Advancing a paradigm shift in evaluation of forensic evidence: The rise of forensic data science

Author and affiliations:

Geoffrey Stewart Morrison [1,2,*]

[1] Forensic Data Science Laboratory, Aston University, Birmingham, UK

[2] Forensic Evaluation Ltd, Birmingham, UK

*Corresponding author: G.S. Morrison, e-mail: geoff-morrison@forensic-evaluation.net

ORCID:

Geoffrey Stewart Morrison 0000-0001-8608-8207

Disclaimer:

All opinions expressed in the present paper are those of the author, and, unless explicitly stated otherwise, should not be construed as representing the policies or positions of any organizations with which the author is associated.

Declaration of competing interest:

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

Acknowledgements:

This research was supported by Research England’s Expanding Excellence in England Fund as part of funding for the Aston Institute for Forensic Linguistics 2019–2023.

I dedicate this paper in memory of Prof Terrance M Nearey, my former PhD supervisor, who taught me how to be a scientist. Terry passed away on December 18, 2021.

Advancing a paradigm shift in evaluation of forensic evidence: The rise of forensic data science

Abstract

Widespread practice across the majority of branches of forensic science uses analytical methods based on human perception, and interpretive methods based on subjective judgement. These methods are non-transparent and are susceptible to cognitive bias, interpretation is often logically flawed, and forensic-evaluation systems are often not empirically validated. I describe a paradigm shift in which existing methods are replaced by methods based on relevant data, quantitative measurements, and statistical models; methods that are transparent and reproducible, are intrinsically resistant to cognitive bias, use the logically correct framework for interpretation of evidence (the likelihood-ratio framework), and are empirically validated under casework conditions.

Keywords

forensic science; forensic data science; likelihood ratio; paradigm shift; validation

1 Introduction

The present paper is a written version of a keynote presentation given at the European Academy of Forensic Science 2022 conference. It discusses an ongoing paradigm shift in evaluation of forensic evidence. It describes:

- the current state of affairs (*staus quo*)
- the new paradigm (*quo vadis?*)

- obstacles to the advancement of the paradigm shift (*impedimenta*)
- a strategy to advance the paradigm shift (*via progredi*)

2 A paradigm shift in evaluation of forensic evidence

2.1 Status quo

Curran [1]:

> Is forensic science the last bastion of resistance against statistics?

UK House of Lords Science and Technology Committee (HoL) [2]:

> In regard to pattern comparison methods, ... "the comparison of fingerprints, toolmarks, footwear, tire marks and ballistics" [are] "spot-the-difference" techniques in which "there is little, if any, robust science involved in the analytical or comparative processes used and as a consequence there have been questions raised around the reproducibility, repeatability, accuracy and error rates of such analysis." (§155)

In forensic science, the process of *evaluation of strength of evidence* consists of: *analysis*, i.e., extraction of information from items of interest (the evidence);[1] and *interpretation*, i.e., drawing inferences with respect to the meaning of the information extracted by the analysis. Items of interest may be, for example:

- a fingermark of questioned source recovered from a crime scene and a fingerprint

[1] In the forensic-inference-and-statistics literature "evidence" is the term commonly used to refer to the items of interest (i.e., the input to the analysis) or to the information output by the analysis (i.e., the input to the interpretation). Usage is somewhat fluid, but, either way, this is evidence from the perspective of the forensic practitioner. From the perspective of the court, evidence is the oral testimony or written submission of the forensic practitioner. In wider forensic-science literature, the term "trace" is often used to refer to items of interest (Roux et al. [3]).

collected from a known individual

- a recording of a speaker of questioned identity on an intercepted telephone call and a recording of a police interview with a speaker of known identity

- a fired cartridge case recovered from a crime scene and cartridge cases fired in a forensic laboratory from a gun found in the possession of a suspected shooter

Forensic practitioners conduct evaluations in order to assist legal-decision makers to make decisions with respect to questions of legal concern such as:[2]

- Do the fingermark and fingerprint originate from the same finger?

- Is the speaker of questioned identity on the intercepted recording the same as the speaker of known identity?

- Was the cartridge case recovered from the crime scene fired from the suspect's gun?

Currently, across the majority of branches of forensic science, widespread practice is that analysis is conducted using *human perception*, and interpretation is conducted using *subjective judgement*. Even in branches of forensic science in which analysis is conducted using instrumental measurement, interpretation is commonly based on subjective judgement, e.g., by eyeballing graphical representations of the measured values. Human-perception-based analysis methods and subjective-judgement-based interpretation methods are non-transparent and are susceptible to cognitive bias.

Currently, across the majority of branches of forensic science, even branches of forensic science in which interpretation is conducted using statistical models, interpretation of evidence is often logically flawed, and forensic-evaluation systems (the end-to-end combination of analysis and interpretation methods) are often not

[2] The present paper restricts its scope to the problem of source attribution.

empirically validated or not adequately empirically validated.[3]

2.2 Quo vadis?

2.2.1 Introduction

Saks & Koehler [4]:

> we envision a paradigm shift in the traditional forensic identification sciences in which untested assumptions and semi-informed guesswork are replaced by a sound scientific foundation and justifiable protocols. Although obstacles exist both inside and outside forensic science, the time is ripe for the traditional forensic sciences to replace antiquated assumptions of uniqueness and perfection with a more defensible empirical and probabilistic foundation. (p. 895)

US President’s Council of Advisors on Science and Technology (PCAST [5]):

> neither experience, nor judgment, nor good professional practice ... can substitute for actual evidence of foundational validity and reliability. The frequency with which a particular pattern or set of features will be observed in different samples, which is an essential element in drawing conclusions, is not a matter of “judgment.” It is an empirical matter for which only empirical evidence is relevant. (p. 6)

> Objective methods are, in general, preferable to subjective methods. Analyses that depend on human judgment (rather than a quantitative measure ...) are obviously more susceptible to human error, bias, and performance variability across examiners. In contrast, objective, quantified methods tend to yield greater accuracy, repeatability and reliability, including reducing variation in results among examiners. Subjective methods can evolve into or be replaced by objective

[3] Claims made in §2.1 are supported by details and references provided in §2.2 and in §2.3.

methods. (p. 47)

A *paradigm shift* in evaluation of forensic evidence is ongoing in which methods based on human perception and subjective judgement are being replaced by methods based on *relevant data*, *quantitative measurements*, and *statistical models / machine-learning algorithms*; methods that:

- are *transparent and reproducible* (§2.2.2);
- are *intrinsically resistant to cognitive bias* (§2.2.3);
- use the *logically correct framework for interpretation of evidence* (the *likelihood-ratio framework*) (§2.2.4); and
- are *empirically validated under casework conditions* (§2.2.5).

I address each of these elements of the new paradigm in the following four subsections. They are, in turn, followed by a subsection which discusses the applicability to this paradigm shift of Kuhn's [6] description of a paradigm shift, and the implications thereof (§2.2.6).

2.2.2 Transparency and reproducibility

Methods dependent on human perception and subjective judgement are intrinsically non-transparent and therefore not reproducible by others. Human introspection is often mistaken, hence a forensic practitioner's explanation of how they reached their conclusion may not reflect how they actually reached that conclusion (Edmond et al. [7]).

In contrast, methods based on data, quantitative measurement, and statistical models are transparent and reproducible: measurement (feature-extraction) and statistical-modelling / machine-learning methods can be described in detail, and data and software tools can potentially be shared with others.

2.2.3 Cognitive bias

There has been a great deal of concern about cognitive bias in forensic science (National Research Council [8]; Expert Working Group on Human Factors in Latent Print Analysis, EWG [9]; Found [10]; Stoel et al. [11]; PCAST [5]; Edmond et al. [7]; Cooper & Meterko [12]; Expert Working Group on Human Factors in Handwriting Examination [13]; Spellman et al. [14]). Cognitive bias is subconscious bias, it cannot be controlled by strength of will.

Forensic practitioners are susceptible to cognitive bias when making perceptual observations: their degree of belief in the probability that a hypothesis is true can affect their analysis of the evidence. Since the output of the analysis is the input to the interpretation, bias in the former affects the latter. Forensic practitioners are susceptible to cognitive bias when they are making subjective judgements and are exposed to information that could influence their degree of belief in the probability that a hypothesis is true but that would not logically affect the probability of obtaining the evidence conditional on whether the hypothesis were true.

Some potentially biasing information is task-irrelevant and should be withheld from practitioners, but some potentially biasing information is task-relevant and practitioners employing human-perception and subjective-judgement methods will need to be exposed to it at some point in the evaluation process, e.g., practitioners who visually compare known-source fingerprints and questioned-source fingermarks must be exposed to both, but exposure to a higher-quality print may bias their analysis of ambiguous details in a lower-quality mark.

Systems in which the strength-of-evidence conclusion is directly the result of subjective judgement are particularly susceptible to cognitive bias.

Systems based on quantitative measurements and statistical models require subjective judgements on decisions such as whether the data used for training the system and the data used for validating the system are sufficiently representative of the relevant population for the case and sufficiently reflective of the conditions of the items of interest in the case so that the output of the system will be a meaningful answer to the question posed in the case and so that the results of the validation will provide a meaningful indication of the performance of the systems under the conditions of the case. These decisions, however, are made at the beginning of the process before the practitioner has analyzed the items of interest, hence the practitioner cannot know what effect these decisions will have on the strength-of-evidence conclusion. The remainder of the evaluation process is automated, hence not susceptible to cognitive bias.

2.2.4 Likelihood-ratio framework

In current practice, interpretation of evidence is often logically flawed, e.g., it is based on the uniqueness or the individualization fallacy (Saks & Koehler [15]; Cole [16], [17]), and conclusions are often expressed categorically, e.g., “identification”, “inconclusive”, “exclusion” (i.e., posterior probability of 1 or 0 with respect to the same-source hypothesis, with “inconclusive” meaning no conclusion rather than an intermediate probability), or using some form of uncalibrated verbal posterior-probability scale, e.g., “identification”, “probable identification”, “inconclusive”, “probable exclusion”, “exclusion”. Jackson [18] and Kaye [19] review these and other commonly used but logically flawed conclusions.

In contrast, the likelihood-ratio framework is advocated as the logically correct framework for evaluation of evidence by the vast majority of experts in forensic inference and statistics (including: Aitken et al. [20]; Morrison et al. [21]; Morrison et al. [22]; with 31, 19 and 20 authors and supporters respectively), and by key organizations including:

- Association of Forensic Science Providers of the United Kingdom and of the Republic of Ireland [23]

- Royal Statistical Society (Aitken et al. [24])
- European Network of Forensic Science Institutes (Willis et al. [25])
- National Institute of Forensic Science of the Australia New Zealand Policing Advisory Agency (Ballantyne et al. [26])
- American Statistical Association (Kafadar et al. [27])
- Forensic Science Regulator for England & Wales (FSR) [28]

The likelihood-ratio framework requires assessment of:

- the probability of obtaining the evidence if one hypothesis were true

  versus

- the probability of obtaining the evidence if an alternative hypothesis were true

The two hypotheses must be mutually exclusive. One hypothesis should represent the position of the prosecution in the case, and the other the position of the defence, e.g., the fingermark of questioned origin was deposited by a finger of a particular known individual, versus the fingermark of questioned origin was deposited by a finger of some other individual selected at random from the relevant population. In this example:

- the numerator of the likelihood ratio quantifies the *similarity* between the mark and the print
- the denominator quantifies the *typicality* of the mark with respect to the relevant population

For continuously-valued data, likelihood ratios can be calculated as the ratio of two probability-density functions evaluated at the value of the evidence.[4]

2.2.5 Empirical validation

Empirical validation under conditions reflecting those of the case to which a forensic-evaluation system is to be applied is the only way to know how well that system performs under the conditions of the case. The need for validation under casework conditions has been emphasized by FSR [31], and by PCAST [5]:

> Without appropriate estimates of accuracy, an examiner's statement that two samples are similar—or even indistinguishable—is scientifically meaningless: it has no probative value, and considerable potential for prejudicial impact. Nothing—not training, personal experience nor professional practices—can substitute for adequate empirical demonstration of accuracy. (p. 46)

Protocols for validating systems that output likelihood ratios have been developed, including metrics and graphics appropriate for representing the results of such validations (Meuwly [32]; Brümmer & du Preez [33]; Morrison [34]; Meuwly et al. [35]; Ramos et al. [36]; Morrison et al. [22]). Much of the latter work has been conducted in the context of forensic voice comparison, but the results are applicable across forensic science in general.

2.2.6 A Kuhnian paradigm shift

[4] In the forensic inference and statistics literature "likelihood ratio" is commonly used as a cover term for both likelihood ratios based only on sample data and for Bayes' factors based on sample data and prior distributions for model parameters. The present discussion is intended to be neutral with respect to likelihoodist or Bayesian approaches. For simplicity, the present paper glosses over the differences between likelihoodist and Bayesian approaches. For readers particularly concerned with the differences between these two approaches, or with the nuances of one or the other of them, the present paper may be frustrating vague and such readers may feel that it conflates concepts that they considered to be importantly different. Addressing the concerns of such readers would, however, likely lead to greater confusion for the broader audience of forensic scientists and lawyers to whom the present paper is primarily addressed. Readers interested in the differences between likelihoodist and Bayesian approaches may wish to consult Ommen & Saunders [29], [30].

The idea that evaluation of forensic evidence is undergoing a paradigm shift is not new. The most famous article heralding a paradigm shift is Saks & Koehler [4]. Allowing for differences in wording and level of detail and the passage of time, I believe that Saks & Koehler [4] and the present paper describe the same paradigm shift. In contrast to Saks & Koehler's [4] statement that they intended "paradigm shift" as a metaphor, however, I view the paradigm shift in evaluation of forensic evidence as a true *Kuhnian paradigm shift* (Kuhn [6]) in the sense that:[5]

- it requires rejection of existing methods and the ways of thinking that underpin them,
- and rejection of the idea that progress can be made by incremental improvements to existing methods.
- Instead, it requires the wholesale adoption of an entire constellation of new methods and new ways of thinking.

That a paradigm shift requires the wholesale adoption of an entire constellation of new methods and new ways of thinking remains the case irrespective of whether one considers the shift to be from one paradigm to another or to be from a pre-paradigm to a paradigm period of science. As suggested in Saks & Koehler [4], a pre-paradigm period would seem to be a more accurate description of the *status quo*, with multiple traditions of evaluation of evidence used both within individual branches of forensic science and across different branches of forensic science, and hence there being no established widely-accepted overarching paradigm in use.

Some authors have used the term "paradigm shift" in relation to a single element or a subset of the elements of the paradigm shift as I have outlined it above, but I believe

[5] The only aspects in which I think the current paradigm shift in evaluation of forensic evidence deviates from Kuhn's [6] description of paradigm shifts relate to forensic science being an applied science which is not isolated from societal pressures.

that all of these elements are required as part of the constellation. This may be viewed as a radical stance, and it faces resistance, but, over the last decade and a half, my colleagues and I have had substantial success in contributing to advancing this paradigm shift in forensic voice comparison.

2.3 Impedimenta

2.3.1 Introduction

The paradigm shift in evaluation of forensic evidence is ongoing, but progress is slow or stalling for multiple reasons including the following:

- The new paradigm has only been adopted in a few branches of forensic science, and only by a minority of researchers and practitioners (§2.3.2).
- Only some elements of the new paradigm have been adopted as part of incremental change (§2.3.3).
- There is misunderstanding of the new paradigm and resistance to its adoption (§2.3.4).
- Research is often not informed by practice and has no impact on practice (§2.3.5).
- It is difficult to obtain funding for evidential-forensic-science research (§2.3.6).
- There are genuine practical impediments to implementing the new paradigm (§2.3.7).

I discuss each of these impediments in the following six subsections.

2.3.2 The new paradigm has only been adopted in a few branches of forensic sciences, and only by a minority of researchers and practitioners

In the 1990s, the new paradigm was widely adopted for **forensic evaluation of DNA**

(Foreman et al. [37]). Although the volume and importance of casework in this branch of forensic science makes it influential, single-source DNA profiles are invariant and discrete. They therefore have a very different structure from the continuously-valued data with within-source variability that results from analyses in most other branches of forensic science. The situation is more complex for low-template DNA and for DNA mixtures, but there is still a difference in data structure. Interpretation of DNA profiles is also dependent on well-developed theory of genetic inheritance, whereas interpretation in most branches of forensic science will have to be data driven (as is the case in machine learning in general, including in biometric applications). The potential for transfer of new-paradigm knowledge and methods from DNA to other branches of forensic science is therefore limited.

Since around 2000, a growing number of researchers and practitioners in **forensic voice comparison** have developed and adopted methods for calculation of likelihood ratios based on acoustic measurements and statistical models / machine-learning algorithms, and have developed and adopted methods for calibration and validation of likelihood-ratio systems under casework conditions. This has included adoption of state-of-the-art machine-learning approaches to automatic speaker recognition (Lee et al. [38]; Matějka et al. [39]; Villalba et al. [40]; Morrison et al. [41]; Morrison et al. [42]; Weber et al. [43]). At present, however, only a minority of practitioners have adopted the new paradigm. Survey results published in 2011, 2016, and 2019 (Gold & French [44]; Morrison et al. [45]; Gold & French [46]) suggest that, although the proportion of practitioners who have adopted human-supervised-automatic approaches and numeric likelihood ratios is growing, they are still a minority.[6] In addition, inconsistent with the

[6] Percentages of respondents who reported that they used human-supervised-automatic approaches and numeric likelihood ratios were 20% and 9% respectively in the first survey, 33% and 23% respectively in the second, and 41% and 13% respectively in the third. Respondents in the first and third surveys included private practitioners and practitioners in government labs, and the results of these two surveys should be directly comparable. Respondents in the second survey were from a different population, law-enforcement agencies in INTERPOL member countries, which likely accounts for some of the difference between the results of this survey and those of the other two.

new paradigm, most respondents in the most recent survey who reported using a human-supervised-automatic approach also reported that they combined it with human-perception-based analysis and subjective-judgement-based interpretation.

Data in human-supervised-automatic approaches are continuously valued and have intrinsic within-source variability, a data structure shared with many other branches of forensic science. Compared to DNA, new-paradigm knowledge and methods from forensic voice comparison, including statistical models and calibration and validation methods, should therefore be easier to transfer to and adapt for use other branches of forensic science. For an example of such transfer and adaptation, see Basu et al. [47]. Forensic voice comparison is, however, a relatively niche branch of forensic science, which limits the extent to which developments in forensic voice comparison are noticed and adopted by researchers and practitioners in other branches of forensic science.

Curran [1] lamented that only 13% of laboratories surveyed used the likelihood-ratio framework for **glass evidence**, but this may be one of the highest rates of adoption after DNA. In many **other branches of forensic science** the rate of adoption of the likelihood-ratio framework by practitioners is near zero (Bali et al. [48]; Cole & Barno [49]).

2.3.3 Only some elements of the new paradigm have been adopted as part of incremental change

Although in the short term adopting some elements of the new paradigm as part of incremental change may be a viewed as a step in the right direction, in the long term it may actually impede a paradigm shift.

> Just because it is a transition between incommensurables, the transition between competing paradigms cannot be made a step at a time, ... Like the gestalt switch, it must occur all at once (though not necessarily in an instant) or not at all. (Kuhn [6], p. 149)

Some practitioners **assign likelihood-ratio values based on subjective judgement**, and the values they assign are not subject to empirical calibration or empirical validation (see Risinger [50]; Morrison & Thompson [51]; Morrison et al. [52]). Some authors emphasize the logic of the likelihood-ratio framework and consider subjective assignment of likelihood ratio an acceptable end goal or consider it a step in the right direction, but such incremental steps potentially impede a paradigm shift which requires the abandonment of interpretation methods that are entirely dependent on subjective judgement.[7] In addition, placing an emphasis on subjectivist concepts of probability is detrimental to attempts to encourage practitioners to adopt methods based on relevant data, quantitative measurements, and statistical models, and to adopt empirical validation under casework conditions (Morrison [54]).

The majority of **proposals to address cognitive bias** in forensic science (e.g., EWG [9]; Stoel et al. [11]; Thompson et al. [55]; FSR [56]) **assume the continued use of human-perception- and subjective-judgement-based methods**. Although this may be necessary in the short term, it potentially impedes a paradigm shift to quantitative-measurement- and statistical-model-based methods.

Some practitioners **make use of systems based on quantitative measurements and statistical models, but do not empirically calibrate or validate** their system using data that reflect the relevant population and the conditions for the case, and/or **rather than directly report the output of the system, they use it as input to a subjective-judgement process** that also considers other information including from human-perception-based analyses (see Morrison & Thompson [51]; Morrison [57], [58]). Such approaches are pernicious in that use of technology may give the false impression of scientific validity, and reaction against this may impede a paradigm shift that includes adoption of quantitative measurements and statistical models.

[7] My stance on where in the forensic-evaluation process use of subjective judgement is acceptable is more restrictive than that of some leaders in the field of forensic inference and statistics, e.g., Evett et al. [53].

2.3.4 There is misunderstanding of the new paradigm and resistance to its adoption

As with all Kuhnian paradigm shifts, there is misunderstanding of the new paradigm and resistance to its adoption.

Some resistance is cultural. The cultures of some branches of forensic science seem to be especially resistant to the adoption of statistical-model-based methods and of validation (see Mnookin et al. [59]; Curran [1]; Morrison [60]; Morrison & Stoel [61]; Swofford et al. [62]). Practitioners in multiple branches of forensic science often claim that training and experience provide sufficient warrant for their conclusions (see Mnookin et al. [59]; Risinger [50]; PCAST [5]; Morrison & Thompson [51]), and deny or obfuscate about the need for validation (see Cole [63]; Morrison [60]; PCAST [5]; Koehler [64]; Morrison et al. [52]), or propose lax validation protocols that do not require demonstration of performance under casework conditions (see Morrison et al. [65], [66]).

People in general tend to prefer methods which involve greater human input even when validation results indicate that data-driven methods lead to better results. Over time, however, people can come to accept data-driven methods (Swofford & Champod [67]).

There is a **belief that likelihood ratios are difficult to understand** (see Bali et al., 2020; Swofford et al. [62]; Swofford & Champod [68]). Commonly occurring misunderstandings have even been given names, e.g., the “prosecutor’s fallacy” and the “defense attorney’s fallacy” (Thompson & Schurmann [69]). There are many examples of legal rulings in which judges have misunderstood the meaning of a likelihood ratio (the England & Wales Court of Appeal 2010 ruling in *R v T* is an infamous example; see, e.g., Berger et al. [70]; Redmayne et al. [71]; Morrison [72]; Thompson [73]). Results of empirical research on lay understanding of expressions of strength of evidence are mixed (Eldridge [74]; Martire & Edmond [75]).

Despite legal rulings and recommendations concerning the need for validation, **courts often do not understand empirical validation and its necessity**, and accept testimony based on forensic-evaluation methods that have not been validated under conditions reflecting those of the case under consideration, or even that have not been empirically validated at all (see Bernstein [76]; Morrison [60], [57], [58]; Cooper [77]; Edmond [78]).

2.3.5 Research is often not informed by practice and has no impact on practice

Research that appears to be about forensic science may not actually be about solving real forensic-science problems. For example, research may really be about a method in statistics or machine learning with an apparent forensic-science problem being used as an example of the application of that method.

> Research in forensic science is sorely needed, but it should address primarily forensic science questions—not questions relating to the application of chemistry, biology, statistics, or psychology. (Margot [79], p. 801)

> it is critical that researchers and funding bodies understand the importance of conducting research that is informed by practice and can be translated into practical applications. (Roux & Weyermann [80], p. 2)

Research that is divorced from forensic practice may lead to academic papers but nothing else. Even research that does address real forensic-science problems will fail to have impact unless it involves genuine collaboration in which researchers understand the demands of practice, and in which practitioners are willing to embrace research-informed change (Curran [1]).

2.3.6 It is difficult to obtain funding for evidential-forensic-science research

Few funding agencies have sustained funding targeted at forensic science, and funding agencies seldom have panels of reviewers knowledgeable about evidential forensic science. Applications for funding for evidential-forensic-science research made to non-forensic-science-targeted calls are often rejected because reviewers do not understand the epistemology or value of forensic-science research. Applications are often rejected because their goals are to improve forensic science, which is an applied science, and **funding-agency criteria or reviewers often do not value applied science**.

> The larger scientific community must now come to the aid of our forensic colleagues in advocating both for: (i) the research and financial support that is so clearly needed to advance the field and (ii) the requirement for empirical testing that is so clearly needed to advance the cause of justice. ... Forensic scientists have long complained that their work is not always valued by their scientific colleagues because of its applied nature; it is time for the scientific community to move beyond that conceit. (Bell et al. [81])

At the other extreme, when there are calls specific to forensic science, they usually **focus exclusively or primarily on short-term goals related to law-enforcement investigative applications rather than on courtroom-evidential applications** (investigative and evidential applications have very different requirements), and they usually **focus on technology rather than on forensic inference**.

> technology-oriented development ... often overrul[es] the importance of appropriate scientific reasoning to solve actual problems (Roux et al. [82], p. 679)

Research calls requiring deliverables with a high technology readiness level (TRL) are not sources of funding for paradigm-shifting research.

In 2018, United Kingdom Research and Innovation (UKRI) informed the UK House of Lords Science and Technology Select Committee that UKRI had invested GBP 56M over 10 years in forensic-science research (less than 0.1% of UKRI’s total budget), but on closer inspection most of the funding counted to obtain that figure was not for research projects that actually focused on (or even made any contribution to) forensic science: only about GBP 17M went to forensic-science focussed research, and GBP 15M of that went to TRL research, only GBP 2M to foundational research (HoL [2]; Morgan & Levin [83]). HoL [2] recommended that UKRI “urgently and substantially increase the amount of dedicated funding allocated to forensic science” (§187), but, more than 3 years after the publication of the HoL report, this has not (yet) happened.

2.3.7 There are genuine practical impediments to implementing the new paradigm

Even if practitioners want to adopt the new paradigm, they will be unable to do so unless they are provided with the quantitative-measurement and statistical-modelling / machine-learning tools and the case-relevant data necessary to calculate likelihood ratios and validate systems under the conditions of the cases on which they work. Practitioners will also not be able to adopt the new paradigm unless they are provided with training on understanding the new paradigm in general and on how to implement it for the types of cases they work on.

2.4 Via progredi

Kuhn [6]:

> The transfer of allegiance from paradigm to paradigm is a conversion experience that cannot be forced. ... a generation is sometimes required to effect the change ... Conversions will occur a few at a time until, after the last holdouts have died, the whole profession will again be practicing under a single, but now a different, paradigm. (pp. 150–151)

Kuhnian paradigm shifts are not rapid and individuals cannot be forced to embrace the new paradigm, but my aim is to facilitate and thereby advance the adoption of the new paradigm. My **strategy** is to work with researchers and practitioners who want to adopt the new paradigm, to work with them on addressing practical impediments to applying the new paradigm in casework, i.e.:

1. to provide researchers, practitioners, and lawyers with training leading to understanding of the new paradigm;

2. to collaborate with researchers and practitioners on building relevant databases and on developing and validating statistical models applicable in their particular branches of forensic science; and

3. to conduct research on how to present likelihood ratios and validation results so as to maximize understanding by laypeople, and thereby provide guidance to forensic practitioners on how to communicate forensic-evaluation results to legal-decision makers.

I invite others to consider whether this is a strategy that they would also be interested in adopting, either in part or in whole.

Element 2 of the strategy requires collaboration between researchers with expertise in forensic data science and researchers and practitioners with expertise in particular branches of forensic science. Academic publications are unlikely to convince practitioners to adopt the new paradigm, but other practitioners successfully applying the new paradigm are potentially convincing. In any branch of forensic science, the number of practitioners who initially want to adopt the new paradigm and who want to collaborate on this endeavour will almost certainly be a very small minority, but it will be more productive to work with a small minority on developing practical solutions than to try to convince the majority of practitioners without providing practical solutions. Once the practical solutions are being used by the small minority, use of the new paradigm has the potential to spread. Even then, I do not expect adoption of the new paradigm to be rapid, but I do expect higher rates of adoption among newer practitioners and trainees, leading to a generational shift.

3 Conclusion

A paradigm shift in evaluation of forensic evidence is ongoing. The shift is away from methods based on human perception and subjective judgement, to methods based on:

- relevant data,
- quantitative measurements,
- and statistical models.

New paradigm methods:

- are transparent and reproducible
- are intrinsically resistant to cognitive bias
- use the logically correct framework for interpretation of evidence (the likelihood-ratio framework)
- are empirically validated under casework conditions

This is a Kuhnian paradigm shift, which requires:

- rejection of existing methods and the ways of thinking that underpin them
- rejection of the idea that progress can be made by incremental improvements to existing methods
- the wholesale adoption of an entire constellation of new methods and new ways of thinking

Some branches of forensic science, such as forensic voice comparison, are more advanced in the paradigm shift than others. Knowledge gained in advancing the paradigm shift in forensic voice comparison can assist in advancing the paradigm shift in other branches of forensic science. Validation protocols, metrics, and graphics developed in the context of forensic voice comparison are immediately applicable in other branches of forensic science. Statistical models / machine-learning algorithms used in forensic voice comparison can even be transferred and adapted for use in other branches of forensic science.

My strategy for advancing the paradigm shift requires collaboration between researchers with expertise in forensic data science and researchers and practitioners with expertise in particular branches of forensic science. My strategy is to work with researchers and practitioners who want to adopt the new paradigm. My strategy is to work with them on addressing practical impediments to applying the new paradigm in casework, i.e.:

1. to provide researchers, practitioners, and lawyers with training leading to understanding of the new paradigm;

2. to collaborate with researchers and practitioners on building relevant databases and on developing and validating statistical models applicable in their particular branches of forensic science; and

3. to conduct research on how to present likelihood ratios and validation results so as to maximize understanding by laypeople, and thereby provide guidance to forensic practitioners on how to communicate forensic-evaluation results to legal-decision makers.

In any branch of forensic science, the number of practitioners who initially want to adopt the new paradigm and who want to collaborate on this endeavour will almost certainly be very small, but it will be more productive to work with a small minority on developing practical solutions than to try to convince the majority of practitioners without providing practical solutions. Once the practical solutions are being used by the small minority, use of the new paradigm has the potential to spread. Even then, I do not expect adoption of the new paradigm to be rapid, but I do expect higher rates of adoption among newer practitioners and trainees, leading to a generational shift.

I have been asked several times over the years whether I could suggest a name for the new paradigm other than "new". Here, I propose that the new paradigm could be called "forensic data science". My hope is that, after the paradigm shift is complete, it will simply be called "forensic science".

4 References

[1] Curran J.M. (2013). Is forensic science the last bastion of resistance against statistics? *Science & Justice*, 53, 251–252. http://dx.doi.org/10.1016/j.scijus.2013.07.001

[2] House of Lords Science and Technology Select Committee (2019). *Forensic Science and the Criminal Justice System: A Blueprint for Change*. https://publications.parliament.uk/pa/ld201719/ldselect/ldsctech/333/333.pdf

[3] Roux C., Bucht R., Crispino F., De Forest P., Lennard C., Margot P., Miranda M.D., NicDaéid N., Ribaux O., Ross A., Willis S. (2022). The Sydney declaration – Revisiting the essence of forensic science through its fundamental principles. *Forensic Science International,* 332, article 111182. https://doi.org/10.1016/j.forsciint.2022.111182

[4] Saks M.J., Koehler J.J. (2005). The coming paradigm shift in forensic identification science. *Science,* 309, 892–895. https://doi.org/10.1126/science.1111565

[5] President's Council of Advisors on Science and Technology (2016). *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*.

https://obamawhitehouse.archives.gov/administration/eop/ostp/pcast/docsreports/

[6] Kuhn T.S. (1962). *The Structure of Scientific Revolutions*. Chicago IL: University of Chicago Press.

[7] Edmond G., Towler A., Growns B., Ribeiro G., Found B., White D., Ballantyne K., Searston R.A., Thompson M.B., Tangen J.M., Kemp R.I., Martire K.A. (2017). Thinking forensics: Cognitive science for forensic practitioners. *Science & Justice*, 57, 144–154. http://dx.doi.org/10.1016/j.scijus.2016.11.005

[8] National Research Council of the National Academies (2009). *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: National Academies Press. https://doi.org/10.17226/12589

[9] Expert Working Group on Human Factors in Latent Print Analysis (2012). *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*. Gaithersburg, MD: National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.7842

[10] Found, B. (2015). Deciphering the human condition: The rise of cognitive forensics. *Australian Journal of Forensic Sciences*, 47, 386–401. http://dx.doi.org/10.1080/00450618.2014.965204

[11] Stoel R.D., Berger C.E.H., Kerkhoff W., Mattijssen E.J.A.T., Dror E.I. (2015) Minimizing contextual bias in forensic casework. In Strom K.J., Hickman M.J. (Eds.), *Forensic Science and the Administration of Justice: Critical Issues and Directions* (pp. 67–86). Thousand Oaks CA: Sage. http://dx.doi.org/10.4135/9781483368740.n5

[12] Cooper G.S., Meterko V. (2019). Cognitive bias research in forensic science: A systematic review. *Forensic Science International*, 297,35–46.

https://doi.org/10.1016/j.forsciint.2019.01.016

[13] Expert Working Group on Human Factors in Handwriting Examination (2020). *Forensic Handwriting Examination and Human Factors: Improving the Practice Through a Systems Approach*. Gaithersburg, MD: National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8282

[14] Spellman B.A., Eldridge H., Bieber P. (2022). Challenges to reasoning in forensic science decisions. *Forensic Science International: Synergy*. https://doi.org/10.1016/j.fsisyn.2021.100200

[15] Saks M.J., Koehler J.J. (2008). The individualization fallacy in forensic science. *Vanderbilt Law Review*, 61 199–219. Preprint: https://ssrn.com/abstract=1432516

[16] Cole S.A. (2009). Forensics without uniqueness, conclusions without individualization: The new epistemology of forensic identification. *Law, Probability and Risk*, 8, 233−255. https://doi.org/10.1093/lpr/mgp016

[17] Cole S.A. (2014). Individualization is dead, long live individualization! Reforms of reporting practices for fingerprint analysis in the United States. *Law, Probability and Risk*, 13, 117−150. https://doi.org/10.1093/lpr/mgt014

[18] Jackson G. (2009). Understanding forensic science opinions. In Fraser J., Williams R. (Eds.), *Handbook of Forensic Science* (pp. 419–445). Cullompton, UK: Willan. https://doi.org/10.4324/9781843927327

[19] Kaye D.H. (2015). Presenting forensic identification findings: The current situation. In Neumann C., Ranadive A., Kaye D.H. (eds.), *Communicating the Results of Forensic Science Examinations*. Final Technical Report for NIST Award Number 70NANB12H014 (pp. 12–30). https://ssrn.com/abstract=2690899

[20] Aitken C.G.G., Berger C.E.H., Buckleton J.S., Champod C., Curran J.M., Dawid A.P., Evett I.W., Gill P., González-Rodríguez J., Jackson G., Kloosterman A., Lovelock T., Lucy D., Margot P., McKenna L., Meuwly D., Neumann C., Nic Daéid N., Nordgaard A., Puch-Solis R., Rasmusson B., Redmayne M., Roberts P., Robertson B., Roux C., Sjerps M.J., Taroni F., Tjin-A-Tsoi T., Vignaux G.A., Willis S.M., Zadora G. (2011) Expressing evaluative opinions: A position statement. *Science & Justice*, 51, 1–2. http://dx.doi.org/10.1016/j.scijus.2011.01.002

[21] Morrison G.S., Kaye D.H., Balding D.J., Taylor D., Dawid P., Aitken C.G.G., Gittelson S., Zadora G., Robertson B., Willis S.M., Pope S., Neil M., Martire K.A., Hepler A., Gill R.D., Jamieson A., de Zoete J., Ostrum R.B., Caliebe A. (2017). A comment on the PCAST report: Skip the "match"/"non-match" stage. *Forensic Science International*, 272, e7–e9. http://dx.doi.org/10.1016/j.forsciint.2016.10.018

[22] Morrison G.S., Enzinger E., Hughes V., Jessen M., Meuwly D., Neumann C., Planting S., Thompson W.C., van der Vloed D., Ypma R.J.F., Zhang C., Anonymous A., Anonymous B. (2021). Consensus on validation of forensic voice comparison. *Science & Justice*, 61, 229–309. https://doi.org/10.1016/j.scijus.2021.02.002

[23] Association of Forensic Science Providers (2009). Standards for the formulation of evaluative forensic science expert opinion. *Science & Justice*, 49, 161–164. http://dx.doi.org/10.1016/j.scijus.2009.07.004

[24] Aitken C.G.G., Roberts P., Jackson G. (2010). *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses*. London, UK: Royal Statistical Society. https://rss.org.uk/news-publication/publications/law-guides/

[25] Willis S.M., McKenna L., McDermott S., O'Donell G., Barrett A., Rasmusson A., Nordgaard A., Berger C.E.H., Sjerps M.J., Lucena-Molina J.J., Zadora G., Aitken C.G.G., Lunt L., Champod C., Biedermann A., Hicks T.N., Taroni F. (2015). *ENFSI Guideline for Evaluative Reporting in Forensic Science*. http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf

[26] Ballantyne K., Bunford J., Found B., Neville D., Taylor D., Wevers G., Catoggio D. (2017). *An Introductory Guide to Evaluative Reporting*. National Institute of Forensic Science of the Australia New Zealand Policing Advisory Agency. http://www.anzpaa.org.au/forensic-science/our-work/projects/evaluative-reporting

[27] Kafadar K., Stern H., Cuellar M., Curran J.M., Lancaster M., Neumann C., Saunders C., Weir B., Zabell S. (2019). *American Statistical Association Position on Statistical Statements for Forensic Evidence*. https://www.amstat.org/asa/files/pdfs/POL-ForensicScience.pdf

[28] Forensic Science Regulator (2021). *Codes of Practice and Conduct: Development of Evaluative Opinions* (FSR-C-118 Issue 1). https://www.gov.uk/government/publications/development-of-evaluative-opinions

[29] Ommen D.M., Saunders C.P. (2021). A problem in forensic science highlighting the differences between the Bayes factor and likelihood ratio. *Statistical Science*, 36, 344–359. https://doi.org/10.1214/20-STS805

[30] Ommen D.M., Saunders C.P. (2022). Differences between Bayes factors and likelihood ratios for quantifying the forensic value of evidence. In Carriquiry A.L., Tanur J.M., Eddy W.F. (Eds.), *Statistics in the Public Interest* (pp. 169–186). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-030-75460-0_10

[31] Forensic Science Regulator (2020). *Guidance: Validation* (FSR-G-201 Issue 2). https://www.gov.uk/government/publications/forensic-science-providers-validation

[32] Meuwly D. (2001). *Reconnaissance de locuteurs en sciences forensiques: l'apport d'une approche automatique*. Doctoral dissertation, University of Lausanne. https://www.unil.ch/files/live/sites/esc/files/shared/These.Meuwly.pdf

[33] Brümmer N., du Preez J. (2006). Application independent evaluation of speaker detection. *Computer Speech and Language*, 20, 230–275. https://doi.org/10.1016/j.csl.2005.08.001

[34] Morrison G.S. (2011). Measuring the validity and reliability of forensic likelihood-ratio systems. *Science & Justice*, 51, 91–98. http://dx.doi.org/10.1016/j.scijus.2011.03.002

[35] Meuwly D., Ramos D., Haraksim R. (2017). A guideline for the validation of likelihood ratio methods used for forensic evidence evaluation. *Forensic Science International*, 276, 142–153. http://dx.doi.org/10.1016/j.forsciint.2016.03.048

[36] Ramos D., Meuwly D., Haraksim R., Berger C.E.H. (2020). Validation of forensic automatic likelihood ratio methods. In Banks D., Kafadar K., Kaye D.H., Tackett M. (Eds.), *Handbook of Forensic Statistics* (pp. 143–163). Boca Raton, FL: CRC. https://doi.org/10.1201/9780367527709

[37] Foreman L.A., Champod C., Evett I.W., Lambert J.A., Pope S. (2003). Interpreting DNA evidence: A review. *International Statistical Review*, 71, 473–495. http://dx.doi.org/10.1111/j.1751-5823.2003.tb00207.x

[38] Lee K.A., Yamamoto H., Okabe K., Wang Q., Guo L., Koshinaka T., Zhang J., Shinoda, K. (2020). NEC-TT System for mixed-bandwidth and multi-domain speaker recognition. *Computer Speech, Language*, 61, article 101033.

https://doi.org/10.1016/j.csl.2019.101033

[39] Matějka P., Plchot O., Glembek O., Burget L., Rohdin J., Zeinali H., Mošner L., Silnova A., Novotný O., Diez M., Černocký J.H. (2020). 13 years of speaker recognition research at BUT, with longitudinal analysis of NIST SRE. *Computer Speech & Language*, 63, article 101035. https://doi.org/10.1016/j.csl.2019.101035

[40] Villalba J., Chen N., Snyder D., García-Romero D., McCree A., Sell G., Borgstrom J., García-Perera L.P., Richardson F., Dehak R., Torres-Carrasquillo P.A., Dehak N. (2020). State-of-the-art speaker recognition with neural network embeddings in NIST SRE18 and Speakers in the Wild evaluations. *Computer Speech & Language*, 60, article 101026. https://doi.org/10.1016/j.csl.2019.101026

[41] Morrison G.S., Enzinger E., Ramos D., González-Rodríguez J., Lozano-Díez A. (2020). Statistical models in forensic voice comparison. In Banks D., Kafadar K., Kaye D.H., Tackett M. (Eds.), *Handbook of Forensic Statistics* (pp. 451–497). Boca Raton, FL: CRC. https://doi.org/10.1201/9780367527709

[42] Morrison G.S., Weber P., Enzinger E., Labrador B., Lozano-Díez A., Ramos D., González-Rodríguez J. (2022 in press). Forensic voice comparison – Human-supervised-automatic approach. In Houck M., Wilson L., Lewis S., Eldridge H., Reedy P., Lothridge K. (Eds.), *Encyclopedia of Forensic Sciences* (3rd Ed.). Elsevier.

[43] Weber P., Enzinger E., Labrador B., Lozano-Díez A., Ramos D., González-Rodríguez J., Morrison G.S. (2022). Validation of the alpha version of the E^3 Forensic Speech Science System (E^3FS^3) core software tools. *Forensic Science International: Synergy*, 4, article 100223. https://doi.org/10.1016/j.fsisyn.2022.100223

[44] Gold E., French J.P. (2011). International practices in forensic speaker comparison. *International Journal of Speech, Language and the Law*, 18, 143–152. http://dx.doi.org/10.1558/ijsll.v18i2.293

[45] Morrison G.S., Sahito F.H., Jardine G., Djokic D., Clavet S., Berghs S., Goemans Dorny C. (2016). INTERPOL survey of the use of speaker identification by law enforcement agencies. *Forensic Science International*, 263, 92–100. http://dx.doi.org/10.1016/j.forsciint.2016.03.044

[46] Gold E., French J.P. (2019). International practices in forensic speaker comparison: Second survey. *International Journal of Speech, Language and the Law*, 26, 1–20. https://doi.org/10.1558/ijsll.38028

[47] Basu N., Bolton-King R.S., Morrison G.S. (2022 submitted). Forensic comparison of fired cartridge cases: Feature-extraction methods for feature-based calculation of likelihood ratios. Preprint: http://firearms.forensic-data-science.net/

[48] Bali A.S., Edmond G., Ballantyne K.N., Kemp R.I., Martire K.A. (2020) Communicating forensic science opinion: An examination of expert reporting practices. *Science & Justice*, 60, 216–224. https://doi.org/10.1016/j.scijus.2019.12.005

[49] Cole S.A., Barno M. (2020). Probabilistic reporting in criminal cases in the United States: A baseline study. *Science & Justice*, 60, 406–414. https://doi.org/10.1016/j.scijus.2020.06.001

[50] Risinger D.M. (2013). Reservations about likelihood ratios (and some other aspects of forensic ‘Bayesianism’). *Law, Probability and Risk*, 12, 63–73, http://dx.doi.org/10.1093/lpr/mgs011

[51] Morrison G.S., Thompson W.C. (2017). Assessing the admissibility of a new

generation of forensic voice comparison testimony. *Columbia Science and Technology Law Review*, 18, 326–434. https://doi.org/10.7916/stlr.v18i2.4022

[52] Morrison G.S., Ballantyne K., Geoghegan P.H. (2018). A response to Marquis et al (2017) What is the error margin of your signature analysis? *Forensic Science International*, 287, e11–e12. https://doi.org/10.1016/j.forsciint.2018.03.009

[53] 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. *Forensic Science International*, 278, 16–23. http://dx.doi.org/10.1016/j.forsciint.2017.06.018

[54] Morrison G.S. (2017). What should a forensic practitioner's likelihood ratio be? II. *Science & Justice*, 57, 472–476. http://dx.doi.org/10.1016/j.scijus.2017.08.004

[55] Thompson W.C., Black J., Jain A., Kadane J. (2017). *Forensic Science Assessments: A Quality and Gap Analysis- Latent Fingerprint Examination*. Washington, DC: American Association for the Advancement of Science. https://doi.org/10.1126/srhrl.aag2874

[56] Forensic Science Regulator (2020). *Guidance: Cognitive Bias Effects Relevant to Forensic Science Examinations* (FSR-G-217 Issue 2). https://www.gov.uk/government/publications/cognitive-bias-effects-relevant-to-forensic-science-examinations

[57] Morrison G.S. (2018). Admissibility of forensic voice comparison testimony in England and Wales. *Criminal Law Review*, 2018(1), 20–33. Preprint: http://geoff-morrison.net/#Admissibility_EW_2018

[58] Morrison G.S. (2018). The impact in forensic voice comparison of lack of calibration and of mismatched conditions between the known-speaker recording

and the relevant-population sample recordings. *Forensic Science International*, 283, e1–e7. http://dx.doi.org/10.1016/j.forsciint.2017.12.024

[59] Mnookin J.L., Cole S.A., Dror I.E., Fisher B.A.J., Houck M.M., Inman K., Kaye D.H., Koehler J.J., Langenburg G., Risinger D.M., Rudin N., Siegel J., Stoney D.A. (2011). The need for a research culture in the forensic sciences. *UCLA Law Review*, 58, 725–777. https://www.uclalawreview.org/the-need-for-a-research-culture-in-the-forensic-sciences-2/

[60] Morrison G.S. (2014). Distinguishing between forensic science and forensic pseudoscience: Testing of validity and reliability, and approaches to forensic voice comparison. *Science & Justice*, 54, 245–256. http://dx.doi.org/10.1016/j.scijus.2013.07.004

[61] Morrison G.S., Stoel R.D. (2014). Forensic strength of evidence statements should preferably be likelihood ratios calculated using relevant data, quantitative measurements, and statistical models – a response to Lennard (2013) Fingerprint identification: How far have we come? *Australian Journal of Forensic Sciences*, 46, 282–292. http://dx.doi.org/10.1080/00450618.2013.833648

[62] Swofford H., Cole S., King V. (2021). Mt. Everest – we are going to lose many: A survey of fingerprint examiners' attitudes towards probabilistic reporting. *Law, Probability and Risk*, 19, 255–291. https://doi.org/10.1093/lpr/mgab003

[63] Cole S.A. (2006). Is fingerprint identification valid? Rhetorics of reliability in fingerprint proponents' discourse. *Law & Policy*, 28, 109–135. https://doi.org/10.1111/j.1467-9930.2005.00219.x

[64] Koehler J.J. (2017). Forensics or fauxrensics? Ascertaining accuracy in the forensic sciences. *Arizona State Law Journal*, 49(4), 1369–1416. https://arizonastatelawjournal.org/2018/02/07/forensics-or-fauxrensics-ascertaining-accuracy-in-the-forensic-sciences/

[65] Morrison G.S., Neumann C., Geoghegan P.H. (2020). Vacuous standards – subversion of the OSAC standards-development process. *Forensic Science International: Synergy*, 2, 206–209. https://doi.org/10.1016/j.fsisyn.2020.06.005

[66] Morrison G.S., Neumann C., Geoghegan P.H., Edmond G., Grant T., Ostrum R.B., Roberts P., Saks M., Syndercombe Court D., Thompson W.C., Zabell S. (2021). Reply to Response to Vacuous standards – subversion of the OSAC standards-development process. *Forensic Science International: Synergy*, 3, article 100149. https://doi.org/10.1016/j.fsisyn.2021.100149

[67] Swofford H., Champod C. (2021). Implementation of algorithms in pattern & impression evidence: A responsible and practical roadmap. *Forensic Science International: Synergy*, 3, article 100142. https://doi.org/10.1016/j.fsisyn.2021.100142

[68] Swofford H., Champod C. (2022). Probabilistic reporting and algorithms in forensic science: Stakeholder perspectives within the American criminal justice system. *Forensic Science International: Synergy*, 4, article 100220. https://doi.org/10.1016/j.fsisyn.2022.100220

[69] Thompson W.C., Schumann E.L. (1987). Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behavior*, 11(3), 167–187. http://www.jstor.org/stable/1393631

[70] Berger C.E.H., Buckleton J., Champod C., Evett I.W., Jackson G. (2011). Evidence evaluation: A response to the Court of Appeal judgment in R v T. *Science & Justice*, 51, 43–49. http://dx.doi.org/10.1016/j.scijus.2011.03.005

[71] Redmayne M., Roberts P., Aitken C.G.G., Jackson G. (2011). Forensic science evidence in question. *Criminal Law Review*, 2011(5), 347–356. Preprint: https://rke.abertay.ac.uk/en/publications/forensic-science-evidence-in-question

[72] Morrison G.S. (2012). The likelihood-ratio framework and forensic evidence in court: A response to R v T. *International Journal of Evidence and Proof*, 16, 1–29. http://dx.doi.org/10.1350/ijep.2012.16.1.390

[73] Thompson W.C. (2012). Discussion paper: Hard cases make bad law – reactions to R v T. *Law, Probability and Risk*, 11, 347–359. https://doi.org/10.1093/lpr/mgs020

[74] Eldridge H. (2019). Juror comprehension of forensic expert testimony: A literature review and gap analysis. *Forensic Science International: Synergy*, 1, 24–34. https://doi.org/10.1016/j.fsisyn.2019.03.001

[75] Martire K.A., Edmond G. (2020). How well do lay people comprehend statistical statements from forensic scientists? In Banks D., Kafadar K., Kaye D.H., Tackett M. (Eds.), *Handbook of Forensic Statistics* (pp. 201–224). Boca Raton, FL: CRC. https://doi.org/10.1201/9780367527709

[76] Bernstein D.E. (2013) The misbegotten judicial resistance to the Daubert revolution. *Notre Dame Law Review*, 89, 27–70. http://ndlawreview.org/publications/archives/volume-89/issue-1/

[77] Cooper S.L. (2016). Forensic science identification evidence: Tensions between law and science. *Journal of Philosophy, Science & Law*, 16(2), 1–35. https://doi.org/10.5840/jpsl20161622

[78] Edmond G. (2018). Re-assessing reliability. In Roberts P., Stockdale M. (Eds.), *Forensic Science Evidence and Expert Witness Testimony* (pp. 71–105). Cheltenham, UK: Elgar.

[79] Margot P. (2011). Commentary on the need for a research culture in the forensic sciences. *UCLA Law Review*, 58, 795–801. https://www.uclalawreview.org/commentary-on-the-need-for-a-research-culture-

in-the-forensic-sciences-3-2/

[80] Roux C., Weyermann C. (2021). From research integrity to research relevance to advance forensic science. *Forensic Sciences Research*. https://doi.org/10.1080/20961790.2021.1977480

[81] Bell S., Sah S., Albright T.D., Gates S.J., Denton M.B., Casadevall A. (2018). A call for more science in forensic science. *Proceedings of the National Academy of Sciences*, 115, 4541–4544; https://doi.org/10.1073/pnas.1712161115

[82] Roux C., Willis S., Weyermann C. (2021). Shifting forensic science focus from means to purpose: A path forward for the discipline? *Science & Justice*, 61, 678–686. https://doi.org/10.1016/j.scijus.2021.08.005

[83] Morgan R.M., Levin E.A. (2019). A crisis for the future of forensic science: Lessons from the UK of the importance of epistemology for funding research and development. *Forensic Science International: Synergy*, 1, 243–252. https://doi.org/10.1016/j.fsisyn.2019.09.002