The right weight in the scales of justice

Charles E.H. Berger

Netherlands Forensic Institute, The Hague, The Netherlands.

‘Measuring is knowing’ is the common expression, and measurement can indeed provide important information. But to draw conclusions from measurement results you will almost always need more than that. Inferences are not only based on those results but also on logic and empirical data. This should indicate to what extent the results are evidence for the truth of relevant hypotheses in a case. When it comes to truth finding, the scientific side of evidence is essential for putting the right weight in the scales.

Lawyers - and they are not alone in it - often confuse science and technology, and scientists with technicians that do measurements. This is unfortunate, because science has a lot more to offer to lawyers than just technology and measurements. Science is also invaluable when it comes to the interpretation of measurements and observations and reasoning with uncertain information. When a lawyer has no knowledge about this and counts on experience and intuition only, he runs the risk of making critical mistakes. This is a major problem, especially since it has been said that: “Judges and lawyers usually react to science with all the enthusiasm of a child about to get a tetanus shot” [1].

The interpretation of evidence is a central theme in criminalistics, but this knowledge is as important for ‘technical’ evidence as for ‘tactical’ evidence, and observations of the court. Interpreting evidence, and reasoning with evidence still often takes place at too low a level.

For example, a judge often expects the expert to come to a decision on issues in the case that are within his area of expertise. The court then ‘anchors’ its final verdict in the (categorical) statements of the expert(s). But in most cases, an expert cannot make such a statement. He does not have the information or the role to make decisions, but to describe the scientific evidential value of his observations. Thus, he enables the lawyer to weigh that part of the totality of the evidence in the case. It is the duty of the court to come to a final verdict, after weighing all the information in the case.

Until that decision, a rational decision maker should take the (limited) evidential value of all the information into account. When uncertainties are elevated to certainties before then, it can lead to a wrong decision. Just like a bunch of grapes would cost nothing if the price of a grape would be rounded to zero first. The standard of proof should not be applied to the separate elements, but to the final total. That also means that the evidence should be combined first. In another article in this issue combining evidence will be discussed [2]. In this article, we will limit ourselves to determining the evidential value of the individual elements.

Science is ahead of the legal world when it comes to the interpretation of examination results in terms of evidential value. It is indeed directly related to the way the whole of scientific knowledge is constantly expanded. In criminalistics, we see that for subjective forms of examination so-called conclusion scales are used: series of verbal terms that describe the amount of evidential value. But sometimes those terms (even within one series) were referring to different things such as certainty, possibility, probability and indication [3].

Traditionally, many verbal conclusion scales consist of terms for the probability of a hypothesis. This is due to the misconception that a forensic examination of a trace could determine the probability that any particular suspect left that trace. But a forensic examination in isolation cannot determine this probability, because this probability is also determined by many other factors. Other forensic traces, but also other information. For example, what about the probability that the accused was present at the crime scene, does he have an alibi? The stronger the alibi is, the smaller the probability that the suspect left the trace will be. Therefore, the examination of the trace alone cannot determine that probability.

But then what is the evidential value of the examination of the trace? That value lies in the extent to which the examination helps to distinguish between hypotheses, in this case hypotheses about who left the traces behind. The evidence does not determine the probability of the truth of a hypothesis, but it does make this probability (or our legitimate degree of belief) increase or decrease. Evidential value in that sense, is a relative term.

Already in 1908, this relativity was pointed out to the renowned forensic expert Alphonse Bertillon in the Dreyfus trial [4]. The similarities he observed in a comparative examination could not determine the probability of the hypothesis of equal origin, as Bertillon had claimed. It was the mathematician Henri Poincaré from the *Academie des Sciences* that pointed out that the evidential value of the perceived similarities could only give the relative change of the probability of a hypothesis, and not that probability itself.

The evidential value makes the ratio of the probabilities of two competing hypotheses (e.g. about the source of a trace) change. The factor by which it changes follows from the laws of probability. It is called the likelihood ratio (LR) and provides a universal measure of evidential value. The LR (and therefore the evidential value) is equal to the probability of the observation when one hypothesis is true, divided by the probability to observe the same when the competing hypothesis is true [5].

Poincaré had such knowledge, but it was not adopted by the forensic world, which did not have strong academic roots and was strongly related to law enforcement. Due to the outbreak of the First World War, forensic science got even less attention and this hampered the development of forensic knowledge. The existing knowledge became a technical, protocoled tool for law enforcement. Poincaré's warnings were forgotten.

Sir Harold Jeffreys came with a scale of verbal terms for evidential value in 1939. He associated the numerical LR with verbal terms such as 'not worth more than a bare mention' (LR: 1-3), 'substantial' (3-10), 'strong' (10-30), and 'very strong' (30-100) [6]. In more recent times, it was Irving John Good [7] who used probability theory for the weighing of evidence, for which he also saw forensic applications. But a serious application of the probabilistic approach in modern forensic science had to wait until the seventies of last century in forensic glass examination [8]. It was especially the advent of DNA testing [9] in the eighties that caused this approach to penetrate the rest of the forensic world [10].

It enabled forensic science to objectify examinations and evidential value, and made clear what methodology should be applied. In comparative examination of traces [11] the question is whether a trace of unknown (or disputed) origin has the same origin as a reference trace of known origin.

To objectify such examinations, the following steps describe a correct methodology:

- First, the observation itself should be objectified, by defining a number of relevant, objective features of the traces (in the broad sense).
- Then an objective measure of the difference between (or similarity of) two sets of features is defined: the score.
- For the interpretation of this score, background data are necessary. How different are traces from the same source, and how different are traces from different sources? Therefore, comparisons are made between traces from the same source (minor differences are expected), for many sources in a collection that is representative of the population. The same is done for traces from many different sources (larger differences are expected).
- With these background data, we can determine how probable an observed degree of difference in a case is under the hypotheses of same or different source. The ratio of these two probabilities is equal to the evidential value (LR) for the hypotheses of equal or different origins of the traces. With that, we have objectified the interpretation of the examination results as well.

Even when the features cannot be quantified, the same logic applies. Without numbers, the probability of the observations under two competing hypotheses still needs to be evaluated to determine the (non-numerical) evidential value. The verbal conclusion scale of the NFI for example reads [12]:

The findings of the examination are...

- about equally likely;
- somewhat more likely;
- more likely;
- much more likely;
- very much more likely

...when hypothesis 1 is correct, as / than when hypothesis 2 is correct.

Because of the advances in forensic interpretation and evaluation of evidence, it also has become clear how complex it really is. In addition, a number of problems that have always been there now clearly emerged: the so-called thinking errors or fallacies. The two main fallacies are named after the prosecution and defense: the so-called prosecutor's fallacy and defense attorney's fallacy [13].

Although the names suggest who would mainly be committing these fallacies, in practice the prosecutor's fallacy turns out to be the fallacy most often made by all parties. This fallacy is also referred to as the 'transposed conditional'. A common example is the transposition by which a conclusion such as "the probability of a matching DNA profile, given that the accused is not the donor of the trace, is 1 in 1 billion" is wrongly understood as "the probability that the suspect was not the donor of the trace, given the matching DNA profile, is 1 in 1 billion".

How wrong this transposition is becomes evident with a simpler example as "the probability that an animal has four legs, given that it is a cow" which is clearly not the same as "the probability that an animal is a cow, given that it has 4 legs". In general, this fallacy confuses "the probability of the observation given the hypothesis" with "the probability of the hypothesis given the observation".

In the case of the defense attorney's fallacy, a conclusion as "the probability of a matching DNA profile, given that the accused is not the donor of the trace is 1 in 1 billion" is dismissed by adding that the suspect (with the current world population) is only one of seven people that could be the donor of the trace. The fallacy here is that it is implicitly and wrongly assumed that the entire world population is equally probable as a donor, an assumption that in reality will never be justified. A similar error is to state that in spite of such strong evidence a suspect is "merely not excluded".

The above is no easy matter, but I hope that this "scientific injection" was not too painful, and that it has become clear that more knowledge about it is important for the weighing of evidence and hence for making a correct decision.

References

[1.] B. Black, F.J. Ayala, C. Saffran-Brinks, 'Science and the Law in the Wake of Daubert: A New Search for Scientific Knowledge', Texas Law Review 1994-72, p. 715-802.

[2.] See the contribution of Sjerps and Kloosterman in this edition of Ars Aequi: M. Sjerps and A. Kloosterman, 'Het gebruik van Bayesiaanse netwerken in de forensische (DNA)-statistiek', AA July / August 2010, p. 502, AA20100502.

[3.] D.A. Rudram, 'Inter-pretation of scientific evidence', Science & Justice 1996-36, p. 133-138. On p. 137 he cites the conclusion scale of the document examiners of the Metropolitan Police Forensic Science Laboratory.

[4.] Henry Mornard (1907), L'affaire Dreyfus: la revision du procès de Rennes, Paris: Ligue française pour la défense des droits de l'homme et du citoyen, p. 334.

[5.] For an easier explanation see: C.E.H. Berger, 'Criminalistiek is terugredeneren', NJB 2010, p. 784.

[6.] Harold Jeffreys (1939), Theory of Probability, Oxford: The Clarendon Press.

[7.] I.J. Good (1950), Probability and Weighing of Evidence, London: Charles Griffin & Company Ltd.

[8.] D.V. Lindley, 'A problem in forensic science', Biometrika 1977-64, p. 207-213. I.W. Evett, 'The Interpretation of refractive index measu-rements', Forensic Sci 1977-9, p. 209-217.

[9.] I.W. Evett & B.S. Weir (1998), Interpreting DNA evidence – statistical genetics for forensic scientists, Sunderland: Sinauer Associates.

[10.] B. Robertson & G.A. Vignaux, Interpreting Evidence: Evaluating Forensic Science in the Courtroom, Chicester: Wiley 1995. C.G.G. Aitken & F. Taroni, Statistics and the Evaluation of Evidence for Forensic Scientists, Statistics in practice, Chicester: Wiley 2004.

[11.] Most forensic examinations are of this type, whether it concerns DNA, bullets, fibers, glass, or finger marks.

[12.] De reeks waarschijnlijkheidstermen van het NFI en het Bayesiaanse model voor interpretatie van bewijs, Vakbijlage NFI, 2008.

[13.] W.C. Thompson & E.L. Schumann, 'Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy', Law and Human Behaviour 1987-11, p. 167-187.