Conclusion scale for shoeprint and toolmarks examinations

Article *in* Journal of Forensic Identification · March 2006

11 authors, including:

Yaron Shor
Israel police

Special Feature

Conclusion Scale for Shoeprint and Toolmarks Examinations

ENFSI [1]
Expert Working Group
Marks Conclusion Scale Committee [2]

Abstract: The Conclusion Scale Committee (CSC) of the ENFSI Expert Working Group Marks (EWG Marks) reached very soon with common consent a harmonized "Six-Level Conclusion Scale" for interpreting findings in proficiency tests and collaborative exercises within ENFSI. The theoretical fundamentals of this harmonized conclusion scale take into account interpretation models based on the Bayes' rule. However, these mathematical background models weren't and aren't reached with common consent. Some members of CSC say that only one of the three parts of the Bayes' rule - the likelihood ratio - is for the forensic experts. And furthermore, this group favours interpretations of the Bayes' rule with formulations such as *the prosecution and the defense hypotheses* and *presumption of innocence* or *adjudicative fact-finder*. The others - and these are the majority of the members of CSC - have the opinion that words like *prosecution, defense, presumption of innocence,* and *adjudicative fact-finder* are non-scientific elements and therefore things for the judges or jury members and normally these aren't issues for the forensic scientists.

The majority of the members of CSC emphasize that in many countries the practice in court is that the judges and jury members need to get answers to questions such as "What is the probability that the questioned shoesole produced the print?" So, the jury requires answers for a given effect in retrograte to the cause. This is an answer to a

1 European Network of Forensic Science Institutes

2 Horst Katterwe, Chairman of the Conclusion Scale Committee (BKA Wiesbaden, Germany); Dave Baldwin (FSS London, UK); Marcel van Beest (NFI Den Haag, Netherlands); Charles Belser (Kantonspolizei Zürich, Switzerland); John Birkett (FSS London, UK); Alexandre Girod (Police Cantonale Vaudoise Lausanne, Switzerland); Isaac Keereweer (NFI Den Haag, Netherlands); Michel Moes (Federal Police Laboratory Eupen, Belgium); Yaron Shor (Israel Police Headquarters Jerusalem, Israel); Gerrit Volckeryck (Laboratorium voor Technische en Wetenschappelijke Politie Brussel, Belgium); Anja Ytti (NBI Vantaa, Finland).

line of reasoning against the causal direction and that is also termed a *diagnostic result* (a posteriori probability): the interest of the court ultimately lies in the posterior odds in forensic scientific experiments. So, the majority of the members of the CSC apply all three parts of the odds form of the Bayes' rule, using only scientific interpretations by means of the Principle of Causality and of the Principle of Maximum Entropy (PME). And furthermore, they show that this interpretation incorporates the "Traditionalists" with the "Classical Approach".

The harmonized conclusion scale (a level scale) of the EWG Marks is demonstrated by a table with three columns and six steps:

[the likelihood ratio steps are here only in extracts]

Level	***Likelihood Ratio*** *(partial Bayes' rule)*	***Probability*** *(full Bayes' rule, classical approach)*
1	Identification	Identification
2	Very strong support for proposition A	Very probably proposition A
3	Moderate support for proposition A	Probably proposition A
4	Inconclusive	Inconclusive
5	Strong support for proposition Ā	Likely Ā
6	Elimination	Elimination

A = hypothesis: the questioned shoesole produced the print
Ā = Not A = alternative hypothesis: the questioned shoesole didn't produce the print
[here assumed: even prior odds]

Introduction

The main aim of the Conclusion Scale Committee (CSC) of the ENFSI Expert Working Group Marks was to produce a conclusion scale that would enable practitioners to understand the meaning of conclusions formed by examiners across Europe and taking into consideration the criminal justice systems that exist. The scale committee of the Expert Working Group Marks was created in 1999 in Sweden at the business meeting of the 3rd European Meeting for Shoeprint/Toolmark Examiners (SPTM). The committee was formed on a voluntary basis, and it consisted of actively working shoeprint examiners. The members of the committee represented seven different countries. The CSC members are the authors of this report.

The Background for Establishing the Scale Committee

The idea to establish the scale committee of the ENFSI Expert Working Group Marks occurred after getting the results of the two international surveys (1995 and 1997) arranged by marks examiners of the National Bureau of Investigation, Finland [1, 2] and of the survey (1999) conducted by examiners of the Israel National Police Headquarters [3].

In all three surveys it was noticed that there were remarkable differences in the conclusions drawn from similar shoeprint cases in different forensic laboratories. Examiners in different countries, in the same country, or occasionally in the same laboratory had differing opinions concerning the conclusion of a shoeprint case. One reason perhaps could be that there are so many different conclusion scales in European forensic science laboratories. The table below shows a small selection of conclusion scales that are currently in use in Europe and in the United States. The translation from the mother language to English may cause some nuances in the terms used. The scales are listed in alphabetical order.

Examples of Conclusion Scales:

Finland (NBI)	**Israel (DIFS)**	**Netherlands (NFI)**	**United States (AFTE)**	**United States (Footwear)**
identification	identification	identification	identification	identification
probably possible	very high probability high probability probably possible	very probably probably possible	inconclusive A: = some agreement	probably could have
inconclusive	inconclusive	inconclusive	inconclusive B: = no agreement / no disagreement	non-conclusive
appears not		probably not	inconclusive C: = some disagreements	probably did not
elimination	elimination	elimination	elimination	elimination

The Framework and Activities of the Scale Committee

The committee started the work by collecting and reading the literature about statistics, existing scales, etc. in relation to shoeprints, toolmarks, firearms, etc. The actions taken by the committee also included discussions about

- a simple probability scale (yes/no/inconclusive)
- a more complex probability scale (how many steps, using levels 1, 2 instead of words like possible, probably)
- harmonizing the theoretical argumentations/criteria for each conclusion level
- harmonizing the criteria for simple and complex structure of the individual characteristics
- the meaning of wear as a class characteristic and as an identifying characteristic
- the need for a guideline

In the CSC meeting held in Wiesbaden in March 2000, the committee ended up proposing a scale with six levels. The very first proposal for a verbal conclusion scale was *identification, very probably, probably, non conclusive, probably not,* and *elimination*, with proposed definitions to each level.

After the agreement in the number of levels, the work of the committee prolonged because the scale committee received a letter in July 2000 from representatives of the Forensic Science Service (C. Champod, I. Evett, G. Jackson, and J. Birkett) (see Appendix). The work of the CSC and the proposal for a conclusion scale had attracted the attention of some representatives of the Forensic Science Service. In their letter, they called for a complete reconsideration of the proposed scale. They also recommended the adoption of a “Bayesian framework” (likelihood ratio approach) that forces the forensic scientist to adopt reporting convention that describes the degree of support for one proposition versus the other. The letter caused many discussions between the CSC members and discussions and correspondence between the CSC and some forensic examiners in England (FSS), in the Netherlands (NFI), and in Switzerland (Lausanne University).

The "hot" discussions in the CSC meetings in Berlin (2001), Cracow (2000), and Brussels (2000) made it clear that the committee should propose a harmonized scale with numbered levels without any wordings. Thus, the scale of numbered levels could be used in different laboratories in the collaborative and proficiency tests arranged within ENFSI EWG Marks in the future. In an effort to create a harmonized scale, one difficult and demanding problem is the criteria for each of the levels. It has been a generally accepted fact that it is difficult to define or dictate the number of individual characteristics needed in a shoeprint, for example, for a definite positive identification in any given case. There were not any common criteria for "a simple accidental characteristic" and for "a complex accidental characteristic". The guideline of the NFI, presented for the first time in Sweden 1999 [4], is the first attempt to define accidental characteristics in shoeprints and to give value to different types of individualizing features [5].

The Bayes' Rule and Approaches in Forensic Science

Bayesian analysis of the interpretation of evidence, here more in the sense of reasonable expectation of hypotheses (non-statistical probability [6]), relies on the Bayesian rule relating dependencies among uncertain events (hypotheses) through conditional probabilities. In forensic science, there are, as mentioned above, two Bayesian models in use: the likelihood approach and the causality approach (full Bayesian approach). These probability models are based on a probabilistic rule, which was deducted the first time by Bayes [7]. This formula shows how uncertainty about an event (A) can be changed by the knowledge (probabilistic knowledge) of another event (E):

$$p(A|E) = p(E|A) * p(A)/p(E).$$

This is the first Bayes' rule. If the event (A) is called *cause* and E is the *effect*, then this formula is called the Bayes' rule for the probability of causes [8]. If you consider two alternative events/hypotheses A and $\bar{A}$, then you get the second Bayes' rule:

$$p(A|E) = p(E|A)\ p(A)\ /\ [p(E|A)\ p(A)\ +\ p(E|\bar{A})\ p(\bar{A})],$$

and for $p(\bar{A}|E)$ follows:

$$p(\bar{A}|E) = p(E|\bar{A})\ p(\bar{A})\ /\ [p(E|A)\ p(A) + p(E|\bar{A})\ p(\bar{A})].$$

By dividing these two formulas, you get the

"Odds Form of the Bayes' Rule":

$$p(A|E)\ /\ p(\bar{A}|E) \ = \ p(E|A)\ /\ p(E|\bar{A}) * p(A)\ /\ p(\bar{A}),$$

which is, in general, used in forensic science. The left-hand-side of the equation is known as the posterior odds; the last quotient of the right-hand-side as the prior odds. The factor that converts prior odds to posterior odds is the quotient

$$p(E|A)\ /\ p(E|\bar{A}),$$

known as the *likelihood ratio* or *Bayes' factor.*

The Likelihood Ratio Approach

An approach of the Bayes' rule is in favour of the following interpretation: A is the hypothesis proposed by the prosecution, and Ā is the hypothesis proposed by the defense [9-12]. If the likelihood ratio $\{p(E|A)\ /\ p(E|\bar{A})\}$ has a value greater than one, it lends support to the prosecution hypothesis A, and a value less than one lends support to the defense hypothesis Ā. Examples of A and Ā include guilt and innocence, contact or no contact with the crime scene.

The scientific evidence is evaluated by determining a value for the LR (likelihood ratio), because it is said that it is the role of the forensic scientist to evaluate the LR, and it is the role of judge and jury to assess the prior and posterior odds on A. The scientist can assess how the prior odds are altered by the evidence but cannot assign a value to the prior or posterior odds. In order to assign such a value, all the evidence in the case has to be considered. So, the advocates of the LR approach say that the mentioned formulations clarify the position of the scientist as well as that of the jurist and define their relationship: the scientist is concerned solely with the likelihood ratio, whereas jurists deal with the odds on A.

Therefore, the forensic expert can only state the *degree of support* given to the hypothesis versus the alternative. The strength of the evidence is given by the probability of observing the evidence under two chosen propositions. This likelihood ratio has a numerator (very often) close to 1 and a denominator equal to the frequency of the shared features in the relevant population.

The advocates of the LR approach emphasize that this approach has the ability to combine and/or to separate the scientific probabilities (likelihood ratio) from the non-scientific probabilities (prior, posterior odds) in one formula.

The Causality Approach (Full Bayesian Approach)

Another approach is based on the principle of causality using the full Bayes' rule [13]. The principle of causality is the relationship between a cause and the effect that it has:

- there is a thing A (a cause A) that produces a result E (an effect E = evidence E)
- there is a shoesole (A) that produces a shoeprint (E)
- there is a working surface of a tool (A) that produces a mark (E)
- there is the crime scene fracture process (A) that produces a fracture surface (E)
- there is not the crime scene fracture process ($\bar{A}$) that produces the fracture surface (E)

The cause A "produces" the effect E. Knowing that the cause happened, it can be forseen that the effect E will occur or might probably occur, too. This is a predictive line of reasoning: direct causality.

In a scene of crime, the following reality happens: a working surface of a tool A (event A = cause) will produce (or might probably produce) a mark E (event E = effect):

$p(E|A)$ = conditional probability of "effect E" for given "cause A".

For example, the probability of the print E, given that the shoe has left this print (given that event A is true).

However, in many countries the judge will require an answer to the questions, What is the probability that the questioned shoe *had caused* the print on the scene of crime? and, What is the probability that the questioned shoe *had not caused* the print on the scene of crime? This demand was also requested during a Marks Conference about "Interpreting Evidence" in Berlin 2003 by a Supreme Judge of the Superior Court Berlin. The answer to these questions is a line of reasoning against the causal direction that can also be termed diagnostic.

In mathematical terms:

$$p(A|E) = ? \text{ and } p(\bar{A}|E) = ?$$

[a-posteriori probabilities: principle of retro-causal probability].

So, at the scene of crime "happens"

a direct-causality: $p(E|A) = p(\text{effect}|\text{cause})$.

But the answer to the question from the court to
the forensic scientist "demands"

a retro-causality: $p(A|E) = p(\text{cause}|\text{effect})$.

So, the conditioned and the conditioning events are reversed (swapped): "transposing the conditional". The mathematical method, which enables this transposition, is the Bayes' rule (see above).

To answer the above question of the court (the posterior odds), the forensic science expert must make scientific estimations of the prior odds and of the likelihood ratio, for example, by considering the principle of causality and perhaps also the principle of maximum entropy [14]. In doing so, the full Bayes' rule will be used. The result will be a probability scale, which is also in practice of the "Traditionalists" using the "Classical" approach.

The Probability Scale Based on the Causality Approach (Full Bayesian Approach)

In most forensic science disciplines, such as the marks discipline, the experts interpret their findings in the form of verbal conclusions. Very often – for example, in *Footwear Impression Evidence* [15] and *The AFTE Journal* [16] – this process involves the use of probability scales which reflect the experts' degree of probability. The experts who are working in this manner represent the classical approach [15-19]. These verbal probability statements, expressed in natural language, but not backed by numerical statements, have a long tradition.

However, this traditional issue is still the focus of heated worldwide debate that has split the forensic scientific community into two camps: the advocates of the Bayesian approaches and the opponents of the Bayesian approaches. The last mentioned are the "Traditionalists", representing the "Classical" approach.

However, within the advocates of the Bayesian approach, there are also two parts [Appendix]: the advocates of only a partial Bayes' rule, who only make estimations of the LR [9 - 12], and the advocates of the full Bayes' rule, who make two kinds of estimations: one of the likelihood ratio and the other one of the prior probabilities [13].

The advocates of the likelihood ratio approach have the opinion that the Bayes' rule collapses in two parts: one part is for the forensic expert, the other is for the judge. However, many of the forensic experts cannot deal this opinion. The advocates of the likelihood ratio approach say, "Only the likelihood ratio is for the forensic examiner".

Questions and Answers

Why is only the LR for the forensic expert? Why must the hypothesis A be the prosecution's hypothesis? For many forensic experts, the formulation of the hypothesis A is a hypothesis of the forensic expert. Why must the hypothesis Not A be the defense hypothesis? For many forensic experts, the formulation of the hypothesis Not A is a hypothesis of the forensic expert, too.

Why shall the numerator of the LR have always to do with the prosecution's hypothesis and the denominator with the defense's hypothesis? For an unbiased forensic scientific expert, the mentioned nomenclature (prosecution hypothesis, defense hypothesis) is not belonging to an unbiased scientific reasoning! An unbiased forensic scientist shall simply use the Bayes' rule as a tool to compare reasonable hypotheses, without considering which side might advance them! The logical idea of all Bayesian models should be that the full Bayes' rule must be incorporated in each one of the special Bayesian models.

On that manner, a separation of the Bayes' rule into two parts - one for the scientists (likelihood ratio) and the other one for the judge (prior, posterior odds) - is illogical. With skepticism it is viewed that the likelihood approach shall be the only logical way for interpreting forensic scientific evidence.

The Evidence Strongly Supported the Hypothesis A?

When looking to the verbal convention of the likelihood ratio, according to many forensic scientists, the evidence strongly supported the hypothesis A, so the answer of many scientists is, "This verbal convention has nothing to do with the LR. This is not a proper interpretation of the LR."

By saying that the evidence supports a hypothesis, then a guidance, written by the forensic expert for the judge, must be given as to how to combine the evidence with the hypothesis. "The evidence supports a hypothesis", crudely speaking, the forensic expert must mention that this "verbal convention" shall be the LR and not the posterior odds. Otherwise, the judge would make a fatal error because of interpreting these words as the posterior odds!

"The evidence supports a hypothesis" (= LR ???). Crudely speaking, the forensic expert must say in a "paper interpretation": the LR is the "probability of the evidence if the hypothesis (A) is true" in relation to the "probability of the evidence if hypothesis (Not A) is true". So, these arguments point out that the mentioned kinds of interpretations of the Bayes' rule, including the restriction to the LR made by many authors, is accompanied by many illogical attempts. One example is the above-mentioned verbal convention, "The evidence supports the hypothesis A". This phrase isn't a proper definition for the LR. This verbal convention "sounds" more to an expression for the posterior probability than for the LR.

Logical Reasons for a Probability Scale

A logical reason for a probability scale is the question of the judge to the forensic expert:

> "What is the probability that the submitted shoesole made the footwear mark?"

The examination process in the forensic science laboratory begins with this question of the judge posed to the forensic expert. With this question, the forensic examiner is able to formulate some hypotheses (propositions): hypothesis (A) and alternative hypothesis (Not A):

> *A: The submitted shoesole made the footwear mark.*
>
> *Not A: The submitted shoesole didn't make the footwear mark.*

How shall a forensic expert answer to these hypotheses? Many forensic scientists think that the answers to these hypotheses is primarily not concerned with exercises of the "true statistical probability", but more with problems of a "non-statistical probability" that might be called "intuitive probability" or "reasonable expectation of hypotheses" or "plausibility" [6, 14].

Prior Probabilities of the Hypotheses p(A) and p(Not A)

Normally, the forensic expert is given no prior probabilities of the hypotheses p(A) and p(Not A) to work with. The question as posed implies "maximum uncertainty" with respect to the two hypotheses. Maximum uncertainty means that the initial probability distribution should be according the *Principle of Maximum Entropy* (PME) [14, 20 - 22] and without constraints: p(A) = p(Not A) = 0.5. These initial probabilities normally represent the starting point for the expert. Non-scientific information pertaining to matter outside the scope of the expert's responsibility should not be taken into considerations by the forensic science expert.

The Judge and the Expert's Posterior Probability

Experience has shown that most judges intuitively respond appropriately in their use of the expert's opinions:

- The judges raise the posterior probability of, say, a "high probability" to an even higher degree of probability, if other evidence points in the same direction.
- The judges respond with considerable caution to a degree of probability proposed by the expert whenever other evidence points in the opposite direction.

Even Prior Odds as a Bridge Between the Approaches

If the hypotheses are rated equally, the posterior probabilities in the two-hypotheses cases depend only on the likelihood ratio. The same applies when an expert restricts himself to the LR. Thus, in these cases, there is a relationship of equivalence between the LR approach, the full Bayesian approach, and the classical approach. So, the idea of "even prior odds" is the "bridge" between the three approaches.

It must be mentioned that all the members of the CSC made a validation test to prove if even prior odds are the bridge between the approaches. The result was that the participants reached the same level, independant of each other, when using the LR approach and/or the probability scale approach.

To sum it up: prior odds = 1 [based on the scientific background of the PME (in a two-hypotheses case)] seems to be a very good and very transparent estimation.

Full Bayesian Model

A logical probability model requires that the hypotheses in the conclusions - expressed in the expert report - must relate to the hypotheses proposed in answer to the initial question. And this is done by the full Bayesian model. Furthermore, the full Bayesian model is compatible with

- the "Traditional Working Methods" of marks experts in Europe and North America [13, 15 - 18, 19],
- the general conditions under which expert examinations are conducted, and gives
- a logical answer to the question of the judge!

In summary, most of the authors are confident that the proposed logical and scientific approach - *full Bayesian approach including classical approach* - reflects the actual practice and circumstances in many countries. Moreover, verbalized statements of probability are standards in everyday practice and have become a part of judical vocabulary [23 - 25].

The Harmonized Conclusion Scale of the ENFSI EWG Marks

The harmonized conclusion scale for interpreting findings in proficiency tests of the ENFSI EWG Marks is demonstrated by a table with three columns and six levels/steps.

Level	***Likelihood Ratio*** *(partial Bayes' rule)*	***Probability*** *(full Bayes' rule, classical approach)*
1	Identification	Identification
2	Very strong support for proposition A Strong support for proposition A	Very probably proposition A
3	Moderately strong support for proposition A Moderate support for proposition A Limited support for proposition A	Probably proposition A
4	Inconclusive	Inconclusive
5	Limited support for proposition Ā Moderate support for proposition Ā Moderately strong support for proposition Ā Strong support for proposition Ā Very strong support for proposition Ā	Likely not proposition A
6	Elimination	Elimination

A = hypothesis: the questioned tool produced the mark
Ā = (Not A) = alternative-hypothesis: the questioned tool didn't produce the mark
(when using the full Bayes' rule, then even prior odds are assumed).

Experts who are in favour of the LR approach shall look at the middle column first. Experts who are in favour of the full Bayesian approach shall look at the right column at first. In doing so, the full Bayesian approach experts shall also mention in the report the assumed values of the prior probabilities of (A) and (Not A): transparent prior probablities. In many two-hypotheses cases, it is assumed, after estimations by means of the PME method, that the prior probabilities of the two hypotheses are equally likely.

After the first look, both expert groups shall look to the *level* column on the left.

The level number of this scale in the corresponding row is the result of the forensic scientific experiment, regardless if it is in favour of the LR approach, or in favour of the full Bayesian-approach, or in favour of the classical approach, which is in cases of even prior odds incorporated in the full Bayesian approach.

Guidelines

It is necessary to interpret findings by using a harmonized conclusion scale, but it is also very important to know when a specific conclusion will be drawn. A first step was done by the Netherlands Forensic Institute (NFI) that started to develop a guideline in 1997 [4], at first for training footwear examiners. After a number of updates, the guideline is now suitable for all footwear experts at the NFI, scene of crime officers, and anyone working in the field of comparative footwear examination in the Netherlands [5].

The NFI guideline consists of several phases. In the first phase, the shoe and the print are examined with respect to manufacture and acquired features. Then the shoe and the print are compared to determine similarities and differences. In the third phase, the similarities and differences are evaluated, and a conclusion is drawn based on a fixed set of verbal conclusion levels. Now a report should be written.

The main focus of the guideline is based on the evaluation of similarities in acquired features between the shoe and the print. For each similarity, a characteristic value is assigned, ranging from high to low in six categories. Furthermore, it is assessed whether the shape of the damage in the shoe coincides more or less complete with the shape in the print (full similarity) or only partially (partial similarity). The number of similarities is determined for each combination of the characteristic value and the degree of similarity. Subsequently, a score is read from a set of curves for all similarities in the same combination. These scores are added to obtain the total score for the total of all similarities. Finally, the verbal conclusion is read from a scale that relates the total score to a fixed set of verbal conclusion levels.

So, when drawing conclusions in footwear examinations the following aspects play a principal part:

- the number of similarities
- the characteristic value of the similarities (depending on complexity)
- the extent of similarity in shape between the damage to the shoe and the irregularities in the impression
- the value of similar wear and tear

It appears that there are good reasons for following guidelines. Hence, when everyone does the same, the same conclusions shall be drawn. When the examiner fills in the conclusion form of the guidelines, it is clear to see what he has counted and which values he has given to similarities. This is easy for another examiner to check and, when necessary, it gives openness to the court.

In the meetings of the CSC, the Dutch proposal was discussed several times. Some members said that this guideline is too severe, since a too great number of very many acquired features are needed to conclude to identification. They wanted to draw higher conclusions than the guidelines recommended. Furthermore, there were discussions about size restrictions. Besides these points of criticism, everyone agreed that this type of guideline is ideal to get everyone in line and that the principle to assign values to each acquired feature, depending on size and complexity, is a good one.

The NFI guideline was approved for collaborative tests by redoing the shoeprint comparisons that had been distributed by the NBI/Finland for the SPTM meetings in 1995 and 1997 [1, 2]. Conclusions for these comparisons should be made and explained using the proposed six steps scale of the CSC and the NFI guideline. The result of these tests was that all participants reached the same level (acceptable error rate of 1 level).

To sum it up, the NFI guideline is a very good instrument that gives instructions about the way to do and to get harmonized conclusions of forensic scientific findings. It is a very good principle that helps one to decide what one should do in the difficult situation of writing the real result of an objective test. So, there are many pros for the guideline, such as

- it is a thorough system
- it is a perfect teaching model
- the principles of footwear comparison are incorporated
- the latitude for improvising is little
- there are remarkably consistent (harmonized) results

Mathematical Models for the Complexity and for the Accidental Re-occurence of Marks

In addition to the guideline, attempts were made to describe more objectively the complexity of irregularities and to estimate the accidental re-occurence of marks with the help of probability theory models. Are there in the literature probability models for the estimation and the calculation of the likelihood ratio? If the answer is, “Yes, there are”, then it would be possible to estimate the value of the posterior probability, provided that an estimation of the prior probability was done (see above). Many discussions were made about simple structures, complex structures, identifying characteristics [13]: is a triangular plate more complex than a circular plate, or a pyramid more complex than a cone? Discussions were made about the characteristic value of a damage and about the complexity of forms/shapes on considering a stereometrical form factor (i.e., quotient of the form’s area divided by the form’s circumference multiplied with a standardization factor $2 * (\pi)^{1/2}$) [13, 26].

There are some other forensic scientists who made probability theory model calculations for the complexity of impressed tool marks. A useful model using “random walks” as a stochastic-geometric process for the estimation of the complexity of the shape of impressed marks is described by Turkowski, Klingenberg, and Katterwe [27 - 29]. A good overview of “How unique are impressed toolmarks?” is given by Stone [30], although he didn’t mention the earlier formulated impressed tool marks models described by Turkowski, Klingenberg, and Katterwe.

In addition to these complexity models of imprints, efforts have been made to obtain objective criteria to determine whether the degree of experimental similarity is sufficient for identification. Keeping this in mind and, additionally, to get an idea of the nature and perhaps of the uniqueness of the features, a number of probability models, often by using the Bayes’ rule, have been formulated by various authors [22, 29, 31 - 39]. Many probability models are based on the interpretation of the denominator $p(E/\bar{A})$ of the likelihood ratio.

Applications of Probability Models for Marks: Calculations of the Likelihood Ratio and of the Posterior Probability Based on the Full Bayes' Rule

Examples for calculations in real forensic science cases (tool marks, fracture marks) are given that are based on probability theory models [22, 29, 31-35]. These models are based on the interpretation of the denominator $p(E/\bar{A})$ of the likelihood ratio, i.e. the probability of the evidence if the alternative hypothesis ($\bar{A}$) of the forensic expert is true. These models are also described as "obtaining a match by chance". Supposing that the match is certain, if the hypothesis (A) of the forensic expert is true [$p(E/A) = 1$], then the calculations of the LR [$LR = 1 / p(E/\bar{A})$], of the prior probability and of the posterior probability can be made. Those results were presented and accepted in German courts. The models are developed for striation marks as well as impression marks (tool marks, striation marks on bullets, impression marks on cartridges, fracture marks). The following example describes an application of a probability model on the comparison of striated tool marks [22, 32, 35].

Example: Striation Marks

In striation marks cases, the relevant forensic information is given by the position, intensity, and the shape of the striations. The model used here is only based on the position of the striation lines [31, 32, 35]. Therefore, the image data were reduced to line diagrams. The positions of the lines of the evidence and test marks were compared (Figure 1). The range of the marks is divided into n intervals (linear cells). Each interval may contain one line at the most. The width of the interval is determined by the maximum distance of corresponding striae of two lines. The result of a comparison is represented in the pattern of m matches. The probability that m matches fall within n intervals is calculated with the aid of the binomial function (binomial function model).

Here is the result of a screwdriver case including marks on a music box [35]. A probability value $p(E/\bar{A}) = 1.2 \; 10^{-15}$ was obtained. With this, a LR-value of $1/1.2 \; 10^{-15} = 8 \times 10^{14}$ is calculated. The equivalence between numbers and words is according to Champod et al. [40] - a matter of policy, but at the present time, the verbal convention adopted by them is as follows:

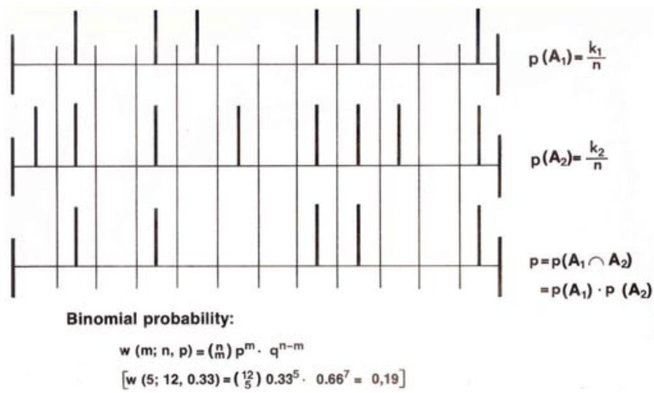

Figure 1

The principle of the binomial function model (linear cells)

LR	**Verbal expression**
1 to 10	Limited evidence to support
10 to 100	Moderate evidence to support
100 to 1000	Moderately strong evidence to support
1000 to 10000	Strong evidence to support
> 10000	Very strong evidence to support
The scale works in directly comparable way for likelihood ratios less than one.	

Here, LR > 10 000 – very strong evidence to support hypothesis A. And furthermore:

$p(A|E) / p(\bar{A}|E) = LR \times p(A) / p(\bar{A}) = 8 \times 10^{14} \times 1 = 8 \times 10^{14}$ (with even prior odds), so

$p(A|E) = 8 \times 10^{14} \times p(\bar{A}|E) = 8 \times 10^{14} \times [1 - p(A|E)]$, so

$p(A|E) = 1$ (identification).

Remarks

It is well known that there are many forensic examiners who disagree with any attempt to determine a degree of similarity, or introduce probability into an identification science. Many feel it is not worthwhile to be concerned with probability theory models and, furthermore, the introduction of them might be harmful. Others feel that if one is to introduce statistical analysis or probability theory, one is tacitly stating that an identification is not an identification but a probability with a possible error factor. However, sometimes requests will come from the court (this often happens in Germany) to calculate numerical values with probability theory models in special cases. Besides these requests, the abstract mathematical models are useful tools and give valuable insights into the properties of mark comparison and into the nature and the uniqueness of a mark. Already this aspect alone makes it worthwhile to be concerned with probability theory models.

Conclusion

When doing a comparative marks examination, the forensic expert comes to a conclusion on the origin of the examined items. The practice until today is that most forensic scientific labs have their own conclusion scale for this kind of examination. This is in most labs not a scale in percentages but in words.

All the conclusion scales have a number of levels, from three to thirteen levels. At the highest level, there is no doubt in the examiner's opinion: the items examined are coming from one source. The middle level is inconclusive: the items can or cannot come from one source. At the lowest level, there's no doubt: the items come from different sources. Between the highest and the middle level, there is an area where specific matching details are found. The same is valid for the level between the middle and the lowest.

The goal of the project "Harmonized Conclusion Scale" was to create a conclusion scale which shall be used by all the participants of proficiency tests created by the ENFSI EWG Marks (work instruction). This work instruction is in accordance with the "Guidance on the conduct of proficiency tests and collaborative exercises within ENFSI" [Standing Committee for Quality and Competence (QCC)].

The project “Harmonized Conclusion Scale” of the ENFSI EWG Marks produced many discussions regarding the way of formulating conclusions not only in proficiency tests and collaborative exercises but also in forensic scientific expert reports. There were many fruitful discussions between advocates and opponents for using the Bayes’ rule on the whole when interpreting evidence. However, there are differences of opinions not only between the traditionalists (classical approach) and the Bayesians, but also between these Bayesians there are different opinions: likelihood ratio approach on the one side and full Bayes’ rule approach on the other side.

The opinion of the members of the Conclusion Scale Committee is that *only the legal systems within the particular countries* (and not a Standing Committee of ENFSI or not the Board of ENFSI) *may decide on approval regarding interpreting evidence of the forensic scientific experts reports.*

The members of the Conclusion Scale Committee are fully aware of the variation in judical systems within Europe. The reporting of scientific findings and the way in which the experts interpret their findings will be driven by the requirements of the legal system within their country. It is therefore not surprising that various countries have developed their own scales. The Conclusion Scale Committee of the ENFSI EWG Marks considers fundamental principles that underpin the evaluation of evidence.

The special requirement for the harmonized conclusion scale for interpreting findings in proficiency tests and collaborative exercises in marks cases is that the proposed scale allows communication of results, whatever the method an examiner uses (classical approach, likelihood ratio approach, full Bayes’ approach).

For further information, please contact:

Horst Katterwe
Bundeskriminalamt
Forensic Science Institute
Materials Technology
D-65173 Wiesbaden / Germany
+49(0)61155-12683
horst.katterwe@bka.bund.de

References

1. Majamaa, H.; Ytti, A. Survey of the Conclusions Drawn of Similar Footwear Cases in Various Crime Laboratories. *For. Sci. Int.* **1996**, *82* (1), 109-120.
2. Ytti, A.; Majamaa, H.; Virtanen, J. Survey of the Conclusions Drawn of Similar Shoeprint Cases, Part II. Proceedings of the European Meeting for Shoeprint and Toolmark Examiners, The Hague, Netherlands, 1997, pp 157-169.
3. Shor, Y.; Weisner, S. A Survey on the Conclusions Drawn on the Same Footwear Marks Obtained in Actual Cases by Several Experts Throughout the World. *J. For. Sci.* **1999**, *44* (2), 380-384.
4. Keereweer, I. Guideline for Drawing Conclusions Regarding Shoeprint Examinations. *Information Bulletin for Shoeprint/ Toolmark Examiners* **2000**, *6*, 47-62.
5. Keereweer, I.; van Beest, M.; van de Velde, J. M. *Guideline for Evaluating and Drawing Conclusions in Comparative Examination of Shoeprints*. Netherlands Forensic Institute: Netherlands, 2005.
6. Cox, R. T. Probability, Frequency and Reasonable Expectation. *Am. J. Physics* **1946**, *14* (1), 1.
7. Bayes, T. An Essay Toward Solving a Problem in the Doctrine of Chances. *Phil. Trans. Roy. Soc.* **1763**, *53*, 330.
8. Feller, W. *An Introduction to Probability Theory and its Application*; John Wiley: New York, 1968; p 124.
9. Taroni, F. M.; Champod, C.; Margot, P. Statistics: A Future in Tool Marks Comparison? *The AFTE J.* **1996**, *28* (4), 222-229.
10. Champod, C.; Evett, I. W. A Probabilistic Approach to Fingerprint Evidences. *J. For. Ident.* **2001** *51* (2), 101-122.
11. Champod, C.; Baldwin, D.; Taroni, F.; Buckleton, J. S. Firearms and Tool Marks Identification: The Bayesian Approach. *The AFTE J.* **2003**, *35* (3), 307-316.
12. Taroni, F.; Biedermann, A. Inadequacies of Posterior Probabilities for the Assessment of Scientific Evidence. *Law, Probability and Risk* **2005**, *4* (1-2), 89-114.
13. Katterwe, H. About the Range of Conclusions. Proceedings of the European Meeting for Shoeprint and Toolmark Examiners, Berlin, Germany, 2001, pp 185-195.
14. Wöger, W. Zu den modernen Grundlagen der Datenauswertung in der Metrologie (Bayes' Theorem, Principle of Maximum Entropy). *PTB-Mitteilungen, Physikalisch-Technische Bundesanstalt* **2001**, *111* (3), 210-225.
15. Bodziak, W. J. *Footwear Impression Evidence*, 2nd ed.; CRC Press: Boca Raton, FL, 2000; pp 372-374.

16. Moran, B. A Report on the AFTE Theory of Identification and Range of Conclusions for Tool Mark Identification and Resulting Approaches to Case Work. *The AFTE J.* **2002**, *34* (2), 2002, 227-235.
17. Bodziak, W. J. Traditional Reporting of Footwear Examination Results by the FBI and Other Laboratories in the United States. Proceedings of the European Meeting for Shoeprint and Toolmark Examiners, Berlin, Germany, 2001, pp 143-148.
18. Wiersema, S. J. Is the Bayesian Approach for You? Proceedings of the European Meeting for Shoeprint and Toolmark Examiners, Berlin, Germany, 2001, pp 149-153.
19. Rodowicz, L. Polish Methodology of Forensic Shoeprint Identification. Proceedings of the European Meeting for Shoeprint and Toolmark Examiners, The Hague, Netherlands, 1997, pp 83-95.
20. Shannon, C. E.; Weaver, W. A Mathematical Theory of Communication (Entropy, Probability and Information), University of Illinois Press: Champaign, IL, reprinted 1999.
21. Jaynes, E. T. Information Theory and Statistical Mechanics. *Physical Review* **1957**, *106* (4), 620.
22. Katterwe, H. Forensic Science and Interpreting Evidence: Probability Conclusions in Expert Opinions on Fracture Matching. Proceedings of Material Testing DVM Berlin, 2005, 437- 444, ISSN 1861-8154.
23. Judgement of the German Federal Supreme Court (BGH Urt. v. 12.8.1992 - 5 StR 239/92 (LG Hannover), Beweiswert der DNA-Analyse: neutrale Anfangswahrscheinlichkeit (unbiased estimation: prior odds = 1), Neue Juristische Woche NJW, 46, 1992, pp 2976- 2977.
24. Rösing, F. W. Standards fü r die Anthropologische Identifikation. *Zeitschrift für Strafrecht* **1999** (5), 230-232.
25. Knußmann, R. Zur Wahrscheinlichkeitsaussage im morphologischen Identitätsgutachten. *Zeitschrift für Strafrecht* **1991** (4), 175-177.
26. Saltykov, S. A. *Stereometrische Metallographie*. VEB Leipzig, 1974, p 334.
27. Turkowski, C. Über die Wahrscheinlichkeit des Auftretens gleicher Muster. *Archiv für Kriminologie* **1973**, *155* (3), 144.
28. Klingenberg, A. Doctoral Dissertation, Tools and Marks: An Identification Problem. Techn. Univ. Lyngby, Danmark, 1980.

29. Katterwe, H. Modern Approaches for the Examination of Toolmarks and Other Surface Marks. *For. Sci. Rev.* **1996**, *8* (1), 45-72.
30. Stone, R. S. How Unique Are Impressed Toolmarks? *The AFTE J.* **2003**, *35* (4), 376-383.
31. Deinet, W. Studies of Models of Striated Marks Generated by Random Processes. *J. For. Sci.* **1981**, *26* (1), 35-50.
32. Deinet, W.; Katterwe, H. The Application of Probability Theory Models on the Comparison of Striations. Proceedings of the Scandinavian Conference on Forensic Science, Linköping, Sweden, 1983, pp 89-94.
33. Katterwe, H. The Matching of Surface Details and its Analysis by Probability Theory Models. *J. For. Sci. Soc.* **1984**, *24* (4), 284.
34. Katterwe, H. Re-occurence of marks. Proceedings of the Probability Methods In Forensic Science. Bundeskriminalamt BKA, Wiesbaden 1986, pp 1-26.
35. Katterwe, H.; Deinet, W. Wahrscheinlichkeitstheoretische Modelle und Werkzeugspuren (Probability Models and Tool Marks). *Archiv für Kriminologie* **1983**, *171* (3), 78-88.
36. Biasotti, A. A. A Statistical Study of Individual Characteristics of Fired Bullets. *J. For. Sci.* **1959**, *4* (1), 34-50.
37. Gardner, G. Y. Computer Identification of Bullets. *The AFTE J.* **1979**, *11* (1), 26-36.
38. Uchiyama, T. The Probability of Corresponding Striae in Toolmarks. *The AFTE J.* **1992**, *24* (4), 273-281.
39. Brackett, J. W. A Study of Idealized Striated Marks and Their Comparisons Using Models. *J. For. Sci. Soc.* **1970**, *10* (1), 27-56.
40. Champod, C.; Evett, I.; Jackson, G.; Birkett, J. Comments on the Scale of Conclusions. *Information Bulletin for Shoeprint/ Tiremark Examiners* **2000**, *6* (3), 15-18.

Appendix

Informative discussions about Bayesian approaches (pro and contra) between experts of the ENFSI EWG Marks and other forensic scientists, published in IBSTE*.

IBSTE	Pages, Authors, Title, Content
Vol. 6, No 3 October 2000	10: H. Majamaa: Harmonising the scale of conclusions – the Bayesian approach? 11-18: Ch. Champod, Ian Evett, Graham Jackson, John Birkett: Comments on the scale of conclusions proposed by the ad hoc committee of the ENFSI Marks Working Group: against the proposed scale and against Charles Belzer' guidelines; and also against Isaac Keereweer's Guideline (2000). Table of LR and Verbal expression: ... "evidence to support"... 19-21: H. Majamaa: Comments to the FSS Comments
Vol. 7, No. 1 January 2001	22: A. Ytti: Report of the second EAFS Meeting in Cracow (2000) Business meeting with Ian Evett, Ch. Champod, G. Jackson about the Bayesian approach (pro and contra)
Vol. 7, No. 2 June 2001	6-28: A. Ytti: Report of the SPTM 2001 in Berlin Panel Session "Range of Conclusions" [see also H. Katterwe, A. Körschgen Proceedings SPTM 2001 Berlin, ISBN 3-00-0093] H. Katterwe: Marks scale committee and the range of conclusion G. Jackson: Principles of evidence interpretation K. Nissen: Alternative hypotheses and Bayes theory W. Bodziak: Traditional reporting of footwear examination results by the FBI and other laboratories in the US S. Wiersema: Is the Bayesian approach for you? B. Moran: A report of the AFTE theory of identification and the range of conclusions for toolmark identification and resulting approaches to casework R. Kennedy: How to express probability in footwear comparison W. McDowell: Report conclusions in Northern Ireland M. Sjerps, I. Keereweer: A Bayesian view on the interpretation of shoeprint evidence H. Katterwe: Principle of causality, Bayes' rule and principle of maximum entropy 37-41: F. Taroni, P. Margot: General comments on the scale of conclusions in shoemarks – the need for a logical framework
Vol. 7, No. 3 November 2001	8-10: H. Katterwe: SPTM 2001 Meeting: Remarks on the session "range of conclusion" 17-18: R. Kennedy: How to express probability in footwear comparison [IAI conference, Miami 2001] 22: H. Katterwe: European marks scale committee and the range of conclusions: Bayesian or Non-Bayesian view [AFTE conference, Newport Beach 2001]

* [IBSTE = Information Bulletin for Shoeprint/Toolmark Examiners published by the Expert Working Group Marks of ENFSI (European Network of Forensic Science Institutes); editors: Anja Ytti, Gerrit Volckeryck; ISSN 1455-4194]

Vol. 8, No. 1 June 2002	17-22: G. Volckeryck: Report of the fifth meeting of the committee on the harmonisation of conclusion scales, London 2002 23-30: H. Katterwe: Comments/objections to reproaches of FSS and University Lausanne 31-32: I. Keereweer: Comments and objections to reproaches
Vol. 8, No. 2 September 2002	11-12 H. Katterwe: European marks examiners wording, the Bayes' rule and a causality model for evidence interpretation [IAI conference, Las Vegas 2002] 15-25: F. Taroni, J. Buckleton: Likelihood ratio as a relevant and logical approach to assess the value of shoeprint evidence
Vol. 8, No. 3 December 2002	17-21: H. Katterwe: Comments to the article of F. Taroni and J. Buckleton 22-25: C. Champod, G. Jackson: Comments on the current debate on the Bayesian approach in marks examinations
Vol. 9, No. 1 June 2003	16-18: H. Katterwe: A letter to C. Champod and G. Jackson – "open-minded dialogue"
Vol. 9, No. 2 September 2003	18-25: H. Katterwe: True or false? – a special forensic science feature – estimation of prior odds in the odds-form of the Bayes' rule
Vol. 10, No. 1 December 2004	12-16: H. Katterwe: Scale committee (principle of maximum entropy, cooperation with judges of superior courts, validation tests, prior odds = 1)
Vol. 11, No. 1 April 2005	10-21: H. Katterwe: Scale committee [six-level-conclusion-scale, Bayes' rule, principle of causality, principle of maximum entropy, LR-approach, estimation of prior odds, DNA and prior probability (prior even odds), guidelines, cooperation with judges, validation tests, models and estimation calculations of the likelihood ratio, references (literature)]