Quaestio facti. Revista Internacional sobre Razonamiento Probatorio
Quaestio facti. International Journal on Evidential Legal Reasoning
Sección: Science for Legal Proceedings
N. 3 | 2022 pp. 481-515
Madrid, 2021
DOI: 10.33115/udg_bib/qf.i3.22599

ISSN: 2604-6202
Recibido: 06/03/2021 | Aceptado: 02/07/2021 | Publicado: 03/11/2021

(UN-)INTERPRETABILITY IN EXPERT EVIDENCE: AN INQUIRY INTO THE FRONTIERS OF EVIDENTIAL ASSESSMENT

Alex Biedermann[*]
Kyriakos N. Kotsoglou[**]

ABSTRACT: Evidence law regimes across several contemporary legal orders provide a host of doctrinal devices designed to probe various sorts and sources of information, especially with respect to their accuracy and reliability. These legal provisions, however, are vulnerable to or even tolerant towards uninterpretable evidence, that is evidence whose probative value cannot be assessed. This article critically examines and discusses the «littering paths» of uninterpretable evidence in legal proceedings. We point out the misinformative character of such evidence as well as the potentially misleading impact on criminal adjudication when using scientific evidence for purposes that, strictly speaking, the evidence cannot help with. We identify common causes and possible remedies and argue that to track uninterpretability, it is necessary to understand what exactly it means to say that a piece of evidence is probative in the procedural space of reasons, and how to draw the line between informative and non-informative items of evidence.

KEYWORDS: scientific evidence, interpretability, probative value, propositions, forensic science

[*] University of Lausanne, Faculty of Law, Criminal Justice and Public Administration, School of Criminal Justice, 1015 Lausanne-Dorigny (Switzerland). The authors gratefully acknowledge the support of the Swiss National Science Foundation through grant BSSGI0_155809, the Fondation pour l'Université de Lausanne and the Société Académique Vaudoise. – alex.biedermann@unil.ch

[**] Northumbria University, Faculty of Law & Business, Newcastle Upon Tyne, NE1 8ST (UK). – kyriakos.kotsoglou@northumbria.ac.uk

RECOMMENDED CITATION: ALEX BIEDERMANN; KYRIAKOS N. KOTSOGLOU, 2021: «(Un-)interpretability in expert evidence: an inquiry into the frontiers of evidential assessment», in *Quaestio facti*, 3: 481-515. Madrid: Marcial Pons Ediciones Jurídicas y Sociales. DOI: http://dx.doi.org/10.33115/udg_bib/qf.i3.22599

1. INTRODUCTION

1.1. Seemingly probative scientific evidence?

It is not an exaggeration to say that the field of evidence & proof has primarily focused—at least for the past few decades—on the *reliability* of various forensic science[1] disciplines. From the ground-breaking *Frye*[2] decision which set the initial standards for scientific validity in the 1920s, widely considered even beyond the American context, over the «DNA wars» of the 1990s[3] to the ongoing discussions over how to define criteria of scientific validity,[4] scholars, criminal courts and scientific organisations have tried to lay down the procedural and methodological architecture for the operation of forensic practitioners in the criminal justice system.

These developments raise the question of what else could and, indeed, should have been discussed, *in addition* to reliability. As we will argue below, the main and usually sole requirement for admissibility in criminal proceedings is a core feature of forensic scientists' operations that, so far, is not adequately illuminated. Whereas reliability—based on «validity» in the US[5] and «sufficiently reliable scientific basis»

[1] For the purpose of this paper, forensic science is broadly understood as «the application of scientific or technical practices to the recognition, collection, analysis and interpretation of evidence for criminal and civil law or regulatory issues». See President's Council of Advisors on Science and Technology (PCAST), 2016: 1, hereinafter cited as PCAST Report, and PCAST as council.

[2] *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).

[3] See Aronson, 2007.

[4] See the PCAST Report for an example of a report that stirred considerable debate in recent years.

[5] *Daubert v. Merrell Dow Pharm., Inc*. (509 U.S. 579 [1993]). The *Daubert* standard is the test currently used on the federal level and in several jurisdictions on the state level. In the federal court

in England and Wales[6]—invokes the exception to the opinion rule which raises a general ban on opinions from any other person than the triers of fact, relevance is the main requirement for *admissibility*, save for any exclusionary rules.

Relevance too seems to be a straightforward topic insofar as it describes a relationship between two facts. A fact *x* is relevant to a fact *y* when the former has some material bearing on the latter. For example, to ascribe criminal liability to or acquit a defendant, we need evidence which is capable of helping to prove or disprove some element of the respective criminal allegation. This is obviously easier said than done. Admitting relevant evidence at trial and excluding irrelevant evidence is similar to the fictional rugby manager's instruction «To win, you need to score more tries than the other side». An expected response to that can only be: «Sure, but *how*?»[7]

In common law, James Fitzjames Stephen's classical definition seems to offer a helpful approach to criminal courts: «The word "relevant" means that any two facts to which it is applied are so related to each other that according to the common course of events on either taken by itself or in connection with other facts proves or renders probable the past, present or future existence or non-existence of the other».[8] Explicit reference to the term «probability» in the above definition has been replicated in modern legislature, most notably in FRE 401 (US) according to which «Evidence is relevant if (a) it has any tendency to make a fact more or less probable that it would be without the evidence; and (b) the fact is of consequence in determining the action». Similarly, in England and Wales Lord Simon of Glaisdale noted that «Evidence is relevant if it is logically probative or disprobative of some matter which requires proof […] It is sufficient to say […] that relevant (i.e. logically probative or disprobative) evidence is evidence which makes the matter which requires proof more or less probable».[9]

But before the critical reader is led to think that judges and jury in criminal trials deploy any sort of probability calculus to determine relevance and weight,[10] it is worth reminding one of the core elements in Stephen's definition. What is necessary, according to the abovementioned determination, is a vast array of background generalisations about the world, contextual information and, often, cultural understanding in order to place the evidence into context. The law predicates admissibility on relevance, but as Thayer put it, «the law furnishes no test for relevance».[11] It is

---

system, it replaced the *Frye* standard, which, nota bene, is still used in some states.

[6] *R v Dlugosz* [2013] EWCA Crim 2.

[7] See Williams, 1995: 239.

[8] Stephen, 1948: art. 1.

[9] *DPP v Kilbourne* [1973] A.C. 729 at 756 HL.

[10] See *R v Adams* [1996] 2 Cr App R 467, 481, 482 (England and Wales), where the Court of Appeal remarked that to «introduce Bayes' Theorem, or any similar method, into a criminal trial plunges the jury into appropriate and unnecessary realms of theory and complexity deflecting them from their proper task».

[11] Thayer, 1898: 265.

important to note here that background knowledge is neither cognitively present nor can it easily be articulated.

We need to distinguish the laissez-faire approach to, let us call it, normal evidence, as outlined above, from expert opinion evidence. The latter is admissible in England and Wales if, inter alia, it is «relevant to a matter in issue in the proceedings».[12] Experts, however, need to do much more than simply claiming that fact *x is* relevant to fact *y*. They need to explain *why* that is the case in their expert opinion. This insight helps us fully grasp the structural difference between relevance for non-expert evidence and relevance for expert evidence: in addition to the *qualitative* assessment which is expressed in a binary mode (relevant-irrelevant), experts need to lay out their underlying assumptions and invoke/present the respective scientific model at play, including the inferential steps which allow them to reach a given conclusion, i.e. the *quantitative* measure of how strongly or weakly (if at all) the evidence favours one proposition over the other. An expert, whose duty in England and Wales is «to help the court to achieve its aims»,[13] thus needs to explain his or her «methods», including the «inferential steps», and the extent and quality of the data on which the expert's opinion is based.[14]

Although this soft-law instrument provides devices with which court judges can decide on the reliability and validity of the proffered method, it is important to note that factual relevance is a function of reliability, not vice versa. A reliable method has a high degree of specificity and sensitivity. In other words, reliability is a precondition for relevance. A method whose results are only slightly better than chance barely deserves the name scientific. In fact, such a pseudo-method would be reminiscent of the diviner's claim that when he holds the rod, a certain sensation in his hand indicates the existence of water five feet under the ground. This fails to establish any relationship between the «feeling in the hand» (method) and the empirical fact of water being present at certain depth under the ground.[15] Similarly, and in a more realistic context, any polygraphic output cannot indicate the truthfulness or falsehood of a statement.[16] Both methods are blatantly unreliable.

Where these requirements cannot be met, we argue here, expert evidence should not be adduced in the criminal process. However, forensic science practice is far from these theoretical understandings. In many instances, forensic science evidence is used to support various strains of argument despite being *uninterpretable* or, at best, only partially interpretable. Such unwarranted uses of evidence, we cannot stress this enough, are a cause of concern and call for an inquiry into the understanding of what

[12] CPD V Evidence 19.A.1
[13] CPD V Evidence 19B.1.
[14] CPD V Evidence 19.A.5(a-c).
[15] See Wittgenstein, 1958: 1-15.
[16] See Kotsoglou, 2021.

exactly it means to say that a piece of expert evidence is informative and, hence, how to probe for evidence interpretability, i.e. the feasibility of assigning probative value.

To illustrate the problem of interpretability, consider a type of forensic science evidence that can be presented in a way that should arouse the interest of prosecutors: no matter what the analyst observes, the respective finding can be said to be in agreement with the prosecution's case. Such evidence may seem too good to be true, yet it actually exists. A well-known example is the discipline of firearms discharge residue (FDR, also called gunshot residue [GSR]) analysis. Suppose a case in which the analyst reports the finding of particles classified as GSR on the hands of a suspected shooter. Suppose also that those particles have the same qualitative composition and morphological features as those found in the head hair of the victim and in spent cartridges recovered from the crime scene. While there are many subtleties and circumstantial aspects to be taken into account in shooting cases, for the illustrative purposes here it shall suffice to mention that, at least in principle, forensic literature considers the presence of such particles «a positive result».[17] This is so, as is widely thought,[18] because «GSR is not common to the average's person daily environment».[19] The exact meaning of the proposition that a finding is «positive» pertains to a discussion in its own right, ranging from broad views such as understanding «positive» merely in a descriptive sense as the presence of target particles (or as a factual statement thereof), to stronger views according to which «positive» means incriminating, i.e. having probative force in relation to someone's alleged activities, or even criminal liability.

But suppose now that a search has been conducted and that *no* particles were found on the hands of a person of interest. Interestingly, such a finding can also be presented in a way that suits the prosecution's case. It suffices to invoke references to specialised literature in the field. For example, scientists have asserted that «[t]he absence of gunshot residue on a person's hands does not eliminate that individual from having discharged a firearm»[20] and that «... negative results [...] are not necessarily exculpatory».[21]

Thus, both the presence and absence of particles seems to be a useful type of evidence for the prosecution. Yet, this seemingly catchall probative property comes at a high price: the same findings can also be «fitted» to the narrative of the defence. On the one hand, the finding of no particles is what we typically expect to see on a person that is *not* related in any way to the discharge of a firearm.[22] On the other

---

[17] Heard, 2008: 255.

[18] We call this a mere thought because, as argued later in §2.2., one cannot directly jump from observations to conclusions. In particular, the mere rarity of a given type of evidence among a certain group of persons by itself does not make an item of evidence probative.

[19] Heard, 2013: 274.

[20] Trimpe, 2011: 29.

[21] Schwoeble & Exline, 2000: 127.

[22] See also *Pole v. Randolph* 570 F.3d 922 (7th Cir. 2009) for an example of a case in which the appellant argued that the retention of negative GSR findings (i.e. no detected GSR particles) implied

hand, any finding of particles raises the question of contamination and, hence, the problem of false positives.[23] In view of these diverging assertions, laypersons may wonder what, if anything, one is logically entitled to conclude without giving undue advantage to one of the competing parties to the matter. Expert witness testimony and forensic science in general thus has to be antithetical to the approach of Pythia, the Oracle of Delphi. For her reply to those going to war («You will go, you will return not in the war shall you die»[24]) could conveniently validate mutually exclusive outcomes. She allowed those asking to hear what they wanted to hear.

We thus see that even a slight scratch on the surface features of a given type of forensic science evidence readily prompts a host of questions. Namely, the questions of what it means to interpret an item of scientific evidence in the first place, what an admissible interpretation of evidence is and, more fundamentally, what it means to say that an item of evidence is interpretable. Moreover, inquiring about whether an item of evidence is interpretable, and to what extent, cannot be separated from the questions of the frontiers of interpretability and how those frontiers ought to be drawn. In this paper, we aim at investigating these questions in the particular context of scientific evidence.

1.2. Aims and scope

The main thesis defended in this paper is that settling questions regarding the nature of (probative) value—a necessary preliminary to the assessment of relevance—and the feasibility of ascribing evidential value, hinges upon our ability to cope with a notion that we will call, for present purposes, *(un-)interpretability.* As a device for drawing the frontiers of the possibility to assign probative value, the notion of interpretability is not understood here as an abstract property of evidence that could be contemplated in the absence of a human reasoner, equipped with knowledge and background information regarding the circumstances of the instant case. Instead, interpretability is understood as a function of a reasoner's assumptions, knowledge and understanding about selected target systems, and the intended purpose for which evidence is being adduced. In a nutshell, we will call evidence *uninterpretable* when

---

that vital evidence was kept from the defence («The exculpatory evidence [...] consisted of two favorable gunshot residue tests. [Defendant] contended that the swabs of his hands and tests of the glove were negative for gunshot residue...», at 930).

[23] See also *Hudson v. Lafler* Civil No. 04-CV-74001-DT (E.D. Mich. Jun. 8, 2006) («the prosecution's expert witness admitted that the presence of gunshot residue on a person did not necessarily indicate that the person had discharged a gun», at 7).

[24] The utterance was successful (and profitable) because its meaning changed depending on whether the comma was put before or after the word «not». If the former was the case (i.e. «you will go, you will return, not in the war shall you die») the person believed that he would survive the battle. If the latter was the case (i.e. «you will go, you will return not, in the war shall you die») the person believed that he would die. See Yeroulanos, 2016: 421.

it is not possible to assert whether—and to what extent—evidence is informative in one way or another with respect to a contested evidential claim. Underlying this definition is the idea of ascertainable meaning with respect to an aspect of *primary* interest in a case. Note that this does not exclude the possibility of evidence being interpretable with respect to some other, but subsidiary or even uncontested aspect of a case.[25] In other words, interpretability is not an intrinsic feature of the evidence, but a relational one. It is predicated on the procedural space of reasons and depends on the parties' specific propositions.

We will investigate the consequences of uninterpretability by pointing out how it can undermine subsequent reasoning processes and lead to unwarranted conclusions. An example for this is evidence presented in support of a particular proposition or claim, but tacitly carried over—in an inferentially flawed way—to support the resolution of a disputed matter about which that evidence, in strict terms, is actually uninformative. This results in a spuriously or a seemingly probative conclusion. Where this happens, evidence is «littering», so to speak, the inferential grounds on which subsequent juridical decision-making needs to be based. Thus, there is more to uninterpretable evidence than it being evidence that doesn't answer the question.

Section 2 starts by delimiting the scope of inquiry to the science-centred view of interpretation. This part provides a working definition of interpretability as applied to scientific evidence and its apprehension in legal proceedings. Section 2 will also explain where the demand of interpretability comes from and why it is not covered by standard measures for evaluating the fitness of forensic science disciplines, such as ground-truth testing and related notions such as error rates, foundational validity and validity as applied.[26] Section 3 focuses on ways to probe and enact interpretability, especially determining its limits. For this purpose, we will argue that the distinction between investigative and evaluative uses of evidence,[27] as well as the notion of hierarchy of propositions[28] is crucial. This section also explains the differences between (un-)interpretability and other commonly encountered notions, such as inconclusiveness and probative value. Though this paper is mainly a theoretical inquiry, Section 3 will include an example of evidence (un-)interpretability in the context of modern DNA profiling, namely probabilistic genotyping. The conclusions, in Section 4, emphasise that uninterpretability is not merely a synonym for evidence that does not answer a question of primary interest, nor that it is an end itself. Instead, we argue, uninterpretability is a means to an end, that is avoiding the use of evidence for conclusions it is not able to support.

---

[25] Uncontested propositions in the adversarial criminal process do not need to be proven.
[26] On foundational validity and validity applied see, e.g. PCAST, 2016: 43.
[27] Jackson, 2000.
[28] Cook et al., 1998a; Jackson & Biedermann, 2019.

2. THEORISING (UN-)INTERPRETABILITY OF SCIENTIFIC EVIDENCE

2.1. Delimiting the scope of inquiry: the model-based view of interpretation

Since the mid-1990s, the term interpretation is «experiencing a steady period of evolution» [29] in forensic science. In a landmark paper, a group of scientists of the now axed Forensic Science Service (FSS) considered interpretation as the «drawing of rational and balanced inferences from observations, test results and measurements». [30] This understanding of interpretation as a cognitive process is the result of an FSS project called Case Assessment and Interpretation (CAI). [31] It focused on scrutinising and structuring forensic science services towards delivering output that meets the needs of the mandating party. These needs, however, are not necessarily defined in a narrow sense, specific to scientific evidence. They are not materially different from the reasons that drive evidence processing in general and inquiries thereinto. As Aitken et al. put it succinctly: «The essential issue is: what does the evidence *mean*?». [32] This question echoes the notion of meaning and the quest for understanding highlighted in Section 1.

These developments have been accompanied by a series of scholarly works centred around the notion of *interpretation* of scientific evidence. [33] They focus on the use of forensic science evidence (e.g. DNA, glass fragments, etc.) to help consumers of expert information discriminate between competing propositions of interest regarding, for example, the source of recovered trace material (i.e. identification), or alleged activities that led to the deposition of trace material, including considerations of phenomena such as transfer and persistence of trace materials. Arguments and principles presented in these works rely on, and invoke, concepts of probability and statistics to deal with uncertainty. [34] The mathematical nature of these works has equally attracted criticism [35] and, from the mid-1980s on, enthusiastic support. [36] It is important to note, however, that these debates over the adequacy of formal methods of rea-

---

[29] Cook et al., 1998b: 152.

[30] Ibid.

[31] For reviews of CAI see, e.g. Jackson et al., 2014; Jackson & Jones, 2009.

[32] Aitken et al., 2010: §2.2.

[33] E.g. Aitken & Stoney, 1991; Aitken, 1995.

[34] Evett and Weir «view the interpretation problem as one of updating uncertainty in the light of new information» (Evett & Weir, 1998: 23). See also Fienberg, 1989, and the two volumes by Gastwirth, 1988.

[35] For a widely cited critique, see Tribe, 1971.

[36] See, e.g. the New Evidence Scholarship movement described by Lempert, 1986, in a contribution to Symposium on Probability and Inference in the Law of Evidence held at Boston University in April 1986.

soning mainly concern questions about evidence & proof in legal adjudication as a whole, rather than specific items of scientific evidence.

At this juncture, it is worthwhile to include a critical distinction regarding the term interpretation that is often misunderstood. As mentioned above, when (forensic) scientists talk about interpretation in the sense outlined above, they mean assessing the value of evidence in terms of its capacity to help discriminate between competing propositions of interest. Interpretation, thus, is conditioned on the assumptions made by the scientist as well as the specific information available to the scientist at the time the interpretation is made.[37] Most importantly, expert witnesses are confined—at least in England and Wales—within the questions asked in relation to live procedural issues. Yet, this interpretation is not the end of the matter because «the ultimate value of any particular piece of evidence, scientific or otherwise, must always be assessed contextually, in the light of its contribution to the case as a whole».[38] This latter dimension of interpretation thus amounts to assessing the value of evidence using possibly higher-level propositions—*different* from those with respect to which the scientist assessed the findings—and using all the information and other items available in the case.

The remainder of this paper will focus on the former, science-centred understanding of interpretation because it is a crucial preliminary to the latter, broader sense of interpretation. While the science-centred view of interpretation is well established in literature and, to some extent, in practice,[39] there is yet no structured account of the frontiers of interpretability, especially of the notion of uninterpretability and its implications for advanced stages of the legal process.

2.2. Principles of scientific interpretation

An early, if not the earliest account of what today is known as the «three precepts for the forensic scientist»[40] can be found in Evett's overview of his «personal odyssey».[41] In this account, he refers to these precepts as the «the principles of our science».[42]

---

[37] Typically, the scientist does not need to know all the information. This is neither practically possible, nor desirable, because it may bias the interpretation. For example, in the case of glass fragments detected on a jumper, task-relevant information includes aspects such as the mode of breaking and the time elapsed between the collection of the jumper and the breaking of the window. Whether the person of interest confessed, or whether she was identified by an eyewitness is not relevant information for the scientist (though it is so for the recipients of expert information). See also National Commission on Forensic Science, Human Factors Subcommittee, 2015.

[38] Aitken et al., 2010: §3.28. On the notion of inferential contextualism, see Kotsoglou, 2013.

[39] See Section 2.2 for references.

[40] Evett & Weir, 1998: 29.

[41] Evett, 1991: 19-20.

[42] Ibid., at 19. Here, by «our science», Evett means forensic science. In later writings, he refers to the precepts as principles of «scientific interpretation» (Evett & Weir, 1998: 29) and the «scientific

In essence, the three precepts can be summarised as follows:[43]

1) The meaning of an item of evidence cannot be assessed in isolation, i.e. with respect to only one proposition.[44]

2) The relevant questions for forensic scientists are of the following kind: *given* a particular proposition, what is the probability of seeing this evidence?[45]

3) In order to assess the strength of the evidence, the scientist must consider not only the probability of the evidence (see above point 2) given each proposition of interest, but also the framework of circumstances. The scientist should ask: given the available information (i.e. framework of circumstances), is the evidence better conceivable with one proposition (i.e. account of the contested event) rather than with the stated alternative? If so, to what extent is the evidence better conceivable with one proposition rather than with the stated alternative?

The rationale underlying these precepts stems from both formal considerations (i.e. inferential logic) and common sense. From a formal point of view, the precepts can be recognised as elements of a (statistical) inference method.[46] It is important, however, to emphasise that this is merely to say that the precepts are fundamentally rooted in formal-logical considerations. There is no suggestion that legal evidence or proof processes, as a whole, are a statistical exercise.[47] Instead, the focus here is on understanding the common-sense rationale underlying these precepts and their role in avoiding logically flawed reasoning. While not commonly presented in legal textbooks,[48] the rationale is briefly exposed in the next paragraphs.

Precept number one is about ensuring a balanced perspective. It does so by requiring the consideration of a pair of contrastive propositions. Precept number two seeks to ensure that scientists focus on the findings (evidence), given propositions, and not the reverse. Indeed, the duty of the expert is not to opine on propositions, but to concentrate on the evidence. This may sound counterintuitive, but it suffices

---

approach to interpretation» (Ibid., 245).

[43] For a more recent formulation of these principles, with an emphasis on the framework of circumstances (i.e. conditioning information) see Evett et al., 2017.

[44] A proposition is a statement or assertion about a contested real-world event, that is one version of an event of interest. For example, a pair of what are referred to as source-level propositions is «The semen came from Mr. B» and «The semen came from some other man» (Cook et al., 1998a: 232). Usually, propositions reflect positions held by parties (i.e. prosecution and defense) though there is no procedural requirement for a party to specify a proposition.

[45] This principle has also been referred to as «the single most important lesson for the logic of evaluative forensic science» (Evett, 2009: 159).

[46] Specifically, in the case here, the precepts conform to the logic of the likelihood ratio, which is a core element of Bayesian statistical inference (Evett, 1991: 19). On the notion of likelihood ratio, see also Friedman, 2017: 70.

[47] E.g. Evett & Weir, 1998, note: «It is not our claim that Bayesian inference is a panacea for all the problems of the legal process. However, we do maintain that it [...] provides insights that are not otherwise possible» (at 29).

[48] For notable exceptions, see Friedman, 2017; Dennis, 2017: Chap. 4.

to consider a reformulation of this precept in the more informal language of causes (propositions) and effects (evidence). Suppose Billy hits a window with a hammer; it is not too difficult, then, to conclude that the probability of a broken window and fragments of glass landing on the floor and, possibly, on Billy's clothing, is large.[49] However, the reverse is not straightforward: given the evidence only (in this case: recovered glass fragments), it is not obvious to infer what the cause was, let alone to extrapolate from this Billy's course of conduct. At least one alternative proposition and more investigation are needed to elucidate the case, a task entrusted to the judiciary.[50] The subtle point here is the difficulty of understanding the structure of conditional probabilities, which is known to cause serious problems.[51] In particular, it is important to stress the difference between, on the one hand, the probability of Billy having smashed the window with a hammer *given* the findings and, on the other hand, the probability of the glass findings *given* that Billy smashed the window.

In turn, precept number three seeks to prevent unwarranted assertions about probative value. For example, the mere fact that the scientist considers the evidence to be well conceivable with a particular proposition (e.g. when saying «if proposition so and so is true, then this evidence is what I would expect to see») does not necessarily mean that the evidence supports that proposition. The reason for this is not only that the evidence may equally well be compatible with the alternative proposition at play, and hence be uninformative.[52] It is also important that the evaluation considers the framework of circumstances of the instant case. Stated otherwise, an item of evidence is not probative in one way or another with respect to a particular proposition *in abstracto*. An item of evidence has value with respect to a proposition A rather than proposition B only insofar as the evidence is better conceivable with proposition A, rather than B, i.e. if the evidence is better accounted for by proposition A rather than proposition B in the light of the task-relevant information (derived from the framework of circumstances).

Though they may seem technical at first sight, these precepts of scientific interpretation can be stated in largely non-technical language,[53] and they are fundamental in the sense that many cases of recorded misuse of scientific evidence can be reconstructed as a violation of one or more of these principles.[54] This is why they

---

[49] This example is adapted from Pearl & Mackenzie, 2018: 99.

[50] Evett et al., 2017: 19.

[51] A typical example is the notorious prosecutor's fallacy (Thompson & Schumann, 1987).

[52] Suppose, for example, that in the above example involving a broken window the alternative proposition is that Johnny smashed the window with a hammer, and Billy was a bystander.

[53] The precepts can be stated without resort to either probability or statistics. Thus, while a formal justification for the three precepts is based on statistical principles and methods, it is not necessary—for practical purposes and ease of communication—to label the precepts openly as statistics or statistics based, let alone Bayesian.

[54] A good example is the case *People v. Collins*, 68 Cal. 2d 319 (1968), discussed in Finkelstein & Fairley, 1970. For a more recent example in England and Wales see, e.g. *R v Barry George* [2007] EWCA Crim 2722.

have been incorporated in standard and guidance documents for forensic examiners and legal practitioners,[55] and are regularly reemphasised in the light of controversial cases.[56] But, as well-intentioned and mildly constraining these precepts are, a realistic assessment of the current reporting practice of (forensic) scientists must conclude that, as a whole, forensic scientists continue to follow a variety of different reporting formats and conclusion scales—some of which do not conform to the three principles exposed above.[57]

While the causes of this are operational constraints and adherence to conventional procedures, which are managerial topics in their own right, there are two conceptual considerations that echo deeper challenges. The first is the widely held view that the «problem of scientific evidence» is, first and foremost, an empirical one, i.e. centred around the question of whether forensic scientists can reliably do what they claim to be able to do.[58] Section 2.3. will briefly address this issue. The second consideration has to do with the question of whether the principles of scientific interpretation are sufficient as a safeguard against uninterpretability, let alone able to define limits of interpretability. This aspect will be addressed in Section 2.4.

2.3. Role, limitations and necessity of empirical measures of scrutiny for forensic science evidence

The discussion so far might have inadvertently alienated legal scholars and practitioners whose perspectives regarding admissibility of scientific testimony are influenced by the quests for (empirical) validity derived from legal provisions, such as FRE 702. The notion of validity has been prominently brought to the broader judiciary's attention in the PCAST Report with its strong emphasis on what the Council termed foundational validity and validity as applied.[59] In a nutshell, the Council views forensic science as a discipline that renders direct opinions about the truth or otherwise of particular propositions[60] (e.g. the proposition that the person of interest rather than an unknown person is the source of a particular trace) which is why, the Council considers, it is necessary to require empirical performance indicators for this mode of operation. Specifically, the Council notes that «scientific

[55] E.g. Association of Forensic Science Providers (AFSP), 2009; European Network of Forensic Science Institutes (ENFSI), 2015 (hereinafter cited as ENFSI Guideline); Aitken et al., 2010.

[56] Aitken et al., 2011.

[57] For a review and discussion see, e.g. Thompson, 2019.

[58] PCAST, 2016, is a prominent example for this.

[59] PCAST, 2016: 42-43, explicitly relates foundational validity and validity as applied to FRE 702 (c, d).

[60] Note that this is contrary to principle no. 2, exposed in Section 2.2., according to which the expert should focus on the probability of the findings, given propositions, *not* the reverse (i.e. focus on the propositions given the evidence).

validity and reliability require that a method has been subjected to empirical testing [...] and the error rates are determined».[61]

Prima facie, this sounds entirely reasonable, for who can really afford to ignore hard data regarding the accuracy of a source of information, i.e. the number of times a source of information—using everyday language—«got it right and wrong»? The point is that this perspective falls short of the scope of interpretability, which is why it is worthwhile to properly distinguish these concepts from one another.

To be clear, there is nothing wrong with false positives and false negatives[62] where the concern is process output that conforms to the Council's description of forensic science. Typical examples are cases in which the examiner's task amounts to applying a diagnostic test (e.g. a drug test) and reporting is limited to the actual outcome (i.e. positive or negative). However, even in these cases, a factual report of a test's outcome, accompanied by an error rate, does *not* directly inform us about whether or not a target proposition (e.g. presence or absence of a detectable amount of a drug) is true.[63] Empirical statements—we cannot stress this enough—are probabilistic in nature, and categorical conclusions are unwarranted.

The main reason for this is that a conclusion about the truth-value of a proposition depends on more than a test outcome alone.[64] Moreover, consumers of expert information need to critically inquire about the extent to which the aggregate performance measure, e.g. the underlying error rate data, refers to types of cases (and related case circumstances) that are comparable to the instant case.[65] Again, this is not to say that error-rate data are useless; indeed, they are helpful to inform discourses regarding admissibility of a method or technique *in general*.[66] In other words, error-rate data, base rates and similar statistics are—at best—informative, but not conclusive, about the single event. This insight reminds us about the conceptual and methodological problems in the process of communication between forensic *scientists* on the one hand and *triers of fact* on the other hand.

Therefore, once a type of examination (or testimony) is deemed to be admissible (unless some exclusionary rule applies), we need to assess the probative value of the test outcome in the light of the task-relevant circumstantial information of the case at hand. The factual report of the test outcome, accompanied by aggregate performance indicators, is an *uninterpreted* finding. It is uninterpreted because it does not

---

61 PCAST, 2016: 46.

62 For a discussion of the topic of error rates see, e.g. Koehler, 2013.

63 For example, there is usually non-scientific data in a case that informs what in more technical accounts is called a prior probability (i.e. a probability prior to considering the scientific evidence). See, e.g. Parmigiani, 2002: Chap. 1.

64 Ibid.

65 For a call of a more detailed scrutiny of the parameters of the studies from which data are drawn and used for assessments in instant cases, see Imwinkelried, 2020.

66 Ibid.

directly inform the consumers of expert evidence *how* they should revise, in the light of the test outcome, their view on the pair of contrasting propositions of interest.

In order to give the test outcome an interpretation in the sense understood in this paper—i.e. the principles of scientific interpretation (Section 2.2.)—one not only needs a probability assignment for observing the test outcome (e.g. a positive result) *given* the target proposition (e.g. presence of a detectable amount of illicit substance). One also needs the probability of obtaining the test result given the appropriate alternative proposition (absence of illicit contents). Common proxies[67] for these probabilities are the sensitivity for the former and the probability of a false positive (i.e. 1-specificity) for the latter.[68] Combined, these two assessments characterise the value of the evidence: the evidence is probative in one way or another to the extent that the method's sensitivity differs from the complement of the specificity (i.e. the probability of a false positive).

Again, this should conform to laypeople's understanding and common intuition. As an example, consider a situation in which the sensitivity is the same as the probability of a false positive. That is, the probability of obtaining a positive test result is the *same*, regardless of whether or not a detectable quantity of illicit substance is present (i.e. the target proposition is true). Such a test result would have *no* probative value. A positive result has probative value if and only if the probability of the positive result is greater given the first rather than the second proposition, or vice-versa. More specifically, the greater the sensitivity compared to the false positive probability, the more probative the positive test result is with respect to the first rather than the second proposition.[69] Stated otherwise, the smaller the false positive probability, compared to the sensitivity, the greater the diagnosticity of the evidence. To summarise, a good test or method should have a high sensitivity and a high specificity (i.e. low false positive probability). Thus, interpreting a finding amounts to addressing points 1 to 3 exposed in Section 2.2., also known in the context here as «the logic of forensic inference».[70]

One question that these considerations raise is whether the PCAST Report depicts an unsuitable approach to the interpretation of forensic science evidence and forensic science as such. Some commentators think so, calling it a «flawed paradigm for forensic evaluation».[71] In a less severe view, however, one could argue that the PCAST Report focused on attempting a minimal first move ahead because, as mentioned above, aggregate performance indicators (e.g. error rates) are an important preliminary consideration (i.e. with respect to admissibility).[72] Empirical data to

---

[67] We use the term proxy here because, strictly speaking, the statistics refer to observations across multiple cases, whereas the problem to be dealt with here is assessment in the individual case.

[68] Kaye, 1987: 360.

[69] Ibid.

[70] Evett et al., 2017: 17.

[71] Ibid, at 18, and Imwinkelried, 2020.

[72] See also Lander, 2017: 367-368.

help assign probabilities for false positive outcomes are also helpful by providing, in certain applications, half of the «equation» stipulated by the three principles of scientific interpretation (Section 2.2.). But, by considering only part of the principles of coherent reasoning, the perspective advocated by the PCAST Report leaves scientific evidence essentially uninterpreted. It provides no guidance regarding what exactly recipients of expert information are logically entitled to infer from a particular test result when merely accompanied by an error rate only.[73] Stated otherwise, the PCAST report did not proceed to address what may be called the micro-level of value of evidence assessment. That is, it stayed at the macro-level of admissibility. In this sense, the perspective advocated by the PCAST report abstains from the case-based assessment of the probative value of scientific evidence.

Despite all scepticism, it is worthwhile to consider a possible argument in favour of the PCAST Report. If one considers that, as mentioned above, black-box-like studies represent an essential, though minimal, component of performance assessment, the PCAST report triggered the hope that the field would agree to conduct more ground truth testing (i.e. the generation of data for empirical performance indicators).[74] Rather disturbingly, however, even that hope set the bar too high for some quarters of forensic science that expressed their discontent with the PCAST Report's insistence on empirical grounding.[75]

2.4. Uninterpretability: conceptualising the frontiers of interpretability

Suppose, for the time being, that we accept the PCAST Report's insistence on empirical foundations as an important preliminary to the consideration and admissibility of expert witness testimony in the legal process. Assume further that, on the micro-level of interpretation in the instant case, it is reasonable to require that the forensic scientist addresses the tripartite question outlined in Section 2.2. These questions seek to determine whether, and to what extent, an item of evidence is probative with respect to particular propositions of interest. According to this view, a probative item of evidence is one that is capable to influence, in one way or another, one's view (or belief) regarding the propositions of interest. In legal scholarship, the

[73] This is not a problem specific to the PCAST Report. Academics remain vague on how to use diagnostic performance measures properly too. For example, Edmond et al., 2014, write «[t]he juror can reason with this information to infer something about the particular case» (at 10), without explaining what exactly this inference should amount to and despite the fact that the answer to this question is largely available in literature.

[74] The authors are grateful to Professor Christophe Champod for this reflection.

[75] For a list of published statements in response to the PCAST Report see https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_2016_public_comments.pdf (last accessed March 4 2021).

latter property is readily recognised as (logical) relevance,[76] a notion that can also be reconstructed in probabilistic terms.[77]

Broadly speaking, assessing evidence thus means assessing whether the evidence is probative with respect to particular propositions of interest. Central to this understanding is that the value of evidence is a property that depends on the intended purpose, the questions of interest in the case at hand (i.e. the litigated matter),[78] the assumptions made and the information available to the interpreter at the time an interpretation is made.[79]

While this working definition of interpretation purports to answer the question of what interpreting an item of scientific evidence means, it does not answer the question of what an *(un-)interpretable* item of evidence is. The former, i.e. interpretation, has to do with the logic of inference whereas the latter, (un-)interpretability, is an ascription made to a piece of evidence by an interpreter. Prima facie, it might be understood as a property of an item of evidence, but ultimately, it hinges upon the interpretative capacities of the interpreter. We can now delve deeper into these intricate aspects.

The definition of interpretation of scientific evidence through the three principles exposed in Section 2.2. allows us to derive a *provisional* definition of interpretability as a substantial property of the relationship between an item of evidence and the latter's interpreter. That is, an item of evidence can be said to be *interpretable* if and only if it is possible to enact the principles of interpretation.[80] In turn, an item of evidence is said to be *uninterpretable* when the interpreter cannot enact the principles of interpretation and, hence, no statement can be given as to the probative value of the evidence (i.e. asserting which of the competing propositions, if any, is favoured by the evidence over the respective alternative). This definition may seem a statement of the obvious, though it is crucial to ask—as emphasised in the remaining parts of this paper—whether the mere enaction of the principles of interpretation is sufficient to ensure that an interpretation is actually purposeful.

The argument made here is that interpretability is a *relational property*, in the sense that an item of evidence may be interpretable with respect to a given pair of propositions, but uninterpretable with respect to another pair of propositions.[81] For example a DNA profile is, strictly speaking, interpretable mainly as regards the source of the biological material, but requires additional considerations when discussed in

---

[76] For a probabilistic understanding of the term relevance, in the forensic sense, see Stoney, 1994.

[77] E.g. Lempert, 1977; Friedman, 2017: 49-84.

[78] Aitken et al., 2010: §2.2.

[79] Jackson, 2000: 85.

[80] Note, however, that an interpretable item of evidence is not necessarily probative. See §I.E. regarding the distinction between interpretability and probative value.

[81] As an aside, note also that interpretability may vary across examiners, depending on their level of expertise for the evidence in the instant case.

the context of alleged transfer mechanisms.[82] Again, this may seem obvious, but the problem is that in practice, this ambivalence is not always properly acknowledged, or even understood, which may result in seemingly probative evidence, leaving vital considerations of uninterpretability unnoticed—possibly to the detriment of the parties.

To account for the ambivalent nature of interpretability, it is helpful to consider a distinction between what we call here *weak* (or, limited) and *strong* (or, strict) interpretability:

— *Weak (or limited) interpretability:* evidence is said to have weak (or, limited) interpretability if, at best, it is interpretable with respect to a pair of propositions at a hierarchically low level, such as source level (i.e. a proposition regarding the source from which a trace or mark comes from).[83]

— *Strong (or strict) interpretability:* evidence is said to have strong interpretability if it is interpretable at an advanced propositional level, in particular with respect to propositions regarding alleged activities.

The pertinence of this distinction is illustrated by the fact that scientists may content themselves with limited interpretability, which is problematic because it may deprive recipients of expert information from the fact that the evidence is actually uninterpretable with respect to higher propositional levels that are closer to the ultimate issue of a case. Worse, when left unassisted, recipients of expert information may—in an unwarranted way—apprehend weakly interpretable evidence as strongly interpretable evidence. At the very least, the notion of weak interpretability emphasises the understanding that «[s]ource level propositions [...] may have limited utility for criminal adjudication».[84]

To illustrate the ambivalent nature of interpretability, consider an example involving DNA recovered on a discarded object, such as headgear, a tool or a weapon, found on a scene of investigation:[85] a small quantity of DNA (also sometimes called trace or touch DNA) is found on the object. The DNA profile corresponds to the DNA profile of a person of interest. Depending on the quality of the DNA profile (which may be a partial or complete, mixed or unmixed), the observed correspondence in DNA profiles may be probative with respect to the propositions according to which the person of interest rather than an unknown person is the source of the

---

[82] For more discussion see e.g. Kotsoglou & McCartney, 2021.

[83] Regarding the notion of hierarchy of propositions, see Cook et al., 1998a. Broadly speaking, the higher a proposition in the hierarchy, the closer it approaches the ultimate issues in a case, and hence the more useful an interpretation is for fact-finders. For a recent overview of the hierarchy of propositions, see also Gittelson et al., 2016.

[84] Aitken et al., 2010: §3.9. For a recent example of this argument in the context of a case involving small quantities of DNA, see *United States v. Gissantaner*, No. 19-2305 (6th Cir. 2021).

[85] For an example of a real case involving a hat found at a crime scene, see Biedermann et al., 2011. Similarly, *United States v. Gissantaner*, supra note 84, involving a weapon.

DNA.[86] Thus, according to our account, the DNA evidence is interpretable with respect to (sub-)source level propositions. This, however, represents only limited interpretability because the source level propositions are distant from the ultimate issues to be decided by the triers of fact. The latter are primarily interested in questions such as whether or not the person of interest is the individual who wore the headgear at the crime scene (or handled the weapon at the relevant time, etc.). Such propositions are at a higher hierarchical level, and interpretability with respect to such propositions would require *strong* interpretability. However, ensuring strong interpretability in this sense is a highly intricate task because it requires more than scientific knowledge regarding the rarity of the genetic features. The scientist would need to discuss scientific knowledge, if available, regarding phenomena such as transfer, persistence and rates of recovery. This requires data from empirical research. When knowledge and data on these topics are scarce, or even missing, the scientist may be unable to assess the probative value of DNA profiling results with respect to the alleged activities. This inability to enact the principles of interpretation will make the evidence *uninterpretable at this advanced propositional level* (i.e. activity level), despite the interpretability at source level. As mentioned above, evidence interpretability is relational insofar as it refers to a specific proposition, not to some intrinsic property of a piece of evidence.

The critical reader may wonder why this poses a problem and why it is not sufficient that scientists fall back to a less ambitious interpretational level (i.e. source level), and thus simply do what they «know best»—after all, this should lead to more robust findings. There are two problems with this intuitive idea. First, the operations of forensic experts is not an anything-goes activity, for this would run contrary to the «good old way» according to which expert witnesses act simply as «helpers of the court».[87] Secondly, a critical problem is that consumers of expert information may be tempted to carry the evaluation given source level propositions over to a conclusion regarding alleged activities—in an argumentatively unwarranted way.[88] This transition may be unwarranted because, as argued above, such a move requires specialised knowledge (e.g. regarding phenomena such as transfer and persistence) which, in case of evidence uninterpretability, is unavailable to the scientist. And, if such specialised knowledge is unavailable to scientists, it is also unavailable for consumers of expert information.

This problem may remain undetected, and thus be aggravated, when the scientist does *not* alert fact-finders that the evidence is *uninterpretable* with respect to higher-level propositions (i.e. alleged activities). The bottleneck, thus, is lacking spe-

---

[86] Note that, strictly speaking, this would be called sub-source level propositions because the question of source only refers to the DNA, not to the trace as such. The reason for this is that because of the low quantity of recovered material, the type of biological material (e.g. saliva) form which the DNA comes from, cannot be determined. See, e.g. Evett et al., 2002: 521.

[87] Thayer, 1892: 665.

[88] E.g. Gill, 2014.

cialised knowledge. This creates the danger that laypeople—unaware of the informational problem—ignore this bottleneck and draw conclusions from the evidence *despite* its uninterpretability with respect to higher-level propositions.[89] Where this happens, uninterpretable evidence can litter the paths of reasoning towards primary matters of litigation (ultimate issues). Such evidence, thus, is only seemingly probative.

It follows from the above that whenever the recipient of expert information cannot be considered to be competent to assess the meaning of evidence with respect to more advanced propositional levels (such as alleged activities), and when these levels reflect the needs in the case, it is not appropriate for the scientist to limit evidence evaluation to only a low propositional level (such as source level).[90] Instead, the scientist should either assess what, if anything, the evidence means with respect to the higher propositional level, provided the necessary specialised knowledge is available, or else alert the fact-finder that the evidence is *uninterpretable* with respect to the advanced propositional level, because of the lack of relevant specialised knowledge.

Notwithstanding these intricacies, the default mode of assessment, especially in forensic genetics, still widely relies on propositions regarding the source of DNA (i.e. so-called sub-source level propositions).[91] This includes applications of modern probabilistic genotyping[92] software that scientists use to evaluate DNA profiling results with respect to propositions according to which the person of interest is (or is not) the (or one of the) contributor(s) to the DNA in the trace.[93]

The sceptical reader may retort that recipients of expert evidence can make disciplined use of evidence and that they do not unduly extend the value of evidence

---

[89] This danger is particularly critical whenever the evidence has a high probative value with respect to source level propositions (i.e. a hierarchically low propositional level), but *not* with respect to higher propositional levels regarding, for example, alleged activities.

[90] This is stipulated, for example, in the ENFSI Guideline: «Propositions are not altered during examination/evaluation unless the key issues in the case and/or the conditioning information have changed. For example, when the issues at hand are at activity level, the absence of data or expert knowledge on transfer, persistence or background level of the trace type under consideration is not a justification to change the set of activity level propositions to a set of source level propositions. In fact, the choice between (sub-) source and activity should not be influenced by the availability of data or expert knowledge but solely from the consideration of factors such as transfer, persistence and background levels that could crucially affect the strength of the findings within the context of the case circumstances» (ENFSI, 2015: 13). Similarly, the Recommendation 4 of the guidelines issued by the DNA Commission of the International Society for Forensic Genetics specifies: «Propositions should be formulated at a level in the hierarchy of propositions, where the forensic scientist brings knowledge that is required, but not readily available to the court. This is crucially important when transfer, persistence and background DNA have a significant impact on the case» (Gill et al., 2020: 5).

[91] On the term sub-source see supra 86.

[92] More detailed discussion on probabilistic genotyping is presented later in Section 3.4.

[93] Bright et al., 2016; Coble & Bright, 2019; Buckleton et al., 2019 (noting that «PG [probabilistic genotyping] software and other interpretation and statistics methods evaluate the DNA results at the subsource level», at 395).

to an advanced proposition level where the evidence is actually uninterpretable. So, what could be wrong then with adhering to an evaluation given source level propositions?

The problem is that an evaluation given source level propositions is not a general solution for all cases. Indeed, in some cases, the question of source may not be disputed at all—but then the parties would not need expert evidence to back their claims. To continue with the DNA example introduced above, suppose that the person of interest concedes that the object (e.g. headgear, weapon, tool etc.) is theirs, but then adds that it was lost or stolen, say, sometime prior to the incident.[94] In such a case, it may not be contested that the corresponding DNA is that of the person of interest.[95] Thus, rather than the issue of source, the competing accounts are whether the person of interest is the man who wore the headgear at the crime scene versus an unknown person of interest wore the headgear (and that the person of interest last wore it, e.g. two weeks ago). Clearly, an expression of the value of the evidence with respect to source level propositions is useless here for there is no need to prove something that is not contested. Instead, of interest here is only the extent to which DNA evidence is capable of discriminating between alleged activities; and, if the evidence cannot be assessed with respect to the alleged activities, this means that it is uninterpretable at this propositional level, and thus has no probative value. It would be unwarranted then to shift attention to a lower, but *irrelevant* propositional level (i.e. source level) for which the evidence may be informative.

2.5. Distinguishing (un-)interpretability from probative value and relevance

Whether or not meaning can be assigned to an item of evidence is, according to the view defended in this paper, at the core of the notion of (un-)interpretability. It is understood here as a property that characterises the relationship between evidence and specific propositions as articulated by the interpreter. It is important, however, to draw a distinction with respect to the notion of probative value. In particular, it should be emphasised that even though a piece of evidence may be interpretable, i.e. is one can enact the principles of interpretation and assign the value of the evidence, this does not imply that the evidence is probative. Indeed, an interpretable item of evidence may be found to be neutral and, hence, irrelevant. That is, the evidence is not probative in one way or another with respect to the competing propositions of interest to the fact-finders.

---

[94] See, e.g. Jackson et al., 2014: 50, for an example of this type of case, involving a balaclava discarded by an offender on the scene of a robbery.

[95] Similarly, for example, it is unsurprising to find a DNA profile corresponding to the profile of a person of interest on an object that has been seized in the apartment where that person lives (or in that person's car, etc.). For a practical example see, e.g. *United States v. Gissantaner*, supra note 84.

Consider Table 1 for a summary of the distinction between (un-) interpretability and probative value. Clearly, when the principles of interpretation cannot be enacted, and hence the evidence is uninterpretable, no statement can be made regarding the probative value. In essence, thus, an uninterpretable item of evidence provides no assistance to anyone who seeks to draw an inference about the competing propositions of interest—such evidence is not relevant.[96] But this may also happen with interpretable pieces of evidence. That is, when the principles of interpretation lead the scientist to conclude that whichever proposition is true, the probability of observing the evidence is the same. Such evidence is interpretable, but not probative in one way or another. Evidence is probative in one way or another if and only if the evidence is better conceivable with one proposition rather than the respective alternative proposition.[97]

(Un-)interpretability	Probative value: the evidence ...		
Weak / strong interpretability (Section 2.4)	is *probative*: it supports the defence proposition over the prosecution's proposition	is *not probative*: does not support one proposition over the other (i.e. the evidence is inferentially neutral)	is *probative*: it supports the prosecution's proposition over the defence proposition
Uninterpretability	No probative value can be assigned		

Table 1: Distinction between (un-)interpretability and probative value.

2.6. (Un-)interpretability vs. inconclusiveness

It is useful to distinguish the concept of (un-)interpretability from other terms commonly used by forensic examiners. One such term is «inconclusiveness». This term is mainly used by forensic scientists working in disciplines that involve comparative examinations (e.g. fingerprint experts). For example, according to a recent directive by the U.S. Department of Justice (DOJ), called Uniform Language for Testimony and Reports for the Forensic Latent Print Discipline (ULTR), inconclusive is (to be) used when «the examiner is unable to identify or exclude the two impressions as originating from the same source».[98] According to this definition,[99] the term in-

[96] This corresponds to the doctrinal understanding of relevance. See also supra note 76.

[97] E.g. Friedman, 2017: 70.

[98] U.S. Department of Justice, 2020: 3 (hereinafter cited as ULTR, short for «Uniform Language for Testimony and Reports for the Forensic Latent Print Discipline»).

[99] The term inconclusive is used in a variety of forensic disciplines, and the reasons which lead examiners to use this conclusion are diverse. For example, it may be that there is simply not enough reference material from a known source to conduct meaningful comparative examinations between materials from known and unknown source.

conclusive means that the findings (observations) are such that the examiner cannot assert whether or not the person of interest is the source of a recovered fingermark. Note, however, that for the logical and procedural reasons explained in Section 2.2., expert witnesses should not opine directly on propositions, i.e. express themselves in terms of whether a person of interest is the source of a mark or trace. When redirecting the focus on the findings, rather than the propositions, inconclusiveness means that the findings are considered not probative in one way or another (i.e. in favour or against the proposition of same versus different source). Inconclusive is, in essence, taken to mean that the findings are neutral. But even this is not particularly helpful because it provides no insight into the deeper reasons for this conclusion.

The notion of (un-)interpretability can help to disentangle the underlying tenets and distinguish itself from inconclusiveness. Table 1 clarifies that there are two situations in which the evidence is uninformative for recipients of expert information. Either the scientist *was able* to assess the probability of the findings given each of two competing propositions, but found that they were the same, and hence the evidence has no value. Note that according to the considerations exposed in Section 2.4., such evidence would be said to be interpretable because the scientist was able to enact the principles of interpretation (defined in Section 2.2.). Or, the scientist cannot enact the principles of interpretation, a situation in which the evidence is uninterpretable and, hence, necessarily uninformative.

More generally, the term inconclusive is internally contradictory because it suggests that one cannot «conclude» anything, yet asserting that the evidence has no probative value is as much a conclusion as asserting that the evidence has (some) probative value. Further, the use of the term inconclusive presupposes that evidence is either categorically conclusive or inconclusive,[100] a view that misconceives the reality that most evidence is only probative to varying degrees. For these reasons, «inconclusive» is a term that should be avoided altogether,[101] not least because it intrigues the consumer of expert information to think that the examiner is presenting an opinion on the propositions, or probabilities thereof, rather than a report on the value of the findings.

3. PROBING FOR (UN-)INTERPRETABILITY

Responsible consumers of expert evidence should take an interest in the topic of (un-)interpretability because it can help them gain a better understanding of the limitations of scientific findings, especially in cases where evidence is only weakly interpretable, i.e. at a propositional level that is far away from the questions (ulti-

[100] E.g. the ULTR, 2020, states: «The examiner may offer any of the following conclusions: 1. Source identification (i.e. came from the same source) 2. Source exclusion (i.e. came from different sources) 3. Inconclusive» (at 2).

[101] E.g. Cole & Biedermann, 2020.

mate issues) that the fact-finder has to determine—and possibly uninterpretable at more advanced propositional levels, closer to the primary matters that fact-finders are concerned with. Exposing the limitations of interpretability should help parties guard against unwarranted carrying-over of evidence to advanced levels of the justificatory chain that leads to the final verdict. This section presents and discusses a series of considerations that are key to probing for uncertainty as part of a critical review of scientific evidence.

3.1. Distinguishing between evaluative, investigative and technical reporting

A useful first step in the review of a report from an expert witness is to understand the type of report insofar as (un-)interpretability—being relational in nature—is not a cause of concern for all report categories. Questions of (un-)interpretability primarily arise in so-called *evaluative* reports. An evaluative report is one that reports on the observations made during comparative examinations conducted on questioned and known materials,[102] and on the probative value the scientist assigns to those observations in the light of the propositions that reflect the views of the parties at various stages in the legal process.[103] Here, the notion of probative value refers to an expression of strength of support for one proposition compared to its alternative, derived by answering the question triad exposed in Section 2.2. It should be emphasised that the focus of a genuine evaluative report is on the observations and findings, and their discriminative capacity with respect to the propositions of interest—the evaluative report must not contain a direct opinion on the propositions, i.e. live issues.

A first urgent remark regarding the requirement outlined above is that the reality of forensic reporting is far removed from this confined methodological space. With the exception of forensic DNA (to some extent),[104] any fair characterisation of current reporting formats reveals that forensic scientists oftentimes opine directly on competing propositions,[105] without being prevented by the parties or the courts. For example, fingerprint analysts commonly state whether or not (they think that) the person of interest is identified as the source of the fingermark found on the crime scene.[106] The latter is a statement about a live procedural issue rather than about the (probative) value of the similarities and differences observed when comparing the

[102] Questioned material is material whose origin or source is unknown. Typical examples are traces (e.g. blood, fibres, etc.) and marks (e.g. toolmarks, fingermarks, etc.) found on a crime scene, victim or a person of interest. Known (or reference) materials are materials whose origin is known (e.g. a named person or object).

[103] E.g. ENFSI, 2015: §1.1 and §2.1.

[104] E.g. Gill et al., 2018.

[105] For a critical analysis and discussion of current reporting language for federal examiners, see Cole and Biedermann, 2020.

[106] E.g. Champod et al., 2016; Thompson, 2018: 775.

questioned fingermark and the reference print of the person of interest. Although one might infer that the scientist must have considered the findings to be of (at least some) value, and hence interpretable, allowing them to opine directly on propositions is tantamount to (*a*) skipping the necessary preliminary stage of assessing the probative value of the observations made in the first place and (*b*) infringing, thus, the fact-finder's decision-making prerogative, which shows disregard for the procedural architecture of modern criminal justice systems, especially—we say this at the cost of oversimplification—the criminal process.[107] The former, (*a*), obstructs any insight into how the scientist assessed the value of the *observations*. Whether that assessment was thorough thus remains hidden, unarticulated and, ultimately, escapes proper scrutiny.

The above considerations clarify that evaluative reports involve material from a known source, i.e. a person or object (tool), that has been compared with some questioned item of unknown source. The reason for emphasising the need of a careful assessment of probative value, according to the principles exposed in Section 2.2., is that such an assessment can result in evidence for or against a person of interest. This needs to be distinguished from reports that do not involve results of comparative examinations with a potential source—and hence do not tend to directly implicate a person of interest. For example, during early stages of an investigation, when only questioned material from the crime scene is available, but no potential source, examiners may examine traces to provide information useful for investigation. For example, based on blood stains allegedly left by the offender at the crime scene, analysts may provide information regarding the DNA profile that the offender must possess.[108] Similarly, in a hit-and-run case, the examination of paint fragments found on the scene of the incident can reveal information regarding the (type of) vehicle that was involved. In such cases, the examiner's report is confined to describing the features of recovered material only. Such reporting is also sometimes called technical or investigative reporting.[109]

The above distinction between evaluative and investigative reporting is not necessarily neat. For example, it may be argued that a report is evaluative even though it involves no direct comparison with a potential source (person or object). For example, in the case of a seized substance, analyses may reveal the presence of a drug which may result in charges against a person of interest. Thus, strictly speaking, measurements and results should be assessed in the light of two competing propositions, e.g. that the material does (or does not) contain a detectable quantity of illicit substance. While this could be seen as an instance that raises the problem of (un-)interpretability, it is also possible to look at this type of case as an example of classification. This

---

[107] See Biedermann & Kotsoglou, 2018.

[108] Stoney, 1994: 18.

[109] It is worth noting, however, that evaluative reporting often involves elements of technical reporting, in particular sections that report on the features of the recovered materials. ENFSI, 2015: §1.1.

is a task that seeks to determine the nature of examined material by assigning it to a particular category (e.g. illicit versus legal substance). The ENSI Guideline, for example, considers this type of reporting (classification) as technical because it is only concerned with ascertaining the nature of recovered material. [110]

The distinction between technical and evaluative reporting is important because it can help guard against the unwarranted use of investigative information for evidence and proof purposes at more advanced stages of the legal process. Suppose, for example, that a technical report asserts that the fragments of paint found on a victim in a hit-and-run case are blue and composed of several distinct layers. While this information may be helpful to direct investigations (i.e. searching a certain type of blue vehicles), it is unsuitable for direct use as evidence against a person of interest who is found to have a blue car. The reasons for this are, first, that detailed comparative examinations would need to be conducted between the recovered fragments and control material from the car of the person of interest. Second, the value of the observed similarities and differences would need to be assessed with respect to a pair of competing propositions (e.g. the recovered fragments come from the car of the person of interest vs. an unknown vehicle). Only such comparative examinations and the assessment of the probative value of their outcomes can lead to a proper evaluative report.

3.2. Uninterpreted and partially interpreted vs. uninterpretable evidence

Having clarified that interpretation in evaluative settings (as opposed to investigative settings, Section 3.1.) requires the forensic expert to consider a bipartite question, that is the probability of the observations (findings) *given* a pair of contrasting and competing propositions and the task-relevant information (Section 2.2.), we can now use this framework to delineate different interpretational limitations.

A common objection against the principles of evaluative reporting is that the requirement for at least one alternative proposition could be misunderstood as a practical obligation. Clearly, to ensure helpfulness, the contrasting set of propositions that the forensic scientist must take for granted when assessing the probative value of the findings maps on the actual procedural claims of the parties. But this is an empirical rather than a doctrinal observation, and does not necessarily map neatly on real litigation tactics in each case. For the defendant has the right to remain silent or can put forward some other claim.

Yet, at the time when forensic experts write their reports, they may not have been provided with an alternative proposition because, strategically, the defence may choose to remain silent until they learn what the scientist has observed (e.g. whether

[110] ENFSI, 2015: §5.0 («Reporting, technical (factual)»).

or not there is a correspondence between the DNA profile of the crime stain and the DNA profile of the defendant, or whether any DNA is found at all). In the absence of an alternative proposition, there are—according to the ENFSI Guideline[111]—three options that the scientist may follow, two of which will leave the findings *uninterpreted*:

1) Adopt an alternative proposition that, given the circumstances of the case, most suitably reflects the position of the party.

2) Adopt no alternative proposition.

3) Present a range of plausible explanations of the findings.

It is important to note that only option 1 can lead to a proper evaluative report, because the evaluation will be based on a pair of competing and mutually exclusive propositions. However, although experts should be prepared to review their evaluation of the findings when they are provided with a proposition other than the one assumed by default, option 1 remains tricky. In fact, it is problematic for a party to leave the choice of an alternative proposition to the scientist because the scientist's choice may result in an evaluation that is less favourable for the party of interest than it could be.[112]

Option 2 implies that the scientist cannot enact the principles of interpretation, thus leaving the findings uninterpreted.[113] When there is only one proposition, the scientist may, at best, assess the extent to which the findings may be expected to occur given the single proposition available. This amounts to working half-way through the principles of interpretation (Section 2.2.), meaning that the findings can be considered only partially interpreted. Most importantly, however well conceivable the findings may be with the single proposition available, *no* statement can be made about the credit (or, support) the findings provide for the single proposition available. For example, when there is a biological stain (evidentiary item) with a DNA profile that corresponds to the DNA profile of a person of interest, this observation—taken on its own—has *no* probative value as long as the scientist has not considered the findings in the light of at least one alternative proposition. In particular, there are no logical grounds to assert that the observed correspondence in DNA profiles somehow supports the proposition according to which the person of interest is the source of the DNA stain, however appealing such a conclusion might appear.

Option 3 is more subtle. It amounts to the scientist exploring a list of explanations for the findings, and perhaps ranking those explanations according to their plausibility. There are two caveats with this option. The first is that explanations,

---

[111] ENFSI, 2015: §4 (Guidance note 1).

[112] It is well known, for example, that in a case where the source of a DNA stain is contested, the default alternative proposition that the DNA comes from an unknown and unrelated person tends to maximize the probative value (Buckleton & Triggs, 2005).

[113] See ENFSI, 2015: §4 (Guidance note 1): «in the absence of an alternative proposition, it is impossible to evaluate the findings».

as understood by scientists, are not the same as formal propositions used in the standard evaluative setting (Sections 2.2. and 3.1.). Propositions are formal, specified by the parties and strictly anchored in the framework of circumstances of the instant case. In turn, explanations produced by scientists may be open-ended and exploratory (or even speculative) without necessarily being rooted in the case-specific circumstances.[114] Often, such explanations are provided almost by default, as a deflecting shield,[115] without even inquiring about whether they are pertinent to the case at hand.[116] While this may be suitable at the investigative stage, i.e. when the questions of interest focus on what happened and how, contemplating explanations cannot lead to an evaluative report understood in the sense explained in Section 3.1. An evaluative report states the value of the findings for discriminating between propositions of interest that specify an individual or object as the source of a mark or trace, or well-defined activities whereby the mark or trace has been generated.[117] Thus, recipients of expert information should keep in mind that an explorative report in the sense of option 3 does not represent an evaluative report in the sense of the criteria of interpretation exposed in Section 2.2. An explorative report leaves the observations and findings uninterpreted. Explaining the findings and interpreting findings are two fundamentally different tasks.

The above situations illustrate practical and procedural obstacles that can prevent scientists from enacting the principles of interpretation, thus leading to uninterpreted or, at best, only partially interpreted evidence. The proposed report categories can help recipients of expert information understand the nature of experts' reports (or testimony), recognize their limitations and identify instances where the probative value of the evidence remains undetermined. This is the case for situations 2 and 3 as outlined above.

Notwithstanding, it is important to remind that even in a case corresponding to situation 1, i.e. when the scientist evaluates the findings with respect to at least two competing propositions, there is no guarantee that the evidence is actually interpretable. As an example, consider again the hypothetical case presented in Section 2.4., where a small quantity of DNA was detected on an item handled by an offender (e.g. headgear, tool, etc.), and discarded on the crime scene. When the DNA profile of the trace corresponds to the DNA profile of the person of interest, it is not an easy task to assess the value of this correspondence with respect to so-called activity level propositions, e.g. the person of interest wore the balaclava at the relevant time vs. the person of interest wore the item a few weeks ago (before it went missing or

---

[114] Evett et al., 2000.

[115] The idea of a deflecting shield is the protection of the scientist against accusations of not providing an exhaustive list of all conceivable explanations of the evidence.

[116] For an example see, e.g. the default statements provided in the context of gunshot residue analyses mentioned in Section 1 of this paper.

[117] Cook et al., 1998a.

got lost).[118] The difficulty of assessing the findings given such propositions is due to the fact that data on phenomena such as transfer and persistence of DNA under the circumstances of the instant case are scarce. Specifically, when the scientist cannot assess the probability of observing the DNA findings given each activity level proposition and the task-relevant case information, this means that the principles of interpretation cannot be enacted, and hence that the evidence must be considered uninterpretable at this propositional level, *despite* possible interpretability at source level. It is also important, at this juncture, to recall the practical implication of uninterpretability and to emphasise the critical need for parties to understand this implication: uninterpretable evidence means that it provides no assistance to anyone asked to decide which of the competing propositions is true.

3.3 The central role of main objects of litigation and propositions in interpretation

Successfully probing evidence for (un-)interpretability hinges upon one's ability to keep track, as pointed out in Section 2.4, of the relational nature of evidence (un-)interpretability. That is, the feasibility of ascertaining evidential value can vary as a function of the propositions with respect to which the evidence is assessed, the available task-relevant information and any additional assumptions made. But, if the value of evidence depends on the propositions being considered, how should one define propositions suitably? This is a crucial question because expert reports are often unclear about the reasons that led to the definition of the propositions that have been retained. Experts have a natural tendency to shape propositions in such a way that their findings will be interpretable with respect to the chosen propositions. However, as emphasised in Section 2.4., interpretability is not tantamount to, nor does it guarantee or imply helpfulness. Interpretability merely means that the expert *can* enact the principles of interpretation. Whether they *should* enact these principles in a given way is contingent on procedural circumstances: the claims that parties bring forward and the claims that remain disputed. Remember that forensic science serves the purposes of criminal justice, not vice versa. For example, a person of interest may not contest that an item (e.g. headgear, tool, etc.) is theirs, and hence them being the source of trace material (e.g. DNA, fingermarks etc.) found on the item. Instead, there may be disagreement over the activities that led to the deposition of DNA or fingermarks on the item of interest. Clearly, there is no need to adduce evidence for an uncontested issue. Thus, as noted by Aitken et al., it is important to understand what the main object of litigation is:

[118] Recall that the question of source (i.e. whether or not the DNA comes from the person of interest) may not be of interest because the person of interest may not contest that the item or object is theirs. Instead, they may contest the way (i.e. activity) that led to the deposition of DNA on the item or object. For further discussion, see Section 3.3.

> One cannot assess whether evidence is successful in proving a matter in issue until one knows what the issue is and how the evidence relates to it. This observation might sound banal; but it is not. In fact, nearly all of the reasoning errors [...] are either variations on, or are at least exacerbated by, an elementary failure to identify, with sufficient care and particularity, the question which the evidence is capable of answering.[119]

A corollary for parties thus is that they should insist on the understanding that, first, the meaning of evidence depends on the intended purpose(s) for which it is adduced (i.e. the primary object of litigation), and hence, second, that it is these objects of litigation—not minimal interpretability—that define the propositions with respect to which evidence is to be evaluated. If that may result in evidence being uninterpretable, then this is worthy of being exposed. This is vital, especially for the defence. Indeed, it would be disingenuous to change propositions away from the principal matters of litigation for the sole purpose of making the evidence interpretable (though with respect to an uncontested aspect of the case). Such evidence, if uninterpretable with respect to the principal questions, will remain so; interpretability with another set of non-principal questions is prone to lead to seemingly probative evidence. That is, it is prone to being carried over to a conclusion that the evidence cannot logically support.

3.4. Exemplifying (un-)interpretability: probabilistic genotyping

Probabilistic genotyping (PG) is the combined use of biological models, statistics and computational methods to help assess complex DNA profiling results,[120] in particular traces with DNA from more than one person (i.e. DNA mixtures). Various software solutions exist to practically implement PG and they are regularly debated in courts.[121] PG aims at quantifying the probative value of DNA profiling results with respect to propositions according to which a person of interest is, or is not, a contributor to the detected DNA together with, in both propositions, a certain number of other individuals (e.g. claimant and/or unknown person(s)).[122] This type of assessment is fundamentally compatible with the principles of interpretation exposed in Section 2.2.: the assessment informs about the probability of obtaining the DNA profiling results *given* competing propositions regarding the source(s) of the detected DNA. Thus, the output of applying PG in an instant case is an assessment of the probative value of the DNA profiling results with respect to particular propositions. More specifically, these are called sub-source level proposi-

---

[119] Aitken et al., 2010: §3.3.

[120] E.g. Scientific Working Group on DNA Analysis Methods (SWGDAM), 2015; AAFS Standards Board, 2020.

[121] E.g. *United States vs. Gissantaner*, 2019 U.S.Dist.LEXIS 178848, 2019 WL 5692183 (W.D.Mich. Oct. 16, 2019) and *United States v. Gissantaner*, supra note 84.

[122] E.g. Buckleton et al., 2019: 395-396.

tions[123] which, according to the framework exposed in Section 2.4., lead at best to weak interpretability. The reason for this is that such propositions are distant from so-called activity-level propositions that more closely approach the principal issues of interest to fact-finders. Most importantly, it is crucial to understand that a brute PG result is *uninformative* with respect to activity level propositions. Recall that, as mentioned in Section 2.4., an evaluation of the findings with respect to more advanced propositional levels requires specialised knowledge regarding aspects such as transfer and persistence of DNA in the light of the circumstances of the instant case. But Buckleton et al. insist that this is beyond the scope of PG: «Discussion of transfer and persistence [...] has nothing to do with PG».[124] Consumers of expert information should thus carefully distinguish between a PG result and an assessment of the findings for higher-level propositions. While the former, PG, may lead to impressive numbers, it is critically important to resist temptations to let such assessments unduly influence reasoning paths concerned with the latter, higher propositional levels regarding alleged activities.

For illustration, consider a hypothetical case in which A was physically assaulted by a man, resulting in serious injury.[125] Part of the uncontested case circumstances is that A fought with the offender, grabbing his clothing and his skin vigorously. Shortly after the assault, A's nails were swabbed for subsequent DNA analyses. Suppose that the DNA profiling analyses led to a DNA mixture showing a major and a minor contributor corresponding to, respectively, the DNA profiles of A and B. The defendant, B, denies any involvement. Further, B asserts that he had been scratched by A a few days ago during a minor argument, resulting in A reacting hysterically and disproportionally. Therefore, it might not be surprising —B says— to find a DNA profile that corresponds (partially) to his DNA, as is the case with a mixture between A's and B's DNA. In this case, both parties acknowledge that some sort of physical aggression *has* indeed occurred. This, in turn, can justify the assumption that the source of the DNA is not contested. It is the timing and, maybe, the intensity and duration of the quarrel over which there is disagreement. Hence, conducting PG would not directly help with a primary object of litigation except, maybe, with helping to narrow the circle of individuals to investigate. Recall that PG would only provide a result with respect to propositions regarding persons who contributed to the DNA trace, yet the question of source is not contested in this case.

Instead, what is contested, are activities, as is reflected by propositions of the following kind: «B has been assaulted by A» vs. «an unknown man has assaulted A (but A scratched B a few days prior to the assault)». Stated otherwise, the reasoning is directed towards discriminating between competing versions regarding the way by which, and the timing, DNA was deposited under/on A's nails. Assessing the

[123] On the concept of sub-source level propositions, see supra note 86.
[124] Buckleton et al., 2019: 395.
[125] The authors are grateful to Dr. Tacha Hicks for suggesting this case example.

probability of seeing the DNA profiling results given each of these propositions, and the case-specific circumstances, requires expert knowledge (e.g. structured data from experiments under controlled conditions) regarding the transfer and persistence of DNA under circumstances comparable to the alleged activities in the instant case. Such expertise may be scarce, or fragmentary, leading to only limited, if any, capacity to discriminate between the two propositions of interest, not least because the two alleged activities are similar in nature, differing only in timing. Should the scientist be unable to assess the findings with respect to the proposition regarding the alleged activities and circumstances, then the DNA profiling result would need to be considered *uninterpretable* at this propositional level. Further, the scientist should not be allowed then to retreat to reporting results of PG because the output of such an analysis would refer to other propositions (i.e. the source of DNA), however probative the findings may be with respect to such propositions. Indeed, there is no need to adduce evidence for propositions that are not contested. Doing otherwise is prone to let PG output unduly affect inference about a propositional level regarding which the evidence is actually lacking interpretability.

4. CONCLUSIONS

This paper started with the observation that evidence, as we perceive it, is not necessarily what the mainstream view in forensic science claims to be. Real-world evidence, irrespective of its nature, does not come with a built-in, self-explanatory account of its probative force, i.e. with meta-rules of assessment. At the same time, the terms «interpretation» and «interpretability», though commonly used in scientific research and judicial practice, remain largely undefined in the context of expert evidence.

As a main strain of argument, this paper advocated a rule-governed, balanced and logically defensible definition of interpretability, based on principles derived from formal methods of reasoning, called here the principles of interpretation (Section 2.2.). We have equated (un-)interpretability with the forensic expert's (in-)ability to enact minimal principles of logical reasoning applied to expert evidence. Central to this view is that, fundamentally, interpretability is not an inherent or, as it were, an axiomatic property of evidence, but a function of the purpose for which it is adduced, and the specialised knowledge and expertise available to the scientist, necessary to relate evidence meaningfully to the circumstances of the instant case. Being contrastive by design, evidence analysis focuses on providing an assessment of the extent—which may well be nil (remember: interpretability does not necessitate relevance)—to which the evidence is able to discriminate between competing propositions of interest, but abstaining from providing a direct opinion on the propositions and litigated matters that are in the fact-finder's area of competence.

Though it can be considered a substantive property of the relationship between evidence and its analyst, interpretability is actually underspecified and underrecog-

nised in practice, thus rendering evidential processes vulnerable to unwarranted inferential reasoning, possibly detrimental to litigants. Specifically, where interpretability with regard to low-level case aspects, such as the source of evidential material, is carelessly extended to higher-level questions, such as alleged activities, *un*interpretability may undermine or—as we argue in this paper—«litter» the fact-finders' inferential paths. Stated otherwise, where informational gaps due to uninterpretability are misconceived, actively ignored or concealed, evidence is prone to become seemingly probative for issues for which it is actually uninformative.

In all, there is more to «uninterpretability» than merely referring to evidence that does not answer a question of primary interest, i.e. an ultimate issue. Probing for uninterpretability is not the ultimate goal, but a means to an end. It is about ensuring the proper elicitation of the meaning of evidence, i.e. the rules for its use, and the avoidance of conclusions that go beyond the realm of what is logically warranted. Sceptical readers might argue that the insistence on the notion of (un-)interpretability is redundant because triers of fact may be able to proceed responsibly without it. This paper has argued that this would be short-sighted insofar as it presupposes that controlling for (un-)interpretability could be achieved in the absence of any inferential framework. Moreover, as we have shown, legal provisions are vulnerable to, or even tolerant towards, evidence whose probative value cannot be assessed.

BIBLIOGRAPHY

American Academy of Forensic Sciences (AAFS) Standards Board, 2020: *Standard for Validation of Probabilistic Genotyping Systems, ASB Standard 018* (1st ed.), https://www.asbstandardsboard.org (last accessed March 7 2021).

Aitken, C., 1995: *Statistics and the Evaluation of Evidence for Forensic Scientists*, Hoboken (New Jersey): John Wiley & Sons.

Aitken, C., Berger, C., Buckleton, J., Champod, C., Curran, J., Dawid, A., Evett, I., Gill, P., González-Rodríguez, J., Jackson, G., Kloosterman, A., Lovelock, T., Lucy, D., Margot, P., McKenna, L., Meuwly, D., Neumann, C., Daéid, N., Nordgaard, A., Puch-Solis, R., Rasmusson, B., Redmayne, M., Roberts, P., Robertson, B., Roux, C., Sjerps, M., Taroni, F., k Tjin-A-Tsoi, T., Vignaux, G., Willis, S. M. &.Zadora G., 2011: «Expressing evaluative opinions: a position statement», in *Science & Justice*, 51: 1-2.

Aitken, C., Roberts, P. & Jackson, G., 2010: *Fundamental of Probability and Statistical Evidence in Criminal Proceedings*, London: Royal Statistical Society's Working Group on Statistics and the Law.

Aitken, C. & Stoney, D. A. (eds.), 1991: *The Use of Statistics in Forensic Science*, New York, London: Ellis Horwood.

Aronson, J. D., 2007: *Genetic Witness: Science, Law, and Controversy in the Making of DNA Profiling*, New Brunswick (New Jersey), London: Rutgers University Press.

Association of Forensic Science Providers (AFSP), 2009: «Standards for the formulation of forensic science expert opinion», in *Science & Justice*, 49: 161-164.

Biedermann, A., Kotsoglou, K. N., 2018: «Decisional dimensions in expert witness testimony—A structural analysis», in *Frontiers in Psychology (Cognition)*, 9: 2073.

Biedermann, A., Taroni, F. & Thompson, W. C., 2011: «Using graphical probability analysis (Bayes nets) to evaluate a conditional DNA inclusion», in *Law, Probability and Risk*, 10: 89-121.

Bright, J.-A., Taylor, D., McGovern, C., Cooper, S., Russell, L., Abarno, D. & Buckleton, J., 2016: «Developmental validation of STRmixTM, expert software for the interpretation of forensic DNA profiles», in *Forensic Science International: Genetics*, 23: 226-239.

Buckleton, J. S., Bright, J.A., Gittelson, S., Moretti, T. R., Onorato, A. J., Bieber, F. R., Budowle, B. & Taylor, D., 2019: «The probabilistic genotyping software STRmix: utility and evidence for its validity», in *Journal of Forensic Sciences*, 64: 393-405.

Buckleton, J. S. & Triggs, C. M., 2005: «Relatedness and DNA: are we taking it seriously enough?», in *Forensic Science International*, 152: 115-119.

Champod, C., Lennard, C., Margot, P. & Stoilovic, M., 2016: *Fingerprints and other Ridge Skin Impressions* (2nd. ed.), Boca Raton, London, New York: CRC Press.

Coble, M. D. & Bright, J.-A., 2019: «Probabilistic genotyping software: an overview», in *Forensic Science International: Genetics*, 38: 219-224.

Cole, S. A. & Biedermann, A., 2020: «How can a forensic result be a "decision"? A critical analysis of ongoing reforms of forensic reporting formats for federal examiners», in *Houston Law Review*, 57: 551-592.

Cook, R., Evett, I. W., Jackson, G., Jones, P. J. & Lambert, J. A., 1998a: «A hierarchy of propositions: deciding which level to address in casework», in *Science & Justice*, 38: 231-239.

——, 1998b: «A model for case assessment and interpretation», in *Science & Justice*, 38: 151-156.

Dennis, I., 2017: *The Law of Evidence* (6th ed.), London: Sweet & Maxwell Ltd.

Edmond, G., Thompson, M. B. & Tangen, J. M., 2014: «A guide to interpreting forensic testimony: scientific approaches to fingerprint evidence», in *Law, Probability and Risk*, 13: 1-25.

European Network of Forensic Science Institutes (ENFSI), 2015: *ENFSI Guideline for Evaluative Reporting in Forensic Science*, http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf (last accessed July 6 2021).

Evett, I. W., 1991: «Interpretation: a personal odyssey», in Aitken, C., Stoney, D. A. (eds.), *The Use of Statistics in Forensic Science*, New York, London: Ellis Horwood, 9-22.

——, 2009: «Evaluation and professionalism», in *Science & Justice*, 49: 159-160.

Evett, I. W., Berger, C. E. H., Buckleton, J. S., Champod, C. & Jackson, G., 2017: «Finding the way forward for forensic science in the US—A commentary on the PCAST Report», in *Forensic Science International*, 278: 16-23.

Evett, I. W., Gill, P. D., Jackson, G., Whitaker, J. & Champod, C., 2002: «Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks», in *Journal of Forensic Sciences*, 47: 520-530.

Evett, I. W., Jackson, G. & Lambert, J. A., 2000: «More on the hierarchy of propositions: exploring the distinction between explanations and propositions», in *Science & Justice*, 40: 3-10.

Evett, I. W. & Weir, B. S., 1998: *Interpreting DNA Evidence*, Sunderland (Massachusetts): Sinauer Associates Inc.

Fienberg, S. E. (ed.), 1989: *The Evolving Role of Statistical Assessments as Evidence in Courts*, New York: Springer.

Finkelstein, M. O. & Fairley, W. B., 1970: «A Bayesian approach to identification evidence», in *Harvard Law Review*, 83: 489-517.

Friedman, R. D., 2017: *The Elements of Evidence*, St. Paul (Minnesota): West Academic Publishing.

Gastwirth, J. L., 1988: *Statistical Reasoning in Law and Public Policy*, Cambridge (Massachusetts): Academic Press.

Gill, P., 2014: *Misleading DNA Evidence: Reasons for Miscarriages of Justice*, Cambridge (Massachusetts): Academic Press.

Gill. P., Hicks, T., Butler, J. M., Connolly, E., Gusmão, L., Kokshoorn, B., Morling, N., van Oorschot, R. A. H., Parson, W., Prinz, M., Schneider, P. M., Sijenj, T. & Taylor, D., 2018: «DNA Commission of the International Society for Forensic Genetics: assessing the value of forensic biology evidence—Guidelines highlighting the importance of propositions, Part I: evaluation of DNA profiling comparisons given (sub-)source propositions», in *Forensic Science International: Genetics*, 36: 189-202.

——, 2020: «DNA Commission of the International Society for Forensic Genetics: assessing the value of forensic biological evidence—Guidelines highlighting the importance of propositions. Part II: evaluation of biological traces considering activity level propositions», in *Forensic Science International: Genetics*, 44: 102186.

Gittelson, S., Kalafut, T., Myers, S., Taylor, D., Hicks, T., Taroni, F., Evett, I. W., Bright, J.-A. & Buckleton, J., 2016: «A practical guide for the formulation of propositions in the Bayesian approach to DNA evidence interpretation in the adversarial environment», in *Journal of Forensic Sciences*, 61: 186-195.

Heard, B. J., 2008: *Handbook of Firearms and Ballistics* (2nd ed.), Hoboken (New Jersey): John Wiley & Sons.

——, 2013: *Forensic Ballistics in Court, Interpretation and Presentation of Firearms Evidence*, Hoboken (New Jersey): John Wiley & Sons.

Imwinkelried, E. J., 2020: «The admissibility of scientific evidence: exploring the significance of the distinction between foundational validity and validity as applied», in *Syracuse Law Review*, 70: 817-849.

Jackson, G., 2000: «The scientist and the scales of justice», in *Science & Justice*, 40: 81-85.

Jackson, G., Aitken, C. & Roberts, P., 2014: *Case Assessment and Interpretation of Expert Evidence*, London: Royal Statistical Society's Working Group on Statistics and the Law.

Jackson, G. & Biedermann, A., 2019: «"Source" or "Activity". What is the level of issue in a criminal trial?», in *Significance*, 16: 36-39.

Jackson, G. & Jones, P. J., 2009: «Case assessment and interpretation», in Jamieson, A. & Moenssens, A. (eds.), *Wiley Encyclopedia of Forensic Science*, Hoboken (New Jersey): John Wiley & Sons, 1-15.

Kaye, D. H., 1987: «The validity of tests: caveant omnes», in *Jurimetrics Journal*, 27: 349-361.

Koehler, J. J., 2013: «Proficiency tests to estimate error rates in the forensic sciences», in *Law, Probability and Risk*, 12: 89-98.

Kotsoglou, K. N., 2013: «How to become an epistemic engineer: what shifts when we change the standard of proof?», in *Law, Probability and Risk*, 12: 275-298.

——, 2021: «Zombie forensics. The use of polygraph and the integrity of the criminal justice system», in *International Journal of Evidence & Proof*, 25: 16-35

Kotsoglou, K. N. & McCartney, C. 2021: «To the exclusion of all others? DNA profile and transfer mechanics—*R* v *Jones (William Francis)* [2020] EWCA Crim 1021 (03 Aug 2020)», in *The International Journal of Evidence & Proof*, 25: 135-140.

Lander, E. S., 2017: «Response to the ANZFSS council statement on the President's Council of Advisors on Science and Technology Report», in *Australian Journal of Forensic Sciences*, 49: 366-368.

Lempert, R. O., 1977: «Modeling relevance», in *Michigan Law Review*, 75: 1021-1057.

——, 1986: «The new evidence scholarship: analyzing the process of proof», in *Boston University Law Review*, 66: 439-477.

National Commission on Forensic Science, Human Factors Subcommittee, 2015: *Views of the Commission Ensuring That Forensic Analysis is Based Upon Task-Relevant Information*, Approved 12/08/2015, https://www.justice.gov/archives/ncfs/file/ 818196/download (last accessed March 12, 2021).

Parmigiani, G., 2002: *Modeling in Medical Decision Making: A Bayesian Approach*, Hoboken (New Jersey): John Wiley & Sons.

Pearl, J. & Mackenzie, D., 2018: *The Book of Why, The New Science of Cause and Effect*, London: Penguin Books.

President's Council of Advisors on Science and Technology (PCAST), 2016: *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*, Washington, D.C.: Executive Office of the President.

Scientific Working Group on DNA Analysis Methods (SWGDAM), 2015: *Guidelines for the Validation of Probabilistic Genotyping Systems*, https://www.swgdam.org/publications (last accessed March 7 2021).

Schwoeble, A. J. & Exline, D. L., 2000: *Current Methods in Forensic Gunshot Residue Analysis*, Boca Raton, London, New York: CRC Press.
Stephen, J. F., 1948: *Digest of the Law of Evidence* (12th ed.), London: Macmillan.
Stoney, D. A., 1994: «Relaxation of the assumption of relevance and an application to one-trace and two-trace problems», in *Journal of the Forensic Science Society*, 34: 17-21.
Thayer, J. B., 1892: *Select Cases on Evidence at the Common Law*, Cambridge: C. W. Sever.
——, 1898: *A Preliminary Treatise on Evidence at the Common Law*, Boston: Little Brown.
Thompson, W. C., 2018: «How should forensic scientists present source conclusions?», in *Seton Hall Law Review*, 48: 773-813.
Thompson, W. C. & Schumann, E. L., 1987: «Interpretation of statistical evidence in criminal trials: the prosecutor's fallacy and the defense attorney's fallacy», in *Law and Human Behavior*, 11: 167-187.
Tribe, L. H., 1971: «Trial by mathematics: precision and ritual in the legal process», in *Harvard Law Review*, 84: 1329-1393.
Trimpe, M. A., 2011: «The current status of GSR examinations», in *FBI Law Enforcement Bulletin*, May: 24-32.
U.S. Department of Justice, 2020: *Uniform Language for Testimony and Reports for the Forensic Latent Print Discipline* (ULTR, vers. 08.15.20), https://www.justice.gov/olp/page/file/1284786/download (last accessed 05 March 2021).
Williams, M., 1995: *Problems of Knowledge*, Oxford: Oxford University Press.
Wittgenstein, L., 1958: *Preliminary Studies for the "Philosophical Investigations". Generally known as The Blue and Brown Books*, London: Blackwell Publishers Ltd.
Yeroulanos, M. (ed.), 2016: *A Dictionary of Classical Greek Quotations*, London: Bloomsbury.