Measuring Uncertainty in Forensic Science

Joëlle Vuille and Franco Taroni

This paper aims at describing the ways in which uncertainty is (or is not) measured in the field of forensic science and how the measure of uncertainty can be used to make better judicial decisions. From the traditional fingerprint comparison to the latest advances in forensic DNA analysis, we discuss the advantages and drawbacks of various ways of reporting forensic science results. We point to past and current controversies, and analyze what remains to be done in the field to ensure that the probative value of forensic science evidence is communicated to judicial fact finders in a way that is scientifically robust, balanced and transparent, to allow them to make coherent decisions.

Forensic Science and Uncertainty

Forensic science practitioners use the natural sciences, such as biology, chemistry and physics, to help a judge or a court establish the facts in a judicial matter. This typically involves the examination and comparative analysis of so-called evidential material, followed by an evaluation of the value of the evidence within the particular context of the event under investigation. For example, in the criminal context, forensic science practitioners analyze traces (such as DNA, fingermarks, controlled substances, cell tower data, etc.) to help determine whether somebody committed an offense, who the offender was, and how the offense was committed. In the civil and administrative contexts, forensic experts are often called upon to help a court decide whether a document was altered or falsified (such as a contract, a passport, or a driver's license), whether a person is a minor or an adult (in the immigration context) or to carry out paternity testing. All of these tasks have one important common feature: they require one to reconstruct a past event on the basis of current observations (on recovered traces) that are incomplete and imperfect. Inherent to the forensic process, therefore, is the measurement of uncertainty, which requires the use of probabilities, and the implementation of a model to make inferences on hypotheses of interest and guide the decision-making process.

Uncertainty is an uncomfortable position. But certainty is an absurd one – Voltaire

(Not) Measuring Uncertainty in the Traditional Forensic Sciences

Fingerprint comparison is the first field of modern forensic science that was developed in law enforcement in the late 19th century. When early fingerprint experts started assisting criminal prosecutions, they took on the habit of expressing the results of their analyses as certainties. After comparing the patterns found on a crime scene fingermark and a reference print taken from a suspect, they would testify either that the suspect was excluded as the source of the trace, or that the two marks were identical, and that the suspect could thus be identified with certainty as the source of the print found at the crime scene. This approach rested on the paradigm of "discernible uniqueness" that postulated that each fingerprint was unique, that each crime scene trace reproduced that uniqueness faithfully, and that experts could recognize matching fingerprints and traces and distinguish non-matching ones.

Most of these assumptions have now been proven wrong, or at least greatly nuanced, and reporting results in this form is problematic for multiple reasons. First, no two objects are ever identical; crime scene traces and reference marks especially are always different, because the crime scene trace is usually incomplete and distorted. Second, saying that two objects are identical, hence they must come from the same source, is an unreasonable inference. In fact, multiple objects in the universe can be indistinguishable to the human eye (or analytical process) and still be different. It is thus of

The authors would like to acknowledge the Swiss National Science Foundation (grants PP00P1_176720 and IZSEZ0_191147) and the Fondation pour l'Université de Lausanne (Fonds Echanges Out_FT) for their financial support.

1094-6969/21/$25.00©2021IEEE

utmost importance to have knowledge as to the rarity of the observed characteristics of the object in the population of interest and the accuracy and observational sensitivity of the scientific method. Third, the expert who states that a given person left a trace opines as to the veracity of a hypothesis, which, from an inferential point of view, is unwarranted, as we shall see. Finally, when an expert testifies that a suspect is the source of a trace, the uncertainty inherent to the comparison process is occulted, so that the judge or court handling the case has no idea that the expert made that leap of faith from her observations to the reported results. Unfortunately, almost all forensic disciplines that were developed during the 20th century imitated that "fingerprinting model" of evaluating evidence and reporting results to the fact finder without acknowledging uncertainty.

The DNA Revolution in Forensic Science

In the mid-1980s, things started to change, however, with the advent of forensic DNA analysis. Because forensic DNA analysis was derived from mainstream biological practice (contrary to fingerprint comparisons that were developed in a law enforcement context) and hence influenced by a scientific culture, forensic practitioners in the field of DNA were from the onset in the habit of assessing the value of evidence through statistics. Like fingerprint experts, DNA experts would isolate the characteristics found in the crime scene trace and a reference sample taken from a suspect and compare them. If they declared them to match, they would then use population genetics data to assign the probability that an unknown person in a given relevant population would match, although they were not the source of the trace. To do so, they would (assuming independence between genetic markers) multiply the occurrences of each genetic characteristic observed in the DNA profile in the considered population, to quantify the so-called random match probability for the whole profile (i.e., often values smaller than 1 in several billion). Then, they would report that probability assignment to the fact finder, who would have to decide whether the probative value of the evidence was sufficient to consider that the suspect was indeed the source of the crime scene trace.

The proper use of probabilities to express the probative value of DNA evidence sparked intense debates in the forensic and judicial communities in the 1990s, so much so that this period in forensic history became known as the "DNA wars." In the meantime, the majority of forensic practitioners (i.e., outside DNA analysis) continued to report their results in the old, deterministic way. Today, there are therefore two models of forensic reporting that co-exist: the tenants of the "old school" who are attached to rendering so-called certainty conclusions (and who have recently faced appeals by the larger forensic community to review the foundations of their disciplines and to adopt a probabilistic framework in assessing the probative value of the evidence); and the proponents of a more progressive model who measure uncertainty explicitly through probabilities. For more information on the general definitional meaning of uncertainty in metrology, see [1], [2].

Assessing the Probative Value of DNA Evidence Through Relative Frequencies

As stated above, for a long time, the value of an observed DNA correspondence was expressed through random match probabilities. Using the relative frequency of a series of genetic characteristics to assign a value for a DNA correspondence is problematic, though, for several reasons. First, the genetic data (i.e., the relative frequency) used as the basis to assign a probability value is derived from population databases of approximately 200 to 300 people; whether such a group has sufficient external validity to stand in the population of large countries in terms of genetic makeup is debatable. Second, relative frequencies lead fact finders to view very small probabilities as factual impossibilities. Yet, just as there are people who win the lottery, there are innocent suspects who match a given DNA profile of which they are not the source. Third, if the multiplication of relative frequencies can be considered correct from a theoretical point of view (given a number of questionable assumptions), there is no data to support its empirical accuracy; in short, it is a matter of faith. In the field of DNA, it is not uncommon, for example, to see forensic practitioners report on random match probabilities in the following manner (from Italian):

> *The identity (...) was established for 21 autosomal markers (…) with a statistical frequency of* 2.33×10^{27}*, which equals certainty. If you consider that there are 7 billion people on Earth, you would need to consider 130 million billions of planets similar to the Earth to find another individual [than the accused] who has the same genetic characteristics* [3]. (Note that the text should read "$2.33\text{x}10^{-27}$".)

But such numbers only exist in the minds of expert witnesses; they have no factual reality, and they present the risk of misleading judicial fact finders. This is why some scholars have recommended that the probative value of the evidence in such cases be capped at 10^{-9}, because smaller values *invoke independence assumptions to a scale of robustness that we cannot demonstrate empirically, given the size of available databases* [4].

The Bayesian Approach

To correct for the shortcoming of relative frequencies to assess the value of items of evidence and their influence on a verdict, scholars in the legal and statistical fields started discussing the formal analysis of rational thinking under uncertainty in the judicial context in the mid-1960s. Some authors proposed to rely on probability and decision theory to approach probabilistic inference and decision-making in a formalized, logical and defensible way.

Bayes' theorem allows one (e.g., a judge) to update initial beliefs about hypotheses of interest (i.e., prior beliefs) in light of newly acquired data, to reach so-called posterior beliefs about said hypotheses of interest. In mathematical form, this can be expressed as follows:

$$\frac{P\left(H_p \mid E,B\right)}{P\left(H_a \mid E,B\right)} = \frac{P\left(E \mid H_p,B\right)}{P(E \mid H_d,B)} \times \frac{P(H_p \mid B)}{P(H_d \mid B)}$$

where:

- P represents probability
- H_p represents the prosecutor's hypothesis
- H_d represents the defense hypothesis
- E refers to the evidence (i.e., observing a correspondence between the profile of the crime scene trace and the reference profile of a suspect)
- the vertical bar $\mid$ characterizes the conditional probability one is interested in (the probability of an event given knowledge of the occurrence of other events), for example the probability to observe an item of evidence, E, given that the prosecutor's hypothesis is considered true, H_p, and given some background knowledge about the case, B.

This equation represents the odds form of Bayes' theorem. In this framework, DNA results are rendered in the form of a Bayes factor, $P\left(E \mid H_p,B\right) \Big/ P\left(E \mid H_d,B\right)$, that expresses the probability of observing the evidence if the suspect, rather than somebody else in the suspect population, is the source of the trace. Bayes factor can thus be conceptualized as a metric that provides a balanced measure of the degree to which particular evidence, independent of its nature, is capable of discriminating among competing hypotheses put forward by opposing parties at trial. Such reasoning is considered normative in the sense that it prescribes a standard that, if followed, allows reasoners to avoid logical fallacies.

Consider, for example, a Bayes factor of 1 million. This means that the expert's observations are 1 million times more likely if the suspect is the source of the trace than if somebody else in the population of interest is the source. The Bayes factor has a value comprised between 0 and infinity, with 1 being the neutral value that expresses the fact that the evidence does not support either of the hypotheses, H_p or H_d, over the other. The evidence can be said to be irrelevant because it does not allow one to discriminate between the hypotheses. The Bayes factor can be combined with the other elements in the case—quantified through the so-called prior odds, $P\left(H_p \mid B\right) \Big/ P\left(H_a \mid B\right)$ —to assess the probability that the hypotheses of interest are true, given all the evidence considered; this is expressed by the posterior odds, $P\left(H_p \mid E,B\right) \Big/ P\left(H_d \mid E,B\right)$.

The advantage of the Bayesian approach is that it is transparent, in that it informs the decision maker of the amount of uncertainty inherent in the measurement of the probative value of the evidence considered; moreover, it allows one to coherently update one's uncertainty on the hypotheses of interest. It also makes it clear what the respective roles of the fact finder (opining as to hypotheses) and the expert (assessing the evidence under competing hypotheses) are. One limitation, however, is that the practical application of this approach is still widely neglected in real casework due to a lack of understanding as to its merits. And one of the main criticisms towards the Bayesian approach is that it is too difficult for lay people such as judges and jurors to understand (which, we would argue, is not a sufficient reason for not using it).

Unknown Error Rates

The Bayesian approach theoretically allows one a scientifically coherent assessment of the evidence and of the case as a whole. But in practice, it still falls short of the mandate. The reason is that there is still very little data related to error rates in the forensic field. For the most part of the 20th century, forensic practitioners managed to convince judges and courts that (contrary to all human beings) they never made any mistakes. In the 1990s, this position became untenable, though: the Innocence Project, initiated by two law professors in New York, Barry Scheck and Peter Neufeld, started reopening cases of people who claimed to have been wrongfully convicted and reinvestigating them with the help of DNA evidence. It quickly became clear that forensic science evidence was a major cause of wrongful convictions in the United States: the criminal justice system was regularly misled because forensic practitioners mislabeled or exchanged specimens, contaminated crime scene or reference samples, made computational errors, used a faulty paradigm, and repeatedly claimed that their results were certain instead of applying a model that dealt explicitly with uncertainty [5].

In the mid-2000s, the reputation of forensic science had been sufficiently damaged by the number of so-called DNA exonerations that the U.S. National Research Council (NRC) decided to conduct a systemic evaluation of the field. The report, published in 2009, revealed that many forensic disciplines were insufficiently validated and produced results that were not scientifically robust [6]. One infamous example was forensic odontology, specifically the comparison of bite marks. For a long time, self-proclaimed bite mark experts would match bite marks found on victims of assault, rape or murder with the dentition of a suspect. Without any data related to the validity and accuracy of their work, they would declare that the suspect had in fact bitten the victim to "a reasonable degree of scientific certainty" or "indeed and without doubt." As it later turned out, many bite mark experts were not able to distinguish a bite from a bruise, or a human bite from an animal bite, let alone identify the individual who was the source of a given bite mark [7].

The publication of the 2009 NRC report had several consequences for the forensic field. Some disciplines, such as bite mark comparisons, were abandoned in certain jurisdictions and are not used in criminal prosecutions anymore. Other forensic fields have heeded the warnings of the NRC and begun large-scale validation studies. Researchers in fingerprint comparisons, for example, have endeavored to develop false positive and false negative error rates, mainly through black box studies [8]. In forensic DNA analysis, some laboratories have started conducting their own studies into errors, such as contamination of crime scene samples [9]. There remain some problems, though [10]. First, efforts to assess error rates are fragmentary and depend entirely on the goodwill of individuals and laboratory managers. Second, the results of such studies are only published if it is convenient to do so, and some incidents suggest that unsatisfactory results are simply hidden away. Third, expert reports do not usually contain an assessment of the probability of a false positive result, i.e., the

probability to declare a correspondence between two profiles when they do not correspond. The Bayes factor is calculated as if any possibility of error in the handling of the evidence has been conclusively excluded; it only expresses one type of "error," which is the possibility that one innocent person will match the trace by chance (i.e., the random match probability). This makes no sense from a scientific point of view, because it is clear that if the possibility of an error were to be taken into account, the value of the Bayes factor would very often decrease by orders of magnitude [11].

From Source to Activity

One additional problem that has emerged recently is due, paradoxically, to technological advances: the limits of detection and the analytical capabilities have made it possible to derive a DNA profile from a single cell. But what does it mean to find a single cell somewhere, with regards to somebody's presence or activity? Focusing one's attention to the analysis stage of the life of the evidence presents the risk of blinding the court to the real meaning of the evidence in the context of the case. This was theorized under the concept of *hierarchy of propositions*, which distinguishes the evaluation of evidence under hypotheses (or propositions) referring (a) to the source of the recovered trace, (b) to the activity committed by the suspect or by other people involved and (c) to the offense under scrutiny [12].

In the last two decades, instances of wrongful convictions have demonstrated in a dramatic manner that a person cannot be convicted of a crime simply because a trace seemingly links her to the offense. First, contaminations can happen. In Australia, for example, Farah Jama was exonerated in 2009 after spending several years in prison for a rape he did not commit. His DNA had been observed in a vaginal sample taken from a woman who had been found passed out in the lavatories of a nightclub. It later turned out that the rape kit had been contaminated with Mr. Jama's genetic material [13]. Second, traces can be transferred from one person or one object to another through one or several vectors. Many studies have now documented the possibility of indirect transfer, which renders the interpretation of evidence very difficult in certain cases [14]. Aware of the importance of assessing evidence at the activity level, the European Network of Forensic Science Institutes published guidelines in 2015 to encourage forensic practitioners to render conclusions under activity-level hypotheses when there is a risk, in the case at hand, that rendering conclusions under source hypotheses could lead the fact finder to give the evidence undue value [15].

Decision Theory in Forensic Science

As previously said, on the basis of observations, one can construct probabilistic arguments in support of hypotheses of interest. As such, probabilistic models provide a coherent environment wherein probabilities about target propositions can be revised upon receipt of newly acquired information. This represents a fundamental requirement for an additional step, that is the coherent measure of probabilities for decision-making under uncertainty.

One makes decisions on the basis of essentially two elements: one's beliefs about uncertain events (the hypotheses of interest, H_p and H_d) and one's assignment of desirability of decision consequences. The latter aspect is formalized by invoking the concept of *utility*. Both concepts, probability and utility, can operate within a general decision framework that involves the practical rule that one should select the decision which has the highest expected utility (or, alternatively, the lowest expected loss) [16].

To formalize a decision problem, decision theory first requires one to draw up an exclusive and exhaustive list of available decisions (from a legal point of view, acquitting or convicting a given suspect, for instance). Secondly, a list of exclusive and exhaustive uncertain events (i.e., the hypotheses) is needed. The decision maker is then faced with a problem of decision-making under uncertainty: because it cannot be known with certainty which state of nature holds, one cannot directly tell which decision leads to a consequence that is, in some sense, optimal. The main problem is thus to choose a decision without knowing which state of nature holds. The combination of a particular decision with a given state of nature will result in a foreseeable consequence: e.g., 1) acquitting an innocent; 2) acquitting a guilty suspect; 3) convicting an innocent, and 4) convicting a guilty suspect. These consequences have each a distinct associated utility. Further development of this argument leads to the so-called expected utility, which assigns a numerical value to a given decision, as a function of the probabilities of outcomes as well as their respective desirability. The result thus consists in a qualifier for the appropriateness of particular decisions, evaluated on the basis of the desirability of outcomes.

The final goal that the justice system pursues is that of making decisions. Not making a decision is not an option; it is even forbidden in most modern justice systems, as a denial of justice. Decision theory aims to define what the best decision is for a rational agent, given the information that she has and her personal values for the possible consequences of her decisions [17]. This is why developing a framework for decision-making in the forensic process is so important.

Conclusion

Judges, attorneys and forensic scientists only have incomplete knowledge as to the states of nature that make up the cases before them. Uncertainty is thus a complication that accompanies all actors of the justice system who face inference and decision-making as core aspects of their activities. Probability, as a measure of uncertainty, can help them make the best decisions on the basis of the knowledge at hand and the preferred values of the stakeholders. The probabilistic model, in turn, can help focus resources into appropriate data gathering strategies and clarify the structure and mechanisms of scientific evidence assessment. Finally, as a tool that informs the rational management of the risk of errors, the use of a probabilistic model increases the accuracy of judicial fact-finding and thus helps materialize the right to a fair trial.

References

[1] *International vocabulary of metrology –basic and general concepts and associated terms*, International Bureau of Weights and Measures, Joint Committee for Guides in Metrology, BIPM, 2012.

[2] *Evaluation of measurement data—guide to the expression of uncertainty in measurement,* International Bureau of Weights and Measures, Joint Committee for Guides in Metrology, BIPM, 2008.

[3] Sentenza del 1/7/2016, Corte d'Assise di Bergamo, N. 1/15 R.g. C.A., N. 7701/14 R.G. N.R., p. 96 (translated from Italian).

[4] A. J. Hopwood, R. Puch-Solis, V. C. Tucker, J. M. Curran, J. Skerrett, S. Pope, and G. Tully, " Consideration of the probative value of single donor 15-plex STR profiles in UK populations and its presentation in UK courts," *Science & Justice*, vol. 52, pp. 185–190, 2012.

[5] B. L. Garrett, and P. J. Neufeld, "Invalid forensic science testimony and wrongful convictions," *Virginia Law Review*, vol. 95, no. 1, pp. 1-97, 2009.

[6] "Strengthening Forensic Science in the United States: A Path Forward," National Research Council, Washington, DC, USA, 2009.

[7] C. M. Bowers, "Review of a forensic pseudoscience: identification of criminals from bitemark patterns," *J. Forensic and Legal Med.*, vol. 61, pp. 34-39, 2019.

[8] B. T. Ulery, R. A. Hicklin, J. Buscaglia, and M. A. Roberts, "Accuracy and reliability of forensic latent fingerprint decisions," *Proc. National Academy of Sci.*, vol. 108, no. 19, pp. 7733-7738, 2011.

[9] P. Basset and V. Castella, "Lessons from a study of DNA contaminations from police services and forensic laboratories in Switzerland," *Forensic Sci. Int.: Genetics*, vol. 33, pp. 147-154, 2018.

[10] J. J. Koehler, "Forensics or fauxrensics? ascertaining accuracy in the forensic sciences," *Arizona State Law J.*, vol. 49, pp. 1369-1416, 2017.

[11] W. C. Thompson, F. Taroni, and C. G. Aitken, "How the probability of a false positive affects the value of DNA evidence," *J. Forensic Sci.*, vol. 48, no. 1, pp. 47-54, 2003.

[12] R. Cook, I. W. Evett, G. Jackson, P. J. Jones, and J. A. Lambert, "A hierarchy of propositions: deciding which level to address in casework," *Sci. Justice*, vol. 38, no. 4, pp. 231-239, 1998.

[13] F. H. R. Vincent, "Inquiry into the circumstances that led to the conviction of Mr. Farah Abdulkadir Jama," Victoria Department of Justice, Victorian Government Printer, Melbourne, Australia, 2010.

[14] R. Van Oorschot, B. Szkuta, G. Meakin, B. Kokshoorn, and M. Goray, "DNA transfer in forensic science: a review," *Forensic Sci. Int.: Genetics*, vol. 38, pp. 140-166, 2019.

[15] "Strengthening the Evaluation of Forensic Results across Europe," ENFSI Guideline for Evaluative Reporting in Forensic Science, 2015, p. 11, (accessed Apr. 2020). [Online]. Available: http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf.

[16] D. V. Lindley, *Making Decisions*, 2nd Ed. London, UK: John Wiley and Sons, 1985.

[17] D. H. Kaye, "What is Bayesianism?" in *Probability and Inference in the Law of Evidence, The Uses and Limits of Bayesianism*, P. Tillers and E. Green, Eds., vol. 109, pp. 1-19. Dordrecht, The Netherlands: Springer, 1988.

Joëlle Vuille, Ph.D., (joelle.vuille@unifr.ch) is Professor of Law in the Department of Criminal Law at the University of Fribourg, Switzerland. She has conducted several research projects and published articles on the use of forensic science evidence in the criminal justice system and is a deputy editor for the *International Journal of Evidence and Proof* (SAGE Publications).

Franco Taroni, Ph.D, is Professor of forensic statistics at the University of Lausanne, School of Criminal Justice, Switzerland. His main areas of research and teaching are forensic statistics, Bayesian networks, and Bayesian decision-theoretic data analyses. He has authored and coauthored several books on these topics and is also an editor for *Law, Probability and Risk* (Oxford University Press).

For Further Reading

C. Aitken and F. Taroni, "Fundamentals of statistical evidence - a primer for legal professionals," Int. J. Evidence and Proof, vol. 12, pp. 181-207, 2008.

B. de Finetti, La Logica Dell'incerto (in Italian). Milano, Italy: Il Saggiatore, 1989.

M. O. Finkelstein and W. B. Fairley, "A Bayesian approach to identification evidence," Harvard Law Review, vol. 83, no. 3, pp. 489-511, 1970.

B. Robertson, G. A. Vignaux, and C. E. Berger, Interpreting Evidence - Evaluating Forensic Science in the Courtroom, 2nd Ed. Chichester, UK: John Wiley and Sons, 2016.

F. Taroni, A. Biedermann, S. Bozza, P. Garbolino, and C. Aitken, Bayesian Networks for Probabilistic Inference and Decision Analysis in Forensic Science. Chichester, UK: John Wiley and Sons, 2014.