IS STATISTICAL EVALUATION OF FORENSIC FIREARMS / TOOLMARKS EXAMINATIONS NECESSARY?

Carlo J. Rosati

Forensic examiner, retired from the FBI Laboratory, and adjunct instructor at Virginia Commonwealth University, Richmond, VA, USA

Abstract*. The paper describes the reflections of a senior forensic examiner about the admissibility of statistical testimony of firearms and tool marks examination in courts of law. Examiners compare items through visual observation and, when possible, state they are identified, inconclusive, or eliminated from comparison. The basic principle discussed in the paper is that of 'sufficient agreement' between items. A really complex one. Then the possibility of improving the effectiveness of examination conclusions through technology, statistical thinking and methods, and examiners training is discussed. The paper concludes that there is an effective need for introducing statistical evaluation of forensic firearms/ toolmarks examinations in juridical activities.*

Keywords*: Ammunition, Firearm, Tool mark, Statistical evaluation.*

1. INTRODUCTION

This paper focuses on the admissibility of forensic firearms and tool marks[1] identification issues and the statistical evaluation of the science of firearms and tool mark identification. In order to do that I first give you an overview of firearms identification and how examiners reach conclusion from examinations. Then I will follow with the question raised regarding the admissibility of the science of firearms identification, and will specifically address the prongs of Daubert followed by the research in the field and the need for statistical evaluation.

The Daubert trilogy of U.S. Supreme Court decisions comprises the Daubert, Joiner, and Kumho Tire cases, which, for the first time in the U.S., established a standard for admissibility of scientific testimony in courts of law. In synthesis, the questions were:

(a) Can and has the theory or technique been empirically tested?,

(b) Has the theory or technique been subjected to peer review and publication and accepted within the scientific community?, and

---

[1] The word "tool" refers to the rifled interior of a barrel and "toolmark" to the impressions on a fired bullet.

(c) Should the court ordinarily consider the known or potential error rate and the maintenance of standards controlling the techniques operation?

What examiners are really claiming is that bullets from the same barrel (this would include cartridge cases fired in a single firearm or a second firearm as well as various tools) often appear highly similar and bullets from different barrels virtually always appear highly dissimilar, and that qualified examiners can reliably distinguish between these two classes of things. Firearms examination requires a firearms examiner to make these distinctions. The firearms examiner is integral to the process of examination. Notice that this proposition can be fairly easily tested. There are other analogous propositions for eliminations, and for other types of examinations.

Consider that what was once widely accepted testimony by pattern based scientists has now been limited by courts in the United States. Courts have limited examiners from using words such as 'uniqueness' and 'practical certainty'. In reality, everything is unique if look closely enough which in this field is under microscopic examination. The key question is: "*Who can state two specimens are fired from the same barrel?*" Of course this brings the examiner right back into the science as an integral part of the process.

Proposition #1

Class and microscopic marks imparted to objects by different tools will rarely if ever display agreement sufficient to lead a qualified examiner to conclude the objects were marked by the same tool. That is, a qualified examiner will rarely if ever commit a false positive error (misidentification).

Proposition #2

Most manufacturing processes involve the transfer of rapidly changing or random microscopic marks onto barrel bores, breech face, firing pins, screwdriver blades, and the working surfaces of other common tools. This is caused principally by the phenomena of tool wear and chip formation, or by electrical/chemical erosion. As a matter of fact, microscopic marks on tools may then continue to change from further wear, corrosion, or abuse.

2. HOW CAN THE SCIENCE OF FIREARMS AND TOOL MARKS BE BETTER DEFENDED?

A hypothesis or technique is testable if it is possible to conduct experiments or controlled observations aimed to support or discredit the hypothesis (technique). To do this it must be possible to check propositions deduced from the hypothesis: if it is valid, that is if it explains something well, then something should occur under these conditions. If doesn't occur, then the hypothesis is discredited. Multiple meanings, but a valid hypothesis (technique) does what it is supposed to do.

Forensic science laboratories started with fingerprints, documents, firearms and tool marks known as the corner stone of forensic science. For more than 100 years examiners in these disciplines have been trained as pattern based scientists. Simply stated, examiners compare items through visual observation to each other and, when possible, they state they are identified, inconclusive, or eliminated from comparison.

Identification is the conclusion when the class characteristics between two specimens seem the same, and there is sufficient agreement in the individual characteristics to conclude that they were fired in the same firearm. For instance, two copper jacketed bullets are identified if they agree. "Agreement of a combination of individual characteristics and all discernible class characteristics is where the extent of agreement exceeds that which can occur in the comparison of tool marks made by different tools and is consistent with the agreement demonstrated by tool marks known to have been produced by the same tool." (see AFTE Glossary, AFTE range of conclusions, courtesy of Jack Dillon: www.afte.org).

Inconclusive agreement of class characteristics is defined as "the outcome of a comparison in which there is some agreement of individual characteristics and all discernible class characteristics, but insufficient for identification, agreement of all discernible class characteristics due to an absence, insufficient, or lack of reproducibility, agreement of all discernible class characteristics and disagreement of individual characteristics but insufficient for an elimination" (AFTE range of conclusions, courtesy of Bill Conrad, www.afte.org).

Elimination, or *exclusion*, from the analysis is caused by a significant disagreement of discernible class characteristics and/or individual characteristics. For fired bullet comparisons, an elimination is normally based on observed differences in anyone of the general rifling characteristics. However, an elimination based on individual characteristics is more complex. If it can be shown that a firearm has not been subjected to significant use or abuse over the period of time following the questioned shooting, the qualitative aspects of the striations (e.g. fineness, coarseness) it produces on the fired bullets should remain the same. A difference in these striations indicates an elimination.

Elimination based on individual characteristics requires a detailed history and treatment of the firearm, as well as documentation to support the history. It is the responsibility of the examiner to provide this historical documentation. This type of elimination should be approached with caution. Many experienced examiners have never made such an elimination and the protocols of many laboratories do not allow it.

Pattern based science has been published in journals and defended in court. However, recently pattern based science is under attack in the courts. The most difficult question asked of a forensic examiner is: *'What would be the statistical probability of two items producing the same pattern?'* In the field of firearms and tool marks the question may be phrased as: *'What is the statistical probability of two tools (firearms) producing the same tool marks?'*.

Examiners have to understand the universal scientific method. Though, first, all the sciences see it differently. Observational, theoretical, experimental, historical, and correlational sciences use different methods to support conclusions. In fact, testing of hypotheses, most of the time may not be done and both observations and hypotheses are fallible. There are many scientific methods, used by scientists some of the time. Second, there exist empiricists, intuitivists, false a fictionists, experimentalists, research programs, and other schools of thought. There can be problems with all of them: do not frame and falsify, and cannot really. Three, social sciences use the method of framing and testing more than the natural sciences do. Though many natural scientists argue social science is not science for various reasons. Line setting only by consensus opinion: astrology, no consensus; sociology, perhaps no consensus. All depends on one's definition, and all proposed so far have been the objects of criticism.

In the field of firearms and tool marks identification examiners have utilized the best equipment available (see also Section 4). The primary tool of the examiner is the comparison microscope (Figure 1).

Figure 1: The comparison microscope is the primary tool of an examiner. Examiners can view two objects at the same time through the optical bridge under magnification.

In order to understand the definition of firearms identification, it is necessary to understand what ammunition is, as well as what components make up ammunition. A single round of ammunition is made up of a bullet, gunpowder, a cartridge case, and a primer, which is located on the base of the cartridge case. Ammunition component fired in a firearm will have individual microscopic markings imparted on to them during the firing process. It is through the examination of the microscopic marks that examiners are able to identify the firearm that the ammunition component was fired in.

When a cartridge is discharged in a firearm, markings will be imparted onto the ammunition components as they come into forceful contact with the interior components of the firearm. When the trigger is pulled, the firing pin will strike the primer, causing the priming compound to explode. This explosion will ignite the gunpowder, which will result in an increase in pressure due to the burning of the propellant. The increase in pressure will forcibly propel the bullet down the barrel, and at the same time force the cartridge case into the breech face. In some firearms, the energy of discharge is utilized in cycling of the firearm, where the spent cartridge case is extracted from the chamber and ejected from the firearm. A new round of ammunition is then chambered and ready to repeat the process. This process imparts a number of markings on the ammunition, including the firing pin and breech face impressions, striations on the bullet, and ejector marks.

Modern rifles and handguns have helical grooves designed into the bore of the barrel so as to impart a gyroscopic spin on a projectile as it travels through the barrel. As with an (American) football, the gyroscopic spin on the bullet will cause it to be more stable in flight, and therefore more accurate. The lands, which are the raised areas in a rifled barrel, and the grooves, which are the lower areas on a rifled barrel, may come into contact with the projectile as it travels through the barrel. This will leave marks on the bullet that may be used to identify that particular firearm.

The lands in the bore will result in land impression on the bullet, and the grooves in the bore will result in groove impressions on the bullet. By observing the number of land and groove impressions, the direction of twist, as well as the widths of the land and groove impressions, we can gain some information as to the type of firearm that may have fired that bullet. A discrepancy in the class characteristics can be used to eliminate a particular firearm; however an examination of the individual characteristics is necessary to reach an identification conclusion.

During the manufacturing process, as well as through use and abuse and wear and tear, individual microscopic markings are imparted onto the surfaces of the firearm that come into contact with the ammunition components. The impression and striated marks imparted onto the head of the cartridge case can be used to identify the particular firearm in which a cartridge case was fired. For example, the hemispherical shape of the firing pin impression, and the parallel breech face marks are class characteristics of the firearm, and these characteristics are consistent within firearms made using the same manufacturing processes. It is the individual markings, such as the striations in the firing pin drag, or the craters near the firing pin impression, that are used for a conclusion.

During the manufacturing process, manufacturing tools continuously wear during use. The interaction between the tool and the work piece generates heat, and the friction generated as the cutting edge moves across the work piece results in the cutting edge to become worn and dull. The metal chips and shavings formed during the metal cutting process interact with the surface of the work piece, the edge of the cutting tool, as well as the interface between the cutting edge and the work surface. This results in further irregularities onto the work surface.

There are also a number of manufacturing processes that are random in nature, and impart random microscopic markings onto the work piece. Some of these processes include sanding, polishing, filing, sand blasting, tumbling, chemical etching, and electrical discharge machining. The continuously changing tool, or the random interaction between the tool and the work surface, results in individual microscopic irregularities.

After a firearm leaves the factory, it continues to change through use and abuse and wear and tear. It is these individual microscopic differences that make it possible for a qualified firearms examiner to identify the firearm in which an ammunition component was fired.

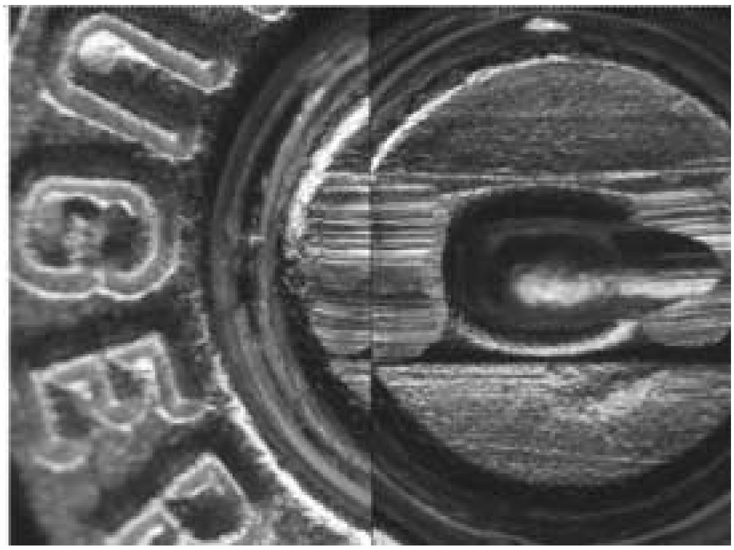

Figure 2: depicts two test-fired cartridge cases from the same firearm compared side-by-side using a comparison microscope. The marks from each cartridge show significant correspondence.

What would you see when looking into a comparison microscope? The image of the specimen on the left stage of the microscope and the specimen positioned on the right stage of the microscope can be seen at the same time, is separated by a hairline. During comparisons, examiners are looking at the primers of two fired cartridge cases. These cartridge cases were fired in the same firearm, and are therefore known matches. As you can see in Figure 2, there is a significant amount of correspondence in the striated firing pin aperture shear marks and also in the impression breech face marks.

Consider the case of two bullets fired through two barrels consecutively made at a factory. It is expected the highest amount of correspondence would be seen between two barrels that were consecutively manufactured, but as you can see in Figure 3, the amount of correspondence in the striated marks is significantly different between the specimen on the left side and the specimen on the right side. There is little, if any, correspondence in the striated firing pin aperture shear marks or the impression breech face marks. This is known as a 'known non-match', or two bullets known to have been fired through different firearm barrels.

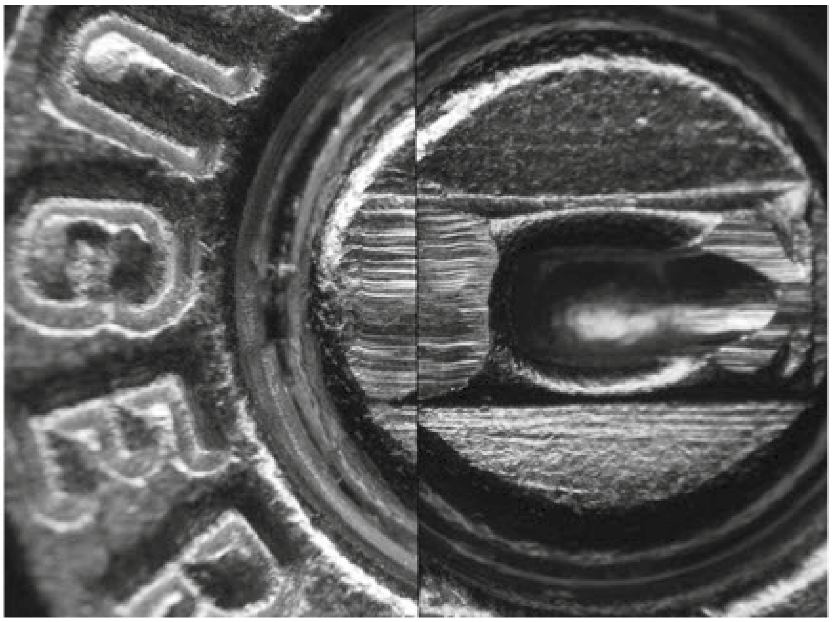

Figure 3 depicts the side-by-side comparison of two test-fired cartridge cases from consecutively manufactured (different) firearms. In contrast to Figure 2, the marks from each cartridge demonstrate a lack of significant correspondence.

It is important to have an understanding of what is meant by *objective* and *subjective*.

We call objective examination one that can be repeated over and over again by equivalent scientists and, if using the same equipment and following the same procedures, results in essentially the same conclusions every time, regardless of who is performing the examination. In firearms and tool marks cases, a number of objective examinations is conducted, including measuring the length of tool marks, widths of lands and grooves on bullets, diameters of bullets, weight of trigger pull, and the weights of bullets and bullet fragments.

As a matter of fact, a subjective examination is an individual's opinion. This does not mean that this type of examination is unreliable or unscientific. Every science has a level of subjectivity in every test, whether it being a doctor diagnosing a patient or a chemist determining the baseline on a scientific instrument. In a firearms and tool marks examination the conclusions as to a common origin depend on an examiner's training and experience. Two examiners having different training and/or experience may come to different conclusions, with one examiner identifying a common origin between two tool marks, and the other coming to a no-conclusion. Two differing opinions will rarely, if ever, claim an identification and an exclusion.

This calls to mind the basic concept of *sufficient, or significant agreement*. All identifications are based on pattern matching. The basic concept refers to the possibility of randomly duplicate the tool marks as evidenced by the correspondence of two or more surface contour patterns: "*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 toolmarks known to have been produced by different tools and is consistent with agreement demonstrated by toolmarks known to have been produced by the same tool. [This] 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*" (AFTE Identification Standard, 1992).

Significance is expressed in statistical terms, refers to scientific principles, though the 'sufficient' level of agreement is left to the subjective interpretation of the examiner, which, in turn, depends on his or her training and experience and the current technology and science. For instance, it is possible to further support an identification referring to the quantifiable consecutive matching striae (CMS) approach.

This is similar to the determination made by firearms and tool marks examiners. During training to become qualified examiners, as well as through working cases, examiners examine and compare numerous tool marks that have been produced by the same tool, as well as tool marks made by different tools. Examiners have established in their minds the amount of agreement that is significant to reach an identification, and can accurately make this determination. Even if they are aware that the amount of agreement may vary according in time and space, examiners have to do examinations *hic et nunc*, here and now.

When an examination is conducted, a 4-step examination method is followed. The first step is the examination of the class characteristics of the tool marks imparted onto the ammunition component. As we already said, the conclusions that can be reached are identification, elimination, and no conclusion.

If the class characteristics are the same, then the individual microscopic marks on the ammunition components are examined using a comparison microscope. From this examination, conclusions can be reached, and if identification is determined (for many years the opinion of the examiner was all that was required) then a confirmation is necessary from another qualified examiner in all case work. This is part of a good quality procedure in each laboratory which is required by American Society of Crime Laboratory Directors – Laboratory Accreditation Board. (ASCLD-LAB).

Hence, a qualified examiner is one who is able to distinguish between random and fixed pattern components and dissimilar patters that make it possible for a qualified examiner to come to conclusions as the identity of an ammunition component to a particular firearm.

The foundation for pattern based science has been well documented through peer review and forensic science journals, such as the Journal of Forensic Science, and the Association of Firearm and Tool mark (AFTE) Journal. Some examiners have wanted to expand the support for the science for years. Among them, Albert Biasotti who worked on individual characteristics of fired bullets (Biasotti, 1959), John Murdock and Bruce Moran who worked on consecutive matching striae (Biasotti and Murdock, 1984; Moran and Murdock, 2003; Grzybowski et al., 2003) may have been ahead of their time. When the before mentioned works were published neither work utilized computers and algorithms. These works only attempted to defend the marks produced as striae. Computer based image systems such NIBIN and Drugfire were added to help solve crimes. These systems allow examiners to search and view images of cartridge cases entered from unsolved cases. However, pattern based science has come under fire in recent years in court as well as in the press where the core question is: "*How can the field guarantee 'statistical quality control' in order to support the empirical science of firearms and tool marks identification?*"

Unfortunately, there have not been any standard methods or protocols for the application of probability and statistics to analysis or compare tool marks till now. DNA profiling has grown to one of the most widely known applied techniques for the identification of biological samples by using modern equipment and statistical databases. The field of firearms and tool marks is now utilizing the newest equipment such as co-focal microscopy along with computer based algorithms which are becoming available to the firearms/toolmarks examiner to do statistical patter recognition. Yet computers and algorithms alone cannot solve the problem facing pattern based science.

Venders want to introduce equipment into laboratories. However, such equipment cannot be used in case work until it is validated. Black box studies are needed to validate the science and satiations from universities are needed to implement the studies. The Alabama State Laboratory system has been involved with research in this field for the past several years. The Federal Bureau of Investigation (FBI) Laboratory and the National Institute of Science and Technology (NIST) have now also begun research in an attempt to answer the questions stated above.

The best approach is the black box research. Special equipment is needed to conduct this testing such as a comparison microscope and any new equipment being evaluated. A proposal could be a mix of researchers and examiners that can split the duty of research and validation testing as well as the work of reviewing the data and publishing the results. So how should we do the research? It is what is called a "Catch 22". It will take training between two and three years to qualify a person in the field of firearms and tool marks. Then laboratories have to struggle with the main question of '*what is more important, case work or research?*'. Examiners in pattern science are needed to work the ever growing high case load. Yet research needs to continue to advance the science.

Science can never have enough research. In the past, research with ten items was considered good research. Today we are talking about making research tests with thousands of items and it is being questioned if it is enough. These issues and the reliance of statistics will be important in the future.

2. VALIDITY *VS.* PROFICIENCY

Proficiency tests are quality assurance devices designed to better test an examiner's competence, or a laboratory's competence. Differences with validity tests (with control examiners) is that they feature such things as control over who's taking the test, anonymity, mandatory returns, controls to ensure blindness. Proficiency tests, whether internal or external, are still relevant for validity purposes, just not as relevant. These two topics overlap, because validity testing involves error rates. We will talk more about it momentarily.

What we are measuring in validity tests and proficiency tests is total error rate, that is errors from any source, which is the one most relevant to expert testimony. Collaborative Testing Service (CTS) error rates are suggestive about true aggregate error rates, but we need to keep in mind that these are uncontrolled tests. Open to all who pay the fee, including attorney's students and others. They are open to any laboratory, accredited or not. Neither examiners nor CTS knows the credentials of the respondents. However, it has been widely discussed that many respondents are trainees. Therefore, one would expect validity tests with qualified examiners to show lower error rates, and in fact that is the case. For casework expect even lower still, owing to confirm of identifications.

There are two types of errors, type I and type II, testing both Propositions #1 and #2. Errors could reflect weaknesses in Proposition #1 or #2, or be quality assurance errors, or combination of each. Different kinds of errors: examiners, administration and the courts are concerned about Type I.

It is impossible to completely disentangle the studies at this time. Brundage (1998) and Hamby et al. (2009) is a well-designed validity test using controls, or to test examiners. DeFrance and VanArsdale (2003) study puts a main emphasis on electrochemical rifling; however, since land impressions were conventionally produced, this also served as basic validity test for examiner ability to identify bullets. Smith (2005) involves both bullets and cartridge cases; Bunch and Murphy (2003) is a very comprehensive test involving cartridge cases useful for control examiners. Murphy drill bit test (not yet published) involves severe circumstances unlikely to be encountered in casework.

The following are error rates from black box testing published in the AFTE Journal. Firearms numbers include both bullets and cartridge cases.

Brundage (1998) and Hamby et al (2009), 0%
DeFrance and VanArsdale (2003), 0%
Smith (2005), 0%
Bunch and Murphy (2003), 0%
Murphy (unpublished) drill bit test
    False positive rate 0.7%
    False negative rate 5.7%.

The tool marks false elimination numbers were higher due to things like eliminating on individual marks on wires, where the cut was made by different areas of the tool blade. In my opinion, this is due to poor training or the fact that many examiners are only trained in firearms yet call themselves tool mark examiners under the definition that a firearm is a tool.

In the United States most forensic laboratories use CTS for proficiency testing. Several studies have been put forth which take time and a willingness of qualified examiners. These numbers do not take into account to properly estimate the probability that a typical examiner misidentified in a particular case, you need to assume a base rate, or prior odds. We should not really be doing that, but for illustration purposes it makes a difference whether 95% of bullet submissions into your laboratory are from the submitted gun, or whether only 5% are. The following are unpublished data utilizing the results of proficiency tests from CTS results.

Firearms

1.05% false positives

1.76% false negatives.

Tool marks

2.63% false positives

5.36% false negatives.

Firearms, Test 92-4

Total participants: 0.83% false positives; 1.28% false negatives

Excluding trainees: 0% false positives; 0.65% false negatives.

This is the only time in which the answer forms allowed participants to indicate whether or not they were trainees. As such, it serves as an indication of how trainees might affect the overall numbers for firearms tests.

Firearms, 1992 – 2000, 2003

1.9% false positives

0.4% false negatives

Tool marks, 1992 - 2000, 2003

2.2% false positives

2.0% false negatives

Most recent years are currently being analyzed for inclusion. For firearms, 1994, 1997 and 1999 provided no results because they were unsuitable for generating error rates or we could not determine how many comparisons were actually effected, nor could we estimate them with any degree of confidence.

Subjective is where a human being is involved and can affect the outcome of an examination or test. The more human is input, then the more subjective it could be. Now, all other things being equal, a less subjective process usually means a more consistent outcome. But seldom are all other things equal. One has to look at the whole picture. Maybe human is more accurate but less precise.

Subjective does not equal untrustworthy. Take, for example, Bachrach's SciClops intelligent machine: it is not a 100% objective process. It would be much closer to objectivity than traditional examinations. Now compare its performance to that of a group of control examiners. Is the former automatically going to be better? We don't know. They both have to be tested and compared. We have to compare their sensitivity, their specificity, their false positive and negative error rates, consistency of results across machines and examiners. All those things matter. Where the objective machine may have an advantage is in consistency over repeated processes.

But the subjective processes are not inherently unreliable or unscientific. There is subjectivity in every science and in every test, in one way or another. Our validity and proficiency tests show that traditional exams are trustworthy. A physician's diagnosis of a head cold certainly involves subjectivity, yet very trustworthy.

3. NEW TECHNOLOGIES

Although several systems are available for purchase, working groups are just forming to establish the foundations that will set the universal guidelines for users (among the others, the ARK Committee of the Scientific Working Group for Firearms and Toolmarks, http://www.swggun.org/). Among the new technologies being evaluated are Gel-Sight, Evofinder, and IBIS BRASSTRAX 3D. IBIS 3D, Evofinder and Gel-Sight are utilizing 3D surface imaging.

Gel-Sight is a system that can capture the impressions that are present in an object. It provides extremely detailed and rapid surface measurements through the Gel-Sight sensor technology. This system has been paired with Cadre Forensics to use an algorithm based on patterns size, shape and distance. Gel-Sight along with Cadre Forensics has showed very good results on impression tool marks such as cartridge casings. The same technology can be used to measure any rigid material: it is currently being used in applications as diverse as forensics analysis and by coinage quality control.

Gel-Sight is a system for capturing microscopic surface geometry. The system extends to the microscopic domain, demonstrating spatial resolution as small as 2 microns. In contrast to existing micro-geometry capture techniques, the system is not affected by the optical characteristics of the surface being measured – it captures the same geometry whether the object is matte, glossy, or transparent. The device can be used as a 2.5D 'scanner' for acquiring surface texture and shape. A camera records an image of this relief, using illumination from red, green, and blue light sources at several different positions. An algorithm is then used to reconstruct the surface. The sensor has no moving parts and can be made into a portable device that can be used 'in the field' to record surface shape and texture.

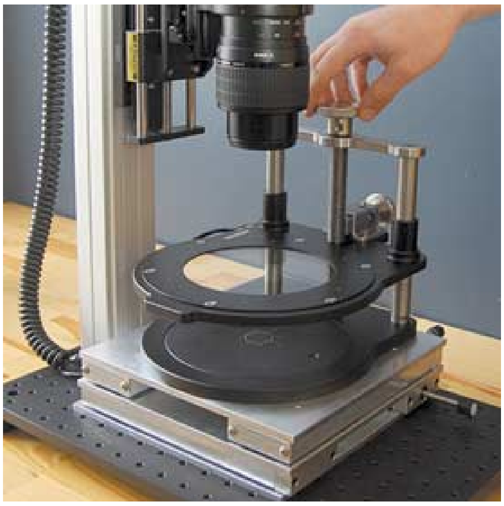

Figure 5: Gel-Sight equipment will photograph an object six times each with light from a different angle. Computer-based systems will incorporate the six photographs together for a 3D image.

The IBIS BRASSTRAX 3D captures high-resolution 2D images and precise 3D topographic information of significant regions of interest. Used for center fire cartridge cases, this includes the breech face and firing pin impressions on the primer, and the ejector mark. For rim fire cartridge cases, this includes the firing pin impression. A detailed image of the complete head is also captured for reference. The majority of the acquisition process is fully automated, so that minimal user input reduces operator variability. The automation of image focus, image lighting, and region of interest outlining ensures consistent image quality for visualization and uniformity for optimal comparison performance.

Through many years of the automated ballistic identification systems project Evofinder development has analyzed thousands of specimens. The number of points on the surface (and in the matrix respectively) is set in accordance with the required resolution, predetermined by the optical system and photo-sensor specific parameters. Efficient automated ballistic expertise requires resolution of 3-5μ or 200-300 points per millimeter, while the dynamic range of intensity (or 'gray scale') comprises approximately 200 levels or 8 bit (Figure 6).

So, is 3D technology better than 2D patter recognition? It might appear that 3D-technology gives certain advantages over the 2D-technology if you apply algorithms of correlation analysis using the data on three-dimensional surface. However, it is not quite evident at this time. Once again the examiners themselves say we need researchers to work with us to analyze the statistical data. With similar

Figure 6: Evofinder automated ballistic identification systems

two-dimensional images, with lands corresponding to lands and grooves to grooves (coincident phases), their amplitudes may differ. It is practically always in evidence for different objects because of different conditions of shooting, which influence on side marks, e.g. different amount of powder in cartridge-cases, initial speeds, objects materials, etc.. That is with the third coordinate (amplitudes) participating in the identification of the correlation degree, the summary result may be worse that in case of two-dimensional analysis.

Figure 4. An example of microscope and computer-based systems which may be used in the future.

4. ARE EXAMINERS SCIENTISTS OR TECHNICIANS?

Examiners are not physicists. However, examiners have a very high skill set for examining evidence and making determinations about the value of the evidence as well as reporting observations made during observations. Whether using the implied Daubert definition of science, then yes examiners are scientists and applied scientists. Examiners do test the foundation propositions, in training, and through research and publication. Also, unlike technicians, examiners have a theoretical understanding of the whole business. Not just following a cookbook recipe without knowing why they do so. The following framework merely sets down in orderly fashion what firearms/toolmarks examiners do.

- Sound training programs
- Sound protocols
- Confirmation of identifications
- Technical review of examination notes
- Administrative review of case file
- Why do it?
    - Reduces chances of error (from all sources)
    - Provides confidence
- Peer reviewed articles, and generally accepted by relevant community. The field of firearms and tool marks has an enormous amount of reference material which continues to be added daily (see the references in http://www.nist.gov/forensics/upload/Annotated-Bibliography-Firearms-Toolmarks.pdf and, of course, the unpublished material for FBI training - Understanding the Science of Firearms/Toolmarks Identification). A combination of the following issues must still be utilized in order to qualify an examiner to conduct forensic firearms and tool marks examinations.
- Examiners operate under basic propositions that have been tested and validated (met tests)
- Known error rates. It is important to note, firearms and tool marks identification has no casework error rate. This is due to the fact that no one knows the answer to the problem in casework. Each case is worked on its own. With that said we go back to the advances in technology which can give us the most reliable conclusions possible
- In practice, casework conclusions are also cleared through a quality assurance process. Conclusions may even be higher reliable
- Examination methods are documented in standard operating procedures (protocols). The core of these is common throughout the examiner community

- Defined by a universal scientific method? Philosophers, historians, sociologists, scientists: no consensus
- Line of demarcation?: If so, by consensus, not method
- For the federal courts: Daubert guidelines that mostly reflect Popper/Hempel theory
- Whether theory or technique has been subjected to peer review and publication; whether the theory or technique has been generally accepted by the relevant scientific community; whether the theory or technique can and has been tested; whether there exists a known or potential rate of error, and whether there are standards controlling the technique's operation
- Readers, referees, editors, before publication. Instead, no peer review of casework. Nothing to do with laboratory quality assurance programs. These are just some of the peer-reviewed journals that publish firearms and tool marks articles: *AFTE Journal*, *Journal of Forensic Sciences* of the American Academy of Forensic Sciences (firearms and tool marks identification is presented under Criminalistics Section); *Journal of the Forensic Science Society; Canadian Society of Forensic Science.*
- 4 levels of analysis:
    - Level 1 analysis- Class characteristics: elimination, but not individualization, can occur here
    - Level 2 analysis - Comparison microscopy: individualization occurs only here
- Conclusions drawn: Identification, inconclusive, elimination,
- Confirmation of identifications

In summary, to answer the question posed in the title of this paper, then *yes a need for statistical evaluation of forensic firearms/toolmarks examinations.*

REFERENCES

Biasotti, A. (1959). A statistical study of the individual characteristics of fired bullets. In *Journal of Forensic Science*. 4(1): 34 – 49.

Biasotti, A. and Murdock, J. (1984). Criteria for identification or state of the art of firearm and toolmark identification. In *AFTE Journal*. 16(4): 16 – 24.

Brundage, D.J. (1998). The identification of bullets fired from 10 consecutively rifled 9mm Ruger pistol barrels: A research project involving 507 participants from 20 countries. In *AFTE Journal*. 30(3): 438–444.

Bunch, S.G. and Murphy D.P. (2003). A comprehensive validity study for the forensic examination of cartridge cases. In *AFTE Journal*. 35(2): 201 – 203.

DeFrance, C.S. and VanArsdale, M. (2003) Validation study of electrochemical rifling, *AFTE Journal*, **35(1)**: 35 – 37.

Grzybowski, R., Miller, J., Moran, B., Murdock, J., Nichols, R. and Thompson R. (2003). Firearm/ toolmark identification: Passing the reliability test under Federal and State evidentiary standards. In *AFTE Journal*. 35(2): 209 – 241.

Hamby, J.E., Brundage, D.J. and Thorpe, J.W. (2009). The identification of consecutively rifled gun barrels. In *AFTE Journal*. 41(2): 99 – 110.

Moran, B. and Murdock J. (2003). Zen and the art of motorcycle maintenance. Contribution to forensic science. An explanation of the scientific method. In *AFTE Journal*. 35(2-Appendix 2): 234 – 240.

Murphy, D. (unpublished). Drill bit test (Toolmarks produced by drill bits).

Smith, E. (2005). Cartridge case and bullet comparison validation study with firearms submitted in casework. In *AFTE Journal*. 37(2): 130 – 135.