

**A CHALLENGE TO THE ADMISSIBILITY OF FIREARMS AND TOOLMARK IDENTIFICATIONS: *AMICUS* BRIEF PREPARED ON BEHALF OF THE DEFENDANT IN UNITED STATES V. KAIN, CRIM. NO. 03-573-1 (E.D. PA. 2004)**<sup>1</sup>

INTRODUCTION

The following *amicus* brief was prepared in connection with a Daubert challenge in federal court to the admission of a firearms and toolmark examiner's testimony that cuts in a fence and grate were made by the defendant's bolt cutters, to the exclusion of all other bolt cutters in the world. Before the brief could be filed with the court, the government offered the defendant a plea bargain that was too good to refuse.

The expert testimony in this case was typical of that offered by firearms and toolmark examiners. The goal of the forensic science discipline of firearms and toolmark identification is to identify particular tools, such as a bolt cutter or the barrel of a particular gun, as the unique source of marks on crime scene evidence, such as a fence or a fired bullet.<sup>2</sup> Although numerous convictions are based on this type of testimony<sup>3</sup>, courts have yet to recognize that adequate statistical and empirical foundations have not been developed for these identifications. The following brief explains the systemic scientific problems that should make firearms and toolmark identifications inadmissible in court.<sup>4</sup>

*AMICUS* BRIEF PREPARED ON BEHALF OF THE DEFENDANT IN UNITED STATES V. KAIN, CRIM. NO. 03-573-1 (E.D. PA. 2004)

STATEMENT OF INTERESTS OF *AMICUS CURIAE*

*Amicus* Adina Schwartz, J.D., Ph.D. (Philosophy), is an Associate Professor in the Department of Law, Police Science and Criminal Justice Administration at John Jay College of Criminal Justice and in the Ph.D. Program in Criminal Justice of The Graduate School and University Center, City University of New York (CUNY). John Jay College is the only liberal arts college in the United States devoted to criminal justice, and the CUNY Criminal Justice Ph.D. Program

is the only Criminal Justice Ph.D. program in the country that has a forensic science track.

*Amicus* regularly teaches evidence law to undergraduates at John Jay College, and has twice taught a course, entitled “Science, Experts and Evidence in the Criminal Justice System,” in the CUNY Criminal Justice Ph.D. Program. Beginning in Spring 2005, she will be teaching evidence law in John Jay College’s newly created M.S. Program in Forensic Computing. As someone who teaches many current and future law enforcement agents and significant numbers of future forensic scientists, she submits this brief in the belief that high standards for the admission of scientific evidence are needed to motivate forensic scientists and law enforcement agents to do the careful scholarly and investigative work of which they are capable. As a scholar who writes on evidence law, forensic identification, and philosophy of science issues, she submits this brief in the belief that this case presents this Court with an important opportunity to apply the Daubert-Kumho standard to exclude unreliable forensic identification testimony.

I. THIS COURT’S RELIABILITY INQUIRY SHOULD FOCUS ON THE REASONING UNDERLYING THE EXPERT’S IDENTIFICATION OF THE BOLT CUTTERS ALLEGEDLY FOUND IN THE DEFENDANT’S CAR AS THE PAIR, TO THE EXCLUSION OF ALL OTHERS, THAT CUT THE GRATE, CHAIN LINK, AND PIECES OF CHAIN LINK FENCE.

A. The Need to Consider the Expert’s Specific Reasoning. Instead of mechanically applying the specific factors listed in Daubert, trial judges are to perform the Daubert-Kumho reliability inquiry by evaluating the reasoning underlying an expert’s testimony. *See, e.g., Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 593 (1993); Kumho Tire Company, Ltd. v. Carmichael, 526 U.S. 137, 151 (1999). As the Advisory Committee explained when Fed. R. Evid. 702 was amended in response to Daubert and its progeny, “The trial judge in all cases of proffered expert testimony must find that it is properly grounded, well-reasoned, and not speculative before it can be admitted. The expert’s testimony must be grounded in an accepted body of learning or experience in the expert’s field, and the expert must explain how the conclusion is so grounded.” (citation omitted). Advisory Committee’s Note to Amendment to Fed. R.Evid. 702 (2000).

*Kumho* explains that judicial gatekeeping must focus on the specific reasoning employed by the particular expert in a case, as opposed to the reasoning generally employed by experts in the field. 526 U.S. at 153-54. The focus must also be on the particular stage of the expert’s reasoning whose reliability is suspect, whether it be the proffered “testimony’s factual basis, data, principles, methods, or their

application [to reach specific conclusions].” *Id.* at 149, 154. See also Barry Scheck, DNA and Daubert, 15 Cardozo L.Rev. 1959, 1959 n.3 (1994) (warning that under Daubert, “judges have to resist the temptation to reach simplistic conclusions about ‘DNA testing’ in general and focus instead on the scientific merits of each application of DNA technology”).

There is no contest in this case that the bolt cutters allegedly found in the defendant’s car can properly be admitted to show that they *could* have made the cuts in the metal grate, chain link, and chain link fence. The issue before this Court is whether there is a reliable foundation for the prosecution expert’s conclusion that these particular bolt cutters, and no others, were the source of the cuts. See Ramirez v. State, 810 So.2d 836, 851-52 (Fla. 2001) (“Ramirez III”) (“We hold that while the knife that was recovered in Ramirez’s constructive possession may be admitted as conventional evidence of guilt, testimony based on [the prosecution expert’s] knife mark identification procedure ... is ... unreliable and inadmissible.”).

B. The Expert’s Testimony. The government’s firearms and toolmark expert<sup>5</sup>, Joseph J. Masson, used the bolt cutters allegedly found in the defendant’s car, introduced as Government Exhibit 31 (“G.E. 31”) and referred to in Mr. Masson’s Laboratory Report (Government Exhibit 15 (“G.E. 15”)) as Exhibit 19, to make test cuts in lead. A comparison microscope was then used to compare the test toolmarks in the lead with the evidence toolmarks found on the metal grate, chain link, and pieces of chain link fence, respectively referred to in the Laboratory Report as Exhibits 3, 4, and 5. Mr. Masson testified that “I make my identification on similarities, not dissimilarities.” Tr. 41. On the basis of his microscopic comparison of the similarities between the test and evidence toolmarks, Mr. Masson concluded that the G.E. 31 boltcutters were the pair, to the exclusion of all others in the world, that cut some of the ends of the grate in Exhibit 3, the two pieces of chain link in Exhibit 4, and “a number of representative samples of the pieces of chain link fence” in Exhibit 5. Tr.54, 57, 59, 60, 63-64; G.E. 15, pp.1-2.

According to Mr. Masson, the class characteristics of the G.E. 31 bolt cutters were “slight[ly] dissimilar” to those of the bolt cutters identified as G.E. 33 and referred to in the Laboratory Report as Exhibit 15. Tr. 48. Mr. Masson took test cuts from the G.E.33 bolt cutter, and eliminated that boltcutter as the source of the cuts “by testing it the same way [he] tested” the G.E. 31 bolt cutter. Tr. 49, 58. There were “[s]ufficient microscopic matching striations to identify it [the G.E. 31 bolt cutter] and to eliminate the other one.” Tr. 60.<sup>6</sup>

The Daubert-Kumho standard requires that “each stage of the expert’s testimony be reliable.” Heller v. Shaw Industries, Inc., 167 F.3<sup>rd</sup> 146, 155 (3<sup>rd</sup> Cir. 1999). See also In re Paoli R.R. Yard PCB Litig., 35 F.3d 717, 745 (3<sup>rd</sup> Cir. 1994)

(stating that “any step that renders the analysis unreliable ... renders the expert’s testimony inadmissible.”). Mr. Masson’s testimony should be excluded because of the unreliability of the procedures he employed for (1) making test toolmarks, (2) eliminating the G. E. 33 bolt cutters as the source of the evidence toolmarks, and (3) concluding that the similarities between the test and evidence toolmarks were so great that the G.E. 31 bolt cutters were the unique source of the evidence toolmarks, to the exclusion of all other bolt cutters in the world.

C. The Invalid Comparison of Toolmarks in Lead and in Harder Materials. In making firearms identifications, Mr. Masson compares test and evidence toolmarks on the same type of ammunition component (e.g., test fired and evidence bullets of the same caliber and make). Tr. 26. By contrast, his toolmark identifications are based on comparisons between test toolmarks and evidence toolmarks that have been made in different media. Tr. 30. Despite admitting that he is “not a metallurgist [and does] not know the consistency of metal and how it’s made and all” (Tr. 28), Mr. Masson testified that the rationale for the use of different media in toolmark (though not firearms) identification is that lead is more impressionable than the harder material in which evidence toolmarks are found. “Lead is a softer material and it leaves the tool marks from the blades, or screw driver or chisel, and it picks them up more distinctly, where a harder material would – might look a little differently.” Tr. 30.

This testimony implies that there will be differences in the toolmarks that the same tool leaves in lead and in harder material. Accordingly, it was incumbent on Mr. Masson to explain why comparing test toolmarks in lead with evidence toolmarks in harder material was not equivalent to comparing apples and oranges. See General Elec. Co. v. Joiner, 522 U.S. 136, 146 (1997) (explaining that an expert’s testimony may fail to pass the Daubert test if “there is simply too great an analytical gap between the data and the opinion proffered”). Since Mr. Masson did nothing to allay the doubts about the validity of comparing toolmarks in different materials that the firearms and toolmarks literature as well as common sense raise (consider, for example, the differences between the marks a knife makes in butter and in steak), the Daubert-Kumho standard precludes the admission of the conclusions that he based on such comparisons. See C. Champod, D. Baldwin, F. Taroni, and J.S. Buckleton, Firearms and Toolmark Identification: The Bayesian Approach, 35(3) AFTE J. 307, 314 (Summer 2003) (stating that because lead rod “is far too good a medium,” it was a serious error for a study (Shirley J. Butcher & P.D. Pugh, A Study of Marks Made by Bolt Cutters, 15 J. Forens. Sci. Soc. 115 (1975)) to use data about bolt cutter marks in lead rod to draw conclusions about “the more common marks [that bolt cutters make] in hardened steel”); J. Hall, Consecutive Cuts by Bolt Cutters and Their Effect on Identification, 24(3) AFTE J. 260 (July 1992) (successive marks that individual boltcutters cut into lead were more similar to each other than successive marks that the same bolt cutters cut into shackles comprised of harder materials).

D. The Questionable Procedure for Excluding the G.E. 33 Bolt Cutters. Mr. Masson's testimony that the G.E.33 and G.E.31 bolt cutters had different class characteristics conflicts with his testimony that he made test cuts with both sets of bolt cutters and used microscopic comparisons to exclude G.E.33, but identify G.E. 31 as the unique source of the evidence toolmarks. Tr. 48-49, 58, 60. A tool can be the source of an evidence toolmark only if the class characteristics of the tool and the evidence toolmark match. See, e.g., Alfred Biasotti & John Murdock, *The Scientific Basis of Firearms and Toolmark Identification* ("The Scientific Basis") in 3 DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE 502 (2002); Bruce Moran, Firearms Examiner Expert Witness Testimony, 32(3) AFTE J. 231, 237-39 (Summer 2000). Hence, unless they first find that the class characteristics of a suspect tool and an evidence toolmark agree, firearms and toolmark examiners neither make test toolmarks with the suspect tool nor compare such test marks with evidence marks under a comparison microscope. See Biasotti & Murdock, *supra*; Moran, *supra*, at 239 ("The firearms examiner relies on the evaluation of these [microscopic, individualized] markings to distinguish a barrel as having fired a bullet to the exclusion of all other barrels *with the same rifling class characteristics.*" (emphasis added)); Jerry Miller, An Introduction to the Forensic Examination of Toolmarks, 33(3) AFTE J. 233, 241 (Summer 2001) (stating that class characteristics "can be used ... to eliminate a tool from having been used.").

Mr. Masson's identification of G.E.31 as the source of the evidence toolmarks was justified only if its class characteristics matched those of the evidence toolmarks. If, as Mr. Masson testified, the class characteristics of the G.E. 33 and G.E.31 bolt cutters differed, G.E. 33 should have been excluded on the basis of class characteristics alone. In accord with the basic principles of toolmark and firearms identifications, Mr. Masson had no reason to make test cuts and microscopic comparisons with both bolt cutters unless, contrary to his testimony, the class characteristics of both were identical and matched the evidence toolmarks. Since, as the Advisory Committee has explained, the Daubert-Kumho standard requires that expert testimony be well-reasoned, Mr. Masson's self-contradictory account of his procedure for excluding the G.E. 33 bolt cutters precludes the admission of his testimony. Advisory Committee's Note to Amendment to Fed. R.Evid. 702 (2000).

E. Mr. Masson's Failure to Base His Identity Conclusions on Objective Criteria.

The preceding doubts about Mr. Masson's reasoning pale beside the more fundamental issue of whether he had a reliable basis for concluding that the similarities between the test and evidence toolmarks were so great that the G.E. 31 bolt cutters must be the source of the evidence toolmarks, to the exclusion of all other bolt cutters in the world. In his testimony, Mr. Masson refused to articulate any objective criteria for how many or what kinds of striae must match in order to determine that two toolmarks must have been made by the same tool. He testified that his subjective judgment was his sole basis for concluding that

the resemblances between the test and evidence toolmarks were so great that the G.E. 31 bolt cutters, to the exclusion of all others in the world, must have made the cuts in the grate, chain link, and pieces of chain link fence. Tr. 68. The basis for his identifications was “pattern recognition, training and experience that comes up with this.” Tr. 91.

Like Mr. Masson, many, though by no means all, toolmark examiners do not rely on any objective criteria as to how many and what kinds of matches between striae are necessary to justify identity conclusions. Instead, they make purely subjective identity determinations, and claim that their identifications are correct because of their experience and training. While acknowledging that “it is something of a stereotype to visualize the distinguished, greying individual on the stand saying, ‘my opinion is based on my many years of experience in the field,’” prominent forensic scientists Christophe Champod and Ian W. Evett deplore this practice on the ground that it conflicts with the basic scientific value of transparency. A Probabilistic Approach to Fingerprint Identification, 51 J. Forens. Identification 101, 106-107 (2001). “[A]s a matter of principle, ... the scientist should, as far as possible, support his/her opinion by reference to logical reasoning and an established corpus of scientific knowledge.” *Id.*

As Champod and Evett recognize, the value of transparency is also implicit in the Daubert-Kumho reliability inquiry. *Id.* To avoid the well-known danger that juries will be awed by expert testimony, Federal Rule of Evidence 702 conditions admission on a trial court’s determination that such testimony “will assist the trier of fact to understand the evidence or determine a fact in issue.” *See also Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. at 595 (“Judge Weinstein has explained: ‘Expert evidence can be both powerful and quite misleading because of the difficulty in evaluating it.’” (citation omitted)); Adina Schwartz, A “Dogma of Empiricism” Revisited: Daubert v. Merrell Dow Pharmaceuticals, Inc. and the Need to Resurrect the Philosophical Insight of Frye v. United States,” 10 Harv. J. L. & Tech. 149, 196-98 (1997) (“A ‘Dogma of Empiricism’”) (explaining the relations between the Frye and Daubert standards and the fear that jurors will be awed by scientific expert testimony). By insisting that his identity conclusions were based solely on his own subjective judgments, Mr. Masson in effect refused to explain how he knew that the resemblances between the test toolmarks made with the G.E.31 bolt cutters and the toolmarks cut into the grate, chain link fence and chain link were so great that no other bolt cutter in the world could possibly have made the cuts. Since the value of transparency is basic to science and to the *Daubert* inquiry, Mr. Masson’s testimony should be excluded on the ground that it is obscure rather than transparent. *See Champod & Evett, supra*, at 107. The proffered testimony will do nothing to help the jury understand whether the G.E. 31 bolt cutters are or are not the only possible source of the cuts on the grate, chain link fence and chain link.

A preference for transparency over obscurity is also implicit in the distinction that Daubert-Kumho draws between the reliability of an expert's testimony and the expert's personal qualifications. By refusing to articulate any criteria for when the resemblances between toolmarks are so great that they must have been made by the same tool, Mr. Masson implied that the jury should accept his identification of G.E.31 because his experience and training make him capable of correctly (if ineffably) judging when the resemblances between toolmarks are sufficient to justify an identification. The Advisory Committee has recognized that the Daubert-Kumho inquiry would be wrongly reduced into an inquiry into experts' qualifications if experts could thus substitute invocations of their experience for explanations of the basis for their conclusions. "If the witness is relying solely or primarily on experience, then the witness must explain how that experience leads to the conclusion reached, why that experience is a sufficient basis for the opinion .... The trial court's gatekeeping function requires more than simply 'taking the expert's word for it'." Advisory Committee's Note to Amendment to Fed. R.Evid. 702 (2000) (citation omitted). In accord with this, Her Honor stated, during the hearing, that even though she had no doubts as to Mr. Masson's personal qualifications, this did not dispose of the question of whether firearms and toolmark identification is a reliable discipline. Tr. 98.

*See also* Joiner, 522 U.S. at 146 (reasoning that "nothing in ... Daubert ... requires a district court to admit opinion evidence which is connected to existing data only by the *ipse dixit* of the expert"); Daubert v. Merrell Dow Pharmaceuticals, Inc., 43 F.3d 1311, 1316 (9<sup>th</sup> Cir. 1995) ("Daubert II") (reasoning that the point of the Daubert standard is lost if "an expert's self-serving assertion that his conclusions were 'derived by the scientific method' [is] deemed conclusive"); Ambrosini v. Labarraque, 101 F.3d 129, 143 (D.C. Cir. 1996) (Henderson, J., dissenting), *cert. denied*, 520 U.S. 1205 (1997) (warning that "if such conclusory statements [as the expert's statement that he employed "the traditional methodology of experts in the field"] must be accepted at face value, ... the Daubert standard becomes meaningless"). Ignoring this body of law, the government's direct examination of Mr. Masson during the hearing and the brief that the government filed before the hearing both dwelt on Mr. Masson's credentials, but made only the slightest attempt to explain the scientific principles underlying firearms and toolmark identification.

The Florida Supreme Court's recent decision in Ramirez III, *supra*, also argues against admitting the identification testimony in this case. In Ramirez III, as in this case, the toolmarks at issue were striated toolmarks. Similarly to Mr. Masson, the prosecution experts in Ramirez III testified that a comparison of the striae on the cast of Ramirez's knife with the striae on the cast of the victim's cartilage enabled them to identify the knife as the murder weapon, to the exclusion of all others, even though they employed no objective criteria for how many or what kinds of striae must match in order to establish identity. The Ramirez III experts averred that an individual toolmark examiner's subjective judgment, gained through experience and

training, suffices for determining whether toolmarks are so similar that they must have come from the same tool.

As defense counsel explained during the hearing, the relevance of Ramirez III to this Court's reliability inquiry is not diminished by the fact that Florida is a Frye state. While ostensibly adhering to Frye in excluding the toolmark experts' testimony (810 So.2d at 843 & 843 n.8), the Ramirez III Court in fact used the factors specifically listed in *Daubert* as surrogates for the Frye general acceptance test. 810 So.2d at 849-51. *See also, The Judicial Response to Firearms and Toolmark Identification Expert Evidence*, in 3 FAIGMAN ET AL., *supra*, at 489 n.30 (2002); David W. Barnes, General Acceptance Versus Scientific Soundness, 31 Fla. St. U. L. Rev. 303, 305 (Winter 2004) (stating that in Ramirez III, "the Florida Supreme Court simultaneously rejected the federal rule and elaborated an approach remarkably similar to that rule, requiring judges to evaluate the scientific basis for novel expert testimony").

As will be argued in Section IV below, *amicus* agrees with the Florida Supreme Court that the specific Daubert factors preclude the admission of toolmark identifications that are not based on objective criteria, such as those proffered by Mr. Masson and by the prosecution experts in Ramirez III. Before considering the specific Daubert factors, however, *amicus* wishes to inform this Court of the basic principles and pitfalls of toolmark and firearms identification and of the cogent arguments that prominent toolmark and firearms examiners have advanced, since the 1930's, to show that objective, statistically-based identification criteria are needed. It is hoped that this will aid this Court to reach a scientifically informed decision on the admissibility of the toolmark expert's testimony in this case. *See Joiner*, 522 U.S. at 147-48 (Breyer, J., concurring) (stating that Daubert's gatekeeping "requirement will sometimes ask judges to make subtle and sophisticated determinations about scientific methodology and its relation to the conclusions an expert witness seeks to offer--particularly when a case arises in an area where the science itself is tentative or uncertain....").

## II. AN UNDERSTANDING OF THE BASIC CHARACTERISTICS OF TOOLMARKS IMPLIES THAT OBJECTIVE, STATISTICALLY-BASED IDENTIFICATION CRITERIA ARE NEEDED.

A. The Distinctions between Class, Subclass and Individual Characteristics of Toolmarks. The distinctions between class, subclass and individual characteristics of toolmarks must be grasped in order to understand the problems with the conclusions of identity that firearms and toolmark examiners draw. As a result of the distinctive designed features of different types of tools, different types of toolmarks, or, in other words, toolmarks with different class characteristics, result when different types of tools are used or applied to materials. For example, the intentionally manufactured differences between

steak and butter knife blades result in different types of marks when the two types of knives are inserted in butter. See Biasotti & Murdock, *The Scientific Basis*, *supra*, at 496 n.3.

Mr. Masson failed to inform this Court that manufacturing processes may also produce subclasses within a type of tool. The tools in each subclass share similarities in appearance, size, or surface finish that are not shared by other tools of the same type. The toolmarks produced by tools of a particular subclass have similarities, or subclass characteristics, that distinguish them from the toolmarks produced by other tools of the type. For example, a study found subclass characteristics among the toolmarks produced by the ram of one, but not another, brand of desk stapler. *Id.* at 500-501; John E. Murdock, The Individuality of Toolmarks Produced by Desk Staplers (“Desk Staplers”), 6 AFTE J. 23 (1974).

The forensic science discipline of toolmark identification is premised on the existence of individual characteristics that, by contrast to class and subclass characteristics, are unique to the toolmarks each individual tool produces. The individual characteristics of a toolmark correspond to random imperfections or irregularities on tool surfaces produced by the manufacturing process and/or subsequent use, corrosion or damage. If the same class characteristics are found on evidence and test toolmarks (for example, the same rifling impressions on a bullet test fired by a gun barrel and an evidence bullet recovered from a crime scene), a toolmark examiner uses a comparison microscope to compare the toolmarks’ individual characteristics (for example, microscopic striations within rifling impressions). The object is to determine whether the individual characteristics are so similar that one and the same tool (for example, a particular gun barrel) must have produced both the test and the evidence toolmarks.

Contrary to Mr. Masson’s testimony that he had never “seen or heard of two different tools creating the same exact tool markings” (Tr. 33, 38-39, 73), a substantial literature argues that only some manufacturing processes make each tool capable of producing toolmarks with individual characteristics from the moment of manufacture. Other manufacturing processes result in batches of tools that are so similar that their toolmarks have the same subclass characteristics, and may or may not also have individual characteristics. See, e.g., Murdock, Desk Staplers, *supra*; Ronald G. Nichols, Firearm and Toolmark Identification Criteria, 42 J. Forensic Sci. 466, 470 (1997) (“Nichols I”); Biasotti & Murdock, *The Scientific Basis*, *supra*, at 500-501; Alfred A. Biasotti & John Murdock, “Criteria for Identification” or “State of the Art” of Firearm and Toolmark Identification, 16 AFTE J. 16, 17 (1984) (“Criteria for Identification”); Bruce Moran, A Report on the AFTE Theory of Identification and Range of Conclusions for Tool Mark Identification and Resulting Approaches to Casework, 34 (2) AFTE

J. 227, 227-28 (Spring 2002) ("A Report"); Kristen A. Tomasetti, Analysis of the Essential Aspects of Striated Toolmark Examination and the Methods for Identification, 34(3) AFTE J. 289, 295 (Summer 2002).

According to the scientific literature, the tools in the uniform batches produce toolmarks with individual characteristics only as they are used, damaged, or corroded. Biasotti & Murdock, *The Scientific Basis, supra*, at 500-501. Even if a tool is capable of producing unique toolmarks from the time of manufacture, the individual characteristics of its toolmarks will change as the tool is used or as damage or corrosion occur. 1 PAUL C. GIANNELLI & EDWARD J. MWINKELRIED, SCIENTIFIC EVIDENCE 633 (3<sup>rd</sup> ed. 1999) (citing Flynn, Tool Mark Identification, 2 J. Forensic Sci. 95, 102 (1957) for the proposition that "the characteristics of a tool will change with use").

B. Central Pitfalls in Toolmark Identification. The foregoing analysis of the distinctions between class, subclass and individual characteristics of toolmarks makes it possible to appreciate three central pitfalls that stand in the way of reliably identifying *one and only one* tool as the source of a particular toolmark(s). Due to their recognition of these pitfalls, many prominent toolmark examiners do not share the prosecution expert's complacency about relying on subjective judgments to make unique identifications.

1. The Individual Characteristics of Toolmarks Are Combinations of Non-Unique Marks. A first barrier in the way of reliably identifying *the* source of an evidence toolmark is that, just as, notwithstanding their uniqueness, parts of each individual's fingerprints and nuclear DNA are the same as other people's, the individual characteristics of toolmarks are comprised of non-unique marks. In 1935, Gunther and Gunther used the analogy of oak leaves to illustrate this point. "No two oak leaves may be exactly alike, but the exact counterpart of a small area of leaf can probably be found in other leaves .... It is probably true that no two firearms with the same class characteristics will produce the same signature, but it is likewise true that each element of a firearm's signature may be found in the signatures of other firearms .... An individual peculiarity of a firearm can, therefore, be established by elements of identity which form a *combination* the coexistence of which is highly improbable in the signature of other firearms with the same class characteristics." JACK D. GUNTHER & C.O. GUNTHER, THE IDENTIFICATION OF FIREARMS 90-91 (1935). *See also* Biasotti and Murdock, "Criteria for Identification", *supra*, at 17 (using the passage from GUNTHER & GUNTHER to explain why toolmark examiners "have come to expect to find small isolated areas of corresponding striae agreement when comparing toolmarks known to have been produced by different working surfaces.") .

Empirical work has shown that a substantial percentage of the striae comprising the individual characteristic of one toolmark can match the striae comprising the individual characteristic of another toolmark.<sup>7</sup> In assessing the expert testimony in this case, this Court should note that up to 29% of the striae were found to match on toolmarks that were made by different bolt cutters of the Record brand 930 centre-cut type. Butcher & Pugh, *supra*, at 122-23. Similarly, in 1942, Burd and Kirk found that up to 25% of the striae matched in comparisons of marks known to have been made by different tools. D.Q. Burd & P.L. Kirk, Tool Marks— Factors Involved in Their Comparison and Use as Evidence, 32 J. Crim. L., Criminology & Police Sci. 679 (1942). See also Eliot Springer, Toolmark Examinations, 40 J. Forensic Sci. 964, 965 (1995) (describing Burd and Kirk’s “important article”); Nichols I, *supra*, at 470 (describing Burd and Kirk’s “often cited study”).

Likewise, in 1955, Biasotti found that 15 to 20% of the striae on bullets fired from different .38 Special Smith & Wesson revolvers (i.e., known non-matches) matched. A.A. Biasotti, Bullet Comparison, A Study of Fired Bullets Statistically Analyzed ( Unpublished Thesis, University of California, Berkeley 1955); A.A. Biasotti, A Statistical Study of the Individual Characteristics of Fired Bullets, 4 J. Forensic Sci. 34 (1959) (summary of his 1955 thesis). See also Springer, *supra*, at 965 (describing the Biasotti study); Biasotti, Principles of Evidence Evaluation, *supra*, at 431 (explaining that his study’s results “corresponded well” to Burd & Kirk’s results). In 1997, Nichols claimed that “[t]o date, [the Biasotti study] stands as the most exhaustive statistical empirical study ever published.” Nichols I, *supra*, at 467.

In the 1990’s, the development of the BATF’s computerized comparison system, IBIS (Integrated Ballistics Information System), enabled investigators to compare the tool marks on vast numbers of bullets and cartridge cases. See Schwartz, *Ballistics*, *supra*, at 7-9. Studies using the IBIS data base support the claim that there can be significant numbers and percentages of matching striae on pairs of bullets fired from different guns. See, e.g., Jerry Miller & Michael McLean, Criteria for Identification of Toolmarks, 30 (1) AFTE J. 15 (Winter 1998); Jerry Miller, Criteria for Identification of Toolmarks Part II, 32 (2) AFTE J. 116 (2000) (“Miller II”).

Although Mr. Masson was unable to recall the paper on direct examination (Tr. 23), in 1997, he published a study that strongly suggests that toolmarks made by different tools may be much more similar to each other than firearms and toolmark examiners currently believe they can be. The study found that as the IBIS data base grew for guns of a particular caliber, increasing similarities were discovered in the individual characteristics of the toolmarks on ammunition components known to have been fired by different guns of that caliber. According to Mr. Masson, “a number of known non-matched testfires from different firearms ... were coming up near the top of the candidate list [for matches with the toolmarks on evidence

ammunition components.] When retrieving these known non-matches on the comparison screen, there were numerous two dimensional similarities. When using a comparison microscope, these similarities are still present and it is difficult to eliminate comparisons even though we know they are from different firearms.” Joseph J. Masson, Confidence Level Variations in Firearms Identification through Computerized Technology, 29 (1) AFTE J. 42 (1997) (DD-1).

Masson urged examiners to avoid misidentifications by using the IBIS database to increase their knowledge of the possible extent of the similarities between non-matching toolmarks. “In the past, best examples of known non-matched agreements were collected from casework and thus, surfaced sporadically. Firearms examiners should take advantage of this current expanded database to fully familiarize themselves with the extent of similarities found in many non-identifications in order to hone their criteria for striae identification.” *Id.* at 43. However, as Mr. Masson acknowledged on cross-examination, there are no databases for bolt cutter toolmarks or toolmarks made by any other type of tool besides firearms. Tr. 67-68. By implying that computerized databases were needed to reveal the extensiveness of the possible similarities between toolmarks made by different firearms, Mr. Masson’s study strongly suggests that, in the absence of computerized databases, toolmark examiners are likely to underestimate the extent of the possible similarities between toolmarks made by different tools of the same type, including bolt cutters. Accordingly, Mr. Masson and other toolmark examiners risk making misidentifications when they base their identity conclusions on their subjective sense, unaided by the use of computerized databases, of how similar two toolmarks can be and yet come from different tools of the same type.

2. The Danger of Confusing Subclass with Individual Characteristics. A difference between fingerprint, nuclear DNA and mitochondrial DNA (mtDNA) identification, on the one hand, and firearms and toolmark identification, on the other, makes firearms and toolmark identification especially difficult. On the one hand, each individual’s fingerprints are unique. With the sole exception of identical twins, the same is true of each individual’s nuclear DNA sequence. Since, by contrast with the nuclear DNA that one inherits both parents, mtDNA is, in theory, inherited only from one’s mother, even the most remote maternal cousins should share the same mtDNA. See Adina Schwartz, Book Review, 3 Punishment and Society 446, 447 (2001) (reviewing BARRY SCHECK, PETER NEUFELD AND JIM DWYER, ACTUAL INNOCENCE: FIVE DAYS TO EXECUTION AND OTHER DISPATCHES FROM THE WRONGLY CONVICTED (2000)) (“Book Review”).

By contrast to these well-established generalizations about the uniqueness of fingerprints and nuclear DNA and the sharing of mtDNA sequences in people descended from the same maternal line, we saw above that only some

manufacturing processes produce individual tools whose surfaces are differentiated enough to produce toolmarks with different individual characteristics. Other manufacturing processes result in batches of tools so similar that their toolmarks have the same subclass characteristics, and may or may not also have individual characteristics. Compounding the absence of any straightforward rule, wear and tear on some tools will cause the subclass characteristics on their toolmarks to be completely replaced by individual characteristics. In other tools, subclass characteristics may persist alongside individual characteristics. See Schwartz, *Ballistics, supra*, at 3.

By failing even to recognize the existence of subclass characteristics, Mr. Masson ignored a major difficulty that bedevils firearms and toolmark identification, and is not analogous to any difficulty scientists face in making fingerprint, nuclear DNA or mtDNA identifications. A particular tool may be wrongly identified as the source of an evidence toolmark if an examiner wrongly concludes that subclass characteristics on test and evidence toolmark(s) are individual characteristics.

This confusion is possible because there are no rules for distinguishing subclass from individual characteristics. To avoid confusing subclass characteristic shared by more than one tool with individual characteristics unique to one and only one tool, examiners can only rely on their personal familiarity with types of forming and finishing processes and their reflections in toolmarks. In accord with this, Biasotti and Murdock explain that “some machining processes are capable of reproducing remarkably similar surface characteristics (i.e., gross contour and/or fine striae, etc.) on the working surfaces of many consecutively produced tools which if not recognized and properly evaluated could lead to a false identification.” “Criteria for Identification”, *supra*, at 17. They go on to warn that “[t]he examiner must ... be familiar with the various forming and finishing processes in order to distinguish those ... surface characteristics that are truly individual from those surface characteristics that may characterize more than one tool.” *Id.* See also Nichols I, *supra*, at 470-72.

In ignoring the possibility of misidentifications resulting from the confusion of subclass with individual characteristics, Mr. Masson failed to inform this Court of a danger that is real, not theoretical. In the 1980's, this type of confusion was discovered to have in fact resulted in misidentifications of striated toolmarks. In response, members of the Association of Firearms and Toolmark Examiners (“AFTE”) formed the Criteria for Identification Committee. The term “subclass characteristics” was coined in 1989 and incorporated in the AFTE glossary definitions in 1992. See Moran, A Report, *supra*, at 227-28 (relating this history and warning that “[c]aution should be exercised in distinguishing subclass characteristics from individual characteristics”). See also Jerry Miller, *An Examination of Two Consecutively Rifled Barrels and a Review of the Literature*,

32(3) AFTE J. 259, 260 (Summer 2000) (describing a study that found that the matching characteristics on the toolmarks of bullets fired from ten consecutively manufactured gang broach barrels were so great that a false identification would have resulted if the matching characteristics had been incorrectly identified as individual, rather than subclass characteristics).

It is particularly relevant to this case that prominent firearms and toolmark examiner John Murdock has claimed that a bolt cutter's tendency to produce toolmarks with individual or subclass characteristics can be expected to vary with the extent to which it has been used. John E. Murdock, Some Suggested Court Questions to Test Criteria for Identification Qualifications, 24(1) AFTE J. 69, 73 (January 1992) ("Court Questions"). Although Hall, *supra*, at 264, states, to the contrary, that "no two bolt cutters manufactured will leave identical marks," his statement is not supported by the two studies he cites.

One of the studies – Billy Hornsby, MCC Bolt Cutters, 21(3) AFTE J. 508 (1989) – is a one-page report on a visit to a bolt cutter manufacturing facility, the Matsuzaka Casting Company (MCC), in Tsu-shi, Mie-Ken, Japan. Hornsby reported that, "Since I was unable to obtain consecutively made bolt cutters, I obtained tests from three pairs of bolt cutters that were made during the same production run. Intercomparisons of these tests disclosed individual characteristics so different that there would be no possibility of misidentification." The only other basis for Hall's denial that *any* bolt cutters can produce toolmarks with subclass, but not individual, characteristics, is the study by Butcher and Pugh, criticized in Section I C above.

In assessing the support that these studies provide for Hall's conclusion, this Court should take account of both the extremely small sample in the Hornsby study (only three bolt cutters) and Butcher and Pugh's problematic procedure of using data on toolmarks in lead to reach conclusions about the marks bolt cutters will make in harder material. See Champod, Baldwin, Taroni, and Buckleton, *supra*, at 314. These problems pale beside the fact that Hornsby only studied MCC bolt cutters and Butcher and Pugh only studied "Record" brand cutters of the 930 centre-cut type manufactured by C & J. Hampton, Ltd. in Sheffield, England.

Murdock's well-regarded study of desk staplers shows that Hall's extrapolation is far too broad. Hornsby's and Butcher and Pugh's studies of *particular brands and types of bolt cutters* cannot provide an adequate foundation for Hall's conclusion that *no two bolt cutters of any brand or type* will ever produce exactly the same toolmarks. Murdock found that when newly manufactured, the surfaces of the rams of one brand of desk stapler were so similar that they produced toolmarks with subclass, but not individual, toolmarks. By contrast, the

process used to manufacture another brand of desk stapler resulted in rams whose unique working surfaces made them capable of leaving toolmarks with unique characteristics. Murdock, *supra*; Biasotti & Murdock, *The Scientific Basis, supra*, at 501 (describing Murdock's study); CRIME LABORATORY MANAGEMENT FORUM 177-78 (R.H. Fox & F.H. Wynbrandt eds. 1976) (favorably evaluating the Murdock study).

Mr. Masson's testimony about different types of bolt cutters makes the results of Murdock's desk stapler study particularly relevant to this case. According to Mr. Masson, the G.E.31 and 33 bolt cutters are "not your standard Stanley bolt cutter, which would be a high class, well made bolt cutter – if you got a cheaper bolt cutter, they're not made to take much abuse and they will – the blades will dull and chip almost immediately after using them." Tr. 35. Together with Murdock's finding of subclass characteristics in the toolmarks produced by some, but not other, brands of stapler rams, Mr. Masson's acknowledgment of major differences among types of bolt cutters shows that even if some bolt cutters produce toolmarks with individual, but not subclass characteristics, this need not be true of all bolt cutters. Indeed, the possibility of subclass characteristics would appear to be particularly high in this case because inexpensive manufacturing methods were presumably used to create the "cheaper bolt cutter[s]" in this case. An article introduced by the government for the March 12 hearing warns that misidentifications of both firearms and other tools may result if subclass characteristics are confused with individual characteristics, and goes on to explain: "As tool manufacturers minimize the steps necessary to produce tools in an effort to become more efficient and economical, the possibility for tools produced with similar characteristics increases." Stephanie J. Eckerman, A Study of Consecutively Manufactured Chisels, 34 (4) AFTE J. 379 (Fall 2002).

For all these reasons, Mr. Masson's identification of the G.E. 31 bolt cutter was unreliable because he failed to rule out the possibility of subclass characteristics. To avoid misidentifications, "[t]he examiner must ..., for any specific tool, be able to: (1) recognize the presence of subclass characteristics and (2) properly evaluate the significance of subclass toolmarks when they are present by determining whether or not they are influencing the nature of any evidence." Biasotti & Murdock, *The Scientific Basis, supra*, at 501.

3. The Individual Characteristics of Toolmarks Change with Time. Firearms and toolmark identification is also difficult because, by contrast to an individual's fingerprints and nuclear DNA, the individual characteristics of the marks made by a particular tool change with time. Studies of the statistical foundations of fingerprint and mtDNA identification bear out the claim that temporal changes in the characteristics of individual tools are a major barrier to developing a reliable method of firearms and toolmark identification. According to prominent statistician Stephen Stigler, "it was only in 1890-95 with the work of Francis Galton that the use

of fingerprints acquired a scientific basis.” Stephen Stigler, Galton and Identification by Fingerprints, 140 *Genetics* 857 (1995). Stigler praises Galton for recognizing that proving that “[a]n individual’s prints [are] persistent over time” was a crucial step in establishing that a single individual can be reliably identified as the source of a particular fingerprint(s). *Id.* (italics omitted). See also Paul C. Giannelli, The DNA Story: An Alternative View (Book Review), 88 *J. Crim. L. & Criminology* 380, 395 (1998) (stating that the fact that fingerprints do not change over time is one reason why “[f]ingerprints are considered the most reliable type of scientific evidence”). By contrast, heteroplasmy, the existence of more than one mtDNA type in a single human being over the course of his or her lifetime, is a major problem in mtDNA identification. See Schwartz, Book Review, *supra*, at 447.

As seen above in Section A, changes in toolmarks occur because the surfaces of a tool change as the tool is used, and/or as damage or corrosion occur. Giannelli and Imwinkelried state that “if the barrel of the firearm has changed significantly, due to erosion or corrosion, a positive identification may be impossible.” *Supra*, at 613. They conclude that toolmark identification “has the same limitations as firearms identification: ‘The characteristics of a tool will change with use.’” *Id.* at 633 (quoting Flynn, *supra*, at 102). Similarly, Mr. Masson agreed that “each time a tool is used, the individual characteristics of that tool may be altered.” Tr. 74.

Biasotti’s well-regarded statistical empirical study, discussed in subsection 1 above, reveals the significant problems that temporal changes in the surfaces of tools and their associated toolmarks create for toolmark identification. Biasotti found that “bullets from the same gun (i.e., known matches) gave only 21 to 38% matching lines [i.e., striae].” Biasotti & Murdock, Criteria for Identification, *supra*, at 20. As Springer stated, this result implies that even between toolmarks created by the same tool, “there is no such thing as a perfect match!” *Supra*, at 965. In a 1997 review of the toolmark and firearms literature, Nichols claimed that this surprising result had held up over time. According to Nichols, one of the results of Biasotti’s study, which is “not particularly news to us now,” is “[t]hat the average percentage of matching lines in jacketed bullets fired from the same gun was 21-24%.” Nichols I, *supra*, at 467.

Similarly, Hall’s 1992 study of bolt cutters was premised on the fact that bolt cutters produce toolmarks whose individual characteristics change over time. Hall reasoned that: “It is known that the condition of the cutting edges of the bolt cutters will change over time during the use of the bolt cutters. The question which arises, however, is how long is it before the individual characteristics have changed sufficiently to prevent a positive identification?” Hall, *supra*, at 261.

In assessing Mr. Masson’s reasoning, it is crucial to note that his testimony that the G.E. 31 and 33 bolt cutters were “cheaper bolt cutters ... whose blades will dull and

chip almost immediately after using them” (Tr. 35), implies that the toolmarks the bolt cutters could produce were likely to have changed between the time the evidence and test toolmarks were made. In turn, this implies that to justify the identification of G.E. 31 as the unique source of the evidence toolmarks, Mr. Masson needed to establish that the differences between the test and evidence toolmarks were small enough to be explained by changes in G.E. 31. The identification would be mistaken if the differences between the test and evidence toolmarks were instead so great that they could only have been made by two different bolt cutters. By testifying that “I make my identification on similarities, not dissimilarities” (Tr. 41), Mr. Masson implied that he had not even considered this issue. Therefore, this Court should exclude his testimony on the ground that “there is simply too great an analytical gap between the data [about the test and evidence toolmarks in this case] and the opinion proffered” identifying G.E.31 as the unique source of the evidence toolmarks. See Joiner, 522 U.S. at 146.

C. The Statistical Nature of Identity Determinations. In sum, substantial resemblances between toolmarks produced by different tools may result from shared subclass characteristics or from similarities between the striae comprising the individual characteristics of the toolmarks. At the same time, because the surfaces of tools change over time, even toolmarks made by the same tool do not perfectly match. Springer, *supra*, at 965. The similarities between toolmarks made by different tools and the differences between toolmarks made by the same tool imply that a statistical question must be answered to determine whether a particular tool was the source of the toolmark on an object recovered from a crime scene. What is the likelihood that the toolmarks made by a randomly selected tool of a given type would do as good a job as the toolmarks made by the suspected tool at matching the characteristics of the questioned toolmark? See Biasotti, The Principles of Evidence Evaluation, *supra*, at 429-30; GUNTHER & GUNTHER, *supra*, at 90-91 (“An individual peculiarity of a firearm can, therefore, be established by elements of identity which form a *combination* the coexistence of which is highly improbable in the signature of other firearms with the same class characteristics.”); Biasotti & Murdock, “Criteria for Identification”, *supra*, at 21 (arguing that “conclusions of identity in firearms and toolmarks ... mean that there is no credible possibility that a gun barrel or tool other than the one identified was used to produce the toolmark in question”). Cf. Brim v. State, 695 So.2d 268, 269 (Fla. 1997) (“Brim I”), 695 So.2d 268, 269 (Fla. 1997) (explaining that “the results obtained through this first step in the DNA testing process simply indicate that two DNA samples look the same. A second statistical step is needed to give significance to a match.”); Murray v. State, 692 So.2d157, 162 (Fla. 1997) (“The fact that a match is found in the first step of the DNA testing process may be ‘meaningless’ without qualitative or quantitative estimates demonstrating the significance of the match.”).

At the hearing in this case, the Court was not informed of a major division among toolmark and firearms examiners. Mr. Masson misleadingly suggested that all

examiners resemble him in relying solely on subjective judgments of when the similarities between the striae of test and evidence toolmarks are so great that they must have been made by the same tool. To the contrary, significant numbers of examiners base their identity conclusions on the objective CMS (consecutive matching striae) criterion that Biasotti and Murdock propounded in 1997 and developed on the basis of statistical empirical studies. See, e.g., Biasotti & Murdock, *The Scientific Basis*, *supra*, at 511-16; Champod, Baldwin, Taroni, & Buckleton, *supra*, at 310-11.

The crucial difference, here, is that when toolmark examiners – such as the expert in this case – insist on relying on inarticulable, mind’s eye criteria to reach conclusions of identity, they evade the task of providing the requisite statistical and empirical foundations for identity claims. In following a subjective approach, examiners implicitly admit that “we lack necessary statistical data which would permit us to formulate precise criteria for distinguishing between identity and nonidentity with a reasonable degree of certainty.” Biasotti, *The Principles of Evidence Evaluation*, *supra*, at 430. By contrast, as even critics admit, the CMS approach is a serious attempt to solve the problem of defining the amounts and types of resemblance between striae necessary to create a vanishingly small probability that the same tool did not produce the evidence and test toolmarks in a case. See, e.g., Champod, Baldwin, Taroni, & Buckleton, *supra*, at 311-15; Stephen G. Bunch, Ph.D., *Consecutive Matching Striation Criteria: A General Critique*, 45 (5) *J. Forens. Sci.* 955, 957-62 (2000).

This contrast between the CMS and the traditional, subjective approach is obscured by the fact that, in accord with the AFTE Range of Conclusions, all firearms and toolmark examiners in the United States testify to only four conclusions. As defense counsel pointed out in the hearing in this case, the only options for examiners are (1) identifying or (2) eliminating a particular tool as *the* source of the mark(s) found on an object, (3) concluding that the comparison of test and evidence toolmarks is inconclusive, or (4) concluding that the evidence toolmark is unsuitable for comparison. Tr. 78-79. See, e.g., Moran, *A Report*, *supra*, at 228-29; Biasotti & Murdock, *The Scientific Basis*, *supra*, at 506-507. This range of conclusions is misleading because it is never possible to know, as the expert in this case claims he does, that a given tool is the source of a particular toolmark, to the exclusion of all other tools in the world. See Champod, Baldwin, Taroni, & Buckleton, *supra*, at 310-11.

Although firearms and toolmark examiners who follow the CMS approach also testify in accord with the AFTE Range of Conclusions, the CMS approach contrasts with the subjective approach in being interpretable in a way that is compatible with the statistical nature of identity claims. The proponents of CMS are best viewed as having used statistical empirical studies to formulate a cut-off point at which the likelihood that another tool of the same type will do as good a job at matching the

evidence toolmark as the suspect tool is so exceedingly small that, for all practical purposes, the suspect tool can be identified as the unique source of the evidence mark. Champod, Baldwin, Taroni, & Buckleton, *supra*, at 311-12; Moran, A Report, *supra*, at 233 (stating that “CMS is a probability model used for toolmark identification”).

The CMS criterion is based on Biasotti’s classic study of .38 Smith & Wesson revolvers, discussed in sections B(1) and (3) above. Follow up studies used IBIS (the BATF’s computerized database for toolmarks on bullets and cartridge cases) to compare numbers of matching striae on ammunition components known to have been fired by the same gun and by different guns of the same type. *See, e.g.*, Biasotti, A Statistical Study of the Individual Characteristics of Fired Bullets, *supra*; Miller & McLean, *supra*; Miller II, *supra*. Other studies made similar comparisons of numbers of matching striae on toolmarks made by chisels and other tools besides firearms. *See, e.g.*, Ronald G. Nichols, Consecutive Matching Striations (CMS), 35(3) AFTE J. 298, 301-02 (Summer 2003); Biasotti & Murdock, The Scientific Basis, *supra*, at 514, 516-17, 516 n.56, 517 n.57.

The CMS criterion is based on these studies’ findings of significant differences between the numbers of consecutive matching striae, but not the percentages or total numbers of matching striae, on pairs of toolmarks known to have been made by the same and different tools. *See, e.g.*, Moran, A Report, *supra*, at 229-232; Biasotti & Murdock, The Scientific Basis, *supra*, at 516 & 516 n.56. The criterion, which is intended to be applied to all firearms and all other types of tools, defines the threshold for identifying a particular tool as the source of a three-dimensional toolmark as a match between evidence and test toolmarks of one group of six consecutive matching striae or two different groups of at least three consecutive matching striae in the same relative position. The threshold for two-dimensional toolmarks is one group of eight consecutive matching striae or two groups of at least five consecutive matching striae in the same relative position. *See* Biasotti & Murdock, The Scientific Basis, *supra*, at 516. However, since CMS requires examiners to compare numbers of striae on individual characteristics of toolmarks, misidentifications will result if, in applying the criterion, examiners mistakenly assume that subclass characteristics on test and evidence toolmarks are individual characteristics. *Id.* *See also* Miller II, *supra*, at 127.

This Court should be aware that increasing numbers of firearms and toolmark examiners rely on the CMS criterion to determine when the match between evidence and test toolmarks is so great that they must have been made by the same tool. *See, e.g.*, Moran, A Report, *supra*, at 229-32; Nichols, Consecutive Matching Striae, *supra*; Biasotti & Murdock, The Scientific Basis, *supra*, at 517 (stating that “approximately 300 members of the ... AFTE ... voluntarily participated in a four-hour workshop [on CMS] at their 1999 annual training seminar.”). At the

same time, CMS is not a definitive solution to the problems of firearms and toolmark identification. Among the unresolved scientific issues is whether the CMS criterion can reliably lead to accurate identifications when different examiners sometimes find different numbers of striae on the same toolmark. Another issue, of particular relevance here, is whether the CMS criterion, which was originally based on studies of .38 Special Smith & Wesson revolvers, can be appropriately applied to all types of tools. See, e.g., Miller II, *supra*, at 116, 130 (stating that while the CMS criteria for two-dimensional and three-dimensional toolmarks were respectively met by only 2% and 0% of pairs of .25 ACP bullets known to have been fired from the same Raven pistol, 0% and 8% of pairs of .380 ACP bullets known to have been fired from the same Lorcin pistol, and less than 2% and 6.5% of pairs of 9mm bullets known to have been fired from the same Stallard pistol, 5% and 14.8% of pairs of .38 special bullets known to have been fired from the same Smith & Wesson revolver respectively met the two- and three-dimensional CMS criteria); Ronald G. Nichols, Firearm and Toolmark Identification Criteria: A Review of the Literature, Part II, 48 (2) J. Forensic Sci. 318 (March 2003) (“Nichols II”); Champod, Baldwin, Taroni, & Buckleton, *supra*, at 313-15; Bunch, *supra*, at 955, 957-62; Moran, Comments and Clarification of Responses from a Member of the AFTE 2001 Criteria for Identification of Toolmarks Discussion Panel, 35(1) AFTE J. 55 (Winter 2003).

This Court need not take sides on the scientific disputes about the CMS criterion in order to decide on the reliability and admissibility of Mr. Masson’s testimony. However, the admissibility decision in this case should be grounded in an awareness that CMS is a serious attempt to develop the necessary statistical and empirical foundations for identity claims. See, e.g., Miller & McLean, *supra* (“The consecutive marks criteria is a solid foundation from which to conduct more research.”). By contrast, Mr. Masson and other adherents of the traditional, subjective approach evade the scientific questions when they insist that their ineffable, mind’s eye judgments are sufficient to determine when the resemblance between toolmarks are so great that they must have come from the same tool. This Court should not countenance this evasion of necessary scientific work.

### III. SUBJECTIVE IDENTITY DETERMINATIONS ARE NOT ERROR- FREE.

Nor should this Court be moved by the plea that “the benefit of the doubt should go to the traditional [subjective] methods” because “with methods such as professional certification and rigorous validation/proficiency testing, the traditional, subjective examination regime can strengthen its scientific grounding.” Bunch, *supra*, at 962. Even assuming that the Daubert-Kumho standard could be satisfied by a method that evades the basic scientific requirement of giving reasons for conclusions, nothing resembling rigorous proficiency testing has been done. See, e.g., Champod, Baldwin, Taroni, & Buckleton, *supra*, at 315; Biasotti & Murdock, *The Scientific Basis*, *supra*, at 508-510. In addition, results from the inadequate proficiency testing that has been done, together with theoretical arguments and the

experience of prominent toolmark examiners, belie the claim that the traditional, subjective procedure results in so few mistakes that objective criteria are simply not needed. As Champod and his colleagues explain: “What would be required [to show that there is no need for objective identification criteria]? First the examiners must often declare a match when the two marks have been made by the same firearm or tool. Next they must NEVER do so when the two marks have been made by differing firearms. How many proficiency tests are required to show that examiners NEVER declare a match when the marks are from differing tools? The standard statistical answer is that an infinite number of tests are required. Examination of CTS proficiency results would suggest that we are not quite there yet.” Champod, Baldwin, Taroni, & Buckleton, *supra*, at 315.

A. Theoretical Arguments for the Possibility of False Positives. Above, we saw that the scientific literature argues that three central difficulties in identifying a tool as the unique source of a toolmark(s) make it necessary to develop objective, statistically-based identification criteria. Two of the difficulties – (i) the danger of confusing subclass with individual characteristics of toolmarks and (ii) the fact that non-unique marks combine to form the individual characteristics of toolmarks –may cause examiners to *overestimate* the significance of matching portions of toolmarks. Consequently, a comparison of striae may lead an examiner to identify a tool as the source of an evidence as well as a test toolmark, even though a mark made by a different tool would do at least as good a job at matching the evidence toolmark.

A danger of false positive identifications also arises from the fact that the individual characteristics of toolmarks change with time. Hence, differences between an evidence toolmark and test toolmark do not necessarily rule out the suspect tool as the source of the evidence mark. It follows that it will sometimes be correct for examiners to attribute differences between evidence and test toolmarks to changes in the surfaces of the suspect tool between the time the evidence and test toolmarks were made. At other times, such an attribution will be wrong; the evidence and test toolmarks differ because the source of the evidence mark was a tool similar, but not identical, to the suspect tool. Thus, it is possible that false positives will occur because examiners *underestimate* the significance of differences between toolmarks.

B. The Experience of Toolmark Examiners. Biasotti and Murdock claim that their own experience as toolmark examiners shows that false positives not only can, but do, occur. “It has been the authors’ experience ... that many of these disagreements [about the identification of toolmarks] stem from one examiner ascribing too much significance to a small amount of matching striae and not appreciating that such agreement is achievable in known non-match comparisons.” Biasotti & Murdock, *The Scientific Basis*, *supra*, at 508-509. See also Biasotti & Murdock, “Criteria for Identification”, *supra*, at 21. Accordingly, Biasotti and Murdock warn that before identifying a single tool as the source of

an evidence toolmark, an examiner needs to compare the evidence toolmark with the marks made by other tools of that type. “We wish to emphasize here that it is essential for the examiner to compare known non-matching toolmarks, especially those made by tools of similar type, size, etc., in order to gain an appreciation of how much agreement can be found in these instances. \*\*\* When comparing questioned and known it is only when agreement is found that exceeds the best known non-match agreement that an identification can be justifiably claimed.” Biasotti & Murdock, “Criteria for Identification”, *supra*, at 19. *See also id.* at 21.<sup>8</sup>

### C. Proficiency Testing.

1. The Current Regime. Nor do current accreditation and proficiency testing requirements in the United States warrant confidence in the accuracy of firearms and toolmark examiners’ conclusions. During the hearing, Mr. Masson testified that he had been employed as a senior firearms and toolmark examiner at the Bureau of Alcohol, Tobacco and Firearms (“BATF”) National Laboratory Center in Rockville, Maryland, and that that laboratory had been accredited by the American Society of Crime Laboratory Directors (ASCLD). Tr. 15,16. However, Mr. Masson did not inform this Court that although the ASCLD bases laboratory accreditation on yearly external proficiency tests, it requires only one examiner in a laboratory to be tested. Laboratories can choose between blind tests and known tests in which test takers are able to distinguish test items from items they are examining as part of their regular case work. *See Schwartz, Ballistics, supra*, at 6-7; ASCLD/Lab, *Proficiency Review Program (2002)*, at [www.asclld-lab.org](http://www.asclld-lab.org).

The only ASCLD-approved provider of proficiency tests for firearms and toolmark examiners is Collaborative Testing Services, Inc. (CTS). *See ASCLD/Lab, Approved Proficiency Test Providers (2003)*, at [www.asclld-lab.org](http://www.asclld-lab.org); CTS, *Test No. 02-526: Firearms Examination. (2002)*, at [http://www.collaborativetesting.com/reports/2226\\_web.pdf](http://www.collaborativetesting.com/reports/2226_web.pdf). (“*Test*”) (DD-2); Tr. 68. In 2002, all firearms examiners who completed the CTS test correctly concluded that the same gun had fired two of the sample “evidence” cartridge cases and the “test” cartridge cases. Of these, 77 percent correctly concluded that the gun had not fired a third “evidence” cartridge case; while 23 percent reported an “inconclusive.” Several test takers commented that the questions were so basic that trainees with one or two weeks of training could answer them.

CTS itself cautions against equating its test results with “an overview of the quality of work performed in the profession.” *See CTS, Test, supra. See also Biasotti& Murdock, The Scientific Basis, supra*, at 510-11. In addition, by stating that its tests are designed to serve laboratories’ interests in demonstrating competence, the CTS website suggests that the tests are biased in favor of proving that examiners are competent. *See CTS, Collaborative Testing Services (2004)*, at [www.collaborativetesting.com](http://www.collaborativetesting.com) (stating that “organizations in more than

55 countries subscribe to our tests to meet their quality assurance objectives including: [d]emonstrating measurement competence to customers ...[and] complying with accreditation and registration requirements.”). See Tr. 69-70.

The relevance of these problems with the CTS testing regime to this case is not diminished by Mr. Masson’s testimony that in the BATF laboratory where he worked, the CTS tests “are handled like a real case”; “the way ATF handled it [the CTS tests] was for the purpose of testing their examiners.” Tr. 37, 70. Like any other test that is designed to aid laboratories to demonstrate proficiency, the CTS test is not a challenging examination, regardless of how a particular toolmark and firearms laboratory may want to use it. Moreover, the key issue here is not Mr. Masson’s competence or even the competence of all the examiners in the laboratory in which he was employed. As Champod and his colleagues recognize, the only demonstration of proficiency that could possibly excuse the firearms and toolmark examiner community from developing objective identification criteria would be a demonstration that NO firearms or toolmarks examiner EVER makes a misidentification, regardless of the laboratory in which he or she is employed.

2. The National Proficiency Study. The only national study of crime laboratory proficiency shows that such proficiency has not been demonstrated. See Joseph L. Peterson, D.Crim. and Penelope N. Markham, Ph.D., Crime Laboratory Proficiency Testing Results, 1978-1991, II, 40 J. Forens. Sci. 1009 (1995) (“Crime Laboratory Proficiency Testing II”) (DD-3). The study reports that on CTS tests administered from 1978-1991, firearms examiners made 12 positive mistakes of concluding that “two or more items shared a common origin when in fact they originated from difference sources.” This compared with 17 negative mistakes of concluding that “two or more items did not share a common source when, in fact, they did.” Peterson & Markham, *supra*, at 1009, 1019. Peterson and Markham report that, “In one exercise, where none of the test fired projectiles matched the evidence projectile, only 29% of the comparisons properly excluded all four test fires.” *Id.* at 1018.

In accord with Mr. Masson’s testimony that “the toolmark aspect of firearms and toolmark examinations” is “the more difficult aspect” of that discipline (Tr. 17), it is unsurprising that the results for toolmark examiners were much worse than those for firearms examiners on the national proficiency study. See *also* GIANNELLI & IMWINKELRIED, *supra*, at 632 & 632 n.146 (explaining that toolmark identification is more difficult than firearms identification because tools, but not firearms, can be used in a variety of ways). On CTS tests administered between 1980 and 1991, 74% of the determinations of common origin or lack thereof by toolmark examiners were correct, as compared with 88% of the determinations by firearms examiners on the 1978-1991 tests. Peterson & Markham, Crime Laboratory Proficiency Testing II, *supra*, at 1010, 1019, 1024. Tr.72-73.

As with firearms examination, false positives comprised a substantial portion of the toolmark examiners' errors on the proficiency study. 41 false negatives compared with 30 false positives. *Id.* at 1024. Peterson and Markham concluded that on one exercise that resulted in 8 false positives and 5 false negatives, "laboratories evidently confused class and individual characteristics [of toolmarks]." *Id.* at 1024-25.

3. Toolmark Examiners' Day-to-Day Performance Is Probably Worse Than Their Performance on the Nationwide Proficiency Study. Toolmark examiners' poor performance on the national proficiency study most likely understates the day-to-day error rates (including false positives) of toolmark laboratories. Peterson and Markham explain that the fact that "these were declared proficiency tests, and examiners knew they were being tested" limits the value of the study's results. Joseph L. Peterson, D.Crim. and Penelope N. Markham, Ph.D., Crime Laboratory Proficiency Testing Results, 1978-1991, I, 40 *J. Forensic Sci.* 994, 997 (1995) ("Crime Laboratory Proficiency Testing I"). "We know ... based on the number of tests and the hours of effort reported by laboratories on several tests, that many laboratories invested more time examining samples than would be expected or required on actual casework." *Id.*

Similarly, Janine Arvizu concludes that the poor results of the national study most likely overestimate the quality of forensic laboratories' work. Forensic laboratories performed badly on the study even though the proficiency testing was not blind; "the participant laboratories knew their reported results would be scored (implying a higher degree of care and attention)." Forensic Labs: Shattering the Myth, *The Champion* (May 2000), *available at* [http://www.nacdl.org/public.nsf/Champion Articles/2000may01](http://www.nacdl.org/public.nsf/Champion%20Articles/2000may01). "Although forensic analysts [in the national proficiency study and other 'open' tests] do not know the 'true value' for a given proficiency sample, they are aware of the fact that a given sample is being used to assess their proficiency. Studies have shown that laboratory performance on this type of 'open' proficiency program is consistently better than on a program where the identification of proficiency samples is blind to the laboratory." *Id.* at n.16. *See also* Biasotti & Murdock, *The Scientific Basis, supra*, at 510 (discussing the superiority of blind proficiency testing).

In addition, the day-to-day error rates of toolmark laboratories are most likely even higher than the error rates on the national proficiency study because participation in the study was voluntary. Peterson and Markham caution that "because the testing was voluntary with about two-thirds of U.S. laboratories subscribing to the program and one-third responding with data, the results do not necessarily represent all laboratories engaged in this type of casework. There

are various possible explanations for the high rate of nonresponses, [including] laboratories' reluctance to have even their anonymous replies recorded and disseminated ...." Peterson & Markham, Crime Laboratory Proficiency Testing I, *supra*, at 997.

D. From the Perspective of the Broader Scientific Community, The Evidence of Inaccurate Toolmark Identifications Is Totally Unsurprising. It is totally unsurprising that firearms and toolmark identifications can turn out to be wrong. Biasotti and Murdock warn that "[m]istakes do occur in forensic science, as in all other professions. All we can do is to try very, very hard to prevent them." *The Scientific Basis*, *supra*, at 518. Their belief in their own and other firearms and toolmark examiners' human fallibility is linked with their commitment to the development of objective identification criteria. Biasotti and Murdock state that, "It is our belief that the continued development of objective criteria and widespread acceptance of criteria for identification will hold mistakes to a minimum ...." *Id.* Janine Arvizu echoes Biasotti and Murdock's warning that human fallibility does not stop at the laboratory door. "Every forensic laboratory makes mistakes." Arvizu, *supra*.

The history of forensic DNA litigation shows that it is crucial for courts to ground their decisions on the reliability of proposed expert testimony in a commonsense awareness that no human enterprise is ever error free. "[F]orensic DNA laboratories maintained for years that the technology was so powerful and foolproof that erroneous results were impossible (one either got the right result or an inconclusive)." Scheck, *supra*, at 1982 (footnote omitted). During the "DNA Wars" of the late 1980's and early to mid-90's, vigorous defense challenges to the admissibility of forensic DNA profiling led to more rigorous judicial scrutiny of forensic laboratories' claims and also sparked concern among academic scientists. Once the broader scientific community became involved, forensic scientists' claims that DNA profiling could not produce false positives were resoundingly rejected. In its 1992 and 1996 reports, the National Research Council warned that "Laboratory errors happen, even in the best laboratories and even when the analyst is certain that every precaution against error was taken;" "No amount of attention to detail, auditing, and proficiency testing can completely eliminate the risk of error." NATIONAL RESEARCH COUNCIL, DNA TECHNOLOGY IN FORENSIC SCIENCE 89 (1992) ("NRC I"); NATIONAL RESEARCH COUNCIL, THE EVALUATION OF FORENSIC DNA EVIDENCE 25 (1996) ("NRC II"). See also Schwartz, Book Review, *supra*, at 447.

#### IV.MR. MASSON'S METHOD FOR REACHING IDENTITY CONCLUSIONS DOES NOT SATISFY THE SPECIFIC DAUBERT FACTORS.

The five specific Daubert factors – testability; error rate; existence and maintenance of standards; peer review and publication; and general acceptance

– are not satisfied by Mr. Masson’s avowed, mind’s eye method for reaching identity conclusions. First, the preceding discussion shows that such proficiency testing as has been performed does not provide any accurate assessment of examiners’ error rate. More fundamentally, it is questionable whether an error rate for an unarticulated technique can even be ascertained. Proficiency tests may indicate particular examiners’ ability to reach correct identity conclusions at a given time. However, unless examiners commit themselves to specific criteria for determining when the resemblances between toolmarks are so great that they must have come from the same tool, a given examiner’s proficiency at a certain time is no guarantee of similar proficiency in the future. Moreover, mind’s eye judgments for when the resemblances between two toolmarks are so great that they must have come from the same tool are, by definition, judgments that cannot be articulated to other people. There is no reason to assume that examiners who possess the ineffable skill of making correct judgments will be able to pass this skill on to future examiners.

In addition, when conclusions are based on inarticulable identity criteria, it is an oxymoron to speak of the existence and maintenance of standards. Only the individual examiner can (ineffably) know whether his or her conclusions follow from his or her personal method. Biasotti and Murdock explain that although “subjective evaluations [of whether a single tool must have been the source of resembling toolmarks] can be valid,” they are of little use to other toolmark examiners. When identifications are based solely on an individual examiner’s subjective judgment, “[t]he basis for forming a pattern recognition conclusion cannot be explained to anyone else.” “Criteria for Identification”, *supra*, at 19. Similarly, Nichols emphasizes that articles that do not explain why an examiner concluded that a particular tool was the unique source of a questioned toolmark, but instead include only subjective comparisons of toolmarks, are “very difficult for other examiners to utilize.” Nichols I, *supra*, at 466. *See also Moran, A Report*, *supra*, at 232 (stating that “[t]he basis for identification is easily communicated between examiners” when the CMS identification criterion, but not, the traditional, subjective approach, is used). As the Florida Supreme Court stated in excluding toolmark identifications based on the traditional, mind’s eye method in Ramirez III, “the record does not show that this method is governed by objective scientific standards. The State’s experts repeatedly testified that the method is entirely subjective and that objective standards would be impractical.” 810 So.2d at 851 (footnote omitted).

Third, the very notion of testability is contravened when an examiner reasons that “I know there’s an identification because my experience and training qualify me to make correct, wholly subjective determinations of when the resemblances between toolmarks are so great that they must have come from the same tool.” Because such an examiner sets forth no theory as to the amounts and kinds of resemblances necessary to show that two toolmarks must have come from the same tool, misidentifications can never raise doubts about the theory on which he

or she relies. Whenever an identification turns out to be erroneous, the fault necessarily lies with the examiner, rather than with the theory of identification. By definition, the subjective judgment of an examiner who makes an erroneous identification has not been honed by adequate experience and training. See Ramirez III, 810 So.2d at 853 (characterizing the traditional, subjective identification procedure of the prosecution experts in that case as “a subjective, untested, unverifiable identification procedure”).

Fourth, the articles published by adherents of the traditional, subjective approach should not be deemed to satisfy the criterion of peer review and publication. Recognizing that “due to the subjective nature of comparisons, ... studies which did not document the examination in other ways were very difficult for other examiners to utilize,” Nichols explains that articles of this type reduce the search for identification criteria to a circle of subjectivity. “Empirical studies have been performed since the early part of the century and easily represent the bulk of the material in quest for identification criteria. Unfortunately, most of these articles are very subjective in nature and as a result, only lend fuel to the ‘subjective’ fire.” Nichols I, *supra* at 466. Similarly, Biasotti claims that this type of literature is not scientific. “From the number of texts devoted exclusively to the subject of firearms and tool mark identification, it might appear that this specialized area of physical comparison is a highly developed science with well defined criteria for evidence evaluation. On the contrary, a review of the literature reveals a very superficial treatment of this basic problem of evaluating results and establishing identity.” Biasotti, The Principles of Evidence Evaluation, *supra*, at 428. See also GIANNELLI & IMWINKELRIED, *supra*, at 614 & 614 n.40 (quoting the above passage from Biasotti for the proposition that “firearms identification is more of an art than a science”).

Fifth, the preceding discussion shows that since the 1930’s, firearms and toolmark examiners have cogently argued that objective, statistical criteria are needed for toolmark identifications to be reliable. The arguments in the toolmark literature are consonant with those that statisticians have advanced in regard to the foundations for scientifically reliable nuclear DNA, mitochondrial DNA, and fingerprint identifications. Cf. Stigler, *supra*, at 857 (explaining how “some of the issues that have arisen in consideration of the forensic use of DNA have striking parallels a century ago” to issues about the statistical basis for fingerprint identification). Accordingly, since Mr. Masson’s traditional, subjective method has been cogently criticized by prominent members of the firearms and toolmark examiner community and also conflicts with views, in the broader scientific community, about the statistical foundations for identity claims, his procedure for drawing identity conclusions cannot be deemed to be generally accepted. See Ramirez III, 810 So.2d at 851 (“In applying the Frye criteria, general scientific recognition requires the testimony of impartial experts or scientists. It is this independent and impartial proof of general scientific acceptability that provides the necessary Frye foundation.”).

## V. THE EXCLUSION OF THE PROSECUTION EXPERT'S TESTIMONY WILL HAVE SALUTORY CONSEQUENCES FOR BOTH SCIENCE AND THE LAW.

A refusal to admit the toolmark expert's testimony in this case is also likely to have salutary consequences for toolmark identification, in particular, and forensic science, more generally. Intrinsic scientific difficulties have not been the main impediment to the development of objective, statistically-based identity criteria. In literature reviews in the 1990's, Springer and Nichols agreed that Biasotti's 1955 bullet comparison study provided strong foundations for developing objective statistical criteria. Springer, *supra*, at 965; Nichols I, *supra*, at 467. Both deplored the fact that neither firearms nor toolmark examiners had built on Biasotti's work by conducting similarly exhaustive, statistical empirical studies. Springer, *supra*, at 966 ("Although the potential for more objective, instrumental methods had been recognized since the late fifties, two decades later, no one had developed any of the methods for proper laboratory use."); Nichols I, *supra*, at 467 ("To date, [Biasotti's] study stands as the most exhaustive statistical empirical study ever published. There are indications Biasotti hoped that this would lead to more studies in an effort to make the criteria more objective."). See also Biasotti, Principles of Evidence Evaluation, *supra*, at 430-32 (calling for further studies).

Deplorably, some of the resistance to developing statistically-based identity criteria appears to stem from opposition to the scientific value of transparency. See Champod & Evett, *supra*, at 106-107. According to one toolmark and firearms examiner, a ground for preferring the traditional subjective method to CMS is that the use of an objective criterion invites questions from judges and juries that may, in turn, cost the prosecution victories. "The final ... difficulty involves explaining and defending in the courtroom conclusions resting on a CMS regime. Examiners schooled in subjective methods may fail to understand or appreciate the research and the logic of interpreting this type of evidence. Thus they may find it difficult to explain them to judge and jury. ... It can be done; DNA examiners successfully wrestle with these difficulties regularly. But if firearms examiners wrestle with them less successfully, it could be a blow to the profession and to the administration of justice." Bunch, *supra*, at 960. By contrast, in responding to criticisms of his decision to publish model cross examination questions, Murdock stated that, "I am aware that some AFTE members will be upset over the publication of these questions. I think they feel that publication amounts to giving ammunition to the enemy. The perceived enemy is, of course, the defense bar. I don't perceive either side as the enemy. I believe that if our profession is to make its maximum contribution to the administration of justice, it must conduct its business in the spirit of openness, which is a hallmark of the scientific method." Murdock, Court Questions, *supra*, at 74. See also Ramirez III, 810 So.2d at 850 n.37 (referring to law review articles deploring the pro-prosecution bias of forensic science in the United States).

Ignorance of statistics and consequent discomfort with probabilistic notions also appears to be a major cause of firearms and toolmark examiners' reluctance to acknowledge the need for statistically-based, objective criteria. See Biasotti, Principles of Evidence Evaluation, *supra*, at 428-30 (explaining why, despite widespread resistance on the part of toolmark examiners, probabilistic notions are central to identity claims); Nichols I, *supra*, at 466 (most of the toolmark literature consists of studies that make only subjective comparisons).

The conclusion that ignorance of statistics on the part of many (but not all) toolmark examiners is a principal cause of the failure to develop objective toolmark identification criteria is supported by Moenssens' more general criticisms of forensic science. Moenssens explains that "[m]any of the witnesses who testify as experts for the prosecution are not truly scientists, but better fit the label of 'technicians.'" Andre A. Moenssens, Novel Scientific Evidence in Criminal Cases, 84 J. Crim. L. & Criminology 1, 5 (1993). Cf. Schwartz, A "Dogma of Empiricism", *supra*, at 208 ("Frye's dictate of judicial deference to scientists implies ... that a relevant scientific community must be composed of 'scientists, not technicians.'" (footnote omitted)). According to Moenssens, as a consequence of the widespread lack of scientific training on the part of so-called forensic "scientists," "[s]ometimes these experts, trained in one forensic discipline, have little or no knowledge of the study of probabilities, and never even had a college level course in statistics." *Id.* at 19.

Judicial tolerance of testimony by experts who do not understand the statistical foundations for identity claims is also likely to have contributed to toolmark examiners' resistance to developing or employing objective identification criteria. Springer explains that although Biasotti's 1955 bullet comparison study was the first properly to address the statistical underpinnings of firearms and toolmark identification, courts had admitted firearms and toolmark identification testimony for many years before the study was done. "[A]fter close to fifty years of firearms/toolmark identification and their use and acceptance by courts, this question [of criteria for identity] had still not been properly addressed [before Biasotti's study]." Springer, *supra*, at 965. This strongly suggests that courts can motivate the toolmark examiner community to develop the requisite objective statistical criteria by excluding firearms and toolmark identification testimony until the proper statistical and empirical foundations are laid. See Nichols I, *supra*, at 473 (warning toolmark examiners that "it is necessary to be able to articulate one's criteria for identification and provide justification of it in a court of law"). Further support for this view is provided by the great amount of attention that the firearms and toolmark examiner community has paid to the Florida Supreme Court's decision in Ramirez III. See, e.g., Nichols II, *supra*, at 324-25; Tomasetti, *supra*, at 294-95. See also Bunch, *supra*, at 955 ("Recently, the debate [over the relative merits of CMS and the traditional, subjective approach] has heated up, in

part owing to the Supreme Court's decision in Daubert v. Merrell Dow Pharmaceuticals, Inc.").

The history of forensic DNA litigation also supports the view that high barriers to the admission of toolmark identification testimony can motivate scientists to develop the necessary scientific foundations for forensic techniques. In a foreward to a symposium on scientific evidence in 1993, Moenssens noted that, "In the early cases, meaningful challenges to prosecution expert testimony on the reliability of 'DNA fingerprinting' were non-existent. Courts held prosecution DNA evidence admissible in state after state. ...[However], a slow ground swell of scientific reservations on use of population statistics resulted in a growing number of more recent court decisions denying admissibility of the evidence." Moenssens, *supra*, at 3. The vigorous defense challenges that fueled the courts' skepticism about forensic DNA evidence also led to widespread academic concern that, in turn, spurred major work in population genetics and statistics. By 1996, firm theoretical foundations had been laid for calculating the statistical significance of nuclear (though not mitochondrial) DNA matches. See NRC II, *supra*, at 25-41; Schwartz, Book Review, *supra*, at 446.

*Amicus* submits this brief in the hope that by excluding the unfounded toolmark identification testimony in this case, this Court will build on the contribution that Ramirez III made to stimulating forensic scientists to develop the requisite statistical and empirical grounding for identity claims. The facts of this case show the importance of laying these foundations. In obligating states to provide expert witnesses to indigent criminal defendants, the United States Supreme Court reasoned that "[t]he private interest in the accuracy of a criminal proceeding that places an individual's life or liberty at risk is almost uniquely compelling." Ake v. Oklahoma, 470 U.S. 68, 78 (1985). The uniquely compelling interests of criminal defendants also argue for especially high barriers to the admission of prosecution expert testimony. See Schwartz, "Dogma of Empiricism", *supra*, at 224-27, 230-31; Moenssens, *supra*, at 4 (stating that "where a person's freedom is at stake, courts ought to be more reluctant to admit evidence based on new, as yet unproven, techniques when such evidence is being offered by the prosecution"); Ramirez III, 810 So.2d at 853 (concluding that "particularly in the face of rising nationwide criticism of forensic evidence in general," "[a]ny doubt as to [the] admissibility [of testimony by forensic scientists] should be resolved in a way that minimizes the chance of a wrongful conviction ...").

## CONCLUSION

For the foregoing reasons, the toolmark identification testimony proffered by the government in this case should be excluded.

## ENDNOTES

---

<sup>1</sup> The pagination of this *Amicus* Brief has been changed for the Internet version of this report, which is otherwise reproduced here in unaltered form. Reproduced by permission of Adina Schwartz.

<sup>2</sup> See, e.g., Alfred Biasotti & John Murdock, *The Scientific Basis of Firearms and Toolmark Identification* (“*The Scientific Basis*”) in 3 DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE 496 (2002) (stating that as part of the broader forensic science discipline of toolmark examination, firearms examination “attempt[s] to identify whether a particular firearm made toolmarks on evidence items, to the exclusion of all other firearms”).

For failures to recognize that firearms and toolmark examiners aim to single out individual tools as the source of crime scene evidence, see *Ramirez v. State*, 810 So.2d 836, 846 (Fla. 2001) (“*Ramirez III*”) (suggesting that firearms and toolmark examiners have traditionally aimed to identify only the type of knife, as opposed to the particular knife, that caused a wound); Simon A. Cole, *Fingerprinting: The First Junk Science?*, 28 Okla. City U.L. Rev. 73, 88 (2003) (erroneously assuming that the toolmark examiner in *Ramirez III* differed from “the profession as [a] whole”) in claiming to be able to identify a unique tool as the only possible source of a toolmark).

<sup>3</sup> At the Daubert hearing in *United States v. Kain*, the judge stated that: “What’s concerning me is that this is a generic issue and I don’t know whether the Government recognizes it. \*\*\* I’ve been a judge for 23 years, nobody has ever challenged this. This is an issue that has great moment for the Department of Justice. ... If I preclude this testimony, it will make ripples all over the country.” She further explained that she had gotten “so agitated” because “there’s rarely a case of any magnitude in ballistics or in arson or anything else that I don’t get some of this testimony.” Transcript of Hearing in *United States v. Kain*, Crim. No. 03-573-1 (E.D. Pa. February 24, 2004) at 87, 101. See also *United States v. Santiago*, 199 F.Supp.2d 101, 111-12 (S.D.N.Y. 2002) (“The Court has not conducted a survey, but it can only imagine the number of convictions that have been based, in part, on expert testimony regarding the match of a particular bullet to a gun seized from a defendant or his apartment.”).

<sup>4</sup> Even courts that have excluded particular firearms and toolmark examiners’ testimony have denied the existence of systemic scientific problems with the discipline. See, e.g., *United States v. Santiago*, 199 F.Supp.2d 101, 111 (“The Court has not found a single case in this Circuit that would suggest that the entire field of ballistics identification is unreliable ... To the extent that [the defendant] asserts that the entire field of ballistics identification is unacceptable ‘pseudo-science,’ the Court disagrees.”), quoted with approval in *United States v. Foster*, 300 F. Supp.2d 375, 376 n.1 (D. Md. 2004); *Sexton v. State*, 93 S.W.3d 96, 101 (Tex. Crim. App. 2002) (concluding, in a firearms identification case, that “the underlying theory of toolmark identification could be reliable in a given case, but ... the State failed to produce evidence of the reliability of the technique in this case”); *Ramirez III*, 810 So.2d at 845 & 852 (accepting “the theory underlying toolmark evidence,” but concluding that the “testimony based on [a prosecution expert’s] knife mark identification procedure, which we find to be new and novel, ... is unreliable and inadmissible”).

Commentators have also failed to realize that there are generic scientific problems with firearms and toolmark identification. See 3 DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE Sec. 29-1.3 (2d ed. updated by the 2003 Pocket Part) (praising both *Ramirez III* and the 1999 decision in *Sexton v. State*, 12 S.W.3d 517 (Tex. Ct. App. 1999) for focusing on the distinctive methods employed by the particular expert in the case, “rather than on global notions of firearms and toolmark theory and practice”).

---

<sup>5</sup> As Mr. Masson testified, firearms identification is part of the forensic science discipline of toolmark identification. Tr. 24, 27-28, 65. (Tr. references are to pages of the hearing before this Court in United States v. Kain, Crim. No. 03-573-1, on February 24, 2004). Firearms examiners deal with the toolmarks that bullets, cartridge cases, and shotshell components acquire by being fired through firearms barrels and also with the toolmarks that unfired cartridge cases and shotshells acquire by being worked through the action of a firearm. Since the basic principles of firearms and toolmark identification are the same, a copy of Adina Schwartz, *Ballistics Recognition/Identification Systems*, in *ENCYCLOPEDIA OF LAW ENFORCEMENT* (forthcoming) ("*Ballistics*"), has been provided to this Court.

<sup>6</sup> The toolmarks at issue in this case are striated toolmarks, or, in other words, patterns of scratches or striae produced by the parallel motion of a tool against an object. Mr. Masson failed to inform this Court that such toolmarks contrast with impression toolmarks resulting from the perpendicular, pressurized impact of a tool on an object. See Alfred Biasotti & John Murdock, *The Scientific Basis*, *supra*, at 496 n.3.

<sup>7</sup> Striae match when they share a "unique character, e.g, width, height, and contour." A. A. Biasotti, The Principles of Evidence Evaluation as Applied to Firearms and Tool Mark Identification, *J. of Forensic Sci.* 428, 430 n.\* (1964) ("Principles of Evidence Evaluation").

<sup>8</sup> Biasotti and Murdock further caution that in order to avoid confusing (i) subclass characteristics common to toolmarks produced by similar tools with (ii) individual characteristics unique to a single tool's toolmarks, "[t]he examiner must be familiar with the various forming and finishing processes" involved in the manufacture of tools. "Criteria for Identification", *supra*, at 17.