JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, May 2014, Vol. 59, No. 3
doi: 10.1111/1556-4029.12382
Available online at: onlinelibrary.wiley.com

PAPER

CRIMINALISTICS

Fabiano Riva,[1] Ph.D.; and Christophe Champod,[1] Ph.D.

Automatic Comparison and Evaluation of Impressions Left by a Firearm on Fired Cartridge Cases

ABSTRACT: Recent years have been characterized by a series of publications in the field of firearms investigation questioning the reliability and objectivity of such examination. This research investigates new solutions to decrease the subjective component affecting the evaluation that follows the comparison of impressions left by a firearm on the surface of spent cartridge cases. An automatic comparison system based on 3D measurements has been developed and coupled to a bivariate evaluative model allowing assigning likelihood ratios. Based on a dataset of 79 pistols (all SIG Sauer 9 mm Luger caliber), the system shows a very high discriminating power and the LRs that it provides are very indicative of the true state under both the prosecution and the defense propositions. For example, likelihood ratios exceeding a billion are predominantly obtained when impressions originating from the same source are compared. The system is also characterized by relatively low rates (≤1%) of misleading evidence depending on the firearm considered.

KEYWORDS: forensic science, firearms identification, likelihood ratio, cartridge cases, 3D topographies, evaluation

In crimes involving firearms, the firearm examiner is often asked to help assess whether or not a questioned cartridge case (typically recovered from a crime scene) had been fired by a given firearm (typically a firearm seized following the inquiry). For that purpose, the firearm examiner compares, by means of an optical comparison macroscope, the impressions left by the firearm on the questioned cartridge case surface with the impressions on known cartridges obtained by test firing the questioned firearm. During this process, the similarities and differences between impressions are observed and recorded. From these observations, the examiner moves into the evaluation phase, where he/she has to make a decision regarding a common source or otherwise between the questioned cartridge and the known cartridges. It is during this evaluation phase that the examiner evaluates and weights the similarities and the differences observed between sets of markings seen on the cartridge cases.

It is commonly agreed that the procedure adopted to reach a conclusion is subjective and based on the training and experience of the firearm examiner (1). In the field of firearms identification, as in many fields of forensic sciences, there is no a widely accepted method describing precisely how the evaluation phase should be conducted.

The last years have been characterized by a series of publications in the field of firearm identification that brought to the front a number of criticisms (2–4). These publications have triggered important discussions and a follow-up rebuttal (5). Some of the raised issues are summarized below:

- The ability to identify a firearm as the source of a questioned cartridge case or bullet is based on two tenets constituting the scientific foundation of the discipline (6). The first assumes the uniqueness of impressions left by the firearms. According to Moran (7) for example, the probability that two different tools would produce the same signature is so low that it is for all practical purposes considered as impossible. The impressions left by different firearms are thus considered – although depending on the quality of the impressions – as unique enough to allow a firearm expert to differentiate them. The second tenet is related to the reproducibility of the impressions left by the firearm. During the shooting process, a firearm will leave on the cartridge cases surface impressions that are reproducible from one fire to another. This principle allows the firearm expert to relate impressions left by the same firearm after several shots. This does not mean that the impressions left by the same firearm are identical; indeed, the components of a firearm can evolve with time, wear, and corrosion (8). Despite the research work done in this field and the literature supporting the scientific foundation of the discipline (9), more research is still called upon to corroborate the validity of the fundamental tenets of uniqueness and reproducibility, which, according to the authors of a recent report of the US National Academy of Science (NAS), have not yet been fully demonstrated (4).
- There is a lack of an objective yardstick in the identification process. It is recognized that the current evaluation of individualization/identification is subjective in nature, founded on scientific principles, but based on the examiner's training and experience (1). Hence, different examiners may evaluate the same set of observations differently. This lack of objective standard is even more noticeable when well-defined evaluation procedures do not exist.

[1]School of Criminal Sciences, Institute of Forensic Science, University of Lausanne, Quartier Sorge, CH-1015 Lausanne, Switzerland.

Received 7 April 2012; and in revised form 11 Feb. 2013; accepted 23 Feb. 2013.

- More research is asked to quantitatively characterize the specificity (or selectivity) of the marks (3 – p. 3). Indeed, the presence of (i) similarities between impressions left by different firearms and (ii) differences observed during the comparison between impressions left by the same firearm implies that a probabilistic question must be tackled to help assess whether a particular firearm is at the source of a mark on a questioned bullet or cartridge case. There is a need to assess the statistical contribution of a correspondence between questioned and known, taking into account the evolution of characteristics (wear of the firearm), the presence of subclass features (that could sometimes be confused with some "individual" characteristics) and the fact that so-named individual characteristics are a combination of nonunique impressions.
- Leaving aside cases where the material is unsuitable or the examination leads to inconclusive statements, firearm examiners tend often to express their conclusions in categorical terms (individualization or exclusion). The authors agree with the statement stated in the study by Champod et al. (10): "there was also a call to refrain from testifying to absolute scientific conclusions often based on the vague concept of individualization or uniqueness, in favor of more modest conclusions recognizing the sometimes imperfect performances measured in these disciplines." Firearm examiners following the AFTE Theory of Identification (1) may state that identifications are made to a "practical certainty", but almost no study has systematically investigated the magnitude of the residual uncertainty associated with such a statement.

Purpose of the Research

The aim of this study is to try to address some of the above-mentioned issues, especially the lack of objectivity and the absence of a robust statistical basis to evaluate the results of a comparison. This study does not pretend to find exhaustive answers to all the criticisms raised in recent years, but wishes to explore the possibilities of bringing conclusions based on an objective process, recognizing the probabilistic nature of this identification discipline.

To do this, an automatic system based on 3D surface measurements allowing comparing impressed impressions on cartridge cases has been developed and used in this study. 3D topographies for comparison purposes have already widely been successfully applied to the field of firearms and toolmarks (11–17). These research efforts have shown that topography measurements in 3D allow good accuracy, while at the same time, maintaining a high level of reproducibility. The main objectives of the statistical analysis carried out in these previous studies were either to measure the global discrimination power of the 3D system, establishing the respective error rates (often used to assess the discrimination power of the analyzed impressions) (14–17), or to measure the efficiency of these systems when they are used in the context of a database search (17). The objective of the present study is different as it aims to obtain an objective assignment of the weight of the evidence derived from a set of observations between two entities measured in 3D. That weight of evidence should be case specific.

Hence, the automatic comparison system developed in this study does not only allow reproducible acquisitions, but also offers a mechanism whereby the strength associated with the similarities and differences observed between two impressions is assessed. The data obtained by the system are used to build and test a statistical model allowing assigning likelihood ratios.

Material and Methods

Marks Used for the Study

Several marks are left on the surface of the cartridge case during the firing process. Some are left before the shot, typically during the introduction of the cartridge in the chamber, others are left by the firearm during the ignition of the powder or when the cartridge case is extracted and ejected from the chamber. The number and the type of marks that could be observed on the surface of a spent cartridge case depend of the type and model of the firearm. For this study, only the firing pin impression and the breech face impression have been taken into account. These two impressions are often used during the identification process not only by the examiner but also by many automated identification systems (such as IBIS®). These impressions have a significant discriminative power and, being located on the primer cup, can be easily observed and used for identification purposes.

Comparison Methodology

An automated comparison system based on 3D technology has been developed. This system first acquires the data using a 3D laser profiler. Then, the resulting measurements undergo algorithmic treatments, namely the aligned overlay of an impression on the other impression, leading to a comparison phase and the computation of scores describing the distance between the two sets of 3D data.

3D Acquisition

A laser profiler coupled with a confocal detector has been used (μScan from Nanofocus AG). The device uses a laser that scans the surface of the cartridge case. At each interval dictated by the spatial resolution, the detector records, using the confocal principle, the height at the point where the laser touches the surface. This profiler allows measuring the topography of an object with a spatial resolution (lateral resolution) of approximately one micrometer and with a vertical resolution less than one micrometer. The depth of field of this technique is limited to one millimeter. The range of this device easily allows measuring a surface like the primer cup of a 9 mm Luger cartridge case. The result of this acquisition is a matrix (analogous to an image) where each point is described by the coordinates X and Y. Unlike a typical 2D digital image, the value of each point (pixel) corresponds to the height of the measurement (Z) at the coordinates X and Y. For this research, the 3D topographies have been scanned with a spatial resolution of 3 μm in both direction (X and Y). This value offers an appropriate compromise between the quality of the measurement and the time needed to acquire the data; the resolution will be decreased in a second step by a factor of 2 to improve the speed of the alignment algorithm used to compare the 3D topographies. Figure 1 shows the measurement performed by the μScan profiler on a 9 mm Luger primer cup. The left image is a 2D picture of the fired primer cup. The right image is a 3D scan of the same subject coded using a grayscale map. Unlike the right image, the differences in shades of gray represent differences in height (Fig. 1).

Preprocessing, Alignment, and Computation of Similarity Scores

Once the measurements have been acquired, the resulting topographies are compared to measure their degree of similarity.

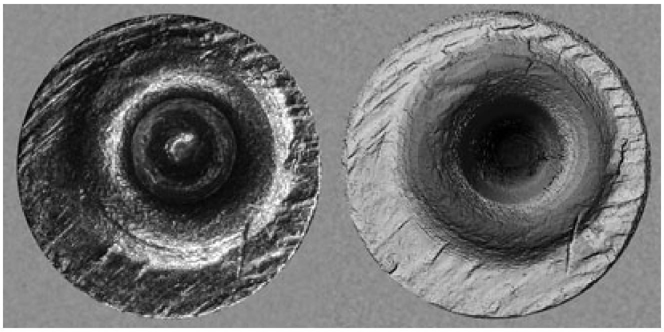

FIG. 1—*Image of a primer cup of a cartridge case fired from a Beretta 92 S 9 mm Luger (left) and its 3D scan (right).*

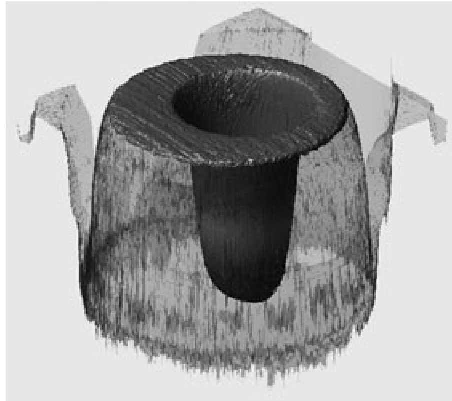

FIG. 2—*Primer cup segmentation. The semitransparent area represents the part of the primer cup that is not relevant to the forthcoming matching process.*

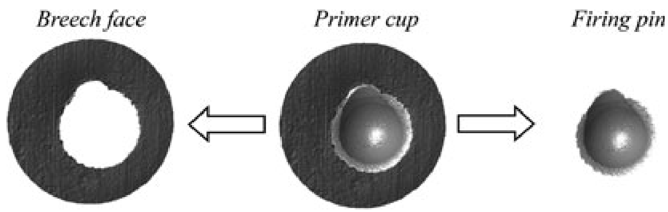

FIG. 3—*Segmentation of the breech face impression (left) from the firing pin impression (right).*

To get to that final stage, there is a requirement (a) to preprocess the data to select the relevant features that will be used for the comparison and (b) to properly align these sets of features to be compared before computing the scores that will describe the degree of similarity between the impressions. Scores will be the metrics of interest; proxies that represent the degree of closeness between two compared impressions. All developments have been carried out in Matlab 7.0® (Mathworks Inc., Natick, MA) and are briefly outlined below.

Preprocessing—The first preprocessing step is the detection of noisy points (or outliers) due to the 3D acquisition process. Outliers are points that do not accurately represent (in its Z axis) the surface of the object; they bring noise to the 3D image that needs to be removed. The criterion to declare a point as an outlier is based on an additional datum gathered by the acquisition system – a measurement of reflectance – associated with each acquired point. The quantity of light reflected to the detector depends on several factors; one of them being the slope of the measured surface. When the slope increases, the quantity of light from the laser reflected vertically by the surface to the detector is diminished. The measure of reflectance is then a good indicator of the quality of the measurements. All points showing reflectance below 10% have been eliminated and replaced by a value obtained by linear interpolation using its neighboring points. The majority of the points considered as outliers are typically situated in the primer cup edge because in this area, the slope of the surface is particularly important.

The second preprocessing step requires the manual positioning of the horizontal plane passing through the plane of the primer cup. This plane adjustment requires a few seconds at the end of the acquisition step. At this stage, the primer cup is not isolated.

Then, an automatic segmentation of the primer cup is performed by taking advantage of normal vectors to the surface. When the angle between the vertical and the normal vector is greater than an adaptative threshold, the correspondent pixels delineate the boundaries of the upper part of the primer cup that have been in contact with the firearm (see Fig. 2).

In Figure 2, the areas of the primer cup that have not been touched by the firearm are represented by a semitransparent texture. These points will not be used in the matching process.

At this stage, the upper part of the primer cup is isolated. The next step leads to the separation between the breech face and firing pin impressions. This step is required because during the firing process, the firing pin impression is partially created before the burning of the powder. When the powder burns and the bullet is pushed through the barrel, and during the ejection of the cartridge case, the morphology of the firing pin impression can be further modified (e.g., firing pin drag mark). In contrast, the breech face impression is created during the increase in pressure in the chamber. The firing pin and the breech face impressions are thus impressed on the surface of the cartridge case in two distinct moments. Their relative position may change, and the depth of the firing pin impression can vary between consecutive shots (18). This means that it is normally not possible to align the breech face impression and the firing pin impression at the same time. These two impressions thus have to be separated before the aligning process. To separate them, the same principle used to segment the primer cup has been used. Figure 3 shows the results obtained after the automatic separation of the firing pin impression and the breech face impression.

Finally, to minimize the influence of the global shape of the impression (including the crater shape of the firing pin impression) and to give more weight to the fine characteristics, typically the features of interest in firearm comparisons, the data are filtered using a Gaussian bandpass filter in the Fourier domain. The result is an image with enhanced fine characteristics of the impressions left by the firearm. Figure 4 shows this final processing step for the firing pin impression. At this point, the preprocessing steps are finalized, and the images coming from different cartridge cases can be aligned to compute similarity scores between them.

Firing Pin Impression Alignment—To perform the alignment of firing pin impression, the Iterative Closest Point (ICP) algorithm has been implemented following the guidelines provided by the works of Jost and Hugli in 2002 (19) and of Rusinkiewicz and Levoy in 2001 (20). This algorithm aligns two 3D datasets minimizing iteratively the distance between the points in one

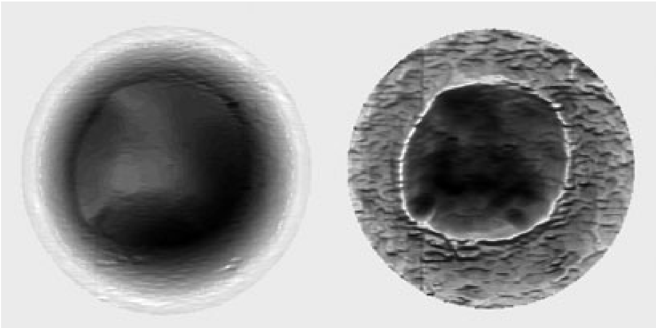

FIG. 4—*Image of a 3D scan of a firing pin impression (left) with resulting image after the data filtering (right). The grayscale values represent the height of each point.*

impression and the closest points in the other. Here, the ICP performs 3D rigid transformation described by 6 degrees of freedom.

Breech Face Impression Alignment—For the alignment of the breech face impressions, another algorithmic method has been adopted. This method approaches the alignment of two impressions as an "optimization problem". It searches the minimum distance between two impressions as a function of the translation and rotation in the plane *XY*. The algorithm used during this research is adapted from the work of Nestares and Heeger (21). The algorithm aligns two images using a multiresolution structure and tries to minimize the smallest differences between the two images, saturating the greatest differences. This procedure allows the alignment to be based on the finest characteristics (such as the striations left by the breech face) and to minimize the influence of the biggest features (such as differences between the shapes of the firing pin drag impressions). The main difference between this algorithm and the ICP is that the rotation and the translation are performed only in the plane *XY* and not in three dimensions; the transformation is thus limited to 3 degrees of freedom. As for firing pin impressions, the results of the alignment can be displayed as side-by-side comparisons in a virtual comparator, with one impression on the left-hand side of the separator (black vertical line) and the second on the right (Fig. 5).

Similarity Scores—Once the impressions have been aligned, one or more scores between them can be calculated. A score represents quantitatively the amount of morphological similarities (or differences) between two impressions. Three scores have been used in this study:

- The first is the median of the Euclidean distances between the points in the first impression and the closest points in the second impression.
- The second score is a correlation coefficient taking into account the number of points superposed between the two impressions.
- The last metric considered is the median of the difference of the angle between normal vectors at each point of the both surfaces.

Each comparison is thus described by three scores. It means that when two cartridge cases are compared, six scores are obtained: three coming from the comparison of the firing pin impressions and three from the comparison of the breech face impressions.

FIG. 5—*Alignment of two breech face impressions fired by the same SIG Sauer P226 9 mm Luger. The black vertical line represents the separation of the impressions as in a comparison macroscope.*

The Evaluative Model

The above comparison system quantifies similarities/differences between two impressions using six scores. The objective of such an automatic comparison system goes further than the scores as we aim to develop a tool that can be used to assist firearm examiners in the evaluation phase. To reach that objective, an evaluative model based on the likelihood ratio (LR) has been developed (22). This model is described in the following paragraphs.

The Likelihood Ratio

The likelihood ratio is a numerical value that expresses the weight of the forensic evidentiary findings (*E,* commonly used for "evidence"). It is obtained by the ratio of the probabilities of the findings under two propositions that we relate, respectively, to the prosecution (H_p) and the defense (H_d) propositions. In cases involving issues of identity of sources in relation to firearm cartridge cases, these propositions can formulated in the following way:

H_p: The questioned cartridge case was fired by the suspect firearm.

H_d: The questioned cartridge case was not fired by the suspect firearm but by another unknown firearm.

The likelihood ratio is the ratio between the probability of observing the evidential findings (*E*, here a set of similarities or differences represented by a series of scores) between the questioned cartridge case and the sample fired by the suspect firearm, respectively, under H_p and H_d conditioned on some background information *I*. Formally, the likelihood ratio is written in the following way:

$$\mathrm{LR} = \frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)}$$

The background information (*I*) represents case-specific information that may help to restrict the nature of the relevant population considered under the proposition H_d. The weight of the forensic findings is essentially a relative and conditional measure that helps to progress a case in one direction or the other depending on the magnitude of the likelihood ratio (23). A LR greater than one supports the prosecutor proposition. If the likelihood ratio is between zero and one, the comparison results support the defense view. Finally, if the LR is close to one, the findings do not bring support for any of the considered propositions; this is the typical "inconclusive" case.

Although for the remainder of this study, we will use the computed similarity scores as the forensic findings (E), nothing in the above framework requires it. Indeed, E could be described by a number such as the number of CMS (consecutive matching striae) between the samples or another value, like the scores generated by another automatic comparator system (22,24).

Computing a Likelihood Ratio Using Scores Obtained from an Automatic Comparison System

To assign a LR from a series of scores, we need to relate E in relation to their underlying distributions corresponding, respectively, to each proposition (H_p and H_d). We call them *within-source* distribution and the *between-sources* distribution. By *within source*, we mean the distribution of the scores when it is actually known that the compared impressions are originating from the same source (H_p). By *between sources*, it is meant the distribution of the scores obtained when cartridge cases fired from different firearms are compared (H_d). These two distributions provide the basis for the assignment of the numerator and the denominator of the likelihood ratio. The scores used on this study are akin to continuous data, hence the LR will be a ratio of probability densities. The probability densities will be estimated either by a nonparametric method such as kernel density estimation (KDE) or by parametric modeling.

In one dimension though, for illustration purposes, assume that the automatic comparison system generates a score (d) for each comparison. The *within* and the *between* distribution could thus be modeled using one-dimensional KDE, as shown in Fig. 6.

These two continuous functions (black line and gray line) represent, respectively, the *within* and the *between* distributions and allow the probability densities $p(d|H_p,I)$ and $p(d|H_d,I)$ to be derived. These values are needed to get the likelihood ratio for any new score generated from the comparison system. In fact, when two cartridge cases are compared in a specific case, the LR for this comparison will correspond to the ratio between the probability densities for the *within* distribution ($p(d|H_p,I)$) and the *between* distribution ($p(d|H_d,I)$) at the point set by the result of the comparison (d – the metric describing the comparison) between the questioned cartridge case and the test fire (for example, in Fig. 6, the comparison is described by a score of 1.31). In this example, the probability density under H_d is greater than under H_p for a score of 1.31. This means that the LR will be less than one and that the result will support the second hypothesis, which is the defense hypothesis (H_d).

Following the above example, it is possible to calculate a likelihood ratio for the one-dimensional case. In the current study, however, six scores are derived, three in relation to the breech face and three relating to the firing pin. How likelihood ratios are derived in such a multivariate environment is described below.

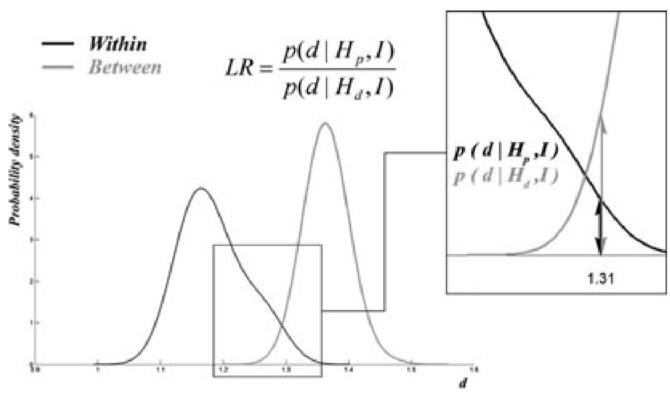

FIG. 6—*Illustration of the mechanism used to calculate a likelihood ratio (LR) for a given comparison. The LR is obtained by the ratio between the probability density* p(d|H$_p$,I) *and the probability density* p(d|H$_d$,I) *at the coordinates dictated by the comparison result d (in this example* d = *1.31).*

Procedure to Handle more than One Score

A procedure to reduce the number of dimensions has been applied. Principal component analysis (PCA) has been chosen to reduce the dimensionality of the problem by selecting the two principal components (PC1 and PC2) that offer the highest contribution to the variability. In this way, the comparison of two impressions is described by these two values, which also act as similarity scores. Figure 7 shows the situation before and after the application of the PCA on a set of data coming from comparisons of breech face impressions.

In the three-dimensional graph on the left, the black cloud (distribution of points) represents the comparisons between breech face impressions left by the same firearm (*within* distribution). The gray cloud represents the comparisons between breech face impressions left by different firearms (*between* distribution). In this example, only the breech face impression has been considered, this means that each point in the Figure is described in three dimensions and represents one comparison between two breech face impressions (Fig. 7 – left). Each score thus represents one dimension (three dimensions represented by the median of Euclidean distances, the correlation index and the median of angle between normal vectors). After the application of the PCA, every case is described by only two values (one for each principal component, PC1 and PC2). Thus, the result is two datasets projected in a plane (Fig. 7 – right).

In addition, still using PCA, it is possible to consider jointly the scores coming from the firing pin impression comparisons and the scores coming from the breech face impression comparisons. In this case, the PCA is applied to all six scores representing a comparison. Regardless of the number of scores describing one comparison before the application of the PCA, the result will be the same: two distributions of points (*within* and *between*) described in two principal components obtained using PCA on all the input scores.

The procedure used to calculate likelihood ratios in this bidimensional case calls for probability densities estimated either by bidimensional KDE or using a bivariate normal distribution. Figure 8 shows the bidimensional KDE (right) starting from a cloud of points in two dimensions (left).

To summarize, the calculation of a LR associated with a comparison between a questioned cartridge case and a known test fire first requires the comparison of the two cartridge cases and the calculation of the scores for the breech face impressions and

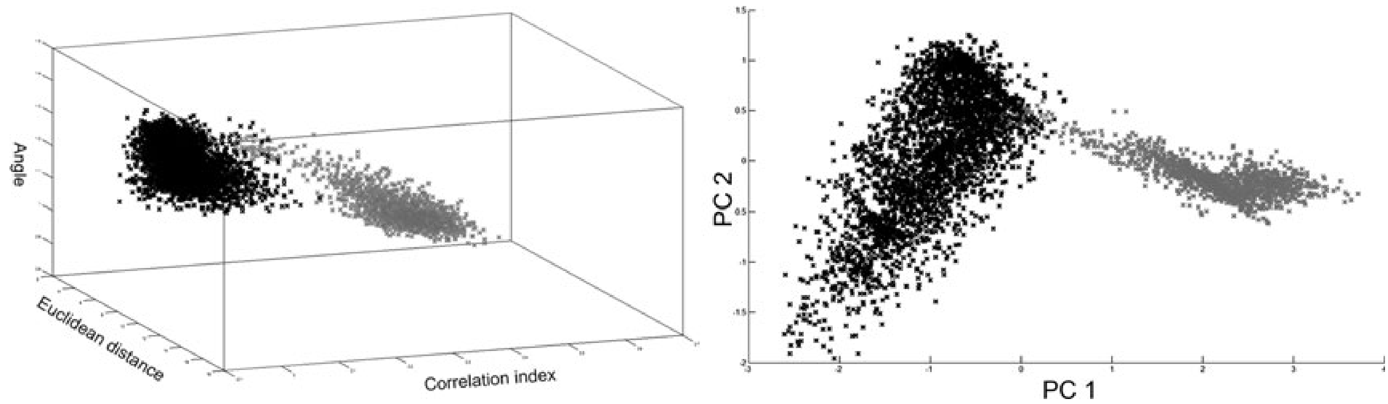

FIG. 7—*Left: 3D plot of three scores (Euclidean distance, correlation index, and difference between angle of normal vectors) characterizing the comparison between breech face impressions left by the same firearm (black) and by different firearms (gray). Right: Plot of the same data according to the two principal components (PC1 and PC2).*

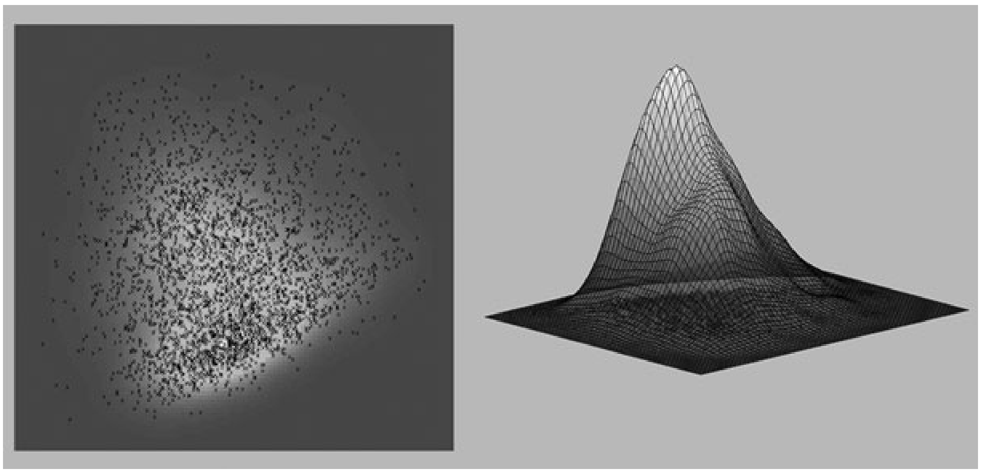

FIG. 8—*2D representation of the points (left) with the density obtained using kernel density estimation (right).*

the firing pin impressions. Then, the transformation dictated by the PCA is applied to the data leading to the transformed scores PC1 and PC2. The LR will correspond to the ratio between the probability densities for the *within* distribution ($p(d|H_p,I)$) and the *between* distribution ($p(d|H_d,I)$) at the coordinates indicated by the result of the comparison between the questioned cartridge case and the test fire. The following image shows this concept for the bidimensional case (Fig. 9).

Diagnostic Graphs: The Tippett Plots

To evaluate the performances of the system and the rates of missing evidence linked to this methodology, an analysis of the distributions of the likelihood ratios in simulated cases with given H_p and H_d is performed. These distributions are studied using a specific plot, called a Tippett plot, which shows one minus the cumulative distribution for the likelihood ratios computed under H_p and the likelihood ratios computed under H_d, respectively (Fig. 10). These plots also allow the study of the proportions of misleading evidence: the percentage of LR < 1 when the prosecution proposition H_p is true (RMED) and the percentage of LR > 1 when the defense proposition H_d is true (RMEP). Figure 10 shows an example of a Tippett plot.

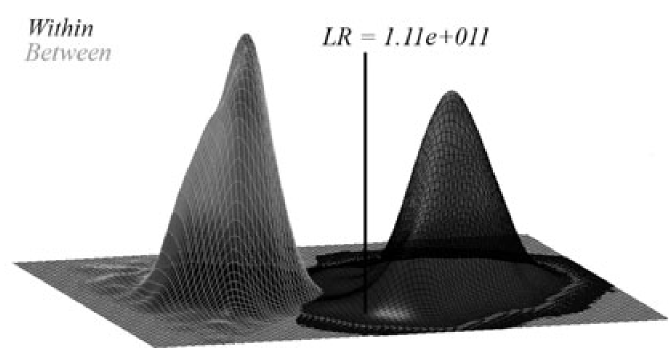

FIG. 9—*Calculation of the likelihood ratio (LR) in bivariate mode at coordinates dictated by the comparison result (black vertical line), exploiting the within (black) and the between (gray) distributions.*

In this study, the same data are used to obtain the *within* and *between* distributions and to compute the likelihood ratios. The Tippett plot in Fig. 10 has been created using samples fired by SIG Sauer 9 mm Luger pistols with the same class characteristics (same general shape of the firing pin). It shows the result for the two principal components, initially computed using all six scores (associated with the firing pin and breech face impressions).

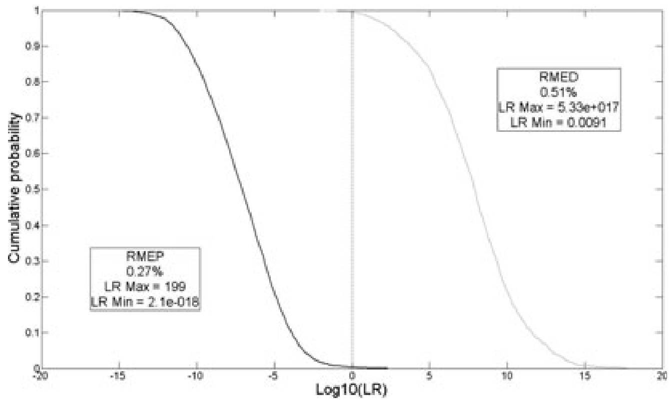

FIG. 10—*Tippett plot showing the distribution of the likelihood ratio (LR) obtained, respectively, when* H_p *(gray line) or* H_d *(black line) is true. The rate of LRs greater than one given* H_d *is 0.27% (RMED). Given* H_p*, the rate of LRs below 1 is 0.51% (RMEP).*

Samples and Computation of Within and Between Distributions

The procedure described above has been applied to various cartridge cases fired from SIG Sauer firearms of the same type (similar models) and having the same class characteristics. The research sample is composed by three distinct groups: one group of 79 cartridge cases fired by different firearms (this group was used to build up the *between* distribution), one group of 60 cartridge cases fired by one particular firearm (W1) and another group of 60 cartridge cases fired by another firearm (W2) (these last two groups were used to build two distinct *within* distributions). The ammunition type used for this research has been the same for each group: Geco Sintox 9 mm Luger. The samples collected for this research are listed in Table 1:

To obtain the data in relation to the *within* distribution, the 60 cartridge cases fired by the same firearm are compared together for each possible pairwise combination. For example, one *within* distribution is composed of $(60^2 - 60)/2 = 1770$ comparisons for the firing pin impression and 1770 comparisons for the breech face impression. The same is applied to obtain the data informing the *between* distribution. In total, the *between* distribution is based on $(79^2 - 79)/2 = 3081$ data points.

From these data, we can obtain 1770 LRs when it is known that H_p is true, and conversely 3081 LRs when is known that H_d is true, when the impressions are taken into account separately or jointly. That serves as a testing environment to assess the forensic performance of the system.

Results

After the application of the PCA, each comparison can be described as a point having two coordinates in the *XY* plane (Fig. 7).

Distributions of Points

First, the respective contribution of the scores associated with the firing pin (Fig. 11) and breech face impressions (Fig. 12) will be explored, followed by the results for the joint contribution of both impressions (Fig. 13). This will be carried out separately for each firearm (W1 and W2) used to construct the *within* distribution.

Each figure shows two distributions of points after the application of the PCA; the gray distribution represents the *within* distribution and the black one the *between* distribution.

The *within* distributions for the both firearms are very similar. The shape of these distributions shows that the two components PC1 and PC2 selected after the application of the PCA are correlated. This means that if only one dimension were used, the discrimination between the *within* and the *between* would not change significantly. Another interesting point is that the overlap between the *within* and *between* distributions. It means that the automatic comparison system cannot easily discriminate firing pin impressions left from different SIG Sauer firearms. This is an expected result: indeed, the firing pin impressions of these SIG Sauer firearms are very smooth and they do not display discriminative features.

For breech face impressions, results show a good separation between the *within* and the *between* distribution with a limited overlap between the two.

When all scores (from breech face and firing pin impressions) are considered jointly, results show a better separation compared with the results obtained considering only breech face impressions or firing pin impressions separately. It shows the firing pin impression increases the discrimination. It is also possible to see that the results are better for the firearm W2 (Fig. 13 – right) than for the firearm W1 (Fig. 13 – left). This means that W2 reproduces impressions on the cartridge cases that are more discriminative than W1. This was confirmed by visual inspection under the comparison macroscope.

Tippett Plots

We restrict the presentation of the Tippett plots to the most efficient situation where all scores are considered jointly in the PCA process. These Tippett plots have been obtained by modeling the data with a bivariate normal distribution that proved to be more efficient than KDE.

Figure 14 and 15 show very low rates of misleading evidence (RMED, RMEP). This means that LRs supporting the wrong proposition are rarely assigned by this system under these conditions. Again, the Tippett plots show that the toolmarks from the first firearm (W1) are less discriminative than these left by the second firearm (W2). The distributions of the LRs are summarized in Table 2:

The system offers very high discriminating power, and the LRs generated are very indicative of the true state under both the prosecution and the defense propositions. The model is able

TABLE 1—*Details of the firearms used for this study to explore the within distribution (firearm W1 and W2) and the between distribution (79 firearms), respectively.*

Firearms	Manufacturer	Model	Ammunition type	Quantity of cartridge cases
Within N°1 (firearm W1)	SIG Sauer	P228	Geco Sintox 9 mm Luger	60 from one firearm
Within N°2 (firearm W2)	SIG Sauer	P226	Geco Sintox 9 mm Luger	60 from one firearm
Between	SIG Sauer	P226 (42), P228 (14), and Sig Pro (23)	Geco Sintox 9 mm Luger	79 from 79 firearms

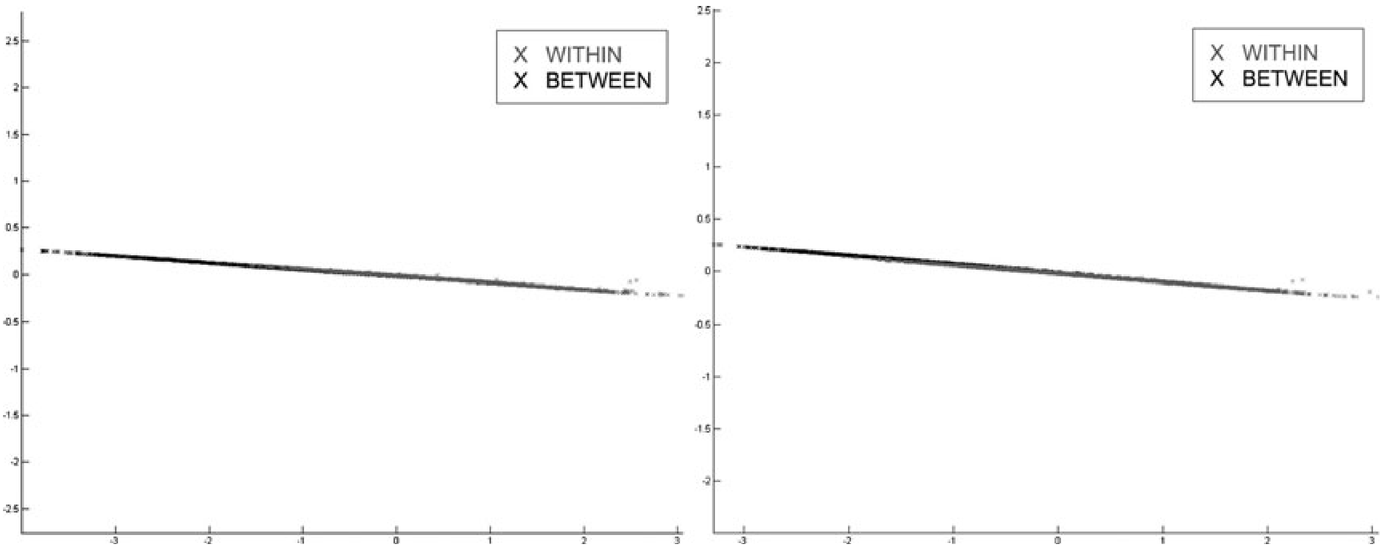

FIG. 11—*Distribution of data as a function of PC1 and PC2 representing the comparison results after the application of the principal component analysis (PCA) for the firing pin impressions. Comparisons between cartridge cases fired by the same firearm are represented using gray dots (within), and comparisons between cartridge cases fired by different firearms are shown with black dots (between). Results are presented for the firearm W1 (left) and the firearm W2 (right).*

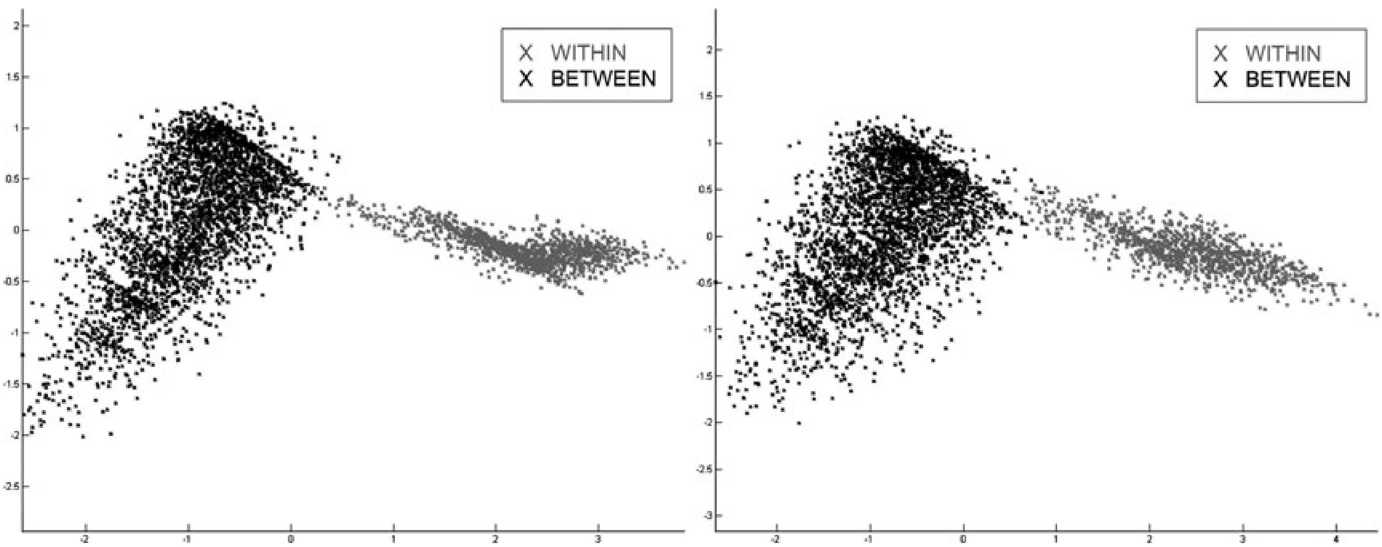

FIG. 12—*Distribution of data as a function of PC1 and PC2 representing the comparison results after the application of the principal component analysis (PCA) for the breech face impressions. Comparisons between cartridge cases fired by the same firearm are represented using gray dots (within), and comparisons between cartridge cases fired by different firearms are shown with black dots (between). Results are presented for the firearm W1 (left) and the firearm W2 (right).*

to support very strongly either hypothesis on a significant proportion of simulated cases. If we focus on the results from firearm W2 (referring to Table 2), on average when the prosecution proposition is true, the evidence will contribute to the issue with a LR of the order of 10^{23}. Conversely, when the defense proposition is true, the evidence will on average brings an increase in favor of its position of the order of 10^{-19}. Under H_p, the 90% of the LRs obtained lies between 10^9 and 10^{37}. Conversely, under H_d, we obtained 90% of the LRs between 10^{-10} and 10^{-19}. Although the rate of misleading evidence in favor of the prosecution proposition is 0.09%, note that the maximum value corresponds to a LR of 13.5. The risk of misleading evidence exists, but with low likelihood ratios that may not even have an impact on the decision-making of the court. Rates of misleading evidence must then be considered in conjunction with the strength by which the evidence is pointing in these situations. For example, for firearm W1, RMED is 4.4%, but these cases span over a larger range of likelihood ratios supporting H_p (the minimum LR is 3.1 10^{-9}). In other words, the overall impact of such misleading evidence may be much larger.

Discussion

From these promising results and with a view to envisage an operational deployment, a few avenues need to be discussed. The

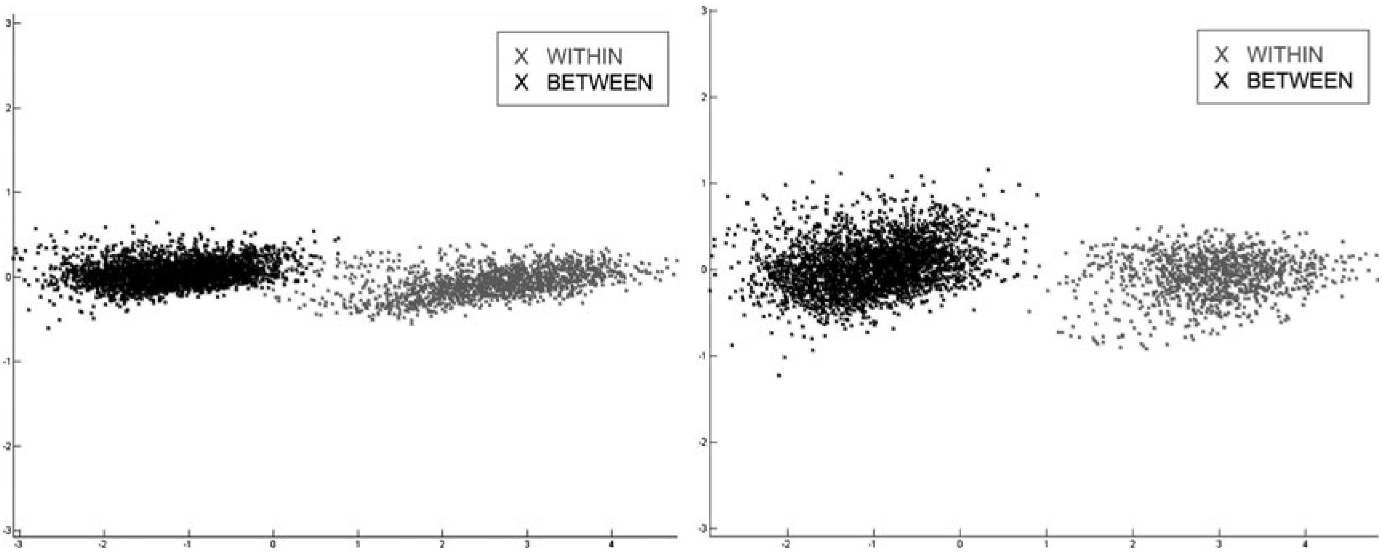

FIG. 13—*Distribution of data as a function of PC1 and PC2 representing the comparison results after the application of the principal component analysis (PCA) for the breech face and firing pin impressions. Comparisons between cartridge cases fired by the same firearm are represented using gray dots (within), and comparisons between cartridge cases fired by different firearms are shown with black dots (between). Results are presented for the firearm W1 (left) and the firearm W2 (right).*

TABLE 2—*Summary of the performance obtained by the system for firearms W1 and W2, respectively.*

	RMEP/ (RMED)	Min $\text{Log}_{10}(\text{LR})$	$\text{Log}_{10}(\text{LR})$ at Q (0.05)	Median $\text{Log}_{10}(\text{LR})$	Mean $\text{Log}_{10}(\text{LR})$	$\text{Log}_{10}(\text{LR})$ at Q (0.95)	Max $\text{Log}_{10}(\text{LR})$
LRs when H_p is true for the firearm (W1)	(4.4%)	−8.5	0.8	21.3	20.6	34.2	19.8
LRs when H_p is true for the firearm (W2)	(0.26%)	−1.4	8.7	23.9	23.6	37.5	22.3
LRs when H_d is true for the firearm (W1)	0%	−19.8	−27.3	−16.5	−17.6	−9.0	−0.7
LRs when H_d is true for the firearm (W2)	0.09%	−16.9	−28.3	−18.6	−18.9	−9.9	1.13

Note that the 0% for RMEP in the Table just means that there were no observations for the number of comparisons carried out. It should not be misconstrued as meaning that the forensic error rate is equal to zero.

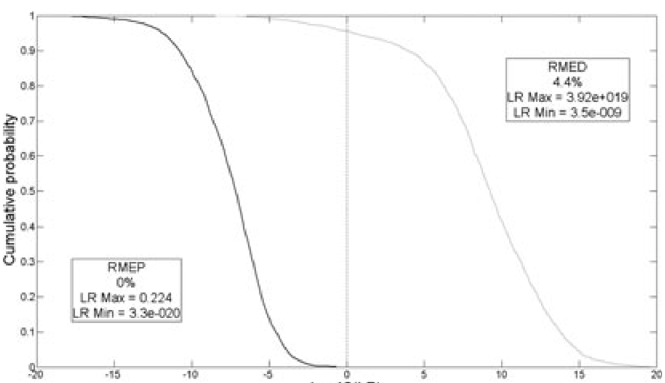

FIG. 14—*Tippett plot of the likelihood ratio (LR)s obtained for firearm W1, when* H_p *(gray line) or* H_d *(black line) is true, respectively. All six scores are taken into account (breech face and firing pin). RMEP obtained is here of 0%, whereas RMED is 4.4%.*

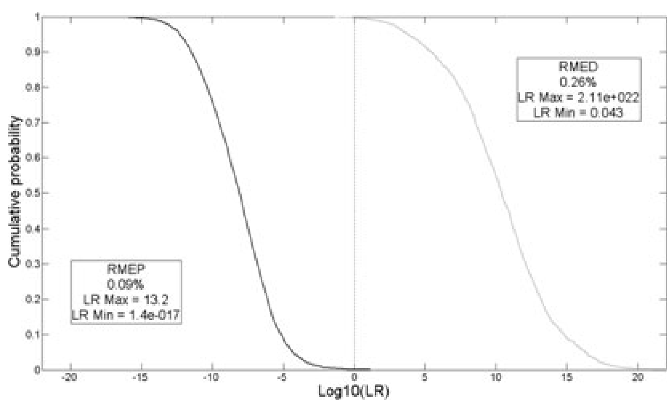

FIG. 15—*Tippett plot of the likelihood ratio (LR)s obtained for firearm W2, when* H_p *(gray line) or* H_d *(black line) is true, respectively. All six scores are taken into account (breech face and firing pin). RMEP obtained is here of 0.09%, whereas RMED is 0.26%.*

first is in relation to the capability of generalization of these results to other types of firearms (within the same class or not) and the number of cartridges needed to construct the *within* distribution. The second deals with the potential effect of the type of ammunition used to produce the test fires. The last point is in relation to the mechanism used in this study to construct the *between* distribution.

Generalization

The results show that the *within* distribution can hardly be generalized to firearms even if that firearm is of the same type. In other words, the system cannot use a default *within* distribution for all cases. Indeed, the performance of the system depends critically on the firearm under examination as shown by the difference reported between W1 and W2. This means that when a comparison between a questioned cartridge case and known cartridges has to be evaluated, the *within* distribution has to be recomputed for the firearm at hand. This observation is valid within a class sharing the same general characteristics (such as our SIG pistols) or between classes, as results obtained on two 7.65 mm Walther PPK and two 9 mm Luger CZ model 75 and 85 have shown (25).

The operational consequence of this is that for each case, a given number of test samples needs to be obtained from the firearm under investigation. As it currently stands, the LRs obtained here represent the somewhat unrealistic situation (given the time required for the 3D acquisition) where 60 cartridges have been used to model that distribution. The question of the appropriate number of test fires to conduct in order to maintain an appropriate balance between acquisition time and performance will be the subject of further research.

Ammunition

In this study, only Geco Sintox ammunition was used. This ammunition has a soft nickel primer allowing a good transfer of fine characteristics. Preliminary tests have been performed with other ammunition types (25). These tests showed the influence of the ammunition type on the quality of the impressions left by a firearm on the primer, which can increase error rates. In other words, performance obtained with the above ammunition may not hold for cases where the questioned cartridge has been fired with different ammunition. The operational consequence would be that both the *within* and the *between* distribution should be obtained on the same type of ammunition as the questioned cartridge. The extent of these effects is currently under investigation.

Mechanisms Used to Obtain the Between Distribution

In its present form, the data used to construct the *between* distribution correspond to all pairwise comparisons allowed by the "between" dataset (a total of 79 firearms, excluding firearms W1 and W2). Hence, the approach uses some sort of an "average" *between* distribution, regardless of the specificity of the markings left on the questioned cartridge case. The assessment here is more suitable to establish the average performance of a system than to deal with a specific case. If casework deployment is foreseen using such a system, in a case involving a questioned cartridge and W1 as putative source, for example, the *between* distribution should be computed by comparing the impressions from the questioned cartridge case against all the impressions from the 79 cartridge cases constituting the *between* database. This is in contrast to the current way of computing the *between* distribution, which does not depend on the specificity of the questioned impressions, that is, there is one distribution for all potential cases. In casework, we would advise computing the *between* distribution on a case-by-case basis, based on the questioned impressions.

Conclusion

An automated technique for the comparison between impressions on cartridge cases based on 3D technology has been developed. This system has been coupled to a bivariate evaluative model allowing assigning likelihood ratios to such comparison results. The system offers a very high discriminating power and the LRs that it provides are very indicative of the true state under both the prosecution and the defense propositions. The model is able to support either hypothesis very strongly for a significant proportion of cases. It is also characterized by relatively low rates of misleading evidence.

In relation to the ongoing debate on the foundation of the discipline (3,4), this research shows that it is feasible to develop an automatic system allowing the objective and reproducible evaluation of evidence arising from similarities/differences between impressed marks found on fired primer cups. This work acts as a first move toward an objective and transparent evaluation process.

Acknowledgments

We would like to thank Jean-Michel Carrier of our institute who provided the main part of the samples used in this research. Also our appreciation goes to the colleagues of the Netherlands Forensic Institute (NFI) for their help in collecting samples used in this study. We also thank the anonymous referees and the editor for insightful and beneficial comments that have so significantly improved our article.

References

1. 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.
2. Schwartz A. A systemic challenge to the reliability and admissibility of firearms and toolmark identification. Columbia Sci Technol Law Rev 2005;6:1–42.
3. National Research Council (NRC) – Committee, Cork DL, Rolph JE, Meieran ES, Petrie CV. Ballistic imaging – committee to assess the feasibility, accuracy and technical capability of a national ballistics database. Washington, DC: National Academies Press, 2008.
4. National Academy of Sciences (US) – Committee on Identifying the Needs of the Forensic Sciences Community. Strengthening forensic science in the United States: a path forward. Washington, DC: National Academies Press, 2009.
5. Nichols RG. Defending the scientific foundations of the firearms and tool mark identification discipline: responding to recent challenges. J Forensic Sci 2007;52(3):586–94.
6. Bunch GS, Smith ED, Giroux BN, Murphy DP. Is a match really a match? A primer on the procedures and validity of firearms and toolmark identification. Forensic Sci Commun 2009;11(3), http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/july2009/review/2009_07_review01.htm (accessed February 11, 2013).
7. Moran B. A report of the AFTE theory of identification and range of conclusions for tool mark identification and resulting approaches to casework. AFTE J 2002;34(2):227–35.
8. Bonfanti M, De Kinder J. The influence of the use of firearms on their characteristic marks. AFTE J 1999;31(3):318–23.
9. SWGGUN and AFTE Committee for the Advancement of the Science of Firearm and Toolmark Identification's response to questions received from the Subcommittee on Forensic Science, Research, Development, Testing, & Evaluation Interagency Working Group, 2011; http://www.afte.org/downloads/RDT&E%20IWG%2025%20Questions%206.14.11%20-%20AFTE%20Response%20w%20cov%20let.pdf (accessed February 11, 2013).
10. Champod C, Vuille J. Scientific evidence in Europe – admissibility, appraisal and equality of arms. Int Comm Evidence 2011;9(1):1–68, http://www.bepress.com/ice/vol9/iss1/art1 (accessed February 11, 2013).

11. Bachrach B. A statistical validation of the individuality of guns using 3D images of bullets. Washington, DC: National Institute of Justice, 2006, March; Document No.: 213674.
12. Bachrach B, Jain A, Jung S, Koons RD. A statistical validation and repeatability of striated tool marks: screwdrivers and tongue and groove pliers. J Forensic Sci 2010;55(2):348–57.
13. Banno A, Masuda T, Ikeuchi K. Three dimensional visualization and comparison of impressions on fired bullets. Forensic Sci Int 2004; 140:233–40.
14. Gambino C, McLaughlin P, Kuo L, Kammerman F, Shenkin P, Diaczuk P, et al. Forensic surface metrology: tool mark evidence. Scanning 2011;33:272–8.
15. Sakarya U, Leloglu UM, Tunali E. Three-dimensional surface reconstruction for cartridge cases using photometric stereo. Forensic Sci Int 2008;175(2):209–17.
16. Senin N, Groppetti R, Garofano L, Fratini P, Pierni M. Three-dimensional surface topography acquisition and analysis for firearm identification. J Forensic Sci 2006;51(2):282–95.
17. Vorburger TV, Yen JH, Bachrach B, Renegar TB, Filliben JJ, Ma L, et al. Surface topography analysis for a feasibility assessment of a national ballistics imaging database. NISTIR 7362: a report prepared for the National Academies Committee to assess the feasibility, accuracy, and technical capability of a national ballistics database under National Institute of Justice Grant 2003-IJ-R-029 with the NIST Office of Law Enforcement Standards, 2007; http://www.nist.gov/manuscript-publication-search.cfm?pub_id=822733 (accessed February 11, 2013).
18. Bonfanti M. Exploitation et interprétation des traces présentes sur les projectiles et les douilles. Can Soc Forensic Sci J 1999;32(1):25–37.
19. Jost T, Hügli H. A multi-resolution scheme ICP algorithm for fast shape registration. Proceedings of the 1st international symposium on 3D data processing, visualization and transmission (3DPVT02) 2002 June 19–21; Padova, Italy. Piscataway, NJ: IEEE Computer Society, 2002; 540–3.
20. Rusinkiewicz S, Levoy M. Efficient variants of the ICP algorithm. Third International Conference on 3-D Digital Imaging and Modeling; 2001 May 28-June 1; Quebec City, Canada; http://www.cs.princeton.edu/~smr/papers/fasticp/fasticp_paper.pdf (accessed February 11, 2013).
21. Nestares O, Heeger DJ. Robust multiresolution alignment of MRI brain volumes. Magn Reson Med 2000;43:705–15.
22. Champod C, Baldwin D, Taroni F, Buckleton JS. Firearm and tool marks: the Bayesian approach. AFTE J 2003;35(3):307–16.
23. Dessimoz D, Champod C. Linkages between biometrics and forensic science. In: Jain AK, Flynn PJ, Ross A, editors. Handbook of biometrics. New York, NY: Springer Verlag, 2008;425–59.
24. Bunch GS. Consecutive matching striation criteria: a general critique. J Forensic Sci 2000;45(5):955–62.
25. Riva F. Etude sur la valeur indicielle des traces présentes sur les douilles [PhD dissertation]. Lausanne, Switzerland: Université de Lausanne, 2011.

Additional information and reprint requests:
Christophe Champod, Ph.D., Professor
Ecole des Sciences Criminelles/Institut de police scientifique
Batochime, Quartier Sorge
CH-1015 Lausanne
Switzerland
E-mail: Christophe.Champod@unil.ch