some internal quotations omitted). Kim may:

> "conclude that casings were fired from the same firearm when all class characteristics are in agreement, and the quality and quantity of corresponding individual characteristics is such that the examiner would not expect to find that same combination of individual characteristics repeated in another source and has found insufficient disagreement of individual characteristics to conclude they originated from a different source."

Ibid.