J Forensic Sci, May 2007, Vol. 52, No. 3
doi:10.1111/j.1556-4029.2007.00422.x
Available online at: www.blackwell-synergy.com

Ronald G. Nichols,[1] *M.Ch.M.*

Defending the Scientific Foundations of the Firearms and Tool Mark Identification Discipline: Responding to Recent Challenges

ABSTRACT: Recent challenges have brought the discipline of firearms and tool mark identification to the forefront in recent court cases. This article reviews those challenges and offers substantial support for the scientific foundations of the firearms and tool mark identification discipline. A careful review of the available literature has revealed that firearms and tool mark identification is rooted in firm scientific foundations, critically studied according to the precepts of the scientific method culminating in the Association of Firearms and Toolmark Examiners' Theory of Identification. Firearms and tool mark identification has been validated in a manner appropriate for evidence of the kind to be expected in firearms and tool mark examinations. Proficiency tests and error rates have been studied and can provide consumers of the disciple with a useful guide as to the frequency with which misidentifications are reported in the community using appropriate methodologies and controls. As a result, the primary issues in recent challenges do not invalidate the firearms and tool mark discipline as a science nor should it detract it from its admissibility in a court of law.

KEYWORDS: forensic science, daubert challenges, firearms and tool mark identification, identification criteria, proficiency testing, scientific foundations, statistics, subclass characteristics, validation

There have been recent challenges that have targeted the discipline of firearms and tool mark identification (1–6). A review of these challenges reveals common themes that include whether tool manufacture results in sufficient individuality and the related issue of subclass characteristics; tool surfaces changing over time affecting individualization; lack of objective criteria; the need for statistics and databases because absolute individualization is not possible; lack of adequate validation; and inadequate proficiency testing. Responses to some of these challenges have been published elsewhere (7).

In addition to the common themes, other points of similarity exist among these challenges. The first is that they have been primarily published in journals offering legal arguments, not scientific research. As a result, such challenges are not necessarily peer reviewed by a subject matter expert. In addition, these challenges consistently use rhetorical arguments in support of their contentions; when scientific articles are used, most are review articles. Therefore, these challenges overlook or ignore primary sources, which weakens their validity.

Despite these shortcomings, it is important not to dismiss these criticisms without giving them due consideration because they provide a gauge on how well the consumers of the discipline understand its fundamentals. In addition, it helps those within the discipline to avoid complacency and maintain a readiness to discuss the basis for what firearms examiners do, why they do it, and why it is reliable. Remaining current with regard to these critical issues, making improvements where necessary, and refining where required are necessary for a scientific discipline.

[1]Bureau of Alcohol, Tobacco, Firearms, and Explosives, Forensic Laboratory Services, San Francisco, 355 North Wiget Lane, Walnut Creek, CA 94598.

Received 7 May 2006; and in revised form 30 Sept. 2006; accepted 11 Nov. 2006; published 13 April 2007.

Dealing with Questions of Sufficiency of Individuality

The primary question when dealing with firearms and tool mark identification is whether sufficient individuality on a tool working surface is present to permit a trained examiner to conclude a tool mark as having been made by that tool. Considering that this question is fundamental to the discipline, it has been and continues to be one of the most-researched areas (8,9).

A careful and thorough review of the literature will demonstrate that the discipline of firearms and tool mark identification is firmly rooted in the scientific method. The Association of Firearms and Toolmark Examiners (AFTE) has defined a theory of identification to assist in the examinations conducted by firearms and tool mark examiners:

> [a] The theory of identification as it pertains to the comparison of tool marks enables opinions of common origin to be made when the unique surface contours of two tool marks are in "sufficient agreement."
> [b] This "sufficient agreement" is related to the significant duplication of random tool marks as evidenced by the correspondence of a pattern or combination of patterns of surface contours. Significance is determined by the comparative examination of two or more sets of surface contour patterns comprised of individual peaks ridges and furrows. Specifically, the relative height or depth, width, curvature and spatial relationship of the individual peaks, ridges and furrows within one set of surface contours are defined and compared to the corresponding features in the second set of surface contours. Agreement is significant when it exceeds the best agreement demonstrated between tool marks known to have been produced by different tools and is consistent with the agreement demonstrated by tool marks known to have been produced by the same tool. The statement that "sufficient agreement" exists between two tool marks means that the agreement is of a quantity and

> quality that the likelihood another tool could have made the mark is so remote as to be considered a practical impossibility.
> [c] Currently the interpretation of individualization/identification is subjective in nature, founded on scientific principles and based on the examiner's training and experience (10).

The AFTE Theory of Identification addresses many important issues for the discipline including issues of common origin, defining observational objectives, defining sufficiency of agreement to establish common origin, what an identification means, and defining the role of subjectivity.

Critics claim that if subclass characteristics are incorrectly ascribed individual character, a misidentification could occur. Subclass characteristics are defined as,

> Discernible surface features of an object which are more restrictive than CLASS CHARACTERISTICS in that they are: (1) Produced incidental to manufacture; (2) Are significant in that they relate to a smaller group source (a subset of the class to which they belong); (3) Can arise from a source which changes over time. Examples would include: bunter marks, extrusion marks on pipe, etc. Caution should be exercised in distinguishing subclass characteristics from INDIVIDUAL CHARACTERISTICS (10).

Other examples can include an imperfection on a rifling tool that imparts similar tool marks on a number of barrels before being modified either through use or refinishing. Another example is a broach that cuts rather small surfaces such that the tool surface does not change markedly between consecutively manufactured pieces. The difficulty of addressing subclass characteristics is not in debate. However, to suggest, as one recent critic has, that despite this knowledge the discipline has done little to deal with the issue (5) displays an ignorance of the large body of work directed at this very issue. It also displays a lack of fundamental knowledge with regard to potential subclass transference and how this affects comparative examinations.

No fewer than 19 references address subclass characteristics. The value of these references is that they allow the examiner to move from a knowledge of manufacturing processes to generalizations with regard to the potential for subclass interference and then onto specific case application.

In 1949, Churchman (11) observed subclass characteristics on bullets that had been fired from consecutively made, broach-cut rifled, rifle barrels. In 1975, Skolrood (12) made similar observations when examining three similar barrels (although now being manufactured by a company different than when Churchman carried out his study). Although not designated as subclass characteristics, Lomoro (13) observed "family characteristics" on bullets fired from different guns. This carryover was only on the groove impressions and attributed to a worn or very poor rifling tool used to cut the grooves.

These three studies linked subclass characteristics on groove impressions with broach or otherwise cut rifling, allowing the examiner to move from generalizations to specific application. In cut rifling, the metal of the barrel (grooves only) is cut away by a bladed tool. If the surface of the cutting tool does not change significantly from barrel to barrel as it cuts, similar markings may be present on multiple barrels manufactured consecutively. If one were to examine a cast of the bore of a firearm, such characteristics would have to exist along the entire length of the cut surface for potential subclass influence to exist. If they did not persist along the entire length, it means the characteristics were either not present on the previously rifled barrel or on the next in sequence. Therefore, the only characteristics capable of being defined as subclass would be those that persist for the entire length of the cut surface. This is why a conscientious examiner will not rely solely on striated correspondence in groove impressions on fired bullets if subclass characteristics cannot be ruled out.

Murdock (14) recognized a significant issue in that some barrels were not formed with a cutting process but a swaging process. In such a process, the barrel is drilled (leaving tool marks perpendicular to the axis of bullet travel) and a button is passed down the barrel. Having a negative impression of the rifling, the button actually pushes metal out of the way, swaging the rifling instead of cutting it. Such a process is significantly different than the cutting approach because in a swaging method no metal is removed.

When a button is passed down a barrel, it does so under a tremendous amount of pressure. As such, it tends to polish tool marks that are already present (from the drilling process) and not impart any other markings except those that appear as imperfections on the portion of the button that comes into actual contact with the bore. This particular issue was observed to be taking place when Matty (15) examined bullets from barrels produced from one button-rifled blank (one long button-rifled barrel sectioned into three smaller barrels).

Biasotti addresses both of these general types of rifling methods (cut and swage). He offers reasoning as to why subclass characteristics are not necessarily common and offers some appropriate words of caution to an examiner:

> Two factors virtually assure that a unique set of individual characteristics will be reproduced in barrels rifled consecutively by the current rifling methods evaluated [hook cutter, scrape cutter, broach cutter, button swage, and hammer swage (forge)]. The first is the random nature and rapidity with which the toolmarks produced by "cut" type rifling methods change within a single barrel, or consecutively rifled barrels. Secondly, the toolmarks remaining in "swage" type rifling are predominately perpendicular to the axis of bullet travel. A possible exception to this generalization is the rare case where barrel blanks, are cut into multiple barrels; or where a swage or broach rifling tool with gross defects is capable of producing axial toolmarks that can be seen to extend the entire length of the bore. This latter case should present a problem to the examiner only where the questioned barrel is not available for examination. In those cases where the barrel is not available for examination, the examiner should use the toolmarks made by the lands or forcing cone to confirm an identification (16).

In a study of the same broach-cut rifled barrels used by Biasotti (17) in his CMS study, Tulleners and Hamiel (18) examined both lead and jacketed bullets specifically for subclass characteristics. In their discussion they report,

> These subclass characteristics were present on some, but not all of the ten sequential barrels and in some but not all of the groove impressions . . . These subclass characteristics were not found on the land impressions of the fired lead bullets or on the land or groove impressions of the copper-jacketed bullets.

Firearm parts other than barrels have also been studied. One of the first was a study performed on consecutively manufactured Smith and Wesson firing pins (19). It was observed that the circumferential tool marks on the surface of the firing pins, caused by a lathe-turning process, displayed remarkable similarity among the firing pins. As a result, firearm and tool mark examiners are aware that such marks are not wholly reliable for identification to a specific firearm.

Breech face marks can be cut, milled, or stamped. In each instance, subclass characteristics may be produced (20–23). As a result of such studies, firearm and tool mark examiners are aware that such processes can result in subclass characteristics. When suspicion of subclass is high and cannot be resolved, conscientious examiners will routinely look to other marks, such as chamber marks, that are not as susceptible to subclass influence.

Advances in technology have included the use of computer numerical-controlled (CNC) machining for more efficient tooling of various tools, including parts of firearms. Such machining has

> "allowed many different tooling operations that might be performed on a single piece by multiple operators to be performed by a single machine equipped with a wide range of various tools operated by a single individual. The concepts of the tooling are the same with the added variable of more precise tool placement from object to object (7)."

Despite observing subclass characteristics on bolt faces that were produced through the use of CNC machining, they were not sufficient to result in a misidentification (24). In a similar study involving anvil marks on .22 caliber cartridge case rims, the author observed significant subclass characteristics to exist on the breech end (not the bore but the rear face of the barrel) on consecutively machined barrels (25).

Ten consecutively made extractors were recently studied for subclass characteristics (7). In this study, it was observed that subclass characteristics persisted on two of the machined surfaces of the extractor. Detailed with photographs, the study demonstrated the importance of not only the presence of subclass characteristics but also the importance of understanding how tools and surfaces interact to determine whether the subclass characteristics present are even relevant. Specifically,

> Two of the extractor surfaces exhibited significant subclass carryover among all ten extractors. One of the surfaces was on the beveled surface on the forward edge of the extractor hook …. The other surface was on the underside of the hook, limited to the area adjacent to the beveled surface at the base of the channel of the extractor hook …. Yet, results demonstrate that the presence of such subclass characteristics did not have any impact on the ability to distinguish between marks produced by each of the ten extractors. One likely reason is the ridge that is formed on the corners to which these surfaces are adjacent. … they [ridges] protrude away from the flat and beveled surfaces of the hook and are the common result of tooling different surfaces that share a common corner. It is apparent that these ridges are having a significant impact on the tool marks produced by the extractor, *so much so that the issue of significant subclass characteristics is negated* [emphasis added] (7).

Tools other than firearms have also been studied. In 1968, Burd and Kirk (26) demonstrated that if the tips of screwdrivers are not subsequently finished, then the stamping or die process used to manufacture them could be a source of subclass characteristics. While subclass characteristics were not observed on the teeth of consecutively broach cut pliers, Cassidy (27) observes that in the normal use of the tool they would not have been relevant anyway.

In some instances, molds are used to produce items; in such instances, it is important to understand the molding process and how such marks may persist across many items from a single mold (28) or across several molds produced from a single master mold (29).

It is important to not only understand the potential of a tool surface to have subclass characteristics but also the action of the tool on an object. Such sentiments were evident in studies performed by Thompson (30) when dealing with stamped and painted breech faces of Lorcin pistols and Moran (31) when dealing with lips on an ammunition magazine. The latter article details manufacture, potential for subclass, and potential for transference of such marks to a cartridge case.

The second variable of importance when dealing with identification criteria is the rate of change of the tool working surface. Simply put, does the tool working surface change too quickly such that individualization is not possible or even practical? It is important to understand that it has never been asserted that characteristics on tool working surfaces would not change. At the same time, a change on a tool working surface does not necessarily negate the potential for a qualified examiner to examine two tool marks and determine that they were produced by the same source. Schwartz (5) asserts otherwise, citing this fact as a "barrier in the way of firearms and toolmark identification's goal of individualization." Others have expressed similar concerns, stating, "Unlike DNA or fingerprints, markings left by an individual gun on ammunition fired through it are neither unique nor permanent (1)."

The surface of a tool may change over time but this does not make identification unreliable. This is true for two reasons. The first is that, through use, a tool will continue to acquire individual characteristics that are vital to the comparative identification process. The second is that were the change of a tool surface so rapid as to change from mark to mark (or bullet to bullet) then any attempt at identification would be impossible. However, with the possible exception of the first series of bullets fired from a newly manufactured barrel, published studies have shown otherwise.

Hamby (32) test fired 501 bullets in a 5.56 NATO caliber M16A1 military rifle. Approximately 40,000 other rounds had been fired previously through this barrel. Every effort was made to make the conditions as deleterious as possible including test firing as rapidly as possible. The first bullet and every 100th after that were collected for comparison. Sufficient similarity of individual markings permited a conclusion that the first and last bullets were fired in the same firearm.

Biasotti (33) performed a limited study that examined the effects of lead build-up in a .22 caliber barrel. He demonstrated that lead buildup in a barrel from successive firings of lead bullets can cause markings to change; cleaning of the barrel with a solvent and brush may be necessary to remove the deleterious leading. He concluded that the best reproducibility for lead bullets was between bullets fired with similar bore conditions.

Shem and Striupaitis (34) performed a study of 501 test fires from a Raven, .25 Auto caliber, semi-automatic pistol. The first and every 10th set of test fires were recovered with comparison between the first and every 50th set of test fires. It was still possible to conclude that the first and last bullets were fired from the same firearm; however, a gradual change in the individual characteristics on the bullets was observed. With regard to the cartridge cases, the individual markings within the breech face markings were sufficient to permit a conclusion that the first and last test-fired cartridge cases were fired in the same firearm.

In a study similar to Biasotti's, Kirby (35) examined the effect of firing 900 cartridges from a .455 caliber Smith and Wesson revolver on individual markings produced on cartridge cases and bullets. Lead bullets were fired through the barrel and the firearm was not cleaned during the test. No significant differences were observed in the firing pin impressions and breech face markings on the first and last test-fired cartridge cases, and it could be concluded that each was fired in the same firearm. The bullets, however, revealed a different situation. The author found no difficulty in determining that the first and 25th bullets were fired in the same firearm. Indeed, some differences were noted by the 50th test-fired bullet but the coarser individual striations showed little to no change. Twenty-five bullets later it could not be concluded that the first and 75th bullets fired were fired from the same gun. There was some similarity but it was insufficient for an unequivocal identification. Further testing showed that test fired bullets #125 and #150 showed sufficient similarity to conclude that they were fired in the same firearm. Like Biasotti's study, those lead bullets fired with similar bore conditions could be compared and a conclusion could be reached that they were fired from the same firearm.

In 1983, several authors collaborated on a study of 5000 full metal jacketed, .45 ACP caliber bullets fired from an M1911A1 semi-automatic pistol (36). Every 10th test-fired bullet and cartridge case were recovered for comparison. The breech face marks showed no significant change with slight form variations of the firing pin and extractor. The ejector marks, however, changed at a relatively rapid rate. With regard to bullets, it was observed that while some land impressions showed a faster relative change of some individual markings than others, a conclusion that the two bullets were fired from the same firearm was possible through all 5000 test-fired bullets.

Interested in ejector marks, Schecter and colleagues performed a study in which they fired 7100 cartridges in a 5.56×45 mm Galil (37) rifle. They observed change within the first several test fires, but once the ejector surface had stabilized, the ejector marks on test-fired cartridge cases nine and 7060 permited a conclusion that the same ejector was responsible for producing the mark.

Most recently, Doelling (38) reported on the persistence of individual markings over the course of 4000 test-fired bullets. He was able to determine that the first and last test-fired bullets could be identified as having been fired from the same firearm.

Hall (39) also addressed this issue when he wished to determine the persistence of tool marks produced by bolt cutters. When the marks were produced in lead, Hall saw no difference in marks produced by any of the bolt cutters to a maximum of 25 cuts. He did notice a difference in markings when the bolt cutters were used to cut lock shackles, but he indicated that this appeared to be more of an issue of the shackle hardness creating reproducibility problems.

It is recognized that a tool surface will change over time. However, the suggestion that individualization to a specific tool is therefore invalid is not an appropriate extension of the concern. The issue has been recognized and studied within the discipline. There will be differences in individual detail from mark to mark produced by the same tool. But, the change is neither rapid enough to devalue firearms and tool marks as an identification science, nor is it necessarily significant enough such that an identification criterion based on similarities cannot be established. Furthermore, the worst possible scenario is that a particular mark will not be able to be associated with the tool from which it was made because the working surface of the tool has changed, thereby not permitting identification.

Criteria for Identification—Objective or Subjective

The answer to the question of whether firearms and tool mark identification is objective or subjective is that it is both. Despite the contention on the part of some that the discipline is entirely subjective, the truth is that it is subjective only in part. Simply put, the observations are objective and the interpretation of those observations is subjective. Where subjectivity comes into play is in the interpretation of the objective observations. For example, consider a series of striated marks on a land impression of a bullet. They have a discrete location on the land impression a certain distance from the shoulder. They have a particular width and depth. They also have a sequence or pattern appearance. Could these be measured? They could but in practice they are not. The reason is that comparative microscopy is an efficient and effective means by which it can be determined whether this series of striated marks on one land impression is in the same relative location and has the same relative physical features as that on another. This is an objective series of observations even though they remain unmeasured. While some might suggest that even these observations are subjective to some extent, well-documented comparison microscopy is extremely effective at minimizing that subjectivity.

Given the reduction of subjectivity in the comparative process, it is now important to consider the subjectivity in the interpretation of those observations. One area that would prove useful in reducing the subjectivity in interpreting observed correspondence would be to better standardize interpretation. The AFTE Theory of Identification states that observed correspondence must exceed that expected between marks produced by different tools and consistent with that expected from marks produced by the same tool. Two issues are important to discuss. The first is that there is no universal agreement as to how much correspondence exceeds the best-known nonmatching situation. The second is that in practice, this limitation is not as significant as critics contend. These will be examined in turn.

An element of an adequate training regimen for a firearms and tool mark examiner will include hundreds of hours comparing marks produced by different tools to determine the amount of coincidental correspondence that can be expected in such situations. It is through such an exercise that an examiner can establish a baseline for identification. While oft-criticized, the concept of "I know a match when I see it" has its basis in such training. The reason is that such comparative examinations store the best coincidental correspondence into one's experiential database. It could be akin to taking the same route to and from work everyday. Before long individuals have difficulty giving street names because they never look at the street names—they simply go in a particular pattern because it is an ingrained experiential element of their daily life.

Experience cannot be ignored because it is vital to an examiner's ability to interpret correctly the correspondence that is observed. While one's definition of their best-known nonmatch may not be easy for others to visualize, routine proficiency testing may help to determine whether an individual's concept of the best-known nonmatch is sufficient to distinguish between tool marks produced by different tools. Exposure to continued research also allows one's concept of a best-known nonmatch to go through further testing and, if necessary, refinement.

A limitation of this concept that many critics point to is that while the concept of correspondence exceeding that observed in a best-known nonmatch situation is a standard ideal, the actual definition of that will be different between examiners because they have different experiences. For example, an examiner in California has access to certain training materials dealing with comparing known nonmatches that establish a baseline correspondence. It is very likely that an examiner in the Northeast has different materials and will therefore develop a different experiential concept of the best-known nonmatch.

These differences in best-known nonmatches are not as detrimental to the discipline as critics would like the court to believe. The first reason is that the majority of comparative casework involves comparisons having correspondence either not closely approaching this level or correspondence far in excess. Support is also found in the overall performance on standardized proficiency tests and other validation studies distributed by researchers within the discipline. To be discussed in detail later, such results show that slight variations in best-known nonmatch concepts are not all that significant.

Given all this, it is important to discuss the attempts that have been made to standardize the concept of the best-known nonmatch discipline-wide. The primary advantage to standardization is that examiners are no longer limited to their own experiential knowledge, as extensive as that may be, but are now able to draw upon the experiential knowledge of others. One attempt that is finding support is the concept of consecutive matching striations (CMS) for striated tool marks. Critics have mistakenly referred to CMS as a more objective method of comparison than the traditional approach of comparing and evaluating the significance of correspondence within a pattern against one's experiential database of known nonmatch correspondence. CMS is not a more objective way of performing comparative examinations but simply a means by which an examiner can describe what he or she is observing in a striated tool mark comparison. Rather than trying to give a coworker directions from work to your home in a relative, descriptive, "take the second right and then turn left at the large tree" manner, one is able to give discrete directions with street names. Is the first possible? Yes, but it is best when actually experienced. The second manner is sufficiently descriptive without having to be experienced first.

The minimum conservative quantitative criterion for a striated tool mark offered by Biasotti and Murdock (40) is nothing more than a description of the correspondence necessary to exceed that expected in a best-known nonmatch situation. It is not a more objective means of identification nor is it a different method than has been practiced throughout the years. It has been argued before though that the ability to draw on the experience and observations of others, easily done through descriptive CMS terminology, has the appearance of greater objectivity than relying upon one's own experience and training, as extensive and as valuable as that is (41).

CMS criterion has two components: one for two-dimensional tool marks and another for three-dimensional tool marks. This has resulted in confusion because in the most critical sense, all tool marks are three-dimensional. While essentially true, not all tool marks are sufficiently deep to allow sufficient interpretation of the vertical (height) aspect of the striated marks; this leads to a greater reliance on the horizontal aspects of position and width. The different components of CMS criteria adjust for the increased conservatism that an examiner should exercise when dealing with such shallow tool marks.

The limitation, of course, is that CMS is restricted to striated tool marks or tool marks representative of a striated tool mark (such as striated tool marks on a breech face impressed onto the head of a cartridge case). Impression tool marks are not so easily standardized. The difficulty is readily apparent: an amorphous, irregular three-dimensional characteristic within a tool mark is not well suited to a numerical description, as unique as that characteristic may be. Regardless of this difficulty, promising models are being developed that might be of assistance in the very near future.

Identifications—Practically Speaking or Absolute

Once an examiner reaches a conclusion of identification, it is important to discern exactly what is meant by that term. A recent criticism cited that absolute individuality is the goal:

> The expert testimony in the case, *United States v. Kain*, was *typical* [emphasis added] of that offered by firearms and toolmark examiners. The goal of the forensic science discipline of firearms and toolmark identifications 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 (5).

The AFTE Theory of Identification, the statement of the relevant scientific community, does not make claims of absolute identification. The AFTE Theory of Identification states, "The statement that 'sufficient agreement' exists between the two toolmarks means that the likelihood another tool could have made the mark is so remote as to be considered a practical impossibility" (10). This is not a statement of absoluteness.

Despite the official published position of the AFTE, Schwartz's point that testimony of firearm and tool mark examiners is *typical* in that claims of absolute identity are made cannot be denied. For the purposes of clarity, examiners should communicate that conclusions of identity are reached because the chances of another tool producing the same mark are so remote that for practical purposes, it can be ignored. Schwartz writes, "Firearms and toolmark examiners' absolute identity conclusions cannot be excused on the ground they are convenient shorthand for well-grounded probabilistic conclusions" (5). This author would agree that language not offering the full meaning and intent of the AFTE Theory of Identification is not appropriate.

The issue at the root of this is not a new one. Kirk (42) recognized this question of absolute identity versus practical identity as a source of much "quibbling of attorneys with expert witnesses." Emphasizing this distinction and the importance of clear articulation, Kirk writes,

> In all matters involved in the examination and interpretation of physical evidence, the term identity must be understood to signify practical and determinable identity only. If necessary, the witness must be very willing to admit that he has not and cannot ever establish absolute identity, and in fact there is no such thing when applied to tangible objects.

Furthermore, Kirk cautions that "accurate identification must rest on a proper basis of training, experience, technical knowledge and skill and an understanding of the fundamental nature of identity itself. It should not be attempted without this kind of background, either by the police officer or the amateur."

In 1991, Stoney (43) discussed individualization as being analogous to a "leap of faith" when addressing statistics in the framework of fingerprints and (at the time) newly emergent DNA analysis. Stoney's claim was that we move from a subjective interpretation of the observed characteristics (in Stoney's example, it was fingerprints) and declare an absolute identity. Stoney writes,

> The conclusions [of a fingerprint examiner] are accepted and supported as subjective; very convincing, undoubtedly valid, but subjective. In fingerprint comparisons, the examiner notes the details in the patterns of the ridges. Beginning with a reference point in one pattern, a corresponding point in a second pattern is sought. From this initial point the examiner then seeks neighboring details that correspond in their form, position, and orientation. These features have an extreme variability that is readily appreciated intuitively, and which becomes objectively obvious upon detailed study. When more and more corresponding features are found between two patterns, scientist and lay person alike become subjectively certain that the patterns could not be possibly duplicated by chance. What has happened here is somewhat analogous to a leap of faith. It is a jump, an extrapolation, based on the observation of highly variable traits among a few characteristics, and then considering the case of many characteristics. Duplication is inconceivable to the rational mind and we conclude that there is an absolute identity.

Stoney moves on to suggest that trying to "prove uniqueness" is a "ridiculous notion." Using the discipline of fingerprints, he comments, "We hold fingerprint specificity and individuality up as our ideal, yet this is achieved only through a subjective process. In fingerprint work, we become subjectively convinced of identity; we do not prove it. And this works just fine. For fingerprints [contrasted with DNA]." He then concludes by saying, "Even without theoretical models and statistics, we can, and do, make absolute identifications. We can apply scientific, critical judgment, expert and informed, to make the subjective determination of identity (or less absolutely, of 'very very rare')."

The logical extension of this issue is a definition of that "likelihood" of a "practical impossibility" in the AFTE Theory of Identification or the more precise definition of, in Stoney's words, "very, very rare." In order to answer this question, it makes sense to examine the question of statistics and its applicability in the discipline of firearms and tool marks identification. Although often cited as an ideal, the model for statistical treatment in DNA, which works well for DNA, is not applicable to all the other forensic science disciplines. First, it is important to remember that the characteristics being compared in DNA profiles are actually subclass characteristics (44). It is the relatively rare frequency with which a combination of subclass characteristics occur in a given population that allows DNA to approach, although not fully achieve, individualization. In firearms and tool mark identification, examiners are not basing their identifications on subclass characteristics or frequencies of combinations thereof, but individual characteristics.

Much confusion in the area of statistical databases for firearms and tool mark identification exists because of the misguided comparisons with DNA analysis so different from firearms and tool mark identification that analogies may be intellectually inappropriate. Furthermore, an examination of the typical arguments proposing analogous databases upon which to draw frequencies demonstrate a lack of fuller understanding of the real relevant issue at hand—it is not necessarily the tool itself, but, rather, the manufacturing process by which the working surface of the tool is formed that is the critical feature in the scientific basis of firearm and tool mark identification.

Many studies have demonstrated that tool marks produced by different tool manufacturing processes can be readily distinguished. Furthermore, research has focused on defining more discretely the identification criteria by which this is done. The implied need for representative statistical databases for each and every tool one might encounter is not founded because the science of firearm and tool mark identification is based on manufacturing methods and an ability to assess and distinguish among the class, subclass, and individual characteristics produced by the tool manufacturing process.

However, that being said it is important not to ignore a pertinent and very relative question. That question is, "Is there a role for statistics in the discipline of firearms and tool mark identification?" This attempts to address a common concern of the critics from a broader perspective. The following discussion will be devoted to examining that very question.

Schwartz (5) claims that "Firearms and toolmark examiners do not even attempt to answer this question." While that may be true of the testimony she has observed, within the published literature it is not. Biasotti (17) made an early attempt in an article published in 1959. In 1970, Brackett (45) examined the use of various models to study "idealized" striated marks. These "idealized" marks consisted of individual elements within a set of striations defined by position only, without the additional defining characteristics of width, contour, or height. The purpose of these models was to examine statistical and probabilistic application to striated tool marks.

Blackwell and Framan (46) ran simulated studies based on Brackett's formulae and models resulting in numbers similar to those produced by Biasotti in 1959. Uchiyama was responsible for another computer simulation granting greater than practical tolerances for striation correspondence and produced numbers similar to those of Biasotti (47). In his article, he developed a probability equation and a significance level based on actual, test-fired bullets. Deinet (48) published a study the purpose of which to "calculate the probability of random occurrence of matches using actual striated tool marks" (8).

There have also been more recent attempts to answer a statistical question. Miller and Neel (49) evaluated the statistical significance of various runs of CMS for 1000 comparisons. Stone (50) entured into a mathematical model to describe the probabilities of impressed tool marks on a theoretical hammer face. Just recently, Collins (51) has offered a follow-up to Stone's model by empirically assessing such marks on 20 hammer faces.

The literature indicates that firearm and tool mark examiners have explored the area of statistics in an effort to determine its potential utility in the discipline. It could be that there was early recognition that an individual examiner, at best, could examine only a small fraction of the firearms that exist. Yet, using probabilities, an examiner could discuss the significance of an identification (52).

Dissenters exist, however. Deschênes et al. (53) cite two objections: The first is that, "statistics never permit to draw conclusions concerning a particular situation." In support of, this they use a weather analogy. "It is not going to rain just because there are 97% chances that it is going to rain. Statistics do not yield a 'cause to effect' relationship." The second is that a firearms and tool mark examiner is in a better position to interpret the meaning of what is being observed.

This article received some relatively rapid criticism (54). The criticism focused on the fact that the science of statistics does have a role to play and that is in the area of uncertainty. They argue that because the examiner does not have a complete set of circumstances regarding a particular tool,

> ". . . the tool mark examiner is never in a position to identify a tool. But when considering the whole population of the world, the expert estimates that the probability of another match is very close to zero then it is common sense to declare an identification."

Bunch's (55) article supports a similar view. He states that firearm and tool mark examiners have the goal of determining the likelihood ratio that a tool mark was made by a particular tool. Use of the word "likelihood" or phrase "likelihood ratio" implies reference to Bayesian inference. Indeed, some favor a Bayesian approach because it allows an assessment of more than just the questions of the comparative results; however, it may not answer the question as discretely as the judicial system may like. It is true that numbers representing a likelihood ratio are generated, but the explanation for what those numbers mean in a real sense leaves the judicial system no closer to a real answer than what is being offered currently.

A different approach is a more routine probabilistic (frequentist) approach. Examples include the aforementioned works by Brackett, Miller and Neel, Stone, and Collins. In order to assess the appropriateness of these statistics, it is important to determine whether or not the individual marks are truly independent of one another. If they are, then it is appropriate to use mathematical and statistical tools. If they are not, then other tools, mostly additive, must be used.

The question of Bayesian versus frequentist statistics has not seen resolution. Two primary articles in support of Bayesian inference used it as a framework to critique the concept of CMS (55,56). Strong responses to those articles suggested that the connections being drawn were not valid but, rather, based in a misunderstanding of the concept of CMS and the practice of those utilizing it (9,57).

The potential role for statistics in the firearms and tool mark discipline has been and continues to be studied contrary to Schwartz's assertion. While statistics may have some utility in its current form, the debate among the scientific community continues and the results published widely.

Testing the Discipline—Proficiency Tests and Validation Studies

While less than ideal, some have suggested that proficiency tests may be of value in providing a general indicator of how often false identifications are made in comparative examinations (58). As recognized, individuality cannot be absolutely proven because it is impossible for an examiner to examine every tool in the world to a tool mark in question. Furthermore, because of the difficulty in assessing the nonquantitative aspects of firearms and tool mark identification, statistics cannot wholly answer the question. In this regard, Gutowski is correct in his assessment that,

> An estimate of the actual or potential error rate is crucial to the probative value of all evidence. This is certainly true of the field and identifications sciences where hard statistics on the frequency of occurrence of a particular pattern are impossible to come by and individuality is assumed but cannot be proven (59).

Therefore, proficiency tests can offer to the court a reliable practical indicator of how often the profession, using accepted procedures, practices, and controls, makes a false identification:

> The statement that the science of firearm and toolmark identification has a "0%" error rate is clearly not responsive to the court when questions of error rate are brought forward. The court is not interested in "theoretical error rate", which assumes everything has been done correctly and the correct answer obtained, but is interested in the real life potential error rate that is reflective of all human endeavors. . ..To proffer that firearm and toolmark identification is "infallible" is simply not true and will be met with immediate suspicion. The court is interested in "known or potential error rate" as a means by which to assign weight to the examiners testimony. The examiner will be more credible by readily discussing the reported error rates in the process of firearm and toolmark identification (*i.e.*, the first half of the *Daubert* element) and then be prepared to discuss what steps have been taken as an individual and through laboratory peer and administrative review processes to eliminate the possibility of error in the work currently being presented in court (i.e., the second half of the element) (58).

Grzybowski et al. (58) recognize that even with their limitations, "Collaborative Testing Service (CTS) is currently the only source of international proficiency testing results in the firearm and toolmark identification discipline from which a source of potential error rate may be inferred." Given that, the authors provide a review of the Peterson and Markham (60) data in addition to CTS data subsequent to that examined by Peterson and Markham with the specification that inconclusive conclusions are not necessarily incorrect or correct. Therefore, such inconclusive conclusions will not be deemed as incorrect responses as was done by Peterson and Markham.

Given this structure of examination, Robert Thompson assessed the CTS data for two time periods, the first 1978 through 1997 (the same as Peterson and Markham) and 1998–2002. The percentage of false identifications for firearms in the time period 1978–2002 was 1.0%. The percentage of false identifications for tool marks in the time period 1981–2002 was 1.3% (58). Based on this evaluation the authors offer the following:

> So, what does this mean for the individual examiner? The examiner must first acknowledge that errors can be made. The examiner must then be prepared to discuss the CTS tests and their limitations, and recognize that, despite their limitations, they may offer the court some indication of error. It does not mean, for example, in the instance of a 1.5% CTS error rate, that every toolmark identification case report is subject to being right only 98.5% of the time, but rather that for all those respondents, 1.5% made an incorrect association. Secondly, assuming that the work <u>has</u> been done thoroughly and the conclusions fully supported by clear and complete notes, it is suggested that examiners advocate that it's his/her opinion that he/ she has made no error in the case at hand. It is easier to

> convince others of this if: 1) he or she has graphically demonstrated the basis for the opinion with the use of photographs; 2) comprehensive notes have been taken that fully support the conclusions in the lab report and; 3) the examiner's work has been technically peer reviewed and administratively reviewed per ASCLD/LAB requirements (whether or not the individual's laboratory participates in this program). Such actions would serve to further minimize any reasonable chance of error in reaching a correct conclusion and will be persuasive to those in court responsible for determining the weight to be accorded the examiner's testimony.

Recent validation studies might also assist in this aspect. In 1992, Brundage (61) reported on a study of 10 consecutively step-broached 9-mm Luger caliber barrels. He provided thirty different laboratories across the country with pairs of test fires from each of the 10 barrels along with 15 unknowns, with at least one from each of the 10 barrels. In each and every instance, the unknowns were properly associated with the barrel from which they were fired. At the 2003 AFTE Training Seminar, Hamby reported that 294 different examiners from 15 countries had examined and compared the bullets without a single instance of a misidentification (62).

A study that involved the 10 consecutively manufactured knives was reported in 2003 (63). The authors obtained 10 consecutively manufactured knives and produced a series of test marks and questioned marks. The final sharpening was accomplished with a 24 in. diameter grinding wheel. One hundred and three examiners provided a total of 1030 results (10 questioned marks per examiner). Of the 1030 results, there were eight errors for a calculated false identification rate of 0.77%.

A third study involved cartridge cases fired using 10 Glock pistols (64). The total number of comparisons conducted was 360 with no errors reported.

A fourth study was directed at assessing the validity of the CMS criteria for two dimensional tool marks (65). If one considers CMS to be a validate representation of a comparative examination of a striated tool mark comparison, then it may be of interest to note that of 1000 known nonmatch comparisons, not one violated the CMS criteria for two dimensional tool marks.

While valuable, the validation studies provide only a part of the picture. Proficiency tests may offer an assessment of laboratory practice, quality assurance, and quality control procedures. In addition, the wide range of proficiency tests offered involves tools and firearms from a variety of manufacturing methods. In combination, the material offered provides a good picture of how often the profession will make an incorrect association. An alternative is offered by Gutowski, who writes,

> A better estimation of error rate in casework would be most rigorously achieved by the re-examination of several thousand cases where each case was examined by a panel of experts to achieve consensus. In the absence of a massive increase in funding, this is unlikely to happen (59).

Conclusions

A careful study of the available literature supports the following:

(1) Firearms and tool mark identification is rooted in firm scientific foundations.
(2) Firearms and tool mark identification has been critically studied according to the precepts of the scientific method culminating in the AFTE Theory of Identification.
(3) If properly accounted for issues of differing identification criteria, subclass characteristics, and the change a tool surface undergoes over time do not invalidate the firearms and tool mark discipline as a science nor should it detract it from its admissibility in a court of law.
(4) The firearms and tool mark identification discipline has been validated in a manner appropriate for evidence of the kind to be expected in firearms and tool mark examinations; and
(5) Proficiency tests and error rates have been studied and can provide the court and community with a useful guide as to the frequency with which misidentifications are reported in the community using appropriate methodologies and controls.

As demonstrated through this response, the material to support the scientific background of the firearms and tool mark identification discipline is available in the relevant professional literature. What remains is for examiners to put this into practice through quality casework and then, just as importantly, learn how to communicate the intricacies of the discipline to a nonscientific audience. Even with the wealth of support, the burden is not on the critics to stop criticizing or the courts simply to accept what is said carte blanche. The burden is on the individual examiner to step up to the plate and be effective not only in their casework but also in their ability to communicate this knowledge.

References

1. Griffin J, LaMagna D. Daubert challenges to forensic evidence: ballistics next on the firing line. Champ 2002;20–23, 58–62.
2. Saks M. Implications of the Daubert test for forensic identification science. Shepard's Expert and Scientific Evidence 1994;1(3):427–32.
3. Saks M, Koehler J. The coming paradigm shift in forensic identification science. Science 2005;892–5.
4. Schwartz A. A challenge to the admissibility of firearms and toolmark identifications: amicus brief prepared on behalf of the defendant in United States v. Kain, Crim. No. 03-573-1 (E.D. PA. 2004). J Philos Sci Law 2004;4:1–32.
5. Schwartz A. A systemic challenge to the reliability and admissibility of firearms and toolmark identification. Col Sci Technol Law Rev 2005;6:1–42.
6. Steele L. All we want you to confirm is what you already know: a Daubert challenge to firearms identification. Crim L Bull 2002;38:466–507.
7. Nichols R. Firearm and tool mark identification: the scientific reliability and validity of the AFTE theory of identification discussed within the framework of a study of ten consecutively manufactured extractors. AFTE J 2004;36(1):67–88.
8. Nichols R. Firearm and toolmark identification criteria: a review of the literature. J Forensic Sci 1997;42(3):466–74.
9. Nichols R. Firearm and toolmark identification criteria: a review of the literature—part 2. J Forensic Sci 2003;48(2):318–27.
10. AFTE Criteria for Identification Committee. Theory of identification, range of striae comparison reports and modified glossary definitions—an AFTE criteria for identification committee report. AFTE J 1992;24(2):336–40.
11. Churchman J. The reproduction of characteristics in signatures of Cooey rifles. RCMP Gaz 1949;11(5):133–40.
12. Skolrood R. Comparison of bullets fired from consecutively rifled Cooey .22 calibre barrels. Can Soc Forensic Sci J 1975;8(2):49–52.
13. Lomoro V. Class characteristics of 32 SWL, F.I.E. Titanic revolvers. AFTE J 1974;6(2):18–21.
14. Murdock J. A general discussion of gun barrel individuality and an empirical assessment of the individuality of consecutively button rifled .22 caliber rifle barrels. AFTE J 1981;13(3):84–111.
15. Matty W. A comparison of three individual barrels produced from one button rifled barrel blank. AFTE J 1985;17(3):64–69.

16. Biasotti A. Rifling methods—a review and assessment of the individual characteristics produced. AFTE J 1981;13(3):34–61.
17. Biasotti A. A statistical study of the individual characteristics of fired bullets. J Forensic Sci 1959;4(1):34–50.
18. Tulleners F, Hamiel J. Sub class characteristics of sequentially rifled .38 special S&W revolver barrels. AFTE J 1999;31(2):117–22.
19. Matty W, Johnson T. A comparison of manufacturing marks on Smith & Wesson firing pins. AFTE J 1984;16(3):51–56.
20. Lardizabal P. Cartridge case study of the Heckler and Koch USP. AFTE J 1995;27(1):49–51.
21. Thompson E. False breech face id's. AFTE J 1996;28(2):95–6.
22. Matty W. Lorcin L9MM and L380 pistol breechface tool mark patterns. AFTE J 1999;31(2):134–7.
23. Lopez L, Grew S. Consecutively machined Ruger bolt faces. AFTE J 2000;32(1):19–24.
24. Coffman B. Computer numerical control (CNC) production tooling and repeatable characteristics on ten Remington model 870 production run breech bolts. AFTE J 2003;35(1):49–54.
25. Nies R. Anvil marks of the Ruger MKII target pistol—an example of subclass characteristics. AFTE J 2003;35(1):75–8.
26. Burd D, Kirk P. Tool marks: factors involved in their comparison and use as evidence. J Crim Law Crimin 1942;32(6):679–86.
27. Cassidy F. Examination of tool marks from sequentially manufactured tongue and groove pliers. J Forensic Sci 1980;25(4):796–809.
28. Kreiser J. Identification of cast bullets and their molds. AFTE J 1985;17(3):88–90.
29. Miller J. An introduction to the forensic examination of tool marks. AFTE J 2001;33(3):233–48.
30. Thompson E. Individual characteristics criteria. AFTE J 1998;30(2):276–9.
31. Moran B. The application of numerical criteria for identification in casework involving magazine marks and land impressions. AFTE J 2001;33(1):41–6.
32. Hamby J. Identification of projectiles. AFTE J 1974;6(5/6):22.
33. Biasotti A. Bullet bearing surface composition and rifling (bore) conditions as variables in the reproduction of individual characteristics on fired bullets. AFTE J 1981;13(2):94–102.
34. Shem R, Striupaitis P. Comparison of 501 consecutively fired bullets and cartridge cases from a .25 caliber Raven pistol. AFTE J 1983;15(3):109–12.
35. Kirby S. Comparison of 900 consecutively fired bullets and cartridge cases from a .455 caliber S&W revolver. AFTE J 1983;15(3):113–26.
36. Ogihara Y, Kubota M, Sanada M, Fukuda K, Uchiyama T, Hamby J. Comparison of 5000 consecutively fired bullets and cartridge cases from a .45 caliber M1911A1 pistol. AFTE J 1983;15(3):127–40.
37. Schecter B, Silverwater H, Etzion M. Extended firing of a Galil assault rifle. AFTE J 1992;24(1):37–45.
38. Doelling B. Comparison of 4000 consecutively fired, steel-jacketed bullets. In: Proceedings of the Annual Meeting of the American Academy of Forensic Sciences; February 19–24, 2001, Seattle, WA. Colorado Springs, CO: American Academy of Forensic Sciences, 2001.
39. Hall J. Consecutive cuts by bolt cutters and their effect on identification. AFTE J 1992;24(3):260–72.
40. Biasotti AA, Murdock J. Firearms and toolmark identification. In: Faigman DL, Kaye DK, Saks MJ, Sanders J, editors. Modern scientific evidence: the law and science of expert testimony, Vol. 2. St. Paul: West, 1997:124–55.
41. Nichols R. Consecutive matching striations (CMS): its definition, study and application in the discipline of firearms and tool mark identification. AFTE J 2003;35(3):298–306.
42. Kirk P. Crime investigation. New York: Interscience Publishers, 1953:14.
43. Stoney D. What made us think we could individualize using statistics? J Forensic Sci Soc 1991;31(2):197–9.
44. Gutowski S. A response to: a systematic challenge to the reliability and admissibility of firearms and toolmark identification, a recently published article by Adina Schwartz. For Bull Winter 2005.
45. Brackett J. A study of idealized striated marks and their comparison using models. J Forensic Sci Soc 1970;10(1):27–56.
46. Blackwell R, Framan E. Automated firearms identification systems AFIDS: phase I. AFTE J 1980;12(4):11–37.
47. Uchiyama T. A criterion for land mark identification. AFTE J 1988;20(3):236–51.
48. Deinet W. Studies of models of striated marks generated by random processes. J Forensic Sci Soc 26(1):35–50.
49. Miller J, Neel M. Criteria for identification of toolmarks part III supporting the conclusion. AFTE J 2004;36(1):7–38.
50. Stone R. How Unique are impressed toolmarks? AFTE J 2003;35(4):376–83.
51. Collins E. How 'unique' are impressed toolmarks? An empirical study of 20 worn hammer faces. AFTE J 2004;37(4):252–95.
52. Hatcher JS. Textbook of firearms investigation, identification and evidence. Marines, NC: Small Arms Technical Publishing Company, 1935:285–7.
53. Deschênes M, Chaltchi A, Desjardins G, Desrochers C, Dion J, Gaulin R. Statistics and toolmarks comparisons. AFTE J 1995;27(2):140–1.
54. Taroni F, Champod C, Margot P. Statistics: a future in tool marks comparison? AFTE J 1996;28(4):222–9.
55. Bunch S. Consecutive matching striation criteria: a general critique. J Forensic Sci 2000;45(5):955–62.
56. Champod C, Baldwin D, Taroni F, Buckleton JS. Firearms and tool marks identification: the Bayesian approach. AFTE J 2003;35(3):307–16.
57. Nichols R. Letter to the editor, Re: 'Firearm and tool marks identification: the Bayesian approach' Champod C. et al. AFTE J 2003;35(4):354–5.
58. Grzybowski R, Miller J, Moran B, Murdock J, Nichols R, Thompson R. Firearm/toolmark identification: passing the reliability test under federal and state evidentiary standards. AFTE J 2003;35(2):209–41.
59. Gutowski S. Error rates in the identification sciences. Forensic Bull Summer 2005.
60. Peterson J, Markham P. Crime laboratory testing results, 1978–1991, II: resolving questions of common origin. J Forensic Sci 1995;40(6):1009–29.
61. Brundage D. The identification of consecutively rifled gun barrels. In: Proceedings of the Annual Training Seminar of the Association of Firearm and Toolmark Examiners, June 1994, Indianapolis, IN. Chicago, IL: Association of Firearm and Toolmark Examiners, 1994.
62. Hamby J. An update on the identification of bullets fired from consecutively rifled 9 mm Ruger pistol barrels. In: Proceedings of the Annual Training Seminar of the Association of Firearm and Toolmark Examiners, May 25–30 2003, Philadelphia, PA. Chicago, IL: Association of Firearm and Toolmark Examiners, 2003.
63. Thompson E, Wyant R. KIP (Knife Identification Project). In: Proceedings of the Annual Training Seminar of the Association of Firearm and Toolmark Examiners, May 25–30 2003, Philadelphia, PA. Chicago, IL: Association of Firearm and Toolmark Examiners, 2003.
64. Bunch S, Murphy D. A comprehensive validity study for the forensic examination of cartridge cases. AFTE J 2003;35(2):201–3.
65. Miller J, Neel M. Criteria for identification of toolmarks part III supporting the conclusion. AFTE J 2004;36(1):7–38.

Additional information and reprint requests:
Ronald Nichols, M.Ch.M.
Bureau of Alcohol, Tobacco, Firearms and Explosives
Forensic Laboratory Services—San Francisco
355 North Wiget Lane
Walnut Creek, CA 94598
E-mail: ronald.nichols@atf.gov