Authors' post-print:
Biedermann A., Kotsoglou K.N., Scientific imperialism: "The judge made me do it!",
Science & Justice, https://doi.org/10.1016/j.scijus.2025.01.003
Published online 25 January 2025

Scientific imperialism: "The judge made me do it!"

Alex Biedermann,[1,*] Kyriakos N. Kotsoglou[2]

[1] *University of Lausanne, Faculty of Law, Criminal Justice and Public Administration, School of Criminal Justice, 1015 Lausanne–Dorigny (Switzerland)*
[2] *University of Northumbria, School of Law, Newcastle Upon Tyne, NE1 8ST (United Kingdom)*

Abstract

This commentary critically reviews a recently published discussion between Hahn et al. [17] and Berger et al. [3] regarding recommendations for the use of probabilistic genotyping systems in criminal proceedings, in particular the proper understanding for the evidentiary use – across legal systems – of results produced by such systems and the communication of system outputs to the judiciary. We find that the exchange between Hahn et al. [17] and Berger et al. [3] reveals a profound divide between diametrically opposed positions, which is symptomatic of a lack of clarity in some quarters of forensic science about the role of expert witnesses and, in particular, novel forms of so-called *machine-generated evidence* in the legal process. We argue that in order to prevent scientific practices from inappropriately invading judicial territory, lawyers should take a more active role in scrutinising recommendations and position statements published by members of the forensic science community.

We need somehow to value and celebrate scientific knowledge without being dragooned into accepting propaganda which suggests it is the only thing that matters.
Mary Midgley

1 Introduction

Partly unnoticed by some sectors of the judiciary, legal practitioners and the public, forensic science has undergone fundamental changes that have accelerated over the past decade. While the common practice in several traditional branches of forensic science has involved, and continues to involve, direct observation, such as the visual inspection of the features (minutiae) of fingermarks, the striations on a fired bullet, or the details of the strokes of a signature, some areas of forensic science are increasingly relying on machine output with little or no additional input from human examiners. A prime example of this trend is the emergence of powerful computational approaches to interpreting and assigning probative value to the results of forensic DNA profiling analyses, particularly in the case of complex DNA mixture traces. These computational systems, also known as probabilistic genotyping systems (or software; PGS), have reached such a level of sophistication that some parts of the forensic science community have deemed it necessary to issue discipline-specific recommendations instructing forensic examiners on how to handle and report the results generated by computational methods in their casework. In this commentary we will take a closer look at a recent example of such a recommendation, published in the journal *Rechtsmedizin* [16], in terms of its reception by the forensic science community in different legal systems.

*Corresponding author. Email: alex.biedermann@unil.ch

The example of PGS discussed in this paper is just one of many methods, techniques, systems and combinations thereof, some of which (but not PGS) fall under the fashionable terms of artificial intelligence (AI) and machine learning (ML). These produce what we will refer to in this paper as *machine-generated evidence*.[1] Other examples of machine-generated evidence can easily be found in disciplines outside of forensic genetics, such as forensic voice or facial image comparison.

At first glance, one might be inclined to think that the publication of recommendations by forensic geneticists, a community considered by many to be one of the most scientifically rigorous and advanced forensic disciplines, would be a source of trust and beyond reproach. However, the example we examine leads us to a different conclusion. Our analysis will highlight several methodological and conceptual issues that are symptomatic of a number of misinterpretations, misrepresentations and misunderstandings in the discussion of forensic science and its role in continental jurisdictions and beyond.

This paper is structured as follows. Section 2 provides a brief and non-technical introduction to PGS and outlines the disagreements among participants in a recent published exchange on recommendations for the use of PGS in criminal proceedings [3, 17], in particular the proper understanding of the evidentiary use – across legal systems – of the results produced by such systems, and the communication of system outputs to the judiciary. Sections 3 to 5 review and discuss selected claims made by proponents of discipline-specific recommendations for the use of PGS [16, 17], with a focus on source attribution determination (hereafter SAD) as a reporting category. SADs are widely regarded as retrograde in forensic genetics, in contrast to other fields. Discussion and conclusions are presented in section 6.

2 A recent controversy over recommendations for the use of probabilistic genotyping software (PGS) in criminal proceedings

The example of a computational method chosen here is probabilistic genotyping (PG), an approach used to interpret forensic DNA profiling results. According to one definition, PG is "[t]he use of biological modelling (i.e., statistical modelling informed by biological data), statistical theory, computer algorithms, and/or probability distributions to infer genotypes and/or calculate likelihood ratios" [2, §3.7]. A PG *system*, in turn, can be defined as "[s]oftware, or software and hardware, which utilizes a probabilistic genotyping approach to infer genotypes and/or calculate likelihood ratios" [2, §3.8]. We will not go into the details of these definitions, except to add that the likelihood ratio is a reference concept for quantifying the probative value of (scientific) evidence [1]. Broadly speaking, it is a measure of the relative degree of support that certain findings give to a proposition over a stated alternative proposition; for example, propositions regarding the person(s) from whom the detected DNA trace originated. Different PGS exist and are reported to produce different results due to differences in design and operation [e.g., 41], which has led to intense debate among critics, proponents and developers of PGS [e.g., 21, 31, 42]. Courts have also reviewed PGS.[2]

Assigning value or weight to evidence using likelihood ratios has long been performed by forensic geneticists using well-established biostatistical theory. In simple cases, calculations using published formulae (e.g. to support certain types of kinship analysis or source inference for non-complex DNA traces) may require little more than paper and pencil. However, as the sensitivity of forensic DNA profiling technology has improved, analyses can now be extended to smaller and smaller amounts of DNA, as well as degraded DNA. This regularly results in partial DNA profiles and DNA profiles that are complicated by the presence of artefacts (such as drop-in and drop-out events) or by the fact that the material analysed contains DNA from more than one person. Assigning probative value to such complex DNA profiling results requires advanced biostatistical approaches that, without computational support, are beyond the capabilities of most human examiners. This is where modern PGS come in, or should come in. However, as we will show in later sections, the mere availability of such systems is not sufficient to ensure the sound use of PGS results in forensic geneticists' evaluative reports submitted in

[1] This term is borrowed from Nunn [30] and is inspired by earlier work by Roth [33, 34].

[2] E.g. United States v. Gissantaner 990 F.3d 457, 463 (6th Cir. 2021).

legal proceedings.

To illustrate the difficulties involved in the practical implementation of PGS, in particular the understanding of PGS results and their communication to the judiciary by reporting forensic geneticists, we will take a closer look at a revealing exchange of letters between Hahn et al. [17] and Berger et al. recently published in the *Journal of Forensic Sciences* [3] . This exchange relates to an earlier publication [16] in the journal *Rechtsmedizin* in which the "Biostatistical DNA Calculations" project group and the "Stain Commission on the Biostatistical Evaluation of Forensic DNA Analytical Findings with Fully Continuous Models (FCM)" (hereafter referred to as "Project Group and Stain Commission", PGSC) presented their recommendations regarding the use of probabilistic genotyping systems in forensic practice.

Hahn et al. [17] and Berger et al. [3] disagree on the trustworthiness of value of evidence calculations, focusing on the likelihood ratio, using modern PGS, and on the meaning and understanding of such computational results. In particular, Hahn et al. [17] defend the suggestion of the PGSC [16] that PGS results (i.e., likelihood ratios) should be classified into different categories (likelihood ratios apportionments) depending on the magnitude of the PGS output. As a result of this recommendation, some likelihood ratios should be discarded despite reaching values up to 1 million, while others may be reported as categorical source attribution determinations. Referring to the rationale underlying PGS and the principles of performance evaluation for LR computing systems, Berger et al. [3] explain why the PGSC recommendations [16] lead to a distortion and hence a misrepresentation of the meaning of PGS output.

In the remainder of this paper, we will neither go into nor repeat the technical details of Berger et al.'s [3] expertly conducted challenge to Hahn et al.'s [17] premises. Instead, we will address aspects that go beyond the exchange between Hahn et al. [17] and Berger et al. [3]. Specifically, we will further examine Hahn et al.'s [17] reaffirmation and justification of the recommendation that likelihood ratios above a certain threshold imply a source attribution determination (Sections 3 to 5). In particular, we show that Hahn et al.'s [17] response to Berger et al.'s [3] criticism misrepresents the legal framework in which Hahn et al. operate.

3 Sad but true: source attribution determination (SAD)

Our first concern is conceptual. It relates to the way in which Hahn et al. [17] respond to the challenges of source attribution determination. In their recommendations [16], the PGSC reiterate the importance of a threshold, i.e. a likelihood ratio of 30 billion, which was already introduced in 2016 [44]. The PGSC recommend that when this threshold is reached, i.e. what has been described elsewhere as the crossing of the statistical Rubicon [25], the expert can conclude that:

> "from an expert's point of view, there is therefore no reasonable doubt that the characteristics of the trace originate from person X (and, possibly, ...)" [16, p. 5]

Berger et al. [3] disagree with this threshold and the type of inference suggested by Hahn et al. [16] that leads to a categorical conclusion. In support of their position, Berger et al. [3] refer to Stoney's seminal paper "What made us ever think we could individualize using statistics?" [36], which is widely known for describing source attribution determinations as a "ridiculous notion" [36, p. 198] that requires a "leap of faith" [36, p. 198], i.e. a logical leap that is not justified by the underlying methodology. However, this reference to fundamental literature and first principles did not settle the debate.

Hahn et al. [17] respond to Berger et al.'s [3] gentle reminder that SADs a) go far beyond the domain of the expert, and b) thus invade the territory of the fact-finder, thus exhibiting a mindset of *scientific imperialism*, in an evasive manner that raises new problems rather than solving existing ones. We will analyse, examine and discuss selected statements from Hahn et al. [17] below and in the following sections.

To cut to the chase, consider Hahn et al.'s attempt to justify their recommended likelihood ratio threshold that supposedly justifies SADs. Hahn et al. claim that the recommendations were adopted as

a consequence of the role of court experts in German legal proceedings and the expectations that the judiciary has of these experts. In particular, Hahn et al. state:

> "the role of the court expert in Germany includes not only to produce and describe evidence, but also to help the court understand its meaning and the uncertainties associated with it. This is why it is not enough to report [sic] LR only, but experts are obliged to include an 'expert opinion'. Presenting an LR in court is always followed by the question what this LR means in the expert's opinion. This is why a written statement includes the LR itself, the meaning of the LR and then the expert's opinion, which must be stated as such, for example by using the phrase ‹Aus gutachterlicher Sicht... ›"[3] [17, p. 737]

Hahn et al. thus clearly albeit unwarrantedly state that experts can make factual determinations based solely on the data and information available to them, i.e. in a methodologically controlled manner. They imply that the expert in question would not simply express a personal opinion based on the DNA profiling data, but would make a final assessment of the (pen)ultimate issues. Let us take a closer look at this claim.

Hahn et al. claim that the question to be answered is "what this LR means in the expert's opinion" [17, p. 737]. They claim that the proper response to this question involves/justifies a factual (source attribution) determination (SAD). However, this is scientifically (methodologically) and doctrinally unjustified, regardless of what the judiciary says or expects of experts.

Admittedly, as a matter of fact, German judges are increasingly and worryingly deferring to forensic scientists when it comes to decisions involving scientific evidence. This procedurally improper and methodologically unjustified approach has already been noted and criticised by commentators and senior judges [12]. The type of explanation (excuse?) used by Hahn et al. reveals an attitude that is reminiscent not of professionals, but of recipients of (unlawful) orders: "They made me do it; it's not my fault!". This is at odds with the actual procedural architecture of the German criminal process – and other jurisdictions across Europe – which assigns the decision-making prerogative to the fact-finders, be they lay magistrates (Schöffen) or professional judges. Experts are called in to assist, but only in their strictly limited areas of expertise. The nature of the problem of assessing evidence and applying probabilistic data to individual cases lies outside the competence of the scientist. As discussed further in Section 4, expert evidence, according to mainstream doctrine, invokes scientific, technical or other specialised knowledge that is beyond the general competence and ordinary common sense experience of fact-finders.

Strictly speaking, if the question is 'What does the likelihood ratio mean?' in a particular case where the source of the DNA is disputed, or where different activities are alleged, then the answer is (or should be) a definitionally sound statement of nothing other than the notion of likelihood ratio. This statement should be related to the case-specific pair of opposing propositions, as recommended, for example, by the ENFSI Guideline for Evaluative Reporting in Forensic Science (hereinafter the "ENFSI Guideline") [45] and the recommendations of the International Association for Forensic Genetics (ISFG) [14, 15].

Overall, these guidelines state that the expert should take certain propositions for granted (i.e. assume their truth) in order to assess the probability of the findings *given* these propositions and any task-relevant information. On the other hand, it is *not* within the expert's remit to testify as to whether or not the recovered DNA trace is from a particular person. The testimony will therefore be probabilistic in nature. In other words, the expert cannot make factual determinations, including SADs.[4]

As an aside, it is contradictory to select and promote a probabilistic metric for the value of evidence (the likelihood ratio), but then to override its output with a SAD. This is tantamount to stripping the concept of its essence. The idea of using a probabilistic measure of the value of evidence is to clarify how to help confirm or disconfirm selected propositions to the extent that reflects the available evidence, whatever that evidence may be: positive, negative or neutral [20]. Imposing a threshold as an add-on is

[3] Note that "Aus gutachterlicher Sicht" means "from the expert's point of view".

[4] It should be noted that the ENFSI guideline clearly states that forensic "practitioners will not report on matters outside their own area of expertise" [45, §1.3]. If they nevertheless do so, e.g. on request, this is not part of evaluative reporting and therefore cannot fall under a recommendation for scientific practice as envisaged by Hahn et al. [16].

alien to, and not covered by, the underlying theory.[5] In addition, the use of phrases such as "no reasonable doubt" [16, p. 5] has repeatedly been declared inappropriate, even in jurisdictions that are actually quite permissive of SADs. Examples include the views of the National Commission on Forensic Science (in the USA) on the expression "reasonable scientific certainty" [29] (which is merely another way of conveying the notion of "reasonable doubt" [16, p. 5]), as well as the report of the President's Council of Advisors on Science and Technology [32] ("Statements suggesting or implying greater certainty are not scientifically valid and should not be permitted. In particular, courts should never permit scientifically indefensible claims such as: (...) "to a reasonable degree of scientific certainty" (...)" [p. 19]).

It is also worth noting that the assertion that a SAD is a "conclusion based exclusively on the data and information available to them [the experts]" [17, p. 737] is hardly surprising in an era where buzzwords such as "data science", "AI" and "machine learning" can create an anything-goes attitude when it comes to drawing conclusions based on data.[6] The dubious idea behind such assertions is that all wisdom, as well as the necessary and sufficient grounds for categorical conclusions, lies in the scientist's exclusively data-driven system. This idea is dubious because it requires the contestable assumption that no (amount of) evidence to the contrary can shake the scientist's certainty [10]. Anyone remotely familiar with the reality of legal proceedings knows that such an ex ante assumption is seriously flawed. Therefore, and this brings us to our second concern, SADs and their mainstream defence by the forensic community, using elements of "data science" language, lack a scientific basis and misunderstand the architecture of the (German) criminal process. In the following, we will explain why this is the case, starting with a brief description of the legal framework.

4 Experts and the (German) legal framework

The criminal trial is a densely woven, normatively structured decision-making process under uncertainty. Legal orders are informationally permissive but structurally rigid. For example, there are countless ways of committing murder. But every single way of committing the offence shall be punishable by life imprisonment, see e.g. section 211 GPC (German Penal Code). Moreover, as even philosophers of science emphasise, the resolution of individual cases requires the consideration of external values provided by the legal order [27]. The crux of this argument lies in the insight that these values are neither scientific nor unscientific. To think otherwise is to confuse *empirical questions* with *questions of justice*. Any attempt to avoid the thorny subject of values altogether by concentrating only on "pure facts" is not only intellectually dangerous but also dubious. It would give priority to certain ideologically charged values over others without any need for justification [19]. It would also confuse empirical statements of a general nature with questions of justice that deal with unique historical events.

This does not mean that experts should be excluded from the group of potential decision-makers. However, they would not be considered in their capacity as *experts*. Expertise refers to specific areas of knowledge, not to a person as such. DNA analysts are not omnicompetent or omniscient. The ability to extract DNA from biological material or to perform biostatistical calculations does not mean that forensic geneticists are better equipped to apply the standard of proof in criminal trials or to decide whether this or that is "reasonable", such as whether – as Hahn et al. [16, 17] claim – the DNA recovered is that of a particular person of interest. Legal systems rely heavily on this learning principle, and every attempt to axiomatise reasonableness over the past three centuries has consistently failed.

In addition, the German legal order assigns the prerogative of evidence assessment (Beweiswürdigung als ureigene Pflicht des Tatrichters) on the one hand, and decision-making on the other, to the fact-finders (lay magistrates and professional judges). According to the prevailing view in German criminal procedure scholarship, it is the fact-finder and no one else who is responsible for assessing evidence under section 261 GCCP (German Code of Criminal Procedure), arguably the most central norm defining the

[5] The consideration of a probabilistic threshold requires further concepts, such as the extension to Bayesian decision theory [e.g. 40], but this is not covered by the recommendations of the PGSC [16].

[6] For a discussion of the limitations of standard machine learning templates for forensic SAD, see e.g. [4].

German criminal process.[7]

In this context, it is also important to remember that the primary objective of fact-finders is to make case-specific judgments about the actions or inactions of individuals.[8] Fact-finders do not seek to make any kind of *general* statement about an area under investigation. The focus is on the particular defendant and the particular dispute that the fact-finders are called upon to resolve. To treat criminal cases and their local aspects in a generalised (empirical) way, with a rigid and context-agnostic recommendation, as suggested by the PGSC [16],[9] would be to neglect the crucial insight that each criminal case is a unique historical fact.

The same conclusion can be reached by looking at SADs through a scientific lens, using decision theory [e.g., 5]. In short, if a SAD is viewed as a *decision*, which is a descriptively accurate view of what the expert is doing, decision theory tells us that a SAD would *only* be logically justified if the combination of the likelihood ratio and the prior odds on the propositions of interest exceeded the ratio of the losses associated with the two undesirable decision outcomes, i.e. incorrect source attribution and incorrect source non-attribution (i.e. "missed" source attribution). However, as neither the prior probabilities nor the loss function are within the domain of the expert, experts are not in a position to make SADs [9]. Therefore, experts can only make SADs at the cost of breaking with methodological rigour and scientificity [38].

And yet, to return to the legal framework, adherence to methodological principles and refraining from commenting on matters that go far beyond one's area of expertise is a condition of admissibility. The invasion of judicial territory, even if, as suggested by Hahn et al. [17], presiding judges invite DNA geneticists to do so by asking them to give their opinion on the reasonableness of a matter of interest or on the question of source (i.e. by inviting them to make a judgement under uncertainty), violates the procedural architecture of the German criminal process.

It is true that in other disciplines, such as fingermark examination, there is a historically grown practice of reporting SADs, with which the judiciary is widely familiar. However, this does not detract from our arguments above. Firstly, because an "Is" does not imply an "Ought", and secondly, because *even* in fingermark examination there are now systems that output probabilistic expressions of the value of evidence (broadly analogous to the role of PGS in forensic genetics) that are capable of replacing traditional SAD-based reporting schemes [e.g., 7, 39].

We also recognise that many Western legal systems do, in fact, tolerate specialised witnesses giving opinions on (pen)ultimate issues (see e.g. FRE 704(a) in the USA). However, these systems also impose requirements regarding the reliability of the methods and techniques used. In addition, as noted above, various institutional recommendations strongly discourage language suggesting absolute certainty on behalf of examiners [e.g., 29, 32]. Considering this as part of the current (best) state of knowledge, forensic examiners are well advised and better off limiting their statements to what is scientifically justified, i.e. avoiding the exaggerations associated with SADs. Forensic genetics has a long history of adhering to this view, albeit not without exceptions. Initially, the emphasis was on reporting the rarity of features (e.g., in terms of conditional genotype probabilities), with a gradual shift to likelihood ratios. There is nothing inherently original about PGS concerning these concepts that would justify differential treatment of PGS output at the reporting stage.

5 Reasonableness and logicality

Modern legal systems, including Germany's, prohibit overstepping methodological or evidentiary boundaries. Suffice it to note that only recently, experts who opined on the reasonableness of the defendant's

[7] For further analysis, see [23].

[8] We make a digression to the ultimate decision here, because in the case of relevant crime scene material, i.e. material left by the offender [37], a SAD logically amounts to an identification of the offender.

[9] The recommendation of the PGSC is rigid and context-agnostic because it suggests that one can reach a conclusion about a (source) proposition *regardless* of what other evidence there may be in a case. For related discussion, see also the last paragraph of Section 3.

ability to explain the presence of DNA with a corresponding profile at the scene of the crime received a memorable jurisprudential slap on the wrist. Lord Justice Irwin of the Court of Appeal for England and Wales said that it was "unwise" for the experts in the case to have opined on whether it was "realistic to expect anyone to be able to account for the ways in which their DNA may have been transferred by indirect methods" [26]. Non-British readers should note that "unwise" is a typical British understatement, expressing indignation and strong disapproval. The fact that this example deals with an alleged activity (by which DNA was deposited), which is different from the question of source dealt with in the PGSC recommendations [16], does not detract from our argument. In both cases, the expert is giving an opinion on a proposition, not just on the DNA findings, which is not in line with current recommendations for evaluative reporting from, for example, ENFSI [45] and the ISFG [14, 15].

It is helpful to consider this attitude from a historical perspective. The persistent attempt to approach practical questions (of justice) through the use of formal patterns of reasoning is not new. Since the seventeenth century, scientists have persistently attempted to reduce reasonableness to logicality, practical reason to pure reason, and thus to answer moral questions in an axiomatised, mathematical way. More generally, the British philosopher Stephen Toulmin, in his seminal book "The Uses of Argument", exposed how the strong emphasis on formal deductive techniques by seventeenth-century philosophers, especially through logical empiricism, has done serious damage to the (traditionally synonymous) notions of "rationality" and "reasonableness" in recent centuries. According to Toulmin, this shift – i.e. the mapping of rationality (and hence reasonableness) onto logicality – "did an injury to our commonsense ways of thought" [43, p. 216 et passim] and led to a substantial loss of legitimacy for established decision-making processes [24]. Any model of decision making that does not have or need a formal logical structure, such as the jurisprudential model, is directly challenged by mathematical methods of proof.

At the same time, contemporary forensic scientists are at pains to point out that the idea of ascertaining the impossibility (i.e. a probability of zero) of a single event, such as the occurrence by chance of selected analytical features observed on a crime scene trace, in an abstract sense, is nothing short of "ridiculous" [36, p. 198]. As noted above, courts deal with particular defendants and particular legal disputes. While singularities pose serious challenges to scientific theories, the idea that the concept of generalisation could lead to case-specific decision making is seriously undermined by the infinite number of cases and their complexity.

Indeed, as L.H. Hoffmann points out, "the slightest movement of the kaleidoscope of facts creates a new pattern which must be examined afresh" [18, p. 204]. Thus, formal methods of reasoning cannot solve practical problems in a vacuum, but can at best help the fact-finder to make an informed decision. The authors of Hahn et al. [17], most of whom belong to police organisations that by default lean heavily toward the investigative side, thus provide us with an example of the difficulties and pitfalls associated with extending analytical results for use in investigative procedures to evaluative reporting for purposes of evidence and proof at more advanced stages of the criminal justice process.

Applying these considerations to the current debate leads to a number of conclusions. Hahn et al.'s assertion that a SAD is "what [a] LR means in the expert's opinion" [17, p. 737] is misleading and a misrepresentation of the probabilistic concept of a likelihood ratio,[10] especially when it is promoted as a set of recommendations across the discipline, as in Ulbrich et al. [44] and Hahn et al. [16]. Both the procedural architecture and the requirement of scientificity require experts to be transparent about the limits of science. The natural consequence of this should be to abandon attempts to shape or anticipate questions of justice, let alone to make SADs, which by definition go beyond the limits of science.[11]

Furthermore, the assertion that a SAD "obviously still leaves room for the court to perform their own evaluation" and that the "[c]ourt may or may not adopt the expert's view" [16, p. 737] is wishful thinking, to say the least, and amounts to making unwarranted assumptions about the ability of defendants to challenge SADs. Putting the ball in the defendant's court is tantamount to admitting methodological breakdown. The question is not whether the judge can disagree with the expert. The critical point is that

[10] A similar point is made in [8].

[11] For an example of such a call in the area of fingerprint examination, see, for example, Champod et al. [7].

expert witnesses – formally a category of evidence (Beweismittel) under the GCCP – are only allowed to give an opinion on technical matters. Defence counsel may not be sufficiently informed about or attentive to these matters. Judges may also fail to understand the division of labour between experts and fact-finders, and find it difficult to understand that DNA analysts cannot expound on everything but a small area of human knowledge. In addition, a SAD can easily push thinking too far in one direction, making it even harder for non-experts to build an argument to justify deviating from the inferential path laid out by the expert's SAD. The fact remains that the recommendations of Hahn et al. [16] represent an unwelcome intrusion into the domain of the judiciary; a violation of its decision-making prerogative [24]. An opinion on subject *A* by an expert who specialises in subject *B* tends to affect one of the integral parts of a trial: its fairness (article 6(1) ECHR).

The German legal system empowers judges to assess the evidence – where they can do so without assistance. And judges do *not* need help when it comes to making decisions under uncertainty, including SADs, based on *all* available evidence, not DNA results in isolation. Hahn et al. write that "[t]he expert, in fact, is part of the court and not a ‹witness›" [17, p. 736], which is misleading to say the least. First, this misrepresents the law. Expert witnesses are classified as evidence. The German legal system deals with complexity by introducing expert witness testimony (as defined in sections 72–85 of the GCCP) as a type of evidence alongside witnesses, documents, etc. According to section 72 of the GCCP, "the provisions of chapter 6 concerning witnesses apply accordingly to expert witnesses, unless otherwise provided". There is no such thing as res ipsa loquitur ("the evidence speaks for itself") in law, or in life in general, and the (presiding) judge needs to remind expert witnesses that they are only part of the evidence, an auxiliary subsystem brought in for specific purposes. Experts *assist* fact-finders in their exclusive role of assessing evidence and in their task of making informed decisions under uncertainty. It is, therefore, not insignificant that German scholars unanimously describe experts as *helpers* of the court [47]. The term 'help' carries the conceptual and doctrinal weight in this context because judges are authorised by the legal order to give specific instructions to experts, i.e. to order them to answer particular questions instead of simply giving their opinions (section 78 GCCP). This is, of course, not an idiosyncratic feature of German law, but seems to be a common structure in modern legal orders. More specifically, the expert witness is assigned the role of a helper, while fact-finders have to make up their own mind about the probative value of the evidence.

The apparent overlap of roles dissolves when we recognise that evidence evaluation is not a single methodological package. On the contrary, different actors will perform different tasks: experts will inform fact-finders about task-relevant data and the extent to which outcomes can be observed if one version of the disputed event(s) is true rather than an alternative version. However, experts will refrain from commenting directly on these competing versions of the event(s) in question.

Hahn et al. [17] stress the importance of avoiding the prosecutor's fallacy, but this misses the point. The discussion about logical coherence (and fallacies) ended about 20 years ago, and the hatchet remains buried, even though – in practice – fallacies continue to be committed on a daily basis. The challenge today is not only to avoid logical fallacies (this is part and parcel of reasoning), but also to enable smooth communication between experts and fact-finders, and to confine experts to their specific domain. This is very different from a logical fallacy. It is about rules of authorisation and division of labour. In this context, the term "expert witness fallacy" [24, 25] has been introduced to describe the usurpation of the jury's domain. The criminal justice system is an advanced institutional tool designed for specific types of social activity, not a primitive scientific instrument. Nor is it a cheap substitute or a shoddy process for replacing popular validation with expert routines [24, 28].

The proposal to incorporate scientific approaches into decision-making processes, particularly in criminal adjudication, does not come from science *per se*, but rather from a group of individuals who hold the "somewhat wild" [28, p. 57] idea, as the philosopher Mary Midgley puts it, that (forensic) science occupies the centre stage [24] and that it alone should drive the resolution of practical questions, including questions of justice.[12] The separation of practical and theoretical reason naturally touches on the deep

[12]It is not surprising, therefore, that some commentators have expressed frustration at the widespread tendency to believe

professional interests of those who believe, and want others to believe, that their status as experts entitles them to comment on matters outside their field of expertise [22]. This false sense of omnipotence has led to an unwarranted and counterproductive backlash against (legitimate) forensic research, as well as a misguided attempt to limit legal decisions to empirical findings (conclusions) that appear to be supported by science.

The Austrian-British philosopher Ludwig Wittgenstein made a similar diagnosis some time ago. He observed that our "craving for generality" [46, p. 17] is a synonym for "the contemptuous attitude towards the particular case" [46, p. 18]. Like philosophers – Wittgenstein's target of criticism – legal scholars see the method of science and feel "irresistibly tempted to ask and answer [questions] in the way science does" [46, p. 18], i.e. with an aspiration to generality. It is precisely this "craving for generality", Wittgenstein remarks, this thirst for numerical analysis, that becomes "the real source of metaphysics" and leads us "into complete darkness" [46, p. 18].

The invasive tendency of forensic scientists – a tendency often welcomed with relief by some fact-finders, as it frees them from the need to understand complex issues and allows them to maintain the blur of accountability – contradicts both the procedural architecture of modern legal orders, including the prerogative of decision-making, and the structure of fundamental concepts [24]. The idea that a rigid, scientifically derived (i.e. general) proposition (here: a numerical threshold) could legitimise the attribution of *relevant*[13] items of evidence to specific individuals, and thereby designate the perpetrator and guarantee the factual and normative rectitude of a criminal verdict, is a fallacy. It goes beyond what the underlying technique can rationally justify in terms of inferential stages [5, 6] and represents a disruption of the established mechanisms for attributing criminal liability. Ultimately, this *methodological imperialism* violates the procedural architecture of the legal order as an autonomous normative system (Eigengesetzlichkeit des Rechts) [24], and leads to the "McDonaldization" of criminal justice systems [13].

6 Discussion and conclusions

It is worth remembering that the interpretation of DNA evidence has a long tradition of serving as a beacon of light in the dark chambers of forensic disciplines. Although DNA evidence has gone through, and continues to go through, difficult periods, it has been instrumental in establishing logical and balanced evaluation and reporting practices based on the likelihood ratio, a reference concept for characterising the meaning and strength of DNA profiling results. For this reason, Saks and Koehler, in their seminal paper published in *Science*, described forensic DNA evidence as "a model for a scientifically defensible approach to questions of shared identity" [35, p. 892]. In the light of our brief analysis here, recommendations suggesting that forensic scientists could provide SADs [e.g., 16, 44] are contrary to these axiomatic principles of modern forensic science.

The discussion in this paper has highlighted the deeper problem that some parts of forensic science do not have a clear view of their role in the legal process and of specific frontier topics such as the handling and communication of *machine-generated evidence* in legal contexts. We therefore see scope for lawyers to take a more active role in scrutinising recommendations and position statements published by members of the forensic science community. This could help to prevent scientific practices from inappropriately invading judicial territory. This is all the more important now that forensic science has entered an era in which value of evidence computations can be made with unprecedented speed and complexity. PGS are prime examples of this development. In a sense, validated PGS and similar likelihood ratio calculation systems in other fields, such as forensic voice comparison and fingerprint analysis, reflect and implement the most sophisticated and well-studied procedures used by scientists to quantify uncertainty when assessing the value of evidence. Nevertheless, the results of such procedures

that a case can be solved on the basis of a forensic scientist's findings alone, referring to this tendency as the "Sherlock Holmes syndrome" [11].

[13] Following Stoney [37], an item of evidence is considered relevant here if it has a connection with the offender, i.e. an item or trace left by the offender. In practice, there is often uncertainty about the relevance of (items of) evidence.

are still inevitably affected by our imperfect understanding of the problem domain and, thus, by various sources of uncertainty. Therefore, deliberately suppressing such uncertainty – i.e. by a SAD, as suggested by Hahn et al. [16, 17] – amounts to emptying a supposedly rational procedure of its meaning. Of course, it is possible to draw stronger conclusions than those produced by current computational methods, but this would require additional supporting evidence. At present, however, forensic scientists cannot provide such additional evidence. And even if such evidence did exist, it would still need to be coherently combined with the existing body of evidence, which is far from trivial.

More generally, the implicit claim of Hahn et al. [17] that the field of forensic genetics is inherently different from other forensic disciplines, where forensic practitioners obediently quantify and communicate uncertainty, does not meet the normative requirements, both methodological and procedural, that forensic fields are supposed to meet. The practice of SADs amounts to a questionable pretence that ultimately conveys the message "trust me, I'm an expert", but without being able to justify this stance sufficiently. As important as the parental reason of last resort ("because I say so") may be in certain circumstances, it is anything but helpful in the institutional setting of liberal democratic societies. With a conclusion that stands on its own, we are simply resting our case on a brute assertion, and we are dealing with the logic of force rather than the force of logic. The explicit excuse that "the judge made me do it!" does not solve these problems either.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Aitken, C. G. G., F. Taroni, and S. Bozza (2020). *Statistics and the Evaluation of Evidence for Forensic Scientists* (Third ed.). Chichester: John Wiley & Sons.

[2] American National Standards Institute/Academy Standards Board (2020). ANSI/ASB Standard 018: Standard for Validation of Probabilistic Genotyping Systems, First Edition 2020. www.asbstandardsboard.org.

[3] Berger, C. E. H., M. Kruijver, T. Hicks, C. Champod, D. Taylor, and J. Buckleton (2024). Commentary on: Hahn M, Anslinger K, Eckert M, Fimmers R, Grethe S, Hohoff C, et al. [Joint recommendations of the project group "Biostatistical DNA Calculations" and the Trace Commission on the Biostatistical Evaluation of Forensic DNA Analytical Findings with Fully Continuous Models (FCM)]. Rechtsmedizin. 2023; 33(1):3..12. doi: 10.1007/s00194-022-00599-5. *Journal of Forensic Sciences 69*, 730–735.

[4] Biedermann, A. (2022). The strange persistence of (source) "identification" claims in forensic literature through descriptivism, diagnosticism and machinism. *Forensic Science International: Synergy 4*, 100222.

[5] Biedermann, A., S. Bozza, and F. Taroni (2008). Decision theoretic properties of forensic identification: underlying logic and argumentative implications. *Forensic Science International 177*, 120–132.

[6] Biedermann, A., S. Bozza, F. Taroni, and C. Aitken (2016). Reframing the debate: a question of probability, not of likelihood ratio. *Science & Justice, Special Issue on Measuring and Reporting the Precision of Forensic Likelihood Ratios 56*, 392–396.

[7] Champod, C., C. Lennard, P. Margot, and M. Stoilovic (2016). *Fingerprints and Other Ridge Skin Impressions* (Second ed.). Boca Raton: CRC Press.

[8] Cole, S. A. (2008). The 'opinionization' of fingerprint evidence. *BioSocieties 3*, 105–113.

[9] Cole, S. A. and A. Biedermann (2020). How can a forensic result be a "decision"? A critical analysis of ongoing reforms of forensic reporting formats for federal examiners. *Houston Law Review 57*, 551–592.

[10] Evett, I. W. and B. S. Weir (1998). *Interpreting DNA Evidence*. Sunderland: Sinauer Associates Inc.

[11] Faigman, D. L., N. Scurich, and T. D. Albright (2022). The field of firearms forensics is flawed. *Scientific American* (May 25).

[12] Fischer, T. (2008). Aussagewahrheit und Glaubhaftigkeitsbeurteilung. Anmerkungen zum Beweiswert von Glaubhaftigkeitsgutachten. In H. Schöch, H. Satzger, G. Schäfer, A. Ignor, and C. Knauer (Eds.), *Strafverteidigung, Revision und die gesamten Strafrechtswissenschaften : Festschrift für Gunter Widmaier zum 70. Geburtstag*, pp. 199–222. Köln: Heymanns.

[13] George, R. (2000). *The McDonaldization of Society*. London: Sage.

[14] Gill, P., T. Hicks, J. M. Butler, E. Connolly, L. Gusmão, B. Kokshoorn, N. Morling, R. van Oorschot, W. Parson, M. Prinz, P. M. Schneider, T. Sijen, and D. Taylor (2018). DNA Commission of the International Society for Forensic Genetics: Assessing the value of forensic biological evidence – Guidelines highlighting the importance of propositions. Part I: Evaluation of DNA profiling comparisons given (sub-) source propositions. *Forensic Science International: Genetics 36*, 189–202.

[15] Gill, P., T. Hicks, J. M. Butler, E. Connolly, L. Gusmão, B. Kokshoorn, N. Morling, R. van Oorschot, W. Parson, M. Prinz, P. M. Schneider, T. Sijen, and D. Taylor (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. *Forensic Science International: Genetics 44*, 1–13.

[16] Hahn, M., K. Anslinger, M. Eckert, R. Fimmers, S. Grethe, C. Hohoff, S. Kranz, C. Leuker, C. Oppelt, S. Razbin, T. Rothämel, H. Schneider, M. Templin, M. Vennemann, A. Wächter, V. Weirich, P. Zimmermann, and P. M. Scheider (2023). Joint recommendations of the project group "Biostatistical DNA Calculations" and the Stain Commission on the Biostatistical Evaluation of Forensic DNA Analytical Findings with Fully Continuous Models (FCM). *Rechtsmedizin 33*, 3–12.

[17] Hahn, M., C. Courts, M. Eckert, R. Fimmers, S. Grethe, S. Kranz, C. Leuker, C. Oppelt, S. Razbin, M. Templin, M. Vennemann, P. Zimmermann, and K. Anslinger (2024). Authors' response. *Journal of Forensic Sciences 69*, 736–738.

[18] Hoffmann, L. H. (1974). Similar facts after boardman. *Law Quarterly Review 91*, 193–206.

[19] Hughes, A. L. (2012). The folly of scientism. *The New Atlantis 37*, 32–50.

[20] Jeffrey, R. C. (1975). Probability and falsification: Critique of the popper program. *Synthese 30*, 95–117.

[21] Kalafut, T., J. M. Curran, M. D. Coble, and J. Buckleton (2024). Commentary on: Thompson WC. Uncertainty in probabilistic genotyping of low template DNA: a case study comparing STRmix™ and TrueAllele™. J Forensic Sci. 2023;68 (3):1049–63. doi: 10.1111/1556-4029.15225. *Journal of Forensic Sciences 69*, 371–377.

[22] Kelsen, H. (2008). *Einleitung in die rechtswissenschaftliche Problematik (Studienausgabe der 1. Auflage 1934)*. Tübingen: Mohr Siebeck.

[23] Kogsoglou, K. N. (2015). *Forensische Erkenntnistheorie: Der Inferentielle Kontextualismus und die Funktion der kontextrelevanten Zweifel im Strafverfahren – Zugleich eine analytische Perspektive zur Sachverhaltsfeststellungsdogmatik (Schriften zur Rechtstheorie)*. Berlin: Duncker & Humblot.

[24] Kotsoglou, K. N. and A. Biedermann (2022). Inroads into the ultimate issue rule? structural elements of communication between experts and fact finders. *The Journal of Criminal Law 86*, 223–240.

[25] Kotsoglou, K. N., A. Biedermann, and J. Vuille (2020). DNA und Beweiswürdigung – Der statistische Rubikon und die Dogmatik der Identitätsfeststellung. *Zeitschrift für die gesamte Straftrechtswissenschaft 132*, 891–937.

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

[27] Kuhn, T. S. (1996). *The Structure of Scientific Revolutions* (3rd ed.). Chicago: The University of Chicago Press.

[28] Midgley, M. (2001). *Science and Poetry*. New York: Routledge.

[29] National Commission on Forensic Science (2016). Views of the Commission: Use of the Term "Reasonable Scientific Certainty", Available at: https://www.justice.gov/archives/ncfs/page/file/1004481/ (last accessed september 16th 2024).

[30] Nunn, A. (2019–2020). Machine-generated evidence. *TheSciTechLawyer 16*, 4–7.

[31] Perlin, M. W., N. Butt, and M. R. Wilson (in press). Commentary on: Thompson WC. Uncertainty in probabilistic genotyping of low template DNA: A case study comparing STRmix™ and TrueAllele®. J Forensic Sci. 2023;68(3):1049–63. *Journal of Forensic Sciences*.

[32] 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.

[33] Roth, A. (2016). Trial by machine. *The Georgetown Law Journal 104*, 1245–1305.

[34] Roth, A. (2017). Machine testimony. *The Yale Law Journal 126*, 1972–2053.

[35] Saks, M. J. and J. J. Koehler (2005). The coming paradigm shift in forensic identification science. *Science 309*, 892–895.

[36] Stoney, D. A. (1991). What made us ever think we could individualize using statistics? *Journal of the Forensic Science Society 31*, 197–199.

[37] Stoney, D. A. (1994). Relaxation of the assumption of relevance and an application to one-trace and two-trace problems. *Journal of the Forensic Science Society 34*, 17–21.

[38] Stoney, D. A. (2012). Discussion on the paper by Neumann, Evett and Skerrett. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 175*, 399–400.

[39] Swofford, H. J., A. J. Koertner, F. Zemp, M. Ausdemore, A. Liu, and M. J. Salyards (2018). A method for the statistical interpretation of friction ridge skin impression evidence: Method development and validation. *Forensic Science International 287*, 113–126.

[40] Taroni, F., S. Bozza, A. Biedermann, G. Garbolino, and C. G. G. Aitken (2010). *Data Analysis in Forensic Science: a Bayesian Decision Perspective*. Statistics in Practice. Chichester: John Wiley & Sons.

[41] Thompson, W. C. (2023). Uncertainty in probabilistic genotyping of low template DNA: A case study comparing STRMix^{TM} and TrueAllele^{TM}. *Journal of Forensic Sciences 68*, 1049–1063.

[42] Thompson, W. C. (2024). Authors' response. *Journal of Forensic Sciences 69*, 378–381.

[43] Toulmin, S. (2001). *Return to Reason*. Cambridge, Massachusetts: Harvard University Press.

[44] Ulbrich, W., K. Anslinger, G. Bässler, M. Eckert, R. Fimmers, C. Hohoff, M. Kraft, C. Leuker, G. Molsberger, U. Pich, S. Razbin, H. Schneider, M. Templin, A. Wächter, V. Weirich, H. Zierdt, and P. M. Schneider (2016). Gemeinsame Empfehlungen der Projektgruppe "Biostatistische DNA-Berechnungen" und der Spurenkommission zur biostatistischen Bewertung von DNA analytischen Befunden. *Rechtsmedizin 26*, 291–298.

[45] Willis, S., L. McKenna, S. McDermott, G. O'Donell, A. Barrett, B. Rasmusson, A. Nordgaard, C. Berger, M. Sjerps, J. Lucena-Molina, G. Zadora, C. Aitken, T. Lovelock, L. Lunt, C. Champod, A. Biedermann, T. Hicks, and F. Taroni (2015). ENFSI Guideline for Evaluative Reporting in Forensic Science, Strengthening the Evaluation of Forensic Results Across Europe (STEOFRAE). Dublin.

[46] Wittgenstein, L. (1965). *The Blue and Brown Books*. New York: Harper & Row.

[47] Wohlers, W. (2016). Die Fragestellung bei aussagepsychologischen Gutachten als Instrument zur Abgrenzung der Aufgabenbereiche von Strafjustiz und Sachverständigen. In S. Bernard, R. Studer, M. Graf, E. Habermeyer, M. Heer, A. Joset, H.-L. Kröber, S. Lau, N. Oberholzer, and W. Wohlers (Eds.), *Feststellung des Sachverhalts im Zusammenhang mit der Begutachtung*, pp. 109–117. Bern: Stämpfli.