

A SYSTEMIC CHALLENGE TO THE RELIABILITY AND ADMISSIBILITY OF
FIREARMS AND TOOLMARK IDENTIFICATION

Adina Schwartz¹

Firearms identification, often improperly referred to as "ballistics identification," is part of the forensic science discipline of toolmark identification. Despite widespread faith in "ballistics fingerprinting," this article contends that because of systemic scientific problems, firearms and toolmark identifications should be inadmissible across-the-board. This article explains that similarities between toolmarks made by different tools and 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 an evidence toolmark. What is the likelihood that the toolmarks made by a randomly selected tool of the same type would do as good a job as the toolmarks made by the suspect tool at matching the characteristics of the evidence toolmark? Firearms and toolmark examiners evade this question by claiming to be able to single out a particular firearm or other tool as the source of an evidence toolmark.

The article further explains that the absence of statistical empirical foundations cannot be excused on the ground that, regardless of how they do it, firearms and toolmark examiners reach accurate identity conclusions. Although firearms and toolmark examiners have feared that Daubert would lead courts to exclude their testimony, both before and after Daubert, firearms and toolmark identification testimony has largely been admitted as a matter of course. No court, including the two recent courts that have excluded particular identification testimony, has recognized the systemic scientific problems with the field. Nonetheless, because of the risk that innocent people will be convicted or even sentenced to death on the basis of erroneous identifications, all firearms and toolmark identifications should be excluded until adequate statistical empirical foundations and proficiency testing are developed for the field.

¹ B.A., Oberlin College, 1971; Ph.D., The Rockefeller University, 1976; J.D., Yale Law School, 1985; Associate Professor, John Jay College of Criminal Justice and The Graduate Center, City University of New York.

I. INTRODUCTION

In 2004 in the District Court for the Eastern District of Pennsylvania, a systemic *Daubert* challenge was brought to the admissibility of firearms and toolmark identification. The expert testimony in the case, *United States v. Kain*,² 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.³ In accord with this, the prosecution expert in *Kain* sought to testify that cuts in a fence and grate were made by a pair of bolt cutters found in the defendant's car, to the exclusion of all other bolt cutters in the world. In challenging the admissibility of this testimony, the defense argued that adequate statistical and empirical foundations and proficiency testing do not exist for the discipline of firearms and toolmark identification. Hence, *Daubert* requires the across-the-board exclusion of firearms and toolmark identification testimony.

At the *Daubert* hearing, Judge Anita M. Brody recognized the breathtaking implications of the defense challenge in *Kain*. She stated that, “[w]hat’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.”⁴ 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.”⁵

² *United States v. Kain*, Crim. No. 03-573-1 (E.D. Pa. 2004). The author was employed as an expert on firearms and toolmark identification by the defense in *Kain*.

³ See, e.g., Alfred Biasotti & John Murdock, *The Scientific Basis of Firearms and Toolmark Identification*, 3 *Modern Scientific Evidence: The Law and Science of Expert Testimony* 495, 496 (David L. Faigman et al. eds., 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) [hereinafter *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); *The Judicial Response to Firearms and Toolmark Identification Expert Evidence*, 3 *Modern Scientific Evidence* § 29.10, at 74 (Supp. 2003) [hereinafter *The Judicial Response*] (implying that the *Ramirez III* court was correct in claiming that “such extreme statements [as that Ramirez’s knife was the murder weapon, to the exclusion of all others] were avoided in most toolmark cases”).

⁴ Transcript of Hearing at 87, *United States v. Kain*, (E.D. Pa. Feb. 24, 2004) (Crim. No. 03-573-1).

⁵ *Id.* at 101. See also *United States v. Santiago*, 199 F. Supp. 2d 101, 111-12 (S.D.N.Y. 2002) (referring to the fact that a significant number of convictions depend on ballistics identification: “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.”).

Before the judge could rule on the issue, however, the government offered the defendant a plea bargain that was too good to refuse.⁶

This article seeks to show that the defense position in *Kain* was correct: because of the systemic scientific problems, firearms and toolmark identification testimony should be inadmissible across-the-board.⁷ After explaining the scientific issues in Part II, I survey the case law in Part III and show that no state or federal court – either before or after *Daubert* – has understood the scientific problems with firearms and toolmark identification. I conclude that because of the risk that innocent people will be convicted or even sentenced to death on the basis of erroneous identifications, all firearms and toolmark identification testimony should be excluded until adequate statistical empirical foundations and proficiency testing are developed for the field.

II. THE SCIENTIFIC BASIS OF FIREARMS AND TOOLMARK IDENTIFICATION

The premise underlying the forensic science discipline of firearms and toolmark identification is that individual tools leave unique marks on surfaces. Firearms identification is a subspecies of toolmark identification dealing with the toolmarks that bullets, cartridge cases, and

⁶ The government dismissed the defendant's two-count indictment on a conspiracy charge and an arson charge that carried a five-year mandatory minimum sentence. *Kain* pled guilty to one count of misprision of felony, and received a ten-month sentence, consisting of five months in a halfway house and five months in prison. *United States v. Kain*, Crim. No. 03-573-1 (E.D. Pa. Mar. 15, 2004).

For an amicus brief that was prepared but not submitted because of the plea bargain, see Adina Schwartz, *A Challenge to the Admissibility of Firearms and Toolmark Identifications: An Amicus Brief Prepared on Behalf of the Defendant in United States v. Kain*, Crim. No. 03-573-1 (E.D. Pa. 2004), J. Phil. Sci. & L. (2004), <http://www.psljournal.com/archives/all/kain.cfm>.

⁷ By contrast, Professor D. Michael Risinger has argued that the *Daubert-Kumho* test should be used to exclude only particular expert testimony offered in a particular case; testimony based on a particular field of expertise should never be excluded across-the-field. See Risinger, *Defining the "Task at Hand": Non-Science Forensic Science After Kumho Tire Co. v. Carmichael*, 57 Wash. & Lee L. Rev. 767, 773 (2000) (stating that "what is clearly not consistent with *Kumho Tire* is any attempt to approach an issue of reliability globally. . . . The emphasis on the judgment of reliability as it applies to the individual case, to the 'task at hand,' runs through the opinion. . . ." (footnote omitted)); see also Edward J. Imwinkelried, *The Task at Hand*, Nat'l L.J., Apr. 19, 2004, at 11 (praising Professor Risinger's "insight," and stating that "it is neither necessary nor sufficient for a judge passing on the admissibility of an expert's testimony to make a global judgment about the general reliability of the expert's discipline"); *The Judicial Response*, *supra* note 3, at 72 (endorsing "*Kumho Tire*'s teaching that what is at issue is the admissibility of the task-at-hand in the case at bar; the issue is never the admissibility of 'an entire field' considered globally").

This article aims to show that Professor Risinger's task-at-hand approach is misguided. When, as in the case of toolmark and firearms identification, there are systemic scientific problems with an entire field, a failure to consider these problems is likely to lead judges to write scientifically misinformed opinions that serve as precedents for the admission of unreliable evidence. Part III will show, in particular, that this criticism applies to the two decisions on firearms and toolmark identification that commentators have praised for following the task-at-hand approach: the Florida Supreme Court's decision in *Ramirez III* and the Texas Court of Criminal Appeals' decision in *Sexton v. State*, 93 S.W.3d 96 (Tex. Crim. App. 2002). See text accompanying notes 144-150 and 179-190 *infra* (discussing *Ramirez III* and *Sexton* and criticizing the analysis of those cases in *The Judicial Response*, *supra* note 3, at 492-93 (3d ed. 2002) & at 73 (Supp. 2003)).

shotshell components acquire by being fired and that unfired cartridge cases and shotshells acquire by being worked through the action of a firearm.⁸ Firearms and toolmark examiners use comparison microscopes to compare evidence toolmarks on ammunition components or other evidence found at crime scenes with test toolmarks that they make with tools that are candidates for having made the evidence toolmark. If an examiner determines that the evidence and test toolmarks are sufficiently similar, a firearm or other tool is identified as the one tool to the exclusion of all others that produced the evidence toolmark.

This section's account of the systemic scientific problems with firearms and toolmark identification will proceed, first, by distinguishing between class, subclass and individual characteristics of toolmarks in Part A. On this basis, Part B will explain that there are three major sources of misidentifications by firearms and toolmark examiners: (1) the individual characteristics of toolmarks are comprised of non-unique marks, (2) subclass characteristics shared by more than one tool may be confused with individual characteristics unique to one and only one tool, and (3) the individual characteristics of the marks made by a particular tool change over time. Part C will then show that 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 a toolmark on an object recovered from a crime scene. What is the likelihood that the toolmarks made by a randomly selected tool of the same type would do as good a job as the toolmarks made by the suspect tool at matching the characteristics of the evidence toolmark? A comparison with forensic DNA identification will be used to show that firearms and toolmark examiners have taken only the most minimal steps towards developing the necessary statistical empirical foundations for their identity claims.

A further, fundamental scientific problem will be discussed in Section D. Adequate proficiency testing has not been developed for firearms and toolmark identification. Nonetheless, such proficiency tests as exist show that examiners make both misidentifications and missed identifications. Part E will explain that far from curing the fundamental scientific problems, the development of computerized firearms identification has shown that the possibility of missed and misidentifications by firearms and toolmark examiners is even greater than previously believed.

A. *Types of Toolmarks*

The distinctions between types of toolmarks must be grasped in order to understand the problems with firearms and toolmark identifications. Toolmarks are either striated toolmarks consisting of patterns of scratches or striae produced by the parallel motion of tools against objects, or impression toolmarks produced on objects by the perpendicular, pressurized impact of tools. Both types of toolmarks have class, subclass and individual characteristics.

⁸ Although firearms identification is sometimes called "ballistics identification," the term is improper because ballistics deals with the motion of projectiles within firearms. See Paul C. Giannelli, *Ballistics Evidence: Firearms Identification*, 27 *Crim. L. Bull.* 195, 197 (1991).

The distinctively designed features of tools are reflected in class characteristics. For example, the rifling impressions on bullets are class characteristics reflecting the number, width and direction of twist of the lands and grooves in the types of barrels that fired them.⁹ Likewise, 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.

Subclass characteristics, which are present in only some toolmarks, arise when manufacturing processes create batches of tools with similarities in appearance, size, or surface finish distinguishing them from other tools of the same type. The toolmarks produced by tools in the batch have matching microscopic characteristics, called subclass characteristics, which distinguish them from 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.¹⁰

Firearms and 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 toolmarks are claimed to 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 test fired bullet and an evidence bullet recovered from a crime scene), an 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 toolmark.

B. *Central Pitfalls in Firearms and Toolmark Identification*

The preceding 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 firearms and toolmark examination's goal of individualization, that is, identifying *one and only one* tool as the source of a particular toolmark(s). A tool may be wrongly identified as the source of a toolmark(s) that it did not produce because, as will be seen, (1) the individual characteristics of toolmarks are comprised of non-unique marks, (2) subclass characteristics shared by more than one tool may be confused with individual characteristics unique to one and only one tool, and (3) the individual characteristics of the marks made by a particular tool change over time. The first difficulty is analogous to difficulties with fingerprint and DNA

⁹ *But see* Jan De Kinder et al., *Reference Ballistic Imaging Database Performance*, 140 *Forensic Sci. Int'l* 207, 213 (2004) (finding that the determination of whether ammunition components were fired by the same gun was complicated by "an overlap of class characteristics of . . . firearms from different manufacturers").

¹⁰ John E. Murdock, *The Individuality of Toolmarks Produced by Desk Staplers*, 6(5) *Ass'n Firearms & Tool Mark Examiners J.* 23 (1974).

identification. However, the second and third difficulties make firearms and toolmark identification more problematic than either fingerprint or DNA identification.¹¹

1. The First Difficulty: 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 parts of each individual's fingerprints and nuclear DNA are the same as that of another, 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.¹²

As a result of the overlapping individual characteristics of toolmarks made by different tools, examiners who assume that a certain amount of resemblance proves that the same tool produced both test and evidence toolmarks may be wrong because the same amount of resemblance may exist in toolmarks produced by different tools of that type. While this can lead to misidentifying a tool as the source of evidence that it did not produce, identifications may also be missed because the toolmark on a fragmented ammunition component or other surface is too small to allow any firearm or other tool, including the one that made it, to be identified as the toolmark's source.¹³

¹¹ Michael Saks's statement, *Merlin and Solomon: Lessons from the Law's Formative Encounters with Forensic Identification Science*, 49 *Hastings L.J.* 1069, 1081 (1998), that "[e]ach member of th[e] family [of forensic identification sciences] subscribes to the same assumptions and draws its inferences from the same basic logic" needs to be qualified to take account of the distinctive difficulties with firearms and toolmark identification.

¹² Jack D. Gunther & Charles Gunther, *The Identification of Firearms*, 90-91 (1935); see also Alfred A. Biasotti & John Murdock, "Criteria for Identification" or "State of the Art" of Firearms and Toolmark Identification, 16(4) *Ass'n Firearms & Tool Mark Examiners J.* 16, 17 (1984) (emphasis added) (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.").

¹³ See, e.g., John E. Murdock, *Some Suggested Court Questions to Test Criteria for Identification Qualifications*, 24(1) *Ass'n Firearms & Tool Mark Examiners J.* 69, 73 (1992) (stating that a "considerable amount of agreement" among striated toolmarks made by different tools is especially likely to be found "if the width of the

The significance of these problems is illustrated by findings that up to 25% of the striae in toolmarks made by different screwdrivers of the same brand matched, while the percentage increased to 28% when comparing toolmarks made by different bolt cutters of the same brand.¹⁴ Similarly, in a classic, statistical empirical study in 1955, Alfred A. Biasotti found that 15 to 20% of the striae on bullets fired from different .38 Special Smith & Wesson revolvers matched.¹⁵

In the 1990s, the development of the Bureau of Alcohol, Tobacco and Firearms' (BATF) computerized comparison system, IBIS (Integrated Ballistics Information System), enabled examiners to compare the toolmarks on a vast number of bullets and cartridge cases, whereas their comparisons had previously been limited to toolmarks encountered in their case work and training.¹⁶ In 1997, Joseph J. Masson published a study finding that as the IBIS database grew for guns of a particular caliber, increasing similarities were discovered in the individual characteristics of toolmarks on ammunition components known to have been fired by different guns of that caliber.¹⁷ The similarities between known non-matching toolmarks were sometimes

mark being compared is quite small [say, two millimeters or less]"); Jeffrey Scott Doyle, *Bullet Identification* (revised 2002), at http://www.firearmsid.com/A_BulletId.htm.

¹⁴ 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) (screwdrivers); Shirley J. Butcher & P.D. Pugh, *A Study of Marks Made by Bolt Cutters*, 15 J. Forensic Sci. Soc. 115, 123 (1975). For approving appraisals of the Burd & Kirk study, see Ronald G. Nichols, *Firearms and Toolmark Identification Criteria: A Review of the Literature*, 42 J. Forensic Sci. 466, 470 (1997); Eliot Springer, *Toolmark Examinations*, 40 J. Forensic Sci. 964, 965 (1995).

¹⁵ 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 Nichols, *supra* note 14, at 467 ("To date, [the Biasotti study] stands as the most exhaustive statistical empirical study ever published.").

¹⁶ See Richard E. Tontarski & Robert M. Thompson, *Automated Firearms Evidence Comparison*, 43(3) J. Forensic Sci. 641, 641-42, 647 (1998).

¹⁷ Joseph J. Masson, *Confidence Level Variations in Firearms Identification through Computerized Technology*, 29(1) Ass'n Firearms & Tool Mark Examiners J. 42 (1997). Ironically, Mr. Masson was the prosecution expert in *United States v. Kain*. On direct examination, he did not recall the study. Transcript of Hearing at 23, *United States v. Kain* (E.D. Pa. 2004) (Crim. No. 03-573-1).

As discussed *infra* at notes 122-125 & 129-131 and accompanying text, studies have also found that as the IBIS database was expanded to include increasing numbers of cartridge cases that had been test fired by guns of the same caliber and make, the top ten or even fifteen candidate matches that IBIS listed for a queried cartridge case increasingly did not include the cartridge case known to have been test fired by the same gun. See Frederic Tulleners, *Technical Evaluation: Feasibility of a Ballistics Imaging Database for All New Handgun Sales* [hereinafter *AB1717 Study*], at 1-4, 1-6 (2001), available at <http://www.nssf.org/pdf/technicalevaluation.pdf>; Jan De Kinder, *Review AB1717 Report. Technical Evaluation Feasibility of a Ballistics Imaging Database for All New Handgun Sales*, at 3 (2002), available at <http://www.nssf.org/pdf/dekinder.pdf> (last visited Jan. 21, 2005) (independent review summarizing and supporting the AB1717 Study's findings); De Kinder et al., *supra* note 9, at 212.

See also Jerry Miller, *Criteria for Identification of Toolmarks Part II*, 32(2) Ass'n Firearms & Tool Mark Examiners J. 116 (2000) (using the IBIS database to find significant numbers and percentages of matching striae on pairs of .25 ACP, .380 ACP, and 9 mm. bullets respectively fired from different individual Raven, Lorcin, and Stallard pistols); Jerry Miller & Michael McLean, *Criteria for Identification of Toolmarks*, 30(1) Ass'n Firearms & Tool Mark Examiners J. 15 (1998) (similar findings with regard to .38 special bullets fired from Smith & Wesson revolvers).

so great that even under a comparison microscope, it was difficult to tell the toolmarks apart and not erroneously attribute them to the same gun.¹⁸

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.”¹⁹

Databases do not exist for toolmarks made by any type of tool besides firearms.²⁰ As will be seen, the existing databases of toolmarks on bullets and cartridge cases are radically incomplete.²¹ Masson’s study implies that, due to the absence of non-firearms toolmark databases and the incomplete databases for firearms toolmarks, misidentifications are likely to result because examiners underestimate the possible similarities between the individual characteristics of toolmarks made by different tools.

2. The Second Difficulty: The Danger of Confusing Subclass with Individual Characteristics

A tool may also be wrongly identified as the source of a toolmark it did not produce if an examiner confuses subclass characteristics shared by more than one tool with individual characteristics unique to one and only one tool. This type of mistake is possible because of a major difference between fingerprint, nuclear DNA and mitochondrial DNA (mtDNA) identification, on the one hand, and firearms and toolmark identification, on the other. 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. By contrast with the nuclear DNA that one inherits from both parents, mtDNA is, in theory, inherited only from one’s mother. Therefore, absent heteroplasmy, even the most remote maternal cousins should share the same mtDNA.²²

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, only some manufacturing processes produce individual tools with sufficiently

¹⁸ Masson, *supra* note 17, at 43; *cf.* De Kinder, *supra* note 17, at 14 (reporting that IBIS’s manufacturer, Forensic Technologies Inc., found that when manual methods were employed, at least one of their firearms examiners failed to match eight out of fifty pairs of cartridge cases that were test fired by the same gun).

¹⁹ Masson, *supra* note 17, at 43.

²⁰ Transcript of Hearing at 67-68, *United States v. Kain* (E.D. Pa. 2004) (Crim. No. 03-573-1).

²¹ *See infra* text accompanying notes 122-125, 128-132.

²² Heteroplasmy is the existence of more than one mtDNA type in a single human being. For a brief account of the differences between nuclear and mtDNA identification, including a discussion of the major problems that heteroplasmy creates for mtDNA identification, see Adina Schwartz, *Barry Scheck et al., Actual Innocence*, 3 *Punishment & Soc’y* 446, 447 (2001) (book review).

differentiated surfaces to produce toolmarks with individual characteristics. Other manufacturing processes result in batches of such similar tools that their toolmarks have the same subclass characteristics, and may or may not also have individual characteristics.²³ The distinction between subclass and individual characteristics of toolmarks continues to be important as tools are used. While wear and tear on some tools may cause the subclass characteristics on their toolmarks to be completely replaced by individual characteristics, in other tools, subclass characteristics may persist alongside individual characteristics.²⁴

Despite their knowledge of this variation, firearms and toolmark examiners have not formulated any generalizations or statistics about which types of tools can be expected to produce toolmarks with subclass or individual characteristics when they are newly manufactured.²⁵ Nor have they developed statistics or generalizations about the rate(s) at which subclass characteristics on toolmarks produced by various types of tools can be expected to be replaced and/or joined by individual characteristics.²⁶

Firearms and toolmark examiners have also failed to develop any rules for distinguishing between subclass and individual characteristics. To avoid confusing subclass characteristics 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.

Accordingly, Alfred A. Biasotti and John Murdock explain that a risk of misidentifications arises because “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

²³ See, e.g., Nichols, *supra* note 14, at 470; 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) Ass'n Firearms & Tool Mark Examiners J. 227, 227-28 (2002); Kristen A. Tomasetti, *Analysis of the Essential Aspects of Striated Toolmark Examination and the Methods for Identification*, 34(3) Ass'n Firearms & Tool Mark Examiners J. 289, 295 (2002).

²⁴ See, e.g., Jerry Miller, *An Examination of the Application of the Conservative Criteria for Identification of Striated Toolmarks Using Bullets Fired from Ten Consecutively Rifled Barrels*, 31(2) Ass'n Firearms & Tool Mark Examiners J. 125, 128 (2001) (finding both subclass and individual characteristics on the striated toolmarks on both land and groove impressions of bullets fired by used guns); Biasotti & Murdock, *supra* note 3, at 501.

²⁵ For a study showing that very similar tools may differ with regard to whether their toolmarks have individual or subclass characteristics, see John E. Murdock, *The Individuality of Toolmarks Produced by Desk Staplers*, 6(5) Ass'n Firearms & Toolmark Examiners J. 23 (1974) (finding subclass characteristics among the toolmarks produced by the ram of one, but not another, brand of desk stapler).

²⁶ Even assuming that subclass characteristics are most likely to lead to misidentifications of newly manufactured guns, the size of the problem cannot be estimated. In the absence of a national gun registry, statistics cannot be compiled about the percentage of crimes committed with new and used guns. For discussions about the consequences of opposition to a national gun registry, see, for example, William G. Krouse, *CRS Issue Brief for Congress: Gun Control Legislation in the 108th Congress*, CRS-15, at http://www.usembassy.de/policy/crime/crs_guncontrol090403.pdf (updated Sept. 4, 2003); Fox Butterfield, *Law Bars a National System for Tracing Bullets and Shells*, N.Y. Times, Oct. 7, 2002, at A12; Aaron Zitner, *Sniper Case Puts Focus on Plan to Link Bullets, Guns*, Los Angeles Times, Oct. 14, 2002, at A4.

evaluated could lead to a false identification.”²⁷ 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.”²⁸

The danger that misidentifications will result from confusing subclass with individual characteristics is real, not theoretical. In the 1980s, this type of confusion was discovered to have produced misidentifications of striated toolmarks.²⁹ 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 1992.³⁰

Examiners’ performance has been shaped by the recognition that the possibility of subclass characteristics creates a serious threat of misidentifications. For instance, test takers have criticized recent firearms proficiency tests for not providing them with sample guns but nonetheless asking them to decide whether multiple bullets or cartridge cases were fired by a single gun.³¹ Invoking laboratory policy that identifications cannot be reached unless the suspect firearm is examined to eliminate the possibility of subclass characteristics, test takers have refused to make identifications in the absence of a gun.³²

²⁷ Biasotti & Murdock, *supra* note 12, at 17.

²⁸ *Id.*; see Nichols, *supra* note 14, at 470-72.

²⁹ See Moran, *supra* note 23, at 227-28.

³⁰ *Id.* (relating this history and warning that “[c]aution should be exercised in distinguishing subclass characteristics from individual characteristics”); see Jerry Miller, *An Examination of Two Consecutively Rifled Barrels and a Review of the Literature*, 32(3) Ass’n Firearms & Tool Mark Examiners J. 259, 260 (2000) (reporting that the toolmarks on the groove, but not the land, impressions on bullets fired from ten consecutively manufactured gang broach barrels were so similar that a false identification would have resulted if the characteristics had been incorrectly identified as individual, rather than subclass characteristics).

Some examiners and legal commentators still fail to recognize the existence of subclass characteristics and the consequent risk that misidentifications may result from confusing subclass with individual characteristics. See, e.g., Transcript of Hearing at 33, 38-39, 73, *United States v. Kain* (E.D. Pa. 2004) (Crim. No. 03-573-1) (prosecution expert’s testimony that he had never “seen or heard of two different tools creating the same exact tool markings”); Andre A. Moenssens et al., *Scientific Evidence in Civil and Criminal Cases* 325 (4th ed. 1995) (stating that striae “are not the same for any two [gun] barrels, even though manufactured one right after the other”); Giannelli, *supra* note 8, at 202-03; Lisa J. Steele, “*All We Want You to Do Is Confirm What We Already Know*”: A *Daubert* Challenge to Firearms Identification, 38 *Crim. L. Bull.* 466, 469 (2002) (“In theory, it is not possible to make two machined surfaces that are microscopically identical” (footnote omitted)); Jeffrey Scott Doyle, *Fundamentals of Firearms ID* (2001), at http://www.firearmsid.com/A_FirearmsID.htm (last visited Jan. 21, 2005).

³¹ See Collaborative Testing Services, Inc., *Firearms Examination Test No. 03-526 Summary Report* at 35, 37 (2003), at http://www.collaborativetesting.com/reports/2326_web.pdf (last visited Jan. 21, 2005) [hereinafter *Fire Exam I*]; Collaborative Testing Services, Inc., *Firearms Examination Test No. 03-527 Summary Report* at 20-21 (2004), at http://www.collaborativetesting.com/reports/2327_web.pdf (last visited Jan. 21, 2005) [hereinafter *Fire Exam II*].

³² *Fire Exam I*, *supra* note 31; *Fire Exam II*, *supra* note 31.

Changes in manufacturing processes are likely to increase the risk of misidentifications resulting from the confusion of subclass with individual characteristics. A recent article explains that “[a]s 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.”³³

3. The Third Difficulty: The Individual Characteristics of Toolmarks Change with Time

A further barrier in the way of firearms and toolmark identification’s goal of individualization is that, by contrast to an individual’s fingerprints and nuclear DNA, the individual characteristics of the marks made by a particular tool change with time.³⁴ The changes in toolmarks reflect the changes in a tool’s surfaces that occur as the tool is used, and/or as damage or corrosion occurs.³⁵

There is widespread agreement that the impermanence of toolmarks makes firearms and toolmark identification more difficult than fingerprint or forensic (nuclear) DNA identification. According to Jan De Kinder, “[f]irearms-related marks are much more difficult to interpret and compare than DNA types or fingerprints. Unlike DNA, firearms-related marks from a single gun show variation.”³⁶

Similarly, prominent statistician Stephen Stigler explains that “it was only in 1890-95 with the work of Francis Galton that the use of fingerprints acquired a scientific basis.”³⁷ Stigler praises Galton for recognizing that proving “[a]n individual’s prints [are] persistent over time”

³³ Stephanie J. Eckerman, *A Study of Consecutively Manufactured Chisels*, 34(4) Ass’n Firearms & Tool Mark Examiners J. 379, 380 (2002); see also Biasotti & Murdock, *supra* note 3, at 500-01 (explaining that, consistently with the claim that economies in manufacturing processes tend to increase the numbers of tools whose toolmarks have subclass characteristics, “[t]he manufacturer’s goal is to produce many items of the same shape that are, *within certain tolerances*, the same size. They also want each of these items to have an acceptable surface finish or appearance. . . . The manufacturers are not, however, concerned that many or all of these items may bear toolmarks composed of subclass characteristics. . . .” (emphasis added)).

³⁴ An individual’s fingerprints may change in rare cases of disease or injury. See Moenssens, *supra* note 30, at 502 (“Rare cases of mutilation, or the occurrence of some skin disease, such as leprosy, may partially or totally destroy the epidermal ridges.”).

³⁵ See, e.g., Transcript of Hearing at 74, *United States v. Kain* (E.D. Pa. 2004) (Crim. No. 03-573-1) (prosecution expert’s testimony that “[e]ach time a tool is used, the individual characteristics of that tool may be altered.”); 1 Paul C. Giannelli & Edward J. Imwinkelried, *Scientific Evidence* 613, 633 (3d ed. 1999) (stating that “if the barrel of the firearm has changed significantly, due to erosion or corrosion, a positive identification may be impossible,” and concluding that toolmark identification “has the same limitations as firearms identification: ‘The characteristics of a tool will change with use.’”).

³⁶ De Kinder, *supra* note 17, at 17.

³⁷ Stephen M. Stigler, *Galton and Identification by Fingerprints*, 140 *Genetics* 857, 862 (1995).

was a crucial step in establishing that a single individual can be reliably identified as the source of a particular fingerprint.³⁸

The significance of the possible changes in the toolmarks made by individual tools is shown by Biasotti's finding that only 21% to 38% of the striae on pairs of bullets fired from the same .38 Special Smith & Wesson revolver matched.³⁹ On the basis of a follow-up study that used IBIS and reached similar results, Miller and McLean urge examiners to be wary of the fact that "[e]ven in comparisons of toolmarks made with the same tool, there are differences, and these differences may outnumber the similarities."⁴⁰

As a consequence of the impermanence of toolmarks, differences between evidence and test toolmarks will sometimes be correctly attributed 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. While misidentifications may occur if examiners wrongly attribute differences in test and evidence toolmarks to changes in the same tool over time, identifications may also be missed if examiners fail to realize that differences between test and evidence toolmarks are compatible with their having been produced by the same tool.

C. Firearms and Toolmarks Identification Does Not Rest on Adequate Statistical Empirical Foundations

The preceding discussion implies that the discipline of firearms and toolmark examination has not developed the requisite statistical empirical foundations for identity claims. As explained above, firearms and toolmark identification is *not* a simple binary matter of determining whether test and evidence toolmarks match.⁴¹ On the one hand, shared subclass characteristics and/or similarities between marks comprising the individual characteristics of toolmarks create substantial resemblances between toolmarks produced by different tools. On

³⁸ *Id.* (emphasis omitted).

³⁹ Biasotti & Murdock, *supra* note 12, at 20; *see also* Nichols, *supra* note 14, at 467 (stating that 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%").

⁴⁰ Miller & McLean, *supra* note 17, at 20; *see also* Charles Meyers, *Some Basic Bullet Striae Considerations*, 34(2) Ass'n Firearms & Tool Mark Examiners J. 158, 159 (2002).

⁴¹ On the basis of this misguided, binary view, the trial judge in *Ramirez III* criticized the defense for dwelling on whether there were objective criteria for determining when two toolmarks must have come from the same tool.

Much of the defense's cross-examination focused on the lack of specific numerical quantification in making a match in tool mark examination. Every expert testified that evidence of this nature does not require a specific number of striae to declare a match. Clearly, the science of tool mark examination does not quantify its standards. Put bluntly, the items compared either match or not.

Record on Appeal at 1225-27, *Ramirez III*.

the other hand, because the surfaces of tools change over time, even toolmarks made by the same tool do not perfectly match. 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 a toolmark on an object recovered from a crime scene. What is the likelihood that the toolmarks made by a randomly selected tool of the same type would match, as closely as the toolmarks made by the suspect tool, the characteristics of the evidence toolmark?⁴²

Firearms and toolmark examiners do *not* even attempt to answer this question. Instead, they fundamentally mislead judges and juries (and perhaps themselves) by claiming to be able to single out a particular firearm or other tool as *the* source of an evidence toolmark, to the exclusion of all other tools in the world.⁴³ The denial of the need to determine the statistical significance of “matches” is implicit in the restrictions that the Association of Firearms and Toolmark Examiners has set on examiners’ conclusions. In accordance with the AFTE Range of Conclusions, examiners in the United States may only (1) identify a particular tool as *the* source of the toolmark(s) found on an object, (2) eliminate a particular tool as *the* source, (3) conclude that the comparison of test and evidence toolmarks is inconclusive, or (4) conclude that the evidence toolmark is unsuitable for comparison.⁴⁴

Firearms and toolmark examiners’ absolute identity conclusions cannot be excused on the ground that they are convenient shorthand for well-grounded probabilistic conclusions. The development of forensic DNA analysis belies any claim that when firearms and toolmark examiners single out a particular tool as *the* source of an evidence toolmark, they *know* that there is a vanishingly small probability that toolmarks made by a random tool would do as good a job at providing a match.

Although Sir Alec Jeffreys invented “DNA fingerprinting” in 1985, it took until the mid-1990s for forensic DNA identification to be placed on firm statistical empirical foundations.⁴⁵ To develop the requisite foundations, scientists needed to (1) specify the sites on the human genome (DNA sequence) that would be typed to determine whether a suspect’s DNA “matched”

⁴² See *Brim v. State*, 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.”); Kevan Walsh & Gerhard Wevers, *Toolmark Identification: Can We Determine a Criterion?*, 29 *Interfaces* 4 (Jan.-Mar. 2002), available at <http://www.forensic-science-society.org.uk/inter29.pdf> (last visited Jan. 21, 2005) (using Bayesian likelihood ratios to formulate the question); cf. Schwartz, *supra* note 22, at 446 (explaining why an analogous question must be answered in order to base an inculcation on a match between a suspect’s DNA and crime scene DNA).

⁴³ Some firearms and toolmark examiners outside the United States have recognized that it is *never* possible to know that a given tool is the source of a particular toolmark, to the exclusion of all other tools in the world. See, e.g., C. Champod et al., *Firearm and Tool Marks Identification: The Bayesian Approach*, 35(3) *Ass’n Firearms & Tool Mark Examiners J.* 307, 310-11 (2003); Walsh & Weavers, *supra* note 42, at 5 (“It is important to acknowledge that statistics will never prove an identification, only asymptotically approach it.”).

⁴⁴ See, e.g., *AFTE Glossary. Theory of Identification as It Relates to Toolmarks*, 30(1) *Ass’n Firearms & Tool Mark Examiners J.* 86-88 (1998); Moran, *supra* note 23, at 228-29.

⁴⁵ See, e.g., National Research Council, *The Evaluation of Forensic DNA Evidence* 41 (1996) [hereinafter NRC II]; Schwartz, *supra* note 22, at 446.

crime scene DNA; (2) develop objective and reliable procedures for determining whether the suspect's and crime scene DNA matched on particular sites of the genome; (3) type the DNA of large numbers of unrelated individuals from various human populations in order to develop databases that could be used to calculate the frequencies of variant DNA types (alleles) on particular sites of the genome; (4) do population studies to determine whether the frequencies of alleles on different sites of the genome were statistically independent; and (5) arrive at a mathematical rule that could be used to combine the frequencies of the shared alleles composing a DNA "match" so as to determine the numerical likelihood that a random person's DNA would provide as good a match as the suspect's DNA for the crime scene DNA.⁴⁶

Analogous work has not been done for firearms and toolmark identification. To explain why such work is needed, it is helpful to distinguish the traditional subjective approach to firearms and toolmark identification from the consecutive matching striae (CMS) approach propounded by Biasotti and Murdock in 1997.⁴⁷

1. The Subjective Approach

Under the traditional subjective approach, examiners do not even attempt to articulate criteria for when the resemblances between toolmarks are sufficient to justify identity conclusions. Instead, they rely solely on inarticulable, mind's eye judgments of when the resemblances are sufficient.⁴⁸ British and European Bayesians have rightly criticized the subjective approach for conflicting with the scientific value of "as far as possible, support[ing one's] opinion by reference to logical reasoning and an established corpus of scientific knowledge."⁴⁹ Similarly, proponents of CMS have argued that an alternative to the traditional approach is needed because when identifications are based solely on an individual examiner's

⁴⁶ For brief overviews of the work involved in steps (1)-(5) and accounts of remaining difficulties with forensic DNA identification, see Schwartz, *supra* note 22, *passim*; Peter D'Eustachio, *DNA Typing*, 2 Encyclopedia of Law Enforcement 622-24 (Larry E. Sullivan & Dorothy Moses Schultz eds., 2005).

⁴⁷ Some legal commentators seem unaware of the existence of CMS or of the long-standing, cogent criticisms of the subjective approach by firearms and toolmark examiners. See, e.g., Steele, *supra* note 30, at 472 ("The examiner's conclusion, based on his observations, is a subjective judgment. . . . Each examiner determines for himself 'his own intuitive criteria of identity gained through practical experience.'"); Moenssens, *supra* note 30, at 328-29, 349-51.

⁴⁸ For subjectivists' descriptions of their approach, see, for example, Doyle, *supra* note 13 ("It really comes down to the experience of the firearm examiner and what they [sic] perceive to be the overall uniqueness of the striations that are present."); Transcript of Hearing at 68, 91-92, *United States v. Kain* (E.D. Pa. 2004) (Crim. No. 03-573-1).

⁴⁹ Christopher Champod & Ian W. Evett, *A Probabilistic Approach to Fingerprint Identification Evidence*, 51 J. Forensic Identification 101, 106-07 (2001) (labeling this value "transparency" and arguing that fingerprint identification, as well as other forensic identification sciences, must move away from the "stereotype [of] the distinguished, greying individual on the stand saying, 'my opinion is based on my many years of experience in the field.'"); see also Champod et al., *supra* note 43, at 314 (stating that "[t]he CMS regime is undoubtedly an improvement [over the subjective approach] towards transparency in the decision making process.").

subjective judgment, “[t]he basis for forming a pattern recognition conclusion cannot be explained to anyone else.”⁵⁰

The subjective approach is also tantamount to a denial of the need to develop statistical empirical foundations for identity claims. Forensic DNA analysts are able to search databases to calculate the frequencies of the component features of a match because they specify, in advance, just what alleles they will compare to determine whether a suspect’s DNA and crime scene DNA match. By contrast, since they refuse to specify the features that determine whether toolmarks match, firearms and toolmark examiners who take the subjective approach rule out the possibility of searching databases to determine the frequency of matches.⁵¹

2. An Attempt at Objectivity: The CMS Approach

Although the subjective approach is still preponderant, increasing numbers of firearms and toolmark examiners base their conclusions on the CMS identification criteria propounded by Biasotti and Murdock in 1997.⁵² Like their subjectivist colleagues, examiners who employ the CMS approach fail to calculate the statistical significance of “matches” between test and evidence toolmarks and misleadingly identify particular tools as *the* source of evidence toolmarks, to the exclusion of all other tools.⁵³ Nonetheless, CMS differs from and is scientifically superior to the subjective approach because it is interpretable in a way that is compatible with the probabilistic nature of identity claims.⁵⁴ The proponents of CMS are best

⁵⁰ Biasotti & Murdock, *supra* note 12, at 19; *see also* Moran, *supra* note 23, at 232 (stating that “[t]he basis for identification is easily communicated between examiners” when the CMS identification criteria, but not the traditional subjective approach, is used); Nichols, *supra* note 14, at 466 (emphasizing 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”).

⁵¹ Although, as the next section will show, the CMS attempt to develop statistical empirical foundations for firearms and toolmark identifications is ultimately unsuccessful, the attempt was motivated by Biasotti’s correct recognition that 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.” A. A. Biasotti, *The Principles of Evidence Evaluation as Applied to Firearms and Tool Mark Identification*, 9 J. Forensic Sci. 428, 430 (1964).

⁵² *See, e.g.*, Stephen G. Bunch, *Consecutive Matching Striation Criteria: A General Critique*, 45(5) J. Forensic Sci. 955, 962 (2000); Miller, *supra* note 17, at 130; Moran, *supra* note 23, at 229-32; Ronald Nichols, *Consecutive Matching Striae (CMS)*, 35(3) Ass’n Firearms & Tool Mark Examiners J. 298, 301-02 (2003); Tomasetti, *supra* note 23, at 297-300; *see also* Walsh & Wevers, *supra* note 42, at 5 (New Zealand Bayesians state that they “will look at applying the CMS approach”).

⁵³ For criticisms of the absolute identity conclusions that proponents of CMS draw, *see, for example*, Bunch, *supra* note 52, at 961; Champod et al, *supra* note 43, at 310-11; Walsh & Wevers, *supra* note 42, at 4-5. For a reply to these criticisms, *see* Moran, *supra* note 23, at 233-34.

⁵⁴ *See, e.g.*, Champod et al., *supra* note 43, at 314 (“Ultimately, [CMS] stresses the fact that the process is probabilistic”); Bunch, *supra* note 52, at 958 (conceding that CMS “is inherently more scientific than the subjective regime currently used by the vast majority of examiners”).

viewed as having used statistical empirical studies to formulate a cut-off point at which the likelihood that another tool of the same type would 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.⁵⁵

Under CMS, the threshold for identifying a particular tool as the source of a three-dimensional toolmark is 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.⁵⁶

Two major limitations on the objectivity of the CMS approach are acknowledged even by its proponents. First, since the CMS identification criteria only apply to striated toolmarks, examiners must still make purely subjective determinations of when the resemblances between impression toolmarks are so great that they must have been made by the same tool.⁵⁷ A further, undisputed source of subjectivity is that CMS requires examiners to compare the striae comprising the individual characteristics of toolmarks. Therefore, misidentifications will result if, in applying the criteria, examiners mistakenly assume that subclass characteristics on test and evidence toolmarks are individual characteristics.⁵⁸ As seen above, there are no rules for distinguishing between subclass and individual characteristics; examiners can only rely on their personal familiarity with different processes for forming and finishing tools and their reflections in toolmarks.

3. Major Problems with CMS: The Contrast with Forensic DNA Identification

A comparison with forensic DNA identification shows that the problems with CMS are yet more severe. Even leaving aside the possibility of subclass characteristics, the necessary research has not been done to ensure that the CMS criteria will lead tools to be identified as the source of (striated) evidence toolmarks *only when* there is a vanishingly small probability that a random tool would do as good a job as the tool identified as the source at producing toolmarks that match the evidence toolmarks.

a. The Subjectivity of Striae Counting – Before DNA profiling could be relied on for either identification or exculpation, objective criteria needed to be developed for determining when two

⁵⁵ See Champod et al., *supra* note 43, at 311-12; Walsh & Wevers, *supra* note 42, at 5.

⁵⁶ See, e.g., Miller & McLean, *supra* note 17, at 29.

⁵⁷ See, e.g., Walsh & Wevers, *supra* note 42, at 4 (“When one considers the matching of impressed toolmarks, one realizes how difficult it is to reduce this information to any meaningful data. This becomes the next challenge for those searching for objective criteria.”); Tomasetti, *supra* note 23, at 298; Nichols, *supra* note 52, at 305.

⁵⁸ See, e.g., Miller, *supra* note 17, at 127; Walsh & Wevers, *supra* note 42, at 4.

DNA specimens match on particular sites of the genome. The National Research Council's 1992 report on forensic DNA criticized "the use of subjective matching rules (e.g., comparison by eye) and the failure to adhere to a stated matching rule" in the early restriction fragment length polymorphism (RFLP) version of DNA profiling and emphasized the need for "an objective and quantitative rule for deciding whether two samples match."⁵⁹ This problem was largely eliminated when RFLP was replaced by the polymerase chain reaction (PCR) method of DNA typing, which inherently leaves much less room for subjective judgments of when two samples have the same alleles.⁶⁰

By contrast, firearms and toolmark examiners have not developed any adequate response to the analogous problem that different, well-qualified examiners are likely to count different numbers of striae on the same toolmark. This creates the possibility that different experienced examiners will reach different conclusions about whether the same toolmarks satisfy or do not satisfy the CMS criteria.⁶¹

The likelihood of differences in examiners' counts of striae is enhanced by firearms and toolmark examiners' practice of using the CMS criteria to *check* identifications that they *first* reach through their minds' eye.⁶² This is similar to the procedure that English and Welsh fingerprint examiners employed to apply the United Kingdom's former, sixteen point standard for fingerprint identification.⁶³ In a study that contributed to the United Kingdom's abolition of the standard in 2001, I.W. Evett and R.L. Williams found wide variations in the number of points' resemblance that different experienced examiners in England and Wales found between latent and rolled fingerprints when they applied the United Kingdom's sixteen point standard. There was far less variation in the number of points that different Dutch fingerprint examiners found when they applied the Netherlands's twelve point standard.⁶⁴ Evett and Williams attributed the greater variation in the English and Welsh examiners' counting of points to their tendency to "tease out points" to prove that the sixteen point standard was met *after* they reached "an inner conviction about the correctness of an identification."⁶⁵ By contrast, Dutch examiners employed the more scientific procedure of *first* deciding which features of a latent print were

⁵⁹ National Research Council et al., *DNA Technology in Forensic Science* 54 (1992) [hereinafter NRC I], at 54, available at <http://www.nap.edu/books/0309045878/html/index.html> (last visited Nov. 18, 2004).

⁶⁰ For a discussion of PCR typing, see John M. Butler, *Forensic DNA Typing* 81-98, 191-204 (2001).

⁶¹ Tomasetti, *supra* note 23, at 298; Meyers, *supra* note 40, at 158-59.

⁶² See Bruce Moran, *Comments and Clarification of Responses from a Member of the AFTE 2001 Criteria for Identification of Toolmarks Discussion Panel*, 35(1) *Ass'n Firearms & Tool Mark Examiners J.* 55, 55, 61 (2003); Moran, *supra* note 23, at 230-31.

⁶³ See Stephen G. Bunch, *Consecutive Matching Striation Criteria: A General Critique*, 45(5) *J. Forensic Sci.* 955,959 (2000).

⁶⁴ I.W. Evett & R.L. Williams, *A Review of the Sixteen Points Fingerprint Standard in England and Wales*, 46(1) *J. Forensic Identification* 49, 59-65 (1996); see also *United States v. Llera Plaza*, 188 F. Supp. 2d 549, 567-68 (E.D. Pa. 2002) (discussing impact of the Evett & Williams study).

⁶⁵ Evett & Williams, *supra* note 64, at 68.

suitable for comparison and *then* comparing only those features to decide whether the Netherlands's twelve point identification standard was met.⁶⁶ Evett and Williams concluded that "the precision [of the United Kingdom's sixteen point standard] is illusory because the determination of the individual points is subjective and . . . [English and Welsh fingerprint] experts vary widely in their judgments of individual points."⁶⁷ Because firearms and toolmark examiners use CMS to *check* identifications that they *first* reach through subjective judgments, this criticism seems equally applicable to CMS.⁶⁸

Supporters of CMS have responded with little more than hand waving to the criticism that the absence of an objective procedure for counting striae means that their approach is only seemingly different from the traditional, subjective approach. The reply that proper training can eliminate disparities in examiners' counts has not been backed by concrete proposals for improved training.⁶⁹ There are similar problems with the response that even if examiners differ about the number of striae, disparate identity conclusions are unlikely to result from applying CMS.⁷⁰ This response is based on "findings" that the numbers of striae that an individual examiner counts on different toolmarks are consistent and that the amount of disagreement in different examiners' counts varies inversely with the numbers of striae on toolmarks.⁷¹ The reliability of these findings is called into question by their having been solely based on the number of striae that the authors' students counted on bullets fired by sections of a single gun barrel.⁷² No large scale study of variations in striae counting was done.

b. The CMS Identification Criteria Are Not Based on Relevant and Representative Databases – A further, fundamental scientific problem with CMS is that relevant and representative databases have not been developed to support the claim that there is a vanishingly small probability that a random tool would do as good a job at producing toolmarks that match an evidence toolmark as the tool that the CMS criteria identify as the source. To see why such databases are required, it is first necessary to understand that, by contrast to forensic DNA identification, CMS conflates the steps of defining a match and calculating the statistical significance of a match.

Forensic DNA identification was developed on the basis of knowledge acquired by theoretical and clinical geneticists. In particular, this knowledge included (1) that *most* of the DNA sequence is the same in all people, (2) the identification of sites in the human DNA

⁶⁶ *Id.* at 68.

⁶⁷ *Id.* at 64-65.

⁶⁸ *See* Bunch, *supra* note 63, at 959.

⁶⁹ *See* Moran, *supra* note 62, at 61; Bunch, *supra* note 63, at 959.

⁷⁰ *See* Fred Tulleners et al., *Striae Reproducibility on One Thompson Contender Barrel*, 30(1) Ass'n Firearms & Tool Mark Examiners J. 62, 68, 78 (1998); *see also* Moran, *supra* note 62, at 60-61.

⁷¹ *See* Tulleners et al., *supra* note 70, at 67-68.

⁷² *See id.* at 67-68, 78.

sequence (called polymorphisms) where variant DNA sequences (alleles) occur, and (3) the identification of the alleles that could be present in different people's DNA at particular polymorphisms.⁷³ To apply this knowledge to the forensic context, scientists extended tests that clinicians had developed for determining, under sterile conditions, whether one or another allele was present on a segment of a known individual's DNA. This involved developing procedures and criteria for determining whether there was a match with regard to particular alleles between DNA samples taken from a suspect and recovered from a crime scene, including accounting for environmental injuries to crime scene DNA.

DNA could be relied on for exculpation once scientists defined the polymorphisms that would be typed to determine whether there was a match and developed objective matching procedures. Since, with the exception of identical twins, each person's *entire* DNA sequence is unique, the lack of a match on any particular site suffices to exclude a suspect as the source of crime scene DNA. By contrast, since the whole, but not the parts, of each person's DNA sequence is unique, and since some, but not all, alleles on a suspect's and crime scene DNA are typed, a "match" does not suffice to identify a suspect as the source of crime scene DNA. It is also necessary to calculate the likelihood that a random person's DNA would match the crime scene DNA on all the sites where the suspect's DNA matches.⁷⁴

In developing DNA identification, forensic scientists had the great advantage of being able to base their definition of a match on firm theoretical knowledge of the differences and similarities between different people's DNA. By contrast, CMS was developed in the hope of curing the absence of systemic knowledge of the differences and similarities between both (1) toolmarks produced by different tools of the same type and (2) toolmarks produced by the same tool.⁷⁵ In the 1955 study which is still the principal research underlying CMS, Biasotti sought to develop objective criteria for firearms and toolmark identification by merging the tasks of defining a match and calculating the statistical significance of a match. Comparisons between toolmarks known to be made from the same and different tools were used to see if there was a threshold, of numbers and/or sequences of matching striae, at which it was exceedingly unlikely that two toolmarks could have been made by any but a single tool.⁷⁶ If, as supporters claim, the CMS criteria define such a threshold, the criteria identify a tool as the source of an evidence toolmark if and only if there is a vanishingly small probability that toolmarks made by any other

⁷³ See D'Eustachio, *supra* note 46, at 622-23, for an account of the relations between the knowledge acquired by the Human Genome Project and the current, short tandem repeat polymorphism (STR) version of forensic DNA identification.

⁷⁴ See *id.* at 2-3; William C. Thompson & Dan E. Krane, *DNA in the Courtroom*, in *Psychological and Scientific Evidence in Criminal Trials* §11:35 (Jane Campbell Moriarity ed., 1996), available at <http://bioforensics.com/articles/Chapter11.pdf> (last visited Dec. 2004) ("Evidence of a DNA 'match' between two samples is impossible to understand and interpret without knowing the probability that a match would be declared if the samples are from different individuals.").

⁷⁵ As explained above in Section II (B) (3), firearms and toolmark identification is particularly difficult since, by contrast to the permanence of a person's nuclear DNA and fingerprints, there are changes over time in the toolmarks a particular tool produces.

⁷⁶ See Biasotti, *supra* note 15, *passim*.

tool would do as good a job at matching the evidence toolmark as the toolmarks made by the identified tool.

This claim about the statistical significance of the CMS criteria is undermined by Biasotti's and other researchers' failure to build and use relevant and representative tool databases.⁷⁷ The concepts of relevant and representative databases are illustrated by the example that being blonde and blue-eyed does much less to identify an otherwise unidentified suspect as the perpetrator of a crime committed in Scandinavia than in an isolated village in the Andes. To quantify the likelihood that a random person is as likely as the suspect to be blonde and blue-eyed, it is necessary to determine the frequency of these characteristics within a large, unbiased sample of Scandinavians or people from the Andes village; calculating frequencies on the basis of a selected few people is insufficient.

Developing the requisite large and representative databases was a crucial step in ending scientists' fierce disagreements, during the DNA Wars of the late 1980s and early 1990s, over whether typed alleles could be assumed to be statistically independent. Unless statistical independence could be assumed, the product rule could not validly be used to combine the frequencies of various typed alleles and arrive at the statistical significance of a match. In 1996, the National Research Council ("NRC") reported that "[e]xtensive studies from a wide variety of databases show that there are indeed substantial frequency differences among the major racial and linguistic groups [in the United States] (Black, Hispanic, American Indian, East Asian, and White)."⁷⁸ On this basis, the NRC recommended that "[i]f the race of the person who left the evidence-sample DNA is known, the database for the person's race should be used [to calculate the statistical significance of a match]; if the race is not known, calculations for all racial groups to which possible suspects belong should be made"⁷⁹ According to the NRC, the studies of a wide variety of databases had also ended the dispute about whether, because of the possibility of subpopulations within the major racial and linguistic groups, the frequencies of different alleles within these large populations could not be assumed to be statistically independent. The studies showed that when applied to determine the frequency of alleles within the large population groups, "formulae based on random mating populations are usually quite accurate."⁸⁰ The NRC's conclusion that a modified product rule could be used to calculate the statistical significance of a match effectively ended the DNA Wars.⁸¹

⁷⁷ See Adina Schwartz, *Book Review*, New York L.J., at Journal Today 1, Feb. 7, 2002 (discussing a similar problem with Bertillinage identification as part of a review of Simon A. Cole, *Suspect Identities: A History of Fingerprinting and Criminal Identification*).

⁷⁸ NRC II, *supra* note 45, at 28.

⁷⁹ *Id.* at 122, 126-27; see also Richard Lempert, *After the DNA Wars: Skirmishing with NRC II*, 37 *Jurimetrics* J. 439, 454-462; *People v. Soto*, 981 P.2d 958, 967 n.18, 971 & n.27 (Cal. 1999).

⁸⁰ NRC II, *supra* note 45, at 28.

⁸¹ During the DNA Wars, some courts excluded DNA identification evidence on the basis of concerns about whether the product rule could validly be used to calculate the statistical significance of a match. See, e.g., *Soto*, 981 P.2d at 973-77 (providing a history of the impact of these concerns on California courts' decisions on the admissibility of forensic DNA identifications). William C. Thompson has aptly summarized the effect of judicial scrutiny of the statistical basis for forensic DNA identifications.

Due to the absence of any analogous studies of tool populations, the CMS approach fails to place firearms and toolmark identification on adequate statistical empirical foundations. On the one hand, the CMS identification criteria are intended to apply to *all* striated toolmarks, including those produced (1) by all types of firearms on all types of bullets and all other types of ammunition components and (2) by all types of tools besides firearms on all types of surfaces.⁸² On the other hand, when Biasotti and Murdock propounded the CMS criteria in 1997, Biasotti's 1955 study of .38 special bullets fired from Smith & Wesson revolvers was the only published statistical empirical support for the claim that the criteria could be applied to striated toolmarks with a vanishingly small possibility of misidentifications. To date, the only other statistical empirical support for the claimed absence of any realistic chance that the CMS criteria will produce misidentifications consists of published studies of bullet striae and unpublished studies of chisel and knife toolmarks.⁸³

Deriving identification criteria for *all* tools from statistical empirical studies of *very few types* of tools cannot be justified on the ground that there are overwhelming similarities in the number of consecutive matching striae different types of tools produce on pairs of toolmarks. The number of consecutive matching striae on pairs of toolmarks is known to vary with the size and quality of the working surfaces of tools.⁸⁴ Therefore, just as the statistical significance of a DNA match varies with the frequencies of the matching alleles *within different human populations*, the threshold at which there is a vanishingly small probability that any but a single tool could be the source of two toolmarks can be expected to vary with the frequencies and patterns of matching striae *within different toolmark populations*.⁸⁵

The exclusion of DNA evidence due to concerns about population genetics had important positive effects: it was the prospect of negative admissibility rulings that spurred much-needed research on the problem of population structure, research that might otherwise not have been done. This research was well underway when NRC I appeared. Results of the studies tipped the balance of scientific opinion in favor of the product rule (or something close to it) shortly thereafter.

William C. Thompson, *Accepting Lower Standards: The National Research Council's Second Report on Forensic DNA Evidence*, 37 *Jurimetrics J.* 405, 423 (1997); *see also* Schwartz, *supra* note 22, *passim*.

⁸² *See* Moran, *supra* note 23, at 230-31; Nichols, *supra* note 52, at 301-02.

⁸³ *See* Nichols, *supra* note 52, at 301-02 (summarizing the studies); Miller & McLean, *supra* note 17 (IBIS database used to study .38 special bullets fired from Smith & Wesson revolvers); Miller, *supra* note 17 (IBIS database used to study .25 ACP, .380 ACP, and 9 mm. bullets respectively fired from Raven, Lorcin, and Stallard pistols); Miller, *supra* note 30 (study of bullets worked through two consecutively gang broach rifled .44 caliber barrels); Jerry Miller, *An Examination of the Application of the Conservative Criteria for Identification of Striated Toolmarks Using Bullets Fired from Ten Consecutively Rifled Barrels*, 31(2) *Ass'n Firearms & Tool Mark Examiners J.* 125 (2001) (study of bullets fired from consecutively rifled 9 mm. barrels); *see also* Tulleners et al., *supra* note 70, at 78.

⁸⁴ *See* Miller, *supra* note 17, at 130; Walsh & Wevers, *supra* note 42, at 4.

⁸⁵ *See* Walsh & Wevers, *supra* note 42, at 4. The authors allude to this problem. "Criticisms may now focus on the applicability of transferring the criterion from one set of class features to another. Is the method determined by studies of bullet striae also applicable to screwdriver striae or even within sets of bullets of varying quality toolmarks?" *Id.*

Supporters claim, however, that the CMS identification thresholds have been set so high that misidentifications cannot result from their application to any striated toolmarks, though the cost may be some missed identifications.⁸⁶ The question is whether the CMS identification thresholds can be high enough to avoid misidentifications of tools with large working surfaces without being so high that an unduly high number of identifications of tools with small working surfaces are missed. This question needs to be addressed even if, from the point of view of the legal system, the primary, if not sole, concern is preventing convictions from being based on misidentifications. Since the vast majority of firearms and toolmark examiners are employed by law enforcement, examiners are unlikely to follow CMS, rather than the traditional subjective approach, if the cost is making far fewer identifications.⁸⁷ A related danger is that examiners may pay lip service to CMS, but manipulate the identification criteria so as to avoid missed identifications. The manipulated criteria may then result in misidentifications.

Recent research by a prominent CMS supporter, Jerry Miller, strongly suggests that adherents of CMS engage in such manipulating. In studies using the IBIS database, Miller found that only 5% and 14.8% of the single land impressions on pairs of .38 special bullets known to have been fired from the same Smith & Wesson revolvers met the CMS identification thresholds for two- and three-dimensional toolmarks, respectively.⁸⁸ The application of CMS to single land impressions would accordingly cause many identifications of Smith & Wesson .38 Specials to be missed.

Due to the unusually wide lands in Smith & Wesson barrels, pairs of bullets fired from the same Smith & Wesson revolver are likely to have more consecutive matching striae than pairs fired from a gun of another type. Consistent with this, Miller found even more missed identifications when the CMS criteria were applied to single land impressions on bullets fired from other types of guns.⁸⁹ In particular, 2% and 0% of .25 ACP bullets fired from the same

⁸⁶ See, e.g., Moran, *supra* note 23, at 231; Miller, *supra* note 17, at 130. As indicated above, a caveat to the claimed impossibility of misidentifications is that the examiner must eliminate the possibility of subclass characteristics before applying the criteria.

⁸⁷ For discussions of how forensic science is distorted by law enforcement values, see, for example, Craig M. Cooley, *Reforming the Forensic Science Community to Avoid the Ultimate Injustice*, 15 Stan. L. & Pol'y Rev. 381, 408 (2004); 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). Firearms and toolmark examiner John Murdock has deplored the law enforcement orientation of his fellow examiners, saying:

I am aware that some AFTE members will be upset over the publication of these [model cross-examination] 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.

John E. Murdock, *Some Suggested Court Questions to Test Criteria for Identification Qualifications*, 24(1) Ass'n Firearms & Tool Mark Examiners J. 69, 73 (Jan. 1992).

⁸⁸ Miller, *supra* note 17, at 130.

⁸⁹ *Id.*

Raven pistol counted as “matches” under the two- and three-dimensional CMS criteria, respectively. Additionally, for .380 ACP bullets fired from the same Lorcin pistol, the match rates were 0% and 8%, and for 9 mm. bullets fired from the same Stallard pistol, the match rates were 2% and 6.5%.⁹⁰

In response to these findings, Miller suggested that to avoid missed identifications, examiners should conclude that the same gun fired both test and evidence bullets whenever the CMS criteria are satisfied by the *sum* of the consecutive matching striae on *all* land impressions.⁹¹ However, Biasotti’s study – as well as the follow-up studies of Miller, Miller and McLean, and Tulleners et al. – found that there were too few consecutive matching striae on *single* land impressions for pairs of bullets fired from different guns to satisfy the CMS criteria.⁹² These findings about *single* land impressions are not equivalent to a finding about *all* land impressions. Thus, the research supporting CMS leaves open the possibility that the *total* number of consecutive matching striae on *all* land impressions could be great enough for pairs of bullets fired from different guns to satisfy the CMS criteria.⁹³ Therefore, examiners might erroneously conclude that the same gun fired two bullets that were in fact fired by different guns if, in order to avoid missed identifications, they followed Miller’s suggestion and applied the CMS criteria to *all* land impressions.

4. The Subjective Approach and CMS Both Fail to Provide Adequate Statistical Empirical Foundations for Firearms and Toolmark Identifications

⁹⁰ *Id.*

⁹¹ *Id.*; see also Ronald G. Nichols, *Firearms and Toolmark Identification: A Review of the Literature, Part II*, 48 J. Forensic Sci. 318, 320 (1997) (endorsing Miller’s proposal).

⁹² Some of the studies also compared single groove impressions. See Biasotti, *supra* note 15, at 35; Miller, *supra* note 30, at 127-28 (comparisons of single land and single groove impressions on pairs of bullets fired from consecutively rifled 9 mm. barrels); Tulleners et al., *supra* note 70, at 73 (comparing single land and single groove impressions on pairs of bullets worked through adjacent sections of one Thompson contender barrel).

Champod et al., *supra* note 43, at 313, allude to the problem of basing identification criteria for bullets with multiple land impressions on comparisons of single land impressions: “What we really needed [in order to convert CMS into a rigorous probability model] was the highest number of CMS noted in the whole bullet. We cannot see how to get this data from Mr. Biasotti’s paper. . . .”

⁹³ The published research contains one contrary observation. With regard to bullets worked through two guns, Miller observed that:

If only single land impressions are considered with the [CMS] conservative criteria for identification applied, then no erroneous identifications could be made. Some missed identifications could occur. If all of the available land impressions are considered when applying the conservative criteria for identification, then fewer missed identifications could occur, and no erroneous identifications could be made.

Miller, *supra* note 30, at 262. However, an observation of bullets worked through *two* guns of *one type* cannot eliminate the possibility that misidentifications could result from applying CMS to all the land impressions of bullets fired from, or otherwise worked through, *all types* of guns.

In sum, the example of forensic DNA identification shows that firearms and toolmark examiners have not done the research needed to develop adequate statistical empirical foundations for their identity conclusions. The CMS approach was motivated by the recognition that the traditional, subjective approach was tantamount to a denial of the need to do the necessary scientific work. Nonetheless, the supporters of CMS have yet to do the necessary research to justify their central claim that there is a vanishingly small probability that toolmarks produced by a random tool could do as good a job at satisfying the CMS criteria as test toolmarks produced by the tool that was the source of the evidence toolmarks. Because they have yet to develop objective criteria for counting striae or to base calculations of the frequency of matching numbers and combinations of striae on relevant and representative tool databases, the supporters of CMS cannot possibly know how likely or unlikely it is that their criteria will lead to misidentifications.

D. Proficiency Testing

In a critique of CMS that has garnered much attention, Stephen Bunch contends that “the benefit of the doubt should go to the traditional methods” because “with methods such as professional certification and rigorous validation/proficiency testing, the traditional, subjective examination regime can strengthen its scientific grounding.”⁹⁴ Even if Bunch’s complacency about the subjective approach’s evasion of the scientific requirement of giving reasons for conclusions is justified, his argument for the subjective approach can be criticized on the ground that rigorous proficiency testing has yet to occur.⁹⁵

Most fundamentally, it is questionable whether a meaningful error rate for the subjective method of firearms and toolmark examination can even be calculated.⁹⁶ Proficiency tests may indicate particular examiners’ ability to reach correct identity conclusions at a given time. However, unless examiners commit themselves to specific, articulable 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

⁹⁴ Bunch’s preference for the subjective approach is also based on the view that if adequate statistical empirical foundations are laid for identity claims, firearms and toolmark examiners “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. . . . [This] could be a blow to the profession and to the administration of justice.” Bunch, *supra* note 63, at 960.

For critical responses to Bunch, see Moran, *supra* note 23, at 232-33; Champod et al., *supra* note 43, at 314-15.

⁹⁵ See, e.g., Champod et al., *supra* note 43, at 315; Biasotti & Murdock, *supra* note 3, at 508-510.

⁹⁶ The error rate, or likelihood that examiners will make misidentifications or missed identifications, is distinct from the statistical significance of a match or likelihood that a random object will do as good a job as the suspect object at “matching” the characteristics on crime scene evidence. See Thompson, *supra* note 74, at 417 (“The value of a reported DNA match for proving that two samples have a common source depends on two factors: the probability of a random (coincidental) match and the probability of an erroneous match (i.e., a false match due to an error in the collection, handling or typing of samples”). *But see* Saks, *supra* note 11, at 1088-92 (confusing the concepts of error rate and statistical significance of matches).

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 communicated to other people.⁹⁷ 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 others. Thus, so long as the subjective method is used, proficiency testing can (at most) establish an error rate for the particular people tested, not for firearms and toolmark examination as a whole.

In addition, even if proficiency tests can establish a meaningful error rate, Champod and his colleagues have cogently argued that results on proficiency tests belie Bunch's claim about the dispensability of objective identification criteria.

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 or tools. 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 [Collaborative Testing Services, Inc.] proficiency results would suggest that we are not quite there yet.⁹⁸

Specific problems with the existing proficiency testing regime include that the American Society of Crime Laboratory Directors (ASCLD) conditions laboratory accreditation on yearly proficiency tests for all examiners, but only requires that one examiner in a laboratory take an external proficiency test.⁹⁹ The only ASCLD-approved provider of external proficiency tests for firearms and toolmark examiners, CTS, allows laboratories to use its tests for known or blind proficiency testing, as training exercises, for research and development of new techniques, and

⁹⁷ See, e.g., Biasotti & Murdock, *supra* note 12, at 19 (explaining that 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"); Nichols, *supra* note 14, at 466 (emphasizing 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"); Moran, *supra* note 23, at 232 (stating that "[t]he basis for identification is easily communicated between examiners" when CMS, as opposed to the subjective approach, is used).

⁹⁸ Champod et al., *supra* note 43, at 315; see also Biasotti & Murdock, *supra* note 3, at 509 ("Based on present data, the field is in a poor position to calculate error rates.").

⁹⁹ Am. Soc'y of Crime Laboratory Directors/Laboratory Accreditation Board, *Proficiency Review Program A-4*, at <http://www.asclcd-lab.org/international/pdf/aslabinternproficiencyreviewprogram.pdf>, (Apr. 2003).

Biasotti & Murdock, *supra* note 3, at 510, argue that as a result of the ASCLD's decision to condition laboratory accreditation on proficiency test results, test results are likely to overstate the quality of examiners' day-to-day work. "It is clear that in such a high stakes game, laboratory administration will do everything possible to ensure that the proficiency test results are correct before reporting them."

so forth.¹⁰⁰ CTS cautions that the results for its tests "are not intended to be an overview of the quality of work performed in the profession and cannot be interpreted as such."¹⁰¹

The CTS tests do show, however, that firearms and toolmark examiners make both misidentifications and missed identifications.¹⁰² On CTS firearms identification tests in 2003, 100% and 90% of test takers respectively reached the correct conclusion that the firearm that fired a "known" cartridge case had also fired two "suspect" cartridge cases but not a third; 10% reported inconclusives instead of exclusions.¹⁰³ On a similar exercise with bullets, 92% and 93% of test takers respectively concluded correctly that the same gun had fired both a "known" bullet and each of two "suspect" bullets; 8% and 7% reported inconclusives.¹⁰⁴ While 45% correctly concluded that the gun had not fired a third "suspect" bullet, 52% reached inconclusives and 3% made wrong identifications.¹⁰⁵

On the 2003 CTS toolmark identification tests, 90% of test takers correctly identified a pair of pliers as the source of toolmarks on a section of tubing; 1% made false exclusions; and 9% reported inconclusives.¹⁰⁶ Of those tested, 22% correctly excluded the pliers as the source of the marks on another section of tubing; 3% made misidentifications; and 75% reached inconclusives.¹⁰⁷ On another test, 98% correctly identified one chisel as the source of toolmarks

¹⁰⁰ Am. Soc'y of Crime Laboratory Directors/Laboratory Accreditation Board, *Approved Proficiency Test Providers*, at <http://www.ascl-d-lab.org/international/aslabinternapprovedproviders.html> (last modified Jan 28, 2005); *Fire Exam I*, *supra* note 31, at 1.

¹⁰¹ *Fire Exam I*, *supra* note 31, at 1; *see also* Richard A. Grzybowski & John E. Murdock, *Firearm and Toolmark Identification-Meeting the Daubert Challenge*, 30(1) Ass'n Firearms & Tool Mark Examiners J. 3, 9 (1998) (stating that "how [the CTS] tests are administered in any particular laboratory (blind or known) can influence the results as can their use as training exercises rather than proficiency tests.").

¹⁰² Firearms and toolmark examiners sometimes testify that qualified examiners *never* make misidentifications. *See, e.g., Sexton v. State*, 93 S.W.3d 96, 99, 101 (Tex. Crim. App. 2002); *Ramirez III*, 810 So. 2d at 851, 851 n.44. *But see, e.g., Biasotti & Murdock*, *supra* note 3, at 518 (warning that "mistakes do occur in forensic science, as in all other professions. All we can do is to try very, very hard to prevent them.").

Similarly to some firearms and toolmark examiners, "forensic 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)." Barry Scheck, *DNA and Daubert*, 15 Cardozo L. Rev. 1959, 1982 (1994). In its 1992 report on forensic DNA, the National Research Council rejected this claim and warned that "[l]aboratory errors happen, even in the best laboratories and even when the analyst is certain that every precaution against error was taken." NRC I, *supra* note 59, at 89. In its 1996 report the council again warned that "[n]o amount of attention to detail, auditing, and proficiency testing can completely eliminate the risk of error." NRC II, *supra* note 45, at 25. *See also* Schwartz, *supra* note 22, at 447.

¹⁰³ *Fire Exam I*, *supra* note 31, at 1-3, 8.

¹⁰⁴ *Fire Exam II*, *supra* note 31, at 1-3, 6.

¹⁰⁵ *Id.*

¹⁰⁶ CTS, *Toolmarks Examination Test No. 03-529 Summary Report* 1-3, 5, at http://www.collaborativetesting.com/reports/2329_web.pdf (last modified 2004).

¹⁰⁷ *Id.*

on a padlock; 89% correctly excluded another chisel; and 2% and 5% reported inconclusives.¹⁰⁸ On yet another test, 40% and 38% realized that neither of the chisels had made the mark on another padlock; 2% wrongly identified one or the other chisel as the source; and 57% and 60% reported inconclusives with regard to each chisel.¹⁰⁹

In the only national study of crime laboratory proficiency, Joseph L. Peterson and Penelope N. Markham reported that on CTS tests from 1980 to 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 CTS tests.¹¹⁰ Firearms examiners made twelve misidentifications and seventeen missed identifications, and toolmark examiners made thirty misidentifications and forty-one missed identifications.¹¹¹

These results are likely to have understated day-to-day laboratory error rates because the testing was declared, rather than blind. Peterson and Markham found “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.”¹¹² This is consistent with Janine Arvizu’s argument for the need for blind, rather than declared, proficiency testing.

Although forensic analysts [in the tests in the Peterson and Markham 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.¹¹³

Peterson and Markham further cautioned that their study may have underestimated day-to-day error rates because participation in the testing was voluntary.

¹⁰⁸ CTS, *Toolmarks Examination Test No. 03-528 Summary Report 1-3, 7*, at http://www.collaborativetesting.com/reports/2328_web.pdf (last modified 2003).

¹⁰⁹ *Id.*

¹¹⁰ Joseph L. Peterson and Penelope N. Markham, *Crime Laboratory Proficiency Testing Results, 1978-1991*, 40 *J. Forensic Sci.* 1009, 1010, 1019, 1024 (1995). *But see* Grzybowski & Murdock, *supra* note 101, at 9 (criticizing Peterson & Markham for counting “inconclusives” as errors and thus overestimating firearms and toolmark examiners’ error rates).

¹¹¹ Peterson, *supra* note 110, at 1019, 1024.

¹¹² *See, e.g.*, Joseph L. Peterson & Penelope N. Markham, *Crime Laboratory Proficiency Testing Results, 1978-1991*, 40 *J. Forensic Sci.* 994, 997 (1995). *But see* Grzybowski & Murdock, *supra* note 101, at 9 (reasoning that the Peterson & Markham study may have overstated error rates because “[i]n actual casework . . . , internal quality control procedures of peer and administrative review . . . reduce the error rate”).

¹¹³ Janine Arvizu, *Forensic Labs: Shattering the Myth*, *The Champion*, n.16 (May 2000), at <http://www.nacdl.org/public.nsf/ChampionArticles/2000may01> (last visited Jan. 21, 2005). *See also* Biasotti & Murdock, *supra* note 3, at 510 (arguing that blind, rather than declared or non-blind, proficiency testing is needed).

[W]ith 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 disseminated¹¹⁴

E. Computerized Firearms Identification

A major practical limit on toolmark identification, and also on firearms identification until the 1990s, is that a comparison microscope can be used to compare only two toolmarks at a time. The time consumption of the comparisons makes it feasible to compare the marks on evidence recovered from a crime scene only with test marks made by a tool(s) that has already been linked to the crime. In addition, transportation and chain of custody problems usually preclude comparisons of tools and evidence recovered by different law enforcement agencies.¹¹⁵

In the early 1990s, computerized comparison systems were developed, allowing vast numbers of digital images of bullets and cartridge cases to be quickly scanned into and stored in databases. Computers could rapidly screen the stored images and arrive at short lists of matches for bullets or cartridge cases submitted for identification. Telecommunications made inter-agency comparisons of guns and ammunition components feasible. While toolmark identification and traditional, manual firearms identification are merely tools for verifying investigative leads, computerization created the possibility of using firearms identification to discover links between particular guns and crimes, including linking seemingly unconnected crimes to the same gun.¹¹⁶

As will be seen, however, computerization has not eliminated the risks of misidentifications and missed identifications by firearms as well as toolmark examiners.

1. NIBIN

The National Integrated Ballistics Information Network (NIBIN), formed in 1997, makes the BATF's computerized comparison system, IBIS (Integrated Ballistics Information System), available to federal, state and local law enforcement agencies for inputting, storing, and matching digital images of bullets and cartridge cases that they recover from crime scenes or use crime guns to test fire. Agencies that participate in NIBIN are linked through the FBI's telecommunications network, allowing inter-agency comparisons of digital images of

¹¹⁴ Peterson, *supra* note 112, at 997.

¹¹⁵ See Tontarski & Thompson, *supra* note 16, at 641-42, 647 (suggesting that digital imaging can avoid the difficulties associated with transferring evidence between agents and chain-of-custody procedures); see also De Kinder et al., *supra* note 9, at 213 (estimating that a traditional, manual comparison of cartridge cases, including the transfer of evidence between agencies, would take 4-6 hours).

¹¹⁶ Tontarski & Thompson, *supra* note 16, at 641, 647.

ammunition components.¹¹⁷ By August 2004, examiners had used NIBIN to make ten thousand links between crime investigations that previously were not known to be connected.¹¹⁸

2. Scientific Problems with NIBIN

The BATF is enthusiastic about the crime-solving potential of NIBIN.

Though no investigative tool is perfect or will be effective in every situation, the availability of an open-case file of many thousands of exhibits, searchable in minutes instead of the lifetimes that would be required for an entirely manual search, provides invaluable information to law enforcement agencies.¹¹⁹

Nonetheless, NIBIN has not eliminated the systemic scientific problems with firearms identification.

First, IBIS only generates a short list of the images in its database that most resemble the scanned image of the ammunition component whose provenance is questioned. A firearms examiner then decides whether there is an identification by using a comparison microscope to compare the questioned ammunition component with ammunition components on the short list.¹²⁰ Since people remain responsible for the ultimate conclusions under NIBIN, misidentifications can occur if examiners underestimate how much similarity between toolmarks is needed to prove that the same gun must have fired two ammunition components. Identifications can also be missed if examiners overestimate the amount of similarity needed. These mistakes can be eliminated only if firm statistical empirical foundations and rigorous proficiency testing are developed for firearms identification.¹²¹

Second, questions about the accuracy of IBIS are raised by Masson's finding that bullets of the same caliber that are test fired by different guns can rank very high on IBIS lists of candidate matches.¹²² Additional severe doubts are raised by Frederic Tulleners' AB1717 Study

¹¹⁷ Nat'l Integrated Ballistics Info. Network, *The Missing Link: Ballistics Technology That Helps Solve Crimes* at 8, at <http://www.nibin.gov/missing/briefhistory.pdf>; Nat'l Integrated Ballistics Info. Network, *The Missing Link: Ballistics Technology That Helps Solve Crimes* at 10-11, at <http://www.nibin.gov/missing/techoverview.pdf> (last visited Jan. 21, 2005) [hereinafter *Missing Link*].

¹¹⁸ Nat'l Integrated Ballistics Info. Network, *ATF's NIBIN Program*, at <http://www.nibin.gov/nibin.pdf> (Aug. 2004).

¹¹⁹ *Id.*

¹²⁰ See Tontarski & Thompson, *supra* note 16, at 647; *Missing Link*, *supra* note 117, at 11.

¹²¹ Cf. De Kinder, *supra* note 17, at 14 (reporting that IBIS's manufacturer, Forensic Technologies Inc., found that when manual methods were employed, at least one of their firearms examiners failed to match eight out of fifty pairs of cartridge cases that were test fired by the same gun, and claiming, on this basis, that "[t]he goal of a ballistics fingerprinting system [should] not [be] restricted to [identifying] those cartridge cases that can be [manually] identified by a trained examiner.").

¹²² See Masson, *supra* note 17, at 42.

and a follow-up study by Jan De Kinder, Frederic Tulleners and Hugues Thiebaut.¹²³ In the studies, the IBIS database was expanded to include hundreds of cartridge cases that were test fired by guns of the same caliber and make.¹²⁴ The studies found that as the size of the database increased, IBIS increasingly failed to rank cartridge cases that were known to have been test fired by the same gun within the top ten or even fifteen candidate matches for the queried cartridge case.¹²⁵

De Kinder and Monica Bonfanti argue that computerized comparison systems using three-dimensional images of ammunition components would be more accurate than IBIS or other systems using two-dimensional images.¹²⁶ However, this change is rendered infeasible by the four to five hours currently needed to scan a bullet into a three-dimensional system, as compared to “[five] minutes to complete entry of a cartridge casing, and [twelve] minutes to enter a bullet” into IBIS.¹²⁷

3. A Legal and Scientific Problem: The Limitation of the IBIS Database to Guns Recovered from Crimes

Federal law limits the IBIS database to images of ammunition components recovered from crime scenes or test fired by guns recovered from crimes.¹²⁸ Hence, NIBIN cannot identify the gun that fired a bullet or cartridge case unless a participating agency has already connected the gun to some crime. In addition to causing identifications to be missed, this limitation on the

¹²³ Tulleners, *supra* note 17; De Kinder et al., *supra* note 9.

¹²⁴ Tulleners, *supra* note 17, at 1-4 (IBIS database expanded to include test fired cartridge cases from 792 new Smith & Wesson model 4008 pistols); De Kinder et al., *supra* note 9, at 207-08 (attempt to increase the significance of the AB1717 study’s findings by expanding the IBIS database to include test fired cartridge cases from approximately 600 9 mm. Para Sig Sauer model P 226 series pistols); *see also* De Kinder, *supra* note 17, at 3 (independent review summarizing and supporting the AB1717 Study’s findings).

¹²⁵ *See* Tulleners, *supra* note 17, at 1-4, 1-6; De Kinder, *supra* note 17, at 3; De Kinder et al., *supra* note 9, at 212 (finding of study that entered cartridge cases test fired by approximately 600 9 mm. Para Sig Sauer model P 226 series pistols into the IBIS database).

¹²⁶ *See* Jan De Kinder & Monica Bonfanti, *Automated Comparison of Bullet Striations Based on Topology*, 101 *Forensic Sci. Int’l* 85, 86 (1999); De Kinder, *supra* note 17, at 22 (suggesting that the manufacturer of IBIS “study the possibility of using 3D images”); *cf.* De Kinder, *supra* note 17, at 12 (stating that while “SBC (Russia) sells a system for the automatic comparisons of bullets with a similar discriminative power [to IBIS’s] but providing images of a much higher quality,” IBIS’s manufacturer, Forensic Technologies, Inc., is the only provider of computerized comparison systems for both bullets and cartridge cases).

¹²⁷ De Kinder & Bonfanti, *supra* note 126, at 87; Nat’l Integrated Ballistics Info. Network, *supra* note 117, at 11.

¹²⁸ Nat’l Integrated Ballistics Info. Network, *The Missing Link: Ballistics Technology That Helps Solve Crimes*, at 11, at <http://www.nibin.gov/missing/deployment.pdf> (last visited Jan. 21, 2005). *But see* De Kinder et al., *supra* note 9, at 208 (stating that Maryland and New York both have statewide ballistics imaging databases that include all new handguns sold in-state).

IBIS database may also result in misidentifications. If the crime gun is not in the IBIS database, an examiner cannot realize that it provides a better match for the questioned ammunition component than the gun in the database that he or she wrongly identified as its source. Such misidentifications are especially likely so long as adequate statistical empirical foundations and proficiency testing are not developed for firearms identification.

Opposition to a national gun registry is a major barrier to expanding IBIS to include digital images of test fired cartridge cases and bullets from newly manufactured or imported firearms.¹²⁹ The technical problems are at least as severe. The AB1717 study and De Kinder, Tulleners, and Thiebaut's follow-up study found that IBIS failed to rank between 28% and 79% of known, matching cartridge cases within the top ten candidate matches when its database was expanded to include hundreds of cartridge cases test fired by guns of the same caliber and make.¹³⁰ Moreover, the IBIS rankings became more and more inaccurate as the IBIS database was expanded.¹³¹ Based on these and similar findings, De Kinder, Tulleners, and Thiebaut concluded, in 2004, that "a reference ballistics image database of new guns is currently fraught with too many difficulties to be an effective and efficient law enforcement tool."¹³²

III. THE SYSTEMIC JUDICIAL FAILURE TO UNDERSTAND THE SCIENTIFIC PROBLEMS

In the wake of *Daubert*, legal scholars predicted that judges would scrutinize the scientific bases of all types of expert testimony, regardless of whether a type of testimony had previously been admitted.¹³³ Firearms and toolmark examiners were greatly concerned that increased judicial scrutiny under *Daubert* might result in the exclusion of their testimony, and some examiners attempted to use this concern to convince their colleagues of the need for a more scientific approach.¹³⁴

¹²⁹ See, e.g., Butterfield, *supra* note 26, at A12; Zitner, *supra* note 26, at 14; Krouse, *supra* note 26, at CRS-1511.

¹³⁰ Tulleners, *supra* note 17, at 1-4, 1-5; De Kinder, *supra* note 17, at 3; De Kinder et al., *supra* note 9, at 207, 210, 213-14.

¹³¹ See Tulleners, *supra* note 17, at 1-4, 1-6; De Kinder, *supra* note 17, at 3; De Kinder et al., *supra* note 9, at 207, 211-12.

¹³² De Kinder et al., *supra* note 9, at 207.

¹³³ See, e.g., Saks, *supra* note 11, at 1072 ("Daubert invites so new a look at old scientific evidence . . . that many scientific evidence precedents are now vulnerable to reconsideration and reversal Under *Daubert* review, forensic identification science turns out to be among the most vulnerable." (footnote omitted)).

¹³⁴ For the view that the discipline should respond to *Daubert* by replacing the traditional, subjective approach with CMS, see, for example, Miller & McLean, *supra* note 17, at 20; Nichols, *supra* note 91, at 318; Tomasetti, *supra* note 23, at 299; Grzybowski & Murdock, *supra* note 101, at 7, 11-12; see also Champod & Evett, *supra* note 49, at 106-07 (British authors of article on fingerprint identification claim that "the scientist should, as far as possible, support his/her opinion by reference to logical reasoning and an established corpus of scientific knowledge," and state that "[w]e have no doubt that courts prefer [such] transparency to obscurity: one only has to study a few of the intensive *Daubert* hearings in the USA for evidence of that view."). *But cf.* Bunch, *supra* note 63,

Contrary to these expectations, both before and after *Daubert*, firearms and toolmark identification testimony has largely been admitted as a matter of course. No court, including the two recent courts that have excluded particular identification testimony, has recognized the systemic scientific problems with the field.¹³⁵

A. Firearms Cases

1. Pre-*Daubert* Firearms Cases

In *People v. Berkman* in 1923, the Illinois Supreme Court excluded testimony, finding it “clearly absurd” and “preposterous” for the expert to conclude that a bullet was fired by a particular gun.¹³⁶ Seven years later in *People v. Fisher*, however, the Illinois Supreme Court admitted cartridge case and bullet identifications,¹³⁷ and firearms identification testimony has long been routinely admitted. In 1978 in *Reed v. State*, the Maryland Court of Appeals summarized the legal history, stating that “the accuracy of firearms identification is common knowledge today.”¹³⁸

The Utah Supreme Court’s decision in 1985 in *State v. Schreuder* further illustrates judges’ complacency about firearms identification.¹³⁹ There, the expert did not present photographs or precisely describe the striations on the test and evidence bullets.¹⁴⁰ The court upheld the admission of his identification of the defendant’s gun, reasoning that his vagueness about the striations went to weight, rather than admissibility, and stating that “[t]here is no question that [the expert’s] testimony is within the scope of his special knowledge as a ballistics expert.”¹⁴¹

at 955 & 962 (concluding that “the benefit of the doubt should go to the traditional methods” rather than CMS, despite recognizing that “[a]n objective decision-making regime, which purportedly describes the counting of striations, appears more likely to successfully meet a *Daubert* challenge than does the subjective regime that currently prevails in the discipline”).

¹³⁵ See discussion of the exclusion of firearms identification testimony in *Sexton v. State*, 93 S.W.3d 96 (Tex. Crim. App. 2002) and toolmark identification testimony in *Ramirez III*, 810 So. 2d 836, discussed *infra* at text accompanying notes 144-150 and 179-190.

¹³⁶ *People v. Berkman*, 139 N.E. 91, 94 (Ill. 1923).

¹³⁷ *People v. Fisher*, 172 N.E. 743, 752-55. (Ill. 1930). The *Fisher* court reasoned that the expert in the case was better qualified and gave more detailed and convincing testimony than the *Berkman* expert. *Id.* at 754. Nonetheless, *Fisher* is commonly taken to have overruled *Berkman*. See, e.g., *Reed v. State*, 391 A.2d 364, 422 n.8 (Md. 1978); *The Judicial Response*, *supra* note 3, at 490.

¹³⁸ *Reed*, 391 A.2d at 388.

¹³⁹ *State v. Schreuder*, 712 P.2d 264 (Utah 1985).

¹⁴⁰ *Id.* at 268.

¹⁴¹ *Id.* at 269.

2. Post-*Daubert* Firearms Cases

Despite *Daubert*, no court has recognized the systemic scientific problems with firearms identification. Instead, courts have tended to wave away challenges to the reliability and admissibility of this type of testimony by pointing to its longstanding admission in court.

Epitomizing the trend, in 2002 in *United States v. Santiago*, a federal trial court in the Southern District of New York stated that “[t]he 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,” and ruled, without considering any scientific issues, that “[t]o the extent that [the defendant] asserts that the entire field of ballistics identification is unacceptable ‘pseudo-science,’ the Court disagrees.”¹⁴² In rejecting the possibility of a systemic challenge, the court invoked the unfavorable contrast that the United States Supreme Court had drawn, in *United States v. Scheffer* in 1998, between polygraph experts and “expert witnesses who testify about factual matters outside the jurors’ knowledge, such as the analysis of fingerprints, ballistics, or DNA”¹⁴³ The *Scheffer* Court’s favorable opinion of ballistics was not based, however, on any analysis of the scientific basis for firearms identifications.

In 2002 in *Sexton v. State*, the Texas Court of Criminal Appeals held that “the underlying theory of toolmark examination could be reliable in a given case.”¹⁴⁴ By contrast to all other recent courts, however, the *Sexton* court also held that because of doubts about “the reliability of the technique used in this case,” the particular firearms identification testimony in the case was inadmissible.¹⁴⁵ The examiner in *Sexton* had testified that he could be one hundred percent certain that the marks on cartridge cases had been made by the same magazine(s), even though, because no weapon had been recovered, he had only compared the toolmarks on the cartridge cases and not examined any suspect magazine.¹⁴⁶ In holding that this testimony should have been excluded, the *Sexton* court correctly recognized that doubts about the identification arose from the expert’s failure to examine or make test fires with the suspect magazine(s) or to consider the processes by which they were manufactured.¹⁴⁷

¹⁴² *United States v. Santiago*, 199 F. Supp. 2d 101, 111-12 (S.D.N.Y. 2002), quoted with approval in *United States v. Foster*, 300 F. Supp. 2d 375, 377 n.1 (D. Md. 2004).

¹⁴³ *United States v. Scheffer*, 523 U.S. 303, 313 (1998).

¹⁴⁴ *Sexton v. State*, 93 S.W.3d 96, 101 (Tex. Crim. App. 2002).

¹⁴⁵ *Id.* *Sexton* excluded the testimony by applying the reliability test that *Kelly v. State*, 824 S.W.2d 568 (Tex. Crim. App. 1992), had established for the admission of scientific evidence. As described in *Sexton*, 93 S.W.3d at 98-100, the Texas *Kelly* test is very similar to the *Daubert-Kumho* test.

On remand, the court in *Sexton v. State*, No. 04-98-00598-CR, 2003 WL 21800084, at *2 (Tex. Crim. App. Aug 6, 2003), held that, “Because [the prosecution expert’s] testimony was unequivocal and because the State repeatedly emphasized its ‘definitive’ nature, we cannot conclude the error in admitting the challenged testimony was harmless.”

¹⁴⁶ *Sexton*, 93 S.W.3d at 99.

¹⁴⁷ *Id.* at 101. The *Sexton* opinion does not explain that these steps are needed to prevent confusing subclass with individual characteristics of toolmarks.

Prominent commentators have endorsed the *Sexton* court's decision to focus on the distinctive problems with the identification in the case and not consider the systemic scientific problems with firearms and toolmark identification. "Under a proper *Kumho Tire* task-at-hand analysis, the court is correct to focus on magazine marks – and the question of whether that particular task can be performed sufficiently well for admission, rather than on global notions of firearms and toolmark theory and practice."¹⁴⁸ To the contrary, the *Sexton* opinion illustrates the danger that courts that focus narrowly on the problems with particular expert testimony may fail to understand the systemic scientific problems with a field of expertise and therefore write opinions that set too low a bar for the admission of future expert testimony.

In particular, the *Sexton* opinion's correct conclusion that the absence of the magazine made the identification in the case particularly unreliable is intertwined with uncritical acceptance of testimony that the presence of the magazine would have allowed for a one hundred percent certain identification.¹⁴⁹ To the contrary, there can be no such thing as a one hundred percent certain identification, given the probabilistic nature of firearms and toolmark identifications, results on proficiency tests, and the multiple possible sources of misidentifications. In addition, in erroneously endorsing the analogy that the prosecution expert drew to fingerprint identification, the *Sexton* opinion fails to recognize that the existence of subclass characteristics and the impermanence of toolmarks make firearms and toolmark identification more problematic than fingerprint identification.¹⁵⁰

Due to the possibility of subclass characteristics, significant numbers of examiners have criticized recent CTS proficiency tests for not providing them with a suspect firearm, but nonetheless asking them to determine whether various ammunition components were fired or worked through the same gun. See *Fire Exam I*, *supra* note 31, at 35, 37; *Fire Exam II*, *supra* note 83, at 20- 21; see also Jessica L. Rosenberry, *Firearm/Toolmark Examination and the Daubert Criteria*, 35(1) Ass'n Firearms & Tool Mark Examiners J. 38, 43 (2003) ("The ideal [proficiency] test would be very similar to actual casework, but this was often difficult [for CTS] because of the need to create a large quantity of identical samples to distribute to all the participating laboratories.").

¹⁴⁸ *The Judicial Response*, *supra* note 3, at 493. The above quotation applies to the Texas Court of Appeals' earlier decision in *Sexton v. State*, 12 S.W.3d 517 (Tex. Crim. App. 1999); the authors of *The Judicial Response* have yet to update their discussion to include the Texas Court of Criminal Appeals' decision in *Sexton v. State* in 2002. Nonetheless, the authors' praise of "the task-at-hand analysis" is equally applicable to both *Sexton* decisions.

The above quotation implicitly endorses Professor D. Michael Risinger's view of the proper approach to decisions on the admissibility of expert testimony. See *supra* note 7, for citations to Professor Risinger's task-at-hand approach and a critical discussion.

¹⁴⁹ *Sexton*, 93 S.W.3d at 101.

¹⁵⁰ The *Sexton* court reasoned, *id.* at 100-01, that despite the ultimate inadmissibility of the particular testimony in the case, a factor favoring admission was the clarity of the analogy the expert drew to fingerprint identification. According to the Court:

The clarity with which the underlying theory and technique were explained to the trial court weighs in favor of admission. [The expert], by way of analogy, explained that this technique [of firearms and toolmark identification] was similar to identifying fingerprints. This is a concept that is easily understandable by laypeople.

Id.

No subsequent court has cited *Sexton* or excluded firearms identification testimony, even if the expert's identification was made in the absence of a gun. In 2004, in the federal death penalty case of *United States v. Foster*, a trial court in the District of Maryland admitted testimony matching cartridge cases recovered from different crime scenes to the same gun, even though no weapon had been recovered.¹⁵¹ In concluding that the "general reliability of the science of ballistics, including comparisons of spent cartridge cases even where there is no 'known' weapon recovered" had been established, the court reasoned that "[b]allistics evidence has been accepted in criminal cases for many years" and that "[i]n the years since *Daubert*, numerous cases have confirmed the reliability of ballistics identification."¹⁵² Instead of recognizing the serious scientific problems with the expert's subjective approach to identity determinations, the court expressed satisfaction with his explanation that "'the human ability to recognize a similar pattern and distinguish between dissimilar patterns' makes identification possible."¹⁵³ The defense's failure to present any contrary expert testimony may partially explain the *Foster* court's failure to recognize the scientific problems.¹⁵⁴

In the 1995 capital murder case of *People v. Hawkins*, however, the defense introduced two of Biasotti's articles to question prosecution experts about their mind's eye determinations that the striations on bullet fragments from different crime scenes were so similar that they must have been made by the same (unrecovered) gun.¹⁵⁵ The trial judge then commented to the jury that Sigmund Freud's psychiatric theories were not invalidated by his failure to use the DSM, and approvingly analogized the experts' subjective approach to Justice Stewart's approach to pornography.¹⁵⁶ In upholding the propriety of these interjections, the California Supreme Court reasoned that "[i]f the court expressed any opinion, it was not as to the ultimate question of whether the firearms identification in this case was reliable, but rather that the Biasotti articles did not in themselves undermine that reliability."¹⁵⁷ Biasotti's point, however, was that absent a database and calculations of statistical significance, examiners cannot *know* when the resemblances between toolmarks are so great that they must have come from a single firearm.

In 2004 in *Commonwealth v. Dinkins*, the Massachusetts Supreme Judicial Court found the government highly culpable for the destruction of the suspect gun, but still upheld the admission of expert testimony matching bullets and cartridge cases recovered from the crime

¹⁵¹ *United States v. Foster*, 300 F. Supp. 2d 375, 375 (D. Md. 2004).

¹⁵² *Id.* at 376, 376 n.1.

¹⁵³ *Id.* at 377.

¹⁵⁴ *Id.*

¹⁵⁵ 897 P.2d 574, 587-88 (Cal. 1995).

¹⁵⁶ *Id.* at 588.

¹⁵⁷ *Id.* at 589. The authors of *The Judicial Response*, *supra* note 3., at 494, fail to see that the issue in *Hawkins* was the propriety of the trial judge's response to the defense questions, not the admissibility of the firearms identification.

scene with bullets and cartridge cases in two other cases.¹⁵⁸ The court recognized the possibility of misidentifications by citing the Peterson and Markham study of proficiency testing and a 1977 case, *Commonwealth v. Ellis*, in which experts had disagreed about whether the same gun that fired two bullets also fired a third.¹⁵⁹ Nonetheless, the *Dinkins* court reasoned that the defendant was not prejudiced by the destruction of the gun because he was able to use its absence to cast doubt on the identification.¹⁶⁰ By contrast, according to the court, “the recognized reliability of ballistics analysis . . . make[s] it highly likely that additional tests [made possible by the availability of the gun] would have mirrored the results of those already done and thereby bolstered the Commonwealth’s case.”¹⁶¹ The court failed to explain why the evidence of misidentifications in *Ellis* and in the Peterson and Markham study had not shaken its faith in the reliability of firearms identification.

B. Toolmark Identification Cases

1. Pre-*Daubert* Toolmark Cases

In *State v. Fasick* in 1928, the Washington Supreme Court held that it was inadmissible for an expert to identify a particular knife as the source of cuts on branches.¹⁶² Less than two years later in *State v. Clark*, however, the Washington court upheld the admission of expert testimony that the defendant’s knife had made the cuts on branches.¹⁶³ The *Clark* opinion portrays the admission of forensic identification testimony as a sign of scientific progress, and misleadingly analogizes toolmark to fingerprint identification.¹⁶⁴

Courts are no longer skeptical that, by the aid of scientific appliances, the identity of a person may be established by fingerprints. There is no difference in principle in the utilization of the photomicrograph to determine that the same tool that made one impression is the same instrument that made another impression. The edge on one blade differs as greatly from the edge on another blade as the lines on

¹⁵⁸ 802 N.E.2d 76, 78-79, 79 n.4 (Mass. 2004).

¹⁵⁹ *Id.* at 80-81 (citing *Ellis*, 364 N.E.2d 808 (Mass. 1977); *Peterson*, *supra* note 110, at 1018-19, 1028).

¹⁶⁰ *Dinkins*, 802 N.E.2d at 80; *see also Ellis*, 364 N.E.2d at 812 (reasoning that conflicts between experts are a factor favoring admission because they enable the jury to realize that experts are fallible).

¹⁶¹ *Dinkins*, 802 N.E.2d at 80.

¹⁶² *State v. Fasick*, 270 P. 123 (Wash. 1928), *aff’d*, 274 P. 712 (Wash. 1929).

¹⁶³ *State v. Clark*, 287 P. 18, 20 (Wash. 1930). Saks, *supra* note 11, at 1108, acknowledges that the *Clark* opinion distinguishes *Fasick* away on the facts, but describes the distinctions as “superficial” and criticizes the *Clark* court for “fail[ing] to explain what changed in its understanding of the scientific claims of toolmark identification.”

¹⁶⁴ *Clark*, 287 P. at 20.

one human hand differ from the lines on another. This is a progressive age. The scientific means afforded should be used to apprehend the criminal.¹⁶⁵

The possibility of scientific problems with toolmark identification is also not recognized in *Fletcher v. Lane*, decided in 1978, the only reported federal case on toolmark (as opposed to firearms) identification, either before or after *Daubert*.¹⁶⁶ There, the defendant argued that it was “impossible” for the prosecution expert to know that his screwdriver had made the prymarks on a trailer door. He also requested that additional experts examine the screwdriver.¹⁶⁷ The *Fletcher* court dismissed the defendant’s habeas corpus petition as a challenge to the expert’s credibility.¹⁶⁸

Similarly, in *State v. Churchill* in 1982, the Kansas Supreme Court did not consider any scientific issues, but held that expert testimony that the defendant’s knife had made the cuts in the victim’s breastbone had been properly admitted.¹⁶⁹ The court asserted that the expert “ha[d] the requisite skill and training to perform the tests, and that the methods used were reliable,” and reasoned that it was up to the jury to decide whether his expertise extended to marks in the human body.¹⁷⁰

By contrast to experts in other cases, the toolmark expert did not make a unique identification in *Commonwealth v. Graves*, decided by the Superior Court of Pennsylvania in 1983.¹⁷¹ Instead, he testified that although the defendant’s fingernail “could not be characterized as ‘unique,’ . . . there is probably a fair degree of probability that this nail or any nail of this shape made this kind of [scratch] mark” on the victim’s neck.¹⁷² On further questioning, the toolmark expert testified that there was a “high probability” that the defendant’s fingernail was the source.¹⁷³ In upholding the admissibility of this testimony, the *Graves* court reasoned that the scientifically recognized field of toolmark examination extended to “[w]ound marks here referred to as toolmarks, whether made by a firearm, knife, blunt instrument or fingernails . . . ,” and correctly identified the issue in the case as “the probability that the [defendant’s] nail caused

¹⁶⁵ *Id.*

¹⁶⁶ 446 F. Supp. 729, 731 (S.D. Ill. 1978).

¹⁶⁷ *Id.*

¹⁶⁸ *Id.*

¹⁶⁹ *State v. Churchill*, 646 P.2d 1049, 1052, 1054 (Kan. 1982).

¹⁷⁰ *Id.* at 1054.

¹⁷¹ *Commonwealth v. Graves*, 456 A.2d 561, 565 (Pa. Super. Ct. 1983).

¹⁷² *Id.* In criticizing the *Graves* opinion, Saks, *supra* note 11, at 1109-10, fails to realize that a forensic odontologist in the case, but not the toolmark expert, singled out the defendant’s fingernail as the unique source of the wound. See *Graves*, 456 A.2d at 565-56; see also Moenssens et al., *supra* note 30 at 379-80 (criticizing the forensic odontologist’s testimony).

¹⁷³ *Graves*, 456 A.2d at 565.

the scratch.”¹⁷⁴ However, the court failed to realize that the absence of a fingernail database meant that the expert had no scientific basis for concluding that there was “a fair degree of probability,” a “high probability,” or, indeed, any degree of probability that the defendant’s nail had been used to scratch the victim’s neck.

2. Post-*Daubert* Toolmark Cases

The decision in *People v. Genrich* in 1996 illustrates courts’ continuing failure to understand the scientific issues. There, the Colorado Court of Appeals held that a trial court had not abused its discretion in refusing to hold a pre-trial hearing on the admissibility of testimony that three sets of pliers had been used to manufacture bombs.¹⁷⁵ The court held that the long-standing admission of toolmark identifications in court showed that the *Frye* test was satisfied.¹⁷⁶ In addition, the court reasoned that the fact that two other prosecution experts had reached “inconclusive” results with regard to two of the pliers went to the weight of the testimony, not its admissibility.¹⁷⁷ In dismissing the defendant’s criticisms of the expert’s identification, the court failed to see that the existence of subclass characteristics belied the expert’s testimony that each tool makes unique marks from the moment of manufacture. Nor did the court understand that even if all toolmarks were unique, databases would still be needed to calculate the statistical significance of resemblances and differences between toolmarks.¹⁷⁸

By contrast to all other recent courts, the Florida Supreme Court excluded toolmark identification testimony in its three decisions in the death penalty case of *Ramirez v. State*.¹⁷⁹ In 1989, *Ramirez I* held that a *Frye* hearing was needed before the expert’s identification of the knife found in Ramirez’s car as the murder weapon could be admitted; the expert’s “self-serving statement” as to the reliability of his technique had not been sufficient to warrant the admission

¹⁷⁴ *Id.* at 566-67.

¹⁷⁵ *People v. Genrich*, 928 P.2d 799, 801-02 (Colo. Ct. App. 1996).

¹⁷⁶ *Id.* at 802; see also *id.* at 801 (noting that *Fishback v. People*, 851 P.2d 884 (Colo. 1993), established that, notwithstanding *Daubert*, the *Frye* test still applies in determining admissibility of scientific evidence in Colorado courts).

For the position that scientists’ testimony and writings, but not courts’ admissibility decisions, should be the evidence used to determine whether a theory or technique satisfies the *Frye* general acceptance test, see 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, 196-222 (1997); Paul C. Giannelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States, a Half-Century Later*, 80 Colum. L. Rev. 1197, 1217-19 (1980).

¹⁷⁷ *Genrich*, 928 P.2d at 802.

¹⁷⁸ *Id.*

¹⁷⁹ *Ramirez v. State*, 542 So. 2d 352 (Fla. 1989) [hereinafter “*Ramirez P*”]; *Ramirez v. State*, 651 So. 2d 1164 (Fla. 1995) [hereinafter “*Ramirez II*”]; *Ramirez III*, 810 So. 2d 836.

of his testimony at Ramirez's trial.¹⁸⁰ After Ramirez was tried and convicted a second time, *Ramirez II* held, in 1995, that the trial judge's failure to allow the defense experts to testify at the *Frye* hearing had made the identification testimony inadmissible.¹⁸¹ By contrast to the procedural reversals of *Ramirez I* and *II*, the Florida Supreme Court reached the scientific issues after Ramirez's third trial and conviction, holding in 2001 in *Ramirez III* that the expert's identification of the defendant's knife as the murder weapon, to the exclusion of all others, was unreliable and inadmissible.¹⁸²

Although the *Ramirez III* court ostensibly applied the *Frye* test to exclude the toolmark experts' testimony, the opinion in fact uses the factors specifically listed in *Daubert* as surrogates for the general acceptance test.¹⁸³ Legal commentators have praised both *Ramirez III*'s turn towards *Daubert* and the opinion's task-at-hand approach of "scrutiniz[ing] the particular application and particular variant of what was regarded to be an otherwise generally accepted field of expertise."¹⁸⁴ To the contrary, as with *Sexton v. State*, the *Ramirez III* court's deliberately narrow approach gave rise to serious scientific mistakes.

The *Ramirez III* court failed to understand that its criticisms of the expert testimony in the case were applicable to firearms and toolmarks examination as a whole. The court characterized the expert's conclusion that Ramirez's knife was the only possible murder weapon as an "extraordinarily precise claim of identification," when individualization is *the* goal of firearm and toolmark examination.¹⁸⁵ Similarly, the court criticized the expert for pursuing a novel method, when his identification was based on the traditional, subjective approach.¹⁸⁶

The *Ramirez III* court also seriously underestimated the difficulty of providing adequate scientific foundations for firearms and toolmark identification. The court contrasted the expert's subjective approach with the reliable, objective method of basing identifications on percentages

¹⁸⁰ *Ramirez I*, 542 So. 2d at 354-55. *But see* Carol Henderson Garcia, *Are 'Knife Prints' Reliable Evidence: An Analysis of Tool Mark Evidence*, 25(4) *Ass'n Firearms & Tool Mark Examiners J.* 266, *passim* (Oct. 1993) (criticizing *Ramirez I* for excluding the toolmark identification testimony); Moenssens et al., *supra* note 30, at 379 (same).

¹⁸¹ *Ramirez II*, 651 So. 2d at 1168. *But see* Saks, *supra* note 11, at 1110 n.223 (mistakenly stating that *Ramirez II* reversed *Ramirez I*'s exclusion of the toolmark identification evidence).

¹⁸² *Ramirez III*, 810 So. 2d at 852.

¹⁸³ *Id.* at 843 & n.8, 849-51; *see* David W. Barnes, *General Acceptance Versus Scientific Soundness*, 31 *Fla. St. U. L. Rev.* 303, 305 (Winter 2004); Paul C. Giannelli, *Admissibility of Scientific Evidence*, 28 *Okla. City U. L. Rev.* 1, 11-12 (2003). *Ramirez III*'s interpretation of *Frye* has been followed in *Sybers v. State*, 841 So. 2d 532, 542-43 (Fla. Dist. Ct. App. 2003), and *Bevil v. Florida*, No. 1D02-1495, 2004 WL 1263625, at *2 (Fla. Dist. Ct. App. June 10, 2004).

¹⁸⁴ Faigman et al., *supra* note 3, at 73; *see also* Giannelli, *supra* note 183, at 12 ("*Ramirez III* represents a reinvigorated *Frye* test"); Barnes, *supra* note 183, at 310 (stating that "when the reliability of conventional [as well as novel] scientific expertise is challenged, the *Ramirez III* standards should apply").

¹⁸⁵ *Ramirez III*, 810 So. 2d. at 845, 849. Cole, *supra* note 3, at 88, and Faigman et al., *supra* note 3 at 73, both fail to realize that it was a mistake for the *Ramirez III* court to draw this contrast.

¹⁸⁶ *Ramirez III*, 810 So. 2d at 851.

of matching striae, not realizing that Biasotti and Miller both found only insignificant differences between the percentages of matching striae in pairs of toolmarks made by the same and different tools.¹⁸⁷ The court's ignorance of the firearms and toolmark literature was also betrayed in its failure to recognize that CMS is the only widely accepted alternative to the expert's traditional subjective approach.¹⁸⁸ Nor was the court aware of criticisms of CMS. Similarly, the court claimed that the expert's method did not have an error rate, instead of recognizing that, despite its insufficient rigor, CTS testing belied the expert's claim that toolmark examiners never make misidentifications.¹⁸⁹

Notwithstanding the spurious distinction that *Ramirez III* drew between the testimony in the case and "[t]he theory underlying tool mark evidence, which . . . is generally accepted in the scientific community," the decision has occasioned great concern among firearms and toolmark examiners.¹⁹⁰ In 2002, the *Ramirez III* decision was published in the Association of Firearm and Tool Mark Examiners Journal at the behest of firearms and toolmark examiners Bruce Moran and John Murdock, who wrote that:

This decision, in our opinion, has significant ramifications for [the] future admissibility of firearm and toolmark evidence. It seems to demonstrate that the courts are less apt to accept our more traditional explanations defending our methods and are looking more deeply into the scientific basis for toolmark identifications accompanied with demonstrable proof of such identifications. Although this decision is painful for our community to accept, it has now become a reality. We hope that by providing all members with the following decision, we

¹⁸⁷ *Id.* at 851 & n.46; Biasotti, *supra* note 15; Miller & McLean, *supra* note 17; Miller, *supra* note 17.

¹⁸⁸ *Ramirez III*, 810 So. 2d at 850-51 (praising German medical literature identifying the type of knife that caused a wound, and ignoring firearms and toolmark examiners' criticisms of the subjective approach); *see also* Saks, *supra* note 11, at 1131 (ignoring the development of CMS and the connected debates, and characterizing toolmark identification as a field "largely frozen in time, with little if any fundamental progress since [its] foundational appearances in court").

¹⁸⁹ *Ramirez III*, 810 So. 2d at 851. The *Ramirez III* court followed the mistaken suggestion of *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 594 (1993), that a technique either has or does not have an error rate. Relying on this dichotomous conception, some firearms and toolmark examiners have reasoned that, despite examiners' relatively poor results on CTS tests and questions about the quality of the tests, the very existence of the CTS tests and the Peterson and Markham study shows that their discipline satisfies *Daubert's* "error rate" factor. *See, e.g.*, Rosenberry, *supra* note 147, at 43-44; Grzybowski & Murdock, *supra* note 101, at 9.

For criticism of *Daubert's* conception of error rates, *see, for example*, Grzybowski & Murdock, *supra* note 101, at 9 ("This requirement of the *Daubert* test [of known or potential error rate] has created some controversy because the measurements of the error rate for any particular forensic science specialty has not been truly standardized. The *Daubert* court opinion again fails to provide any help or shed any light in this regard."); Schwartz, *supra* note 176, at 161-62 (discussing the complexity and scientific controversy of calculations of error rates), *cited with approval* in *United States v. Cordoba*, 991 F. Supp. 1199, 1203 n.11 (C.D. Cal. 1998), *aff'd*, 194 F.3d 1058 (9th Cir. 1999), *cert. denied*, 529 U.S. 1081 (2000).

¹⁹⁰ *Ramirez III*, 810 So. 2d at 845; Nichols, *supra* note 91, at 324-25; Tomasetti, *supra* note 23, at 294-95; Walsh & Wevers, *supra* note 42, at 4 (New Zealand firearms and toolmark examiners voice concern about *Ramirez III*).

will, as an association, be motivated to address these new challenges with more vigor.¹⁹¹

Firearms and toolmark examiners' concern has thus far proved unfounded; the *Ramirez III* decision has not been cited in any firearms or toolmark identification case. In the only subsequently published toolmark case, *Commonwealth v. Foreman* in 2002, the Pennsylvania Superior Court summarily concluded that a *Frye* objection to the admission of toolmark identification testimony would not have been meritorious.¹⁹² Without considering any scientific issues, the court intoned that "this court has previously stated that tool mark identification is a scientifically recognized area for expert testimony."¹⁹³ The court then held that, because prejudice had not been shown, counsel was not ineffective for failing to object to testimony that metal punch stamps were the source of altered serial numbers.¹⁹⁴

IV. CONCLUSION

Adequate statistical empirical foundations and proficiency testing do not exist for firearms and toolmark identification. Examiners themselves admit and results on CTS proficiency tests show that misidentifications as well as missed identifications occur. Far from solving the fundamental scientific problems, the development of computerized firearms identification has shown that the possibility of missed identifications and misidentifications by firearms and toolmark examiners is even greater than previously believed. Despite the exclusion of particular firearms and toolmark identification testimony by the Texas Court of Criminal Appeals in *Sexton v. State* in 2002 and the Florida Supreme Court in *Ramirez III* in 2001, neither these two courts, nor the many courts that have admitted toolmark and firearms identifications, have recognized the systemic scientific problems with the field.

Judicial reluctance to rock the prosecutorial boat may partially explain why, despite widespread concern among firearms and toolmark examiners, courts have failed to recognize the inadequacy of the field's scientific foundations.¹⁹⁵ To the contrary, as the Florida Supreme Court

¹⁹¹ *Joseph Ramirez vs. State of Florida – Supreme Court Decision, December 20, 2001*, 34 (2) Ass'n Firearms & Tool Mark Examiners J. 215 (2002).

¹⁹² 797 A.2d 1005, 1018 (Pa. 2002).

¹⁹³ *Id.* at 1018 (citing the *Graves* fingernail identification case, *Commonwealth v. Graves*, 456 A.2d 561 (Pa. Super. Ct. 1983)).

¹⁹⁴ *Id.*

¹⁹⁵ See, e.g., *United States v. Santiago*, 199 F. Supp. 2d 101, 111-12 (S.D.N.Y. 2002) (denying the possibility that "the entire field of ballistics identification is unreliable," and stating that "[t]he 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"); Transcript of Hearing at 101, *United States v. Kain* (E.D. Pa. 2004) (Crim. No. 03-573-1) (judge explains that she had gotten "so agitated" at the government's inadequate response to the systemic defense challenge to the admissibility of firearms and toolmark identification because "there's rarely a case of any magnitude in ballistics or in arson or anything else that

recognized in *Ramirez III*, “[a]ny doubt as to [the] admissibility [of expert testimony] should be resolved in a manner that minimizes the chance of a wrongful conviction, especially in a capital case.”¹⁹⁶ Especially in light of the major role that firearms identifications play in obtaining convictions, all firearms and toolmark identifications should be excluded until the development of firm statistical empirical foundations for identifications and a rigorous regime of blind proficiency testing.

I don’t get some of this testimony”); see also Craig M. Cooley, *Reforming the Forensic Science Community to Avoid the Ultimate Injustice*, 15 Stan. L. & Pol’y Rev. 381, 382 (2004) (“It seems that the only standard the courts are requiring of forensic science is that it be incriminating to the defendant.” (footnote omitted)).

¹⁹⁶ *Ramirez III*, 810 So. 2d at 853; see also Schwartz, *supra* note 176, at 232-35 (proposing a higher standard for the admission of scientific evidence offered by the prosecution than by a criminal defendant or civil litigant).