Criminalistics is reasoning backwards

Logically correct reasoning in forensic reports...
...and in the courtroom

Charles E.H. Berger[1]

Science plays an increasing role in criminal law, and both are rightly held up to higher standards. The awareness that scientists and lawyers will need to find each other more often to reach a higher level is also increasing. Logically correct reasoning and concluding are indispensable at that higher level, especially when uncertainty is involved. This paper describes how logically correct conclusions are given in forensic reports, and how the reader can deal with this.

Introduction

We are faced with the big challenge to reason in the most rational way in criminal cases. Science teaches us that probability theory is essential to that effort, while for most lawyers it's not their favorite subject. On top of that, communication between scientists and lawyers is not always as good as it should be. Even so, awareness that logically correct reasoning in the presence of uncertainty is important is growing internationally, among lawyers and forensic scientists. In the new "Handbook experts for the judge in criminal law" [1] for example, the importance of logic, probability theory and methodology has been recognized very clearly. It also correctly concludes that a judge can generally limit himself to these central elements, and does not need to have intimate knowledge of all the details of all separate forensic disciplines.

[1] Dr. Charles E.H. Berger is principal scientist at the Netherlands Forensic Institute, and member of the Board of the Dutch Forensic Experts Registry (NRGD).

The Netherlands Forensic Institute (NFI) is engaged in a transition to a new logically correct conclusion scale for its forensic reports. An investigation into how well readers understand logically correct conclusions without further explanation [2] shows that they are struggling. So should the NFI make that transition? The answer to this question is a resounding yes.

There is a false choice: a logically incorrect conclusion that's 'understood' is no alternative to a logically correct conclusion which needs explanation. Of course the scientist should do his best for his conclusions not only to be correct, but also written and explained as clearly as possible. Explanation is also offered in courses and lectures for the various actors in the criminal justice system.

For the treatment of logically correct conclusions it is good to first have a closer look at forensic science.

Forensic science is reasoning backwards from effect to cause

Forensic science is often simply defined as the application of science to questions of interest to the court. Criminalistics [3,4] is the part of forensic science where the exact sciences are applied. This definition is certainly practical, but says very little about forensic science from a scientific standpoint.

In the last century, a number of so-called principles of forensic science were proposed. One example is the so-called Locard's principle. This is usually stated as "Every contact leaves a trace", a phrase that was never found in Locard's oeuvre. Apart from the fact that this is not a principle Locard proposed [5], it can also hardly act as such. One could say with the same validity: "every contact wipes a trace".

In the absence of a number of classical principles, one might wonder what is the binding factor between the various disciplines in forensic science, apart from their application to questions of interest to the court. In my view, forensic science is the science of reasoning backwards from a known result to a cause, using the aforementioned three pillars: logic, probability theory, and methodology. This is also makes the common interest of forensic scientists and lawyers clear: how do we reason back from effect to cause, when we are dealing with uncertainty. Consequences exist for example in the form of traces at a crime scene or on an object, and possible causes are presented in the form of hypotheses about how these traces were created (for example about who left the trace). The process of reasoning backwards, i.e. the use of traces (in the most general sense) to find support for these hypotheses, is a common and binding theme for all forensic disciplines.

Reasoning is based on logic. If logic is to be applied to uncertain events, probability theory is required for the most rational outcome. It is even used as the definition of probability theory: the extension of logic to uncertain events [6]. Reasoning in the presence of uncertainty is something we do intuitively in everyday life, but that does not mean that our intuition is always right, or that we are capable of reasoning sufficiently explicit to convince someone else. To reason in a rational and logically correct way, we need to know some principles from probability theory, but that does not mean that a case can be reduced to a calculation. More importantly, logic tells us which conclusions are justified by the information we have, and which are not.

Three forms of reasoning

There are three main forms of reasoning that play an important role in criminalistics. Abduction [7] is the type of reasoning by which we creatively generate hypotheses, which, if true, best explain the initial observations.

Thus, when encountering a person killed in a house, and observing muddy footprints from the window to the victim, we can come to the hypothesis of an intruder that entered through the window. Abduction creates a starting point for two other forms of reasoning: deduction and induction.

Deduction is a type of reasoning that allows you to reach hard conclusions (categorical conclusions), which are necessarily correct. This is the kind of reasoning we use when we can exclude a hypothesis, such as when our observations are absolutely impossible if the hypothesis were correct. One example is a suspect with a watertight alibi: he could not have been at the crime scene during the crime. In most cases however, we can not exclude a number of hypotheses, and for those hypotheses the use of deduction is unfortunately not an option.

When we make observations that form more support for one hypothesis than for the other, but exclude neither of these hypotheses, we use *induction* for our reasoning. With this type of reasoning, our observations support a hypothesis, but they will never prove it with certainty. Adding more observations can increase our belief further, but certainty is never achieved. For example, by continuously observing white swans the belief that all swans are white will increase, but we will not achieve certainty until all swans have been observed (even those in Australia).

The evidential value of an observation

In order to avoid tunnel vision and to make explicit which hypotheses were or were not considered, at least a second, alternative hypothesis is required. The two competing hypotheses that are considered, are to exclude each other: they can not simultaneously be true.

Our knowledge (or degree of belief) concerning the hypotheses can be expressed in odds: the probability that hypothesis H1 is true (all swans are white), divided by the probability that the alternative hypothesis H2 is true (not all swans are white).

A relevant new observation will change the odds of the hypotheses. According to probability theory, the extent to which an observation changes the odds defines the evidential value of that observation. The odds of the hypotheses *prior to* the observation are called *prior odds*, the odds *after* the observation are called *posterior odds*. Bayes' theorem [4] shows that the evidential value is a multiplication factor, and defines it as the *likelihood ratio (LR):*

$$\text{Prior odds} \times \text{likelihood ratio} = \text{posterior odds.}$$

The three terms in this equation are defined as:

$$\frac{\text{probability that H1 is true before observation}}{\text{probability that H2 is true before observation}} \times \frac{\text{probability of observation when H1 is true}}{\text{probability of observation when H2 is true}} =$$
$$= \frac{\text{probability that H1 is true after observation}}{\text{probability that H2 is true after observation}}.$$

The evidential value of an observation is equal to the likelihood ratio: the probability of that observation if hypothesis H1 is true, divided by the probability of the same observation when the alternative hypothesis H2 is true. The theorem shows that the evidential value (LR) only gives the *relative increase* in the odds and not the posterior odds themselves.

The three examples in the boxes can clarify this. In the box "pot belly" you will find an anecdotal example of the likelihood ratio as the evidential value of an observation.

Pot belly (example 1)

In this example we see an acquaintance – that we rarely meet – across the street. The size of her belly suggests that she is pregnant, but we have not heard about her partner or pregnancy. Before we decide to congratulate her, the question arises: is she really pregnant or not? What is the evidential value of our observation of her belly? Probability theory tells us that the evidential value is equal to the ratio of two probabilities: the probability of a big belly during pregnancy, divided by the probability of a big belly without a pregnancy. The evidential value increases as more pregnant women have a big belly, and less non-pregnant women show that same feature. Thus, the evidential value of the observation of a big belly is much greater in a population with only thin women than in a population where a pot belly is the norm.

A numerical example from the medical world is given in the box "HIV test". The medical example shows that the diagnostic test for a disease by itself is not enough for an estimate on the probability that we have the disease.

HIV-test **(example 2)**

If we test someone for HIV, we consider the hypotheses 'has HIV' and 'has no HIV'. The so-called 'enzyme immunoassay' HIV test is known to give a positive result for 99.7% of all people with HIV. People without HIV have a probability of only 1.5% to obtain such results [8].

But what is the probability that a person who tested positive (bad news) actually has HIV? Intuitively, we might think that that probability is very large, and especially the 99.7% number appears to play a role. In reality, we need more information to conclude on the probability of HIV infection.

But we *can* report on the evidential value of the test result: the likelihood ratio is 99.7% / 1.5% = 66. In the medical world, this is known as the diagnostic value. It is 66 times more likely to test positive when you are HIV infected, than when you are not.

If we know the prior odds, we can now determine the posterior odds. The prior odds depend on other evidence and information. If we know for example that the tested person is an adult South African, then the prior odds are known from the prevalence (the relative proportion of the population that is infected) of HIV among adults in this country: 0.22 [9]. The posterior odds are therefore 0.22 × 66 = 14.6. For this person, the result means that the odds are 14.6 against 1 that he has HIV. For an adult Dutch person, the situation is quite different. In the Netherlands, the prior odds of HIV infection are around 0.002, and the posterior odds are 0.002 × 66 = 0.13. So for the Dutchman the same positive test means that the odds of infection are 1 to 7.6, much better!

For the probability that the tested person has HIV, the knowledge prior to the test (prevalence) is as important as the evidential value of the test.

The same principle applies in forensic science: the comparison (observation) of trace and reference material of a suspect is not sufficient to determine the probability that the traces were left behind by the suspect (hypothesis), except when you can exclude the suspect. To determine that probability, prior odds are required which include all other information (motive, alibi, etc.) and evidence (e.g. other types of traces).

In the “red scarf” box, a forensic example is given from which this is clear (even without using numbers): the examiner cannot report the probability that the fibers originated from the red scarf, but he can report the evidential value of his observations.

Red scarf (example 3)

In this example, red fibers were found on a murder victim. The police send these fibers and a red scarf to the NFI, asking whether the fibers originated from the scarf. If that cannot be determined with certainty, they want to know how likely it is that the fibers came from the scarf. This question can not be answered based on the similarities between the fibers of the victim and the scarf alone (at least not unless the scarf can easily be excluded as a source, such as when large differences are observed). Why not?

The reason is that the NFI examiner does not have all other information, and so does not know the prior odds, while those are essential for the probability that the fibers came from the scarf. This is perhaps easier to see using the example. We consider two different scenarios preceding the submission of the fibers and the scarf. In the first scenario, a suspect was arrested while running away from the victim, and he had the red scarf around his neck. In the second scenario, the police found the victim (a student) the next day, and also observed the red fibers. In the cloakroom of a large school in the neighborhood a red scarf was found and sent to the NFI.

It is clear that these scenarios will lead to a quite different probability of the fibers originating from the scarf. But the NFI examiner has no information on such scenarios. The examiner cannot report the probability that the fibers came from the red scarf and has to limit himself to the evidential value of his observation.

Forensic conclusions

There are different types of conclusions in forensic reports. For example, dactyloscopists (fingerprint experts) traditionally report categorical conclusions. This type of conclusion is scientifically problematic because we have seen that inductive reasoning cannot lead to a categorical conclusion (certainty). A positive conclusion explicitly excludes the possibility that a person exists whose fingerprint is as close to the trace as the fingerprint of the current suspect.

A much more common type of conclusion is one for which a probability is assigned to a hypothesis. For example, an assessment of the probability that a trace of the perpetrator and reference material from the suspect originate from the same source. Because such a conclusion is not a categorical statement this might seem quite reasonable, but we have seen that such a conclusion is not justified. The expert does not have all information in the case, which is necessary for an assessment of the prior odds. The judge is the person who does have all this information available to him, and he can use it in weighing the evidence.

Even if the examiner would have had all that information, it is not good to use it for the conclusion. It would mean he would tread outside his own field of expertise, and possibly even express opinion on the ultimate issue. Moreover, experts in various forensic fields could do so, and by combining the reports in the courtroom things would be counted multiple times (and given too much weight). Therefore, the examiner should limit himself in his conclusion to the evidential value of his observations.

The way that conclusions are reported in DNA examination is in line with what logic and probability theory teach us. The evidential value of a found match is given by the probability of the match when the suspect has left the trace, divided by the probability of that match when a random other person left the trace. The first probability is one (barring errors), while the second probability is equal to the frequency with which the profile appears in the population, and that can be calculated. The conclusions in DNA examination are thus logically correct and numerical. For many other areas of forensic expertise the evidential value can not be reported numerically, because of a lack of data. Still, logically correct conclusions can also be reported with a verbal scale, when numerical data are lacking.

Introducing the new verbal conclusion scale

Currently, the NFI is in a transition to a new verbal conclusion scale. The logically correct wording chosen to report the evidential value is [10]:

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.

Of course, this conclusion scale does not apply where only measurement results are reported, where a categorical response is justified (as in exclusion), or when the evidential value can be reported numerically. Where possible, it is desirable to report numerically. Therefore, a research and development program has been established at the NFI, to obtain more empirical data and objective comparison methods. This is a large but exciting challenge for criminalistics in the years ahead.

Other verbal conclusion scales are also possible, but might give rise to more interpretation errors. The evidential value can e.g. also be expressed in how much more support the observations give to one hypothesis relative to the alternative hypothesis. In practice, this is easily misread as: how much more likely one hypothesis is than the other. This kind of thinking errors (also called fallacies) is very common, so we will now look at the main variants of these fallacies.

Misunderstanding conclusions (fallacies)

The reader of a conclusion of a forensic report is looking for the answer to his question, and when no such answer is given, there is a tendency to read the given answer incorrectly as if it were the desired answer. The logical errors thus made are named after the party that supposedly tends to make that error.

The prosecutor's fallacy is an error of reasoning in which conditions are transposed (it is also called the transposed conditional). Thus the conclusion "the probability of a match is 1000 times greater when the accused left the trace than when a random other person left it" can be incorrectly read as "the probability that the accused left the trace is 1000 times greater than that a random other person left it, when there is a match". This is a fallacious transposition of conditions that is more noticeable when using simpler statements like "an animal has four legs when it is a cow" and "an animal is a cow when it has four legs". The desire to know the probability of the hypothesis leads one to read a logically correct conclusion (in terms of evidential value) as if it would directly give the probability of the hypothesis. But as we have seen before, this also requires the prior odds. By ignoring the prior odds these are implicitly assumed equal to one (50% / 50%), while in reality they can be much higher or lower.

In the defense fallacy the prior odds are implicitly assumed to be very small. The above conclusion can for example be interpreted incorrectly by stating: there are 16,000 people in the Netherlands that would match as well, which means the probability that the current suspect left this trace is only 1 in 16,000. Here it is implicitly (and almost always wrongly) assumed that a priori all 16 million Dutch people have the same probability of having left the trace.

This type of reasoning error exists in many different forms, some more subtle than others. Thus, even with strong evidence it can suggestively be stated that the accused at most is not excluded, while when considering the evidential value it is also crucial how many others have been excluded.

Concluding on conclusions

Logically correct reasoning and concluding is not an effort that can't fail when it is based on experience and intuition only. It is not easy, and logic and probability theory play an important role. Perhaps – from an evolutionary point of view – it is more important to predict the future than the past, and we are therefore not very good at reasoning backwards.

Forensic science is in a transition to conclusions that are logically correct, but not always easy. Sometimes you cannot make things easier without straying from the truth. It is clear that a logically incorrect conclusion that is well understood is no alternative to a logically correct conclusion that requires some explanation. I hope to have contributed with this article to the understanding of this vital matter, but for an improved understanding practice and repetition are important.

A good conclusion considers at least two competing hypotheses in order to avoid tunnel vision and make explicit which hypotheses are considered. If neither of these hypotheses can be completely excluded, the conclusion says how much more likely the results of the examination are under one than under the other hypothesis. This factor is the evidential value, and it tells us how much the odds of the hypotheses have increased because of the results. The final probability that the hypothesis is true can not be given by the expert, because that probability is also determined by information that he does not have, and which often falls outside his expertise. The combining of all the forensic evidence with the other information in the case is up to the judge, who has all information available to him.

So if you read such a logically correct conclusion, keep in mind that (in simple terms) the expert can usually not tell if it is so, nor how likely it is, but only what the evidential value of his observations is. Beware of the errors of reasoning. If lawyers and scientists communicate more and common errors of reasoning can be avoided more often, that can be a very important advance.

The NFI

The NFI provides forensic services, mostly to requesters in the criminal justice system, including the prosecution and the police. The NFI in 2008 provided about 53,500 products, especially in DNA examination [11]. Partly in response to the recommendations in the report of the committee Posthumus [12], the forensic reports have greatly improved: they are more uniform, more structured, and equipped with appendices that give more information on the examination methods used. The introduction of a uniform, logically correct conclusion scale is accompanied by many forms of explanation on this matter, including lectures and workshops, publications and an *e*-learning course.

The NFI is internationally at the forefront of this development. It is in contact with colleagues in countries that already make use of logically correct conclusion scales, such as the United Kingdom, Sweden and Switzerland. The United States still has a long way to go, partly due to the poor shape of much of the U.S. forensic system [13].

References

[1] Handboek deskundigen voor de strafrechter, Project Deskundigheids¬bevordering Programma Strafsector 2010, 2009.

[2] J.W. de Keijser, H. Elffers, R.M. Kok, M.J. Sjerps, Bijkans begrepen?, Den Haag: Boom Juridische uitgevers 2009. This study was commissioned by the NFI, and carried out by the NSCR.

[3] A.P.A. Broeders, Op zoek naar de bron: Over de grondslagen van de criminalistiek en de waardering van het forensisch bewijs, Deventer: Kluwer (juridisch) 2003.

[4] B. Robertson, G.A. Vignaux, Interpreting Evidence: Evaluating Forensic Science in the Courtroom, New York: Wiley 1995.

[5] E. Locard, Manuel de technique policière, Paris: Payot 1923.

[6] E.T. Jaynes, Probability Theory: The Logic of Science, Cambridge: Cambridge University Press 2003.

[7] C.S. Peirce, ‘How to make our ideas clear’, Popular Science Monthly 1878, 12, 286-3.

[8] R. Chou, L.H. Huffman, R. Fu, A.K. Smits, P.T. Korthuis, ‘Screening for HIV: a review of the evidence for the U.S. Preventive Services Task Force’, Ann. Intern. Med. 2005, 143, 55-73.

[9] www.cia.gov, the world factbook, data from 2007.

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

[11] Dutch Forensic Institute, Annual Report 2008.

[12] F. Posthumus, Evaluation in Schiedam parkmoord 2005.

[13] Strengthening Forensic Science in the United States: a path forward, National Academies Press 2009.