A guide to interpreting forensic testimony: Scientific approaches to fingerprint evidence

GARY EDMOND[†]

Professor, School of Law, The University of New South Wales

MATTHEW B. THOMPSON

School of Psychology, The University of Queensland & School of Law, UCLA

AND

JASON M. TANGEN

School of Psychology, The University of Queensland

In response to criticism of latent fingerprint evidence from a variety of authoritative extra-legal inquiries and reports, this essay describes the first iteration of a guide designed to assist with the reporting and interpretation of latent fingerprint evidence. Sensitive to the recommendations of these reports, we have endeavoured to incorporate emerging empirical evidence about the matching performance of fingerprint examiners (i.e. indicative error rates) into their testimony. We outline a way of approaching fingerprint evidence that provides a more accurate—in the sense of empirically and theoretically justified—indication of the value of fingerprint evidence than existing practice. It is an approach that could be introduced immediately. The proposal is intended to help non-experts understand the value of the evidence and improve its presentation and assessment in criminal investigations and proceedings. This first iteration accommodates existing empirical evidence and draws attention to the gap between the declaration of a match and positive identification (or individualization). Represented in this way, fingerprint evidence will be more consistent with its known value as well as the aims and conduct of the accusatorial trial.

Keywords: error; validation; expert evidence; identification; report; expertise; forensic science.

1. Reforming the presentation of comparison evidence

Fingerprint examiners have been active in investigations and presented 'identification' evidence in criminal courts for more than a century.[1] Notwithstanding increasing automation, examiners continue to play a central role in the comparison of prints, the interpretation of prints, and in attributing significance to apparent matches. When confronted with an unknown print, usually a part (or fragment) of

[1] S. Cole, *Suspect Identities: A History of Fingerprinting and Criminal Identification* (Cambridge, MA: Harvard University Press, 2001).

GARY EDMOND *ET AL.*

a fingermark recovered from a crime scene (known as a 'latent'), it is the examiner who decides if the latent print provides sufficient information to interpret and, if so, whether it matches a known (i.e. reference) print.[2] Where the examiner is satisfied about the sufficiency of the print and declares a 'match', this is conventionally understood by examiners, and represented to others, as positive identification of the person who supplied the reference print *to the exclusion of all other persons.*[3]

Remarkably, given the interpretive (i.e. subjective) dimensions of comparison and the considerable gap between declaring a match and positive identification (so-called individualization),[4] there have been few scientific investigations of the human capacity to correctly match fingerprints, let alone attach significance to apparent similarities.[5] Nevertheless, for more than a hundred years, and in the absence of experimental support, fingerprint examiners have claimed that fingerprint evidence is basically infallible.[6] These assertions are typically justified by reference to training and experience (and the use of a method such as ACE-V: Analysis, Comparison, Evaluation and Verification),[7] assumptions about the uniqueness of fingerprints, along with legal acceptance and the effectiveness of fingerprint evidence in securing confessions and convictions.[8] In recent decades, however, commentators have questioned uniqueness (and its significance) and dismissed claims about error-free, positive identification as scientifically implausible. In recent years, these doubts have materialized in notorious mistakes, and scholarly criticisms endorsed in independent inquiries and reports—discussed in Section 2.[9]

---

[2] Often computer programs will assist by providing a ranked list of candidate prints that are highly similar: I.E. Dror and J. L. Mnookin, 'The use of technology in human expert domains: challenges and risks arising from the use of automated fingerprint identification systems in forensic science' (2010) 9 *Law, Probability & Risk*, 47 at 53 ('When comparisons get more and more challenging along certain dimensions, AFIS becomes ever less capable and the need for the human fingerprint expert becomes still more acute.').

[3] J.L. Mnookin, 'The validity of latent fingerprint identification: Confessions of a fingerprinting moderate' (2008) 7 *Law, Probability & Risk* 127.

[4] The contention that two prints—or shoe marks or bullets or handwriting exemplars—share a common source to the exclusion of all other possible sources. See S.A. Cole, 'Forensics without uniqueness, conclusions without individualization: the new epistemology of forensic identification' (2009) 8 *Law, Probability & Risk* 233; S.A. Cole. M. Welling, R. Dioso-Villa and R. Carpenter, 'Beyond the individuality of fingerprints: a measure of simulated computer latent print source attribution accuracy' (2008) 7 *Law, Probability & Risk* 165; M. J. Saks and J. Koehler, 'The Individualization Fallacy in Forensic Science Evidence' (2008) 61 *Vand L Rev* 199 (but c.f. D. Kaye, 'Probability, Individualization and Uniqueness in Forensic Science Evidence' (2009–10) 75 *Brook L Rev* 1163 particularly at 1176–7).

[5] National Research Council of the National Academy of Science, *Strengthening Forensic Science in the United States: A Path Forward* (Washington, DC: National Academies Press, 2009) at 139 (hereafter NRC Report); E.F. Loftus and S.A. Cole, 'Contaminated evidence' (2004) 304 *Science* 959.

[6] S.A. Cole, 'More than zero: Accounting for error in latent fingerprint identification' (2005) 95 *J Crim Law Crim* 985–1078; Federal Bureau of Investigation, *The Science of Fingerprints: Classification and Uses* (Washington, DC: DOJ, 1984). See also, Office of the Inspector Gen., U.S. Dep't of Justice, *A Review of the FBI's Handling of the Brandon Mayfield Case* (2006) at 8 ('Latent fingerprint identifications are subject to a standard of 100 percent certainty.').

[7] On experience, see Koehler, 'Proficiency tests to estimate error rates' ('Is not one hundred years of adversarial casework testing proof enough that the risk of error in fingerprint examination is extraordinarily low? No.') at 1086. See also Cole, Welling, Dioso-Villa and Carpenter, 'Beyond the individuality of fingerprints'; Haber and Haber, 'Scientific validation of fingerprint evidence under *Daubert*'; Vokey, Tangen and Cole, 'On the preliminary psychophysics of fingerprint identification'; Thompson, Tangen and McCarthy, 'Expertise in Fingerprint Identification' and more generally D. Kahneman, P. Slovic and A. Tversky (eds), *Judgment under Uncertainty: Heuristics and Biases* (New York: Cambridge University Press, 1982); D. Kahneman and G. Klein, 'Conditions for intuitive expertise: A failure to disagree' (2009) 64 *American Psychologist* 515.

[8] Consider G. P. Alpert, and J. J. Noble, 'Lies, True Lies, and Conscious Deception Police Officers and the Truth' (2009) 12 *Police Quarterly* 237. We note that false confessions are often obtained when suspects are confronted with other 'evidence' and sometimes a (plea) deal or promise. See e.g. B. Garrett, *Convicting the Innocent: Where Criminal Prosecutions Go Wrong* (Cambridge, MA: Harvard University Press, 2011). Ken Alder's work on the history of the polygraph suggests its primary value was in generating confessions: *The lie detectors: The history of an American obsession* (New York: Free Press, 2007).

[9] E.g. S.A. Cole, 'Who speaks for science? A response to the National Academy of Sciences report on forensic science' (2010) 9 *Law, Probability & Risk* 25–46.

This article presents a first iteration of what we envisage could be an evolving response to the vexed issue of the reporting (or expression) of forensic comparison evidence.[10] Conceived as a practical aid to assist with the presentation and interpretation of forensic science evidence, the guide to interpreting forensic science testimony (or Guide) is intended to embody the current state of relevant scientific research in relation to a particular technique (or set of techniques). This empirically predicated guide is designed to assist with the evaluation of evidence by highlighting areas of demonstrated expertise and incorporating an indicative error rate to assist with the assessment of expert opinion.[11] Using the example of fingerprints, it would enable a fingerprint examiner to express an opinion about whether two prints match (or do not match) against the backdrop of an empirically derived error rate and other indicators of expertise and its limitations. Introducing an error rate into the provision of comparison evidence assists with the evaluation of opinions and in delimiting the scope of expertise.[12]

2. Background to the Guide: evidence, expertise, error and advice

2.1 Authoritative reports and recommendations

In a landmark report issued in 2009, a committee of the National Research Council (NRC) of the US National Academy of Sciences (NAS) drew attention to questionable practices and the lack of research in many areas of forensic science. The committee was surprised to discover that many forensic science disciplines are typically not supported by scientific research and that analysts are not necessarily bound by experimentally derived standards to ensure the evidence offered in courts is valid and reliable.[13] In confronting language, the report itself states:

> Often in criminal prosecutions and civil litigation, forensic evidence is offered to support conclusions about "individualization" (sometimes referred to as "matching" a specimen to a particular individual or other source) or about classification of the source of the specimen into one of several categories. With the exception of nuclear DNA analysis, however, *no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source.*[14]

In relation to latent fingerprint comparison, the NRC report explicitly challenged the dominant *method*—that is, ACE-V.[15]

> ACE-V provides a broadly stated framework for conducting friction ridge analyses. However, this framework is not specific enough to qualify as a validated method for this type of analysis. ACE-V does not guard against bias; is too broad to ensure

---

[10] See J.J. Koehler, 'Fingerprint error rates and proficiency tests: What they are and why they matter' (2008) 59 *Hastings Law Journal* 1077; J.J. Koehler, 'Proficiency tests to estimate error rates in the forensic sciences' (2012) 0 *Law, Probability & Risk* 1–10.

[11] See also, G. Edmond, 'Advice for the courts: A multidisciplinary advisory panel?' (2012) 16 *International Journal of Evidence & Proof* 263–297.

[12] Generally, see T. Gieryn, *Cultural Boundaries of Science* (Chicago: University of Chicago Press, 1999).

[13] NRC Report, 39, 87. See comments by the co-chair, Judge H.T. Edwards, 'Solving the Problems that Plague the Forensic Science Community' (2010) 50 *Jurimetrics J* 5. For an earlier influential essay, see M. Saks and J. Koehler, 'The coming paradigm shift in forensic identification science' (2005) 309 *Science* 892.

[14] NRC Report, 7. (italics added).

[15] On ACE-V, see R.A. Huber, 'Expert Witnesses' (1959) 2 *Crim. LQ*, 276.

repeatability and transparency; and does not guarantee that two analysts following it will obtain the same results. For these reasons, merely following the steps of ACE-V does not imply that one is proceeding in a scientific manner or producing reliable results.[16]

The Committee also confronted and rejected the idea that fingerprint comparisons are free from error.

Errors can occur with any judgment-based method, especially when the factors that lead to the ultimate judgment are not documented. Some in the latent print community argue that the method itself, if followed correctly (i.e., by well-trained examiners properly using the method), has a zero error rate. Clearly, this assertion is unrealistic, and, moreover, it does not lead to a process of method improvement. The method, and the performance of those who use it, are inextricably linked, and both involve multiple sources of error (e.g., errors in executing the process steps, as well as errors in human judgment).[17]

The NRC report highlighted the absence of experiments on human expertise in forensic comparison (or pattern matching): 'The simple reality is that the interpretation of forensic evidence is not always based on scientific studies to determine its validity.' Going further, it concluded that: '[t]his is a serious problem'.[18] The Committee recommended that US Congress fund basic research to help the forensic community strengthen their research foundations, develop valid and reliable measures of performance, and establish evidence-based standards for analyzing and reporting results. The Committee placed emphasis on addressing the limited research base, determining error rates, as well as understanding and reducing the effects of bias and human error.[19]

The NRC is not alone in its expressed concerns about forensic science used for the purposes of identification. Subsequent to the NRC report, two prominent inquiries in the US and the UK have released reports focused directly on fingerprint evidence. One report was produced by the Expert Working Group on Human Factors in Latent Print Analysis (EWGHF)—a large multidisciplinary collective jointly sponsored by the United States National Institute of Standards and Technology (NIST) and National Institute of Justice (NIJ)—*Latent Print Examination and Human Factors*.[20] The other emerged from an inquiry conducted by Lord Campbell into problems with fingerprint evidence following the controversial McKie case in Scotland—*The Fingerprint Inquiry* (SFI).[21] These reports, from jurisdictions generally regarded as leading forensic science providers, are again surprisingly critical in their responses to widely accepted identification practices.

Convened in 2008, in the shadow of the NRC inquiry, the EWGHF was tasked with undertaking a scientific assessment of the effects of human factors on latent print analysis.[22] Specifically, the group

---

[16] NRC Report, 142. See also L. Haber and R.N. Haber, 'Scientific validation of fingerprint evidence under *Daubert*' (2008) 7 *Law, Probability and Risk* 87–109.

[17] NRC Report, 143.

[18] NRC Report, 8.

[19] NRC Report, pp. 14–33. See especially Recommendations 3 and 5.

[20] Expert Working Group on Human Factors in Latent Print Analysis (Editor in Chief: David H. Kaye), *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach* (U.S. Department of Commerce, National Institute of Standards and Technology, National Institute of Justice, 2012) (hereafter LPEHF Report).

[21] A. Campbell, *The Fingerprint Inquiry Report* (Edinburgh, Scotland: APS Group Scotland, 2011) (hereafter *The Scottish Fingerprint Inquiry* or SFI).

[22] LPEHF Report, at vi, states that: 'The study of human factors focuses on the interaction between humans and products, decisions, procedures, workspaces, and the overall environment encountered at work and in daily living. Human factors analysis can advance our understanding of the nature of errors in complex work settings. Most preventable, adverse events are not just the

was directed to evaluate current practices, to explain how human factors contribute to errors, and to offer guidance and recommendations. One way to obtain an impression of the Report's main thrust is to consider its recommendations, particularly those relating to the comparison of fingerprints and the expression of results. Relevant recommendations include:

> **Recommendation 3.3**: Procedures should be implemented to protect examiners from exposure to extraneous (domain-irrelevant) information in a case.
>
> **Recommendation 3.7**: Because empirical evidence and statistical reasoning do not support a source attribution to the exclusion of all other individuals in the world, latent print examiners should not report or testify, directly or by implication, to a source attribution to the exclusion of all others in the world.
>
> **Recommendation 3.9**: The federal government should support a research program that aims to:
>
> (a) Develop measures and metrics relevant to the analysis of latent prints;
> (b) Use such metrics to assess the reproducibility, reliability, and validity of various interpretive stages of latent print analysis; and
> (c) Identify key factors related to variations in performance of latent print examiners during the interpretation process.
>
> **Recommendation 6.3**: A testifying expert should be familiar with the literature related to error rates. A testifying expert should be prepared to describe the steps taken in the examination process to reduce the risk of observational and judgmental error. The expert should not state that errors are inherently impossible or that a method inherently has a zero error rate.
>
> **Recommendation 9.1**: Management should foster a culture in which it is understood that some human error is inevitable and that openness about errors leads to improvements in practice.
>
> **Recommendation 9.5**: The latent print community should develop and implement a comprehensive testing program that includes competency testing, certification testing, and proficiency testing.[23]

The Scottish inquiry into the controversy surrounding the mistaken attribution of a latent print collected from a crime scene to Shirley McKie (a police officer) also generated a large, though perhaps

---

result of isolated or idiosyncratic behavior but are in part caused by systemic factors.' See M. Sanders and E. McCormick. *Human Factors in Engineering and Design*, 7th ed. (New York, NY: McGraw-Hill, 1993); National Academy of Sciences, Institute of Medicine, Committee on Quality of Health Care in America, *To Err Is Human: Building A Safer Health System* (Washington, DC: McGraw-Hill Companies, 1999).

[23] LPEHF Report, 207–10. Other recommendations of special relevance include: 3.4, 3.6, 3.8, 4.3, 5.1, 6.2, 8.1, 8.4 and 9.2. The LPEHF Report was also critical of ACE-V as an adequate 'method' at 9, 39, 123–4: 'The focus on ACE-V is not intended as an endorsement of ACE-V as a "methodology." As explained in Chapter 1, ACE-V maps the steps of a process, but it does not provide specific functional guidance on how to implement that process, nor does it detail the substantive content of the various steps. Although ACE-V provides a useful framework for describing the steps taken for interpreting prints, it does not offer specific criteria to guide those interpretations.'

less systemically oriented, report along with a series of recommendations.[24] Under the heading 'The subjective nature of fingerprint evidence', recommendations from the SFI included:

> **Recommendation 1:** Fingerprint evidence should be recognised as opinion evidence, not fact, and those involved in the criminal justice system need to assess it as such on its merits.[25]
>
> **Recommendation 2:** Examiners should receive training which emphasises that their findings are based on their personal opinion; and that this opinion is influenced by the quality of the materials that are examined, their ability to observe detail in mark and print reliably, the subjective interpretation of observed characteristics, the cogency of explanations for any differences and the subjective view of 'sufficiency'.[26]
>
> **Recommendation 3:** Examiners should discontinue reporting conclusions on identification or exclusion with a claim to 100% certainty or on any other basis suggesting that fingerprint evidence is infallible.[27]

Beneath the heading 'Fingerprint methodology,' recommendations were particularly concerned with contextual bias:

> **Recommendation 6:** The SPSA [Scottish Police Services Authority] should review its procedures to reduce the risk of contextual bias.[28]

From these independent inquiries, supported by a range of scientific studies and pre-existing scholarly critiques, several consensus themes emerge that are consistent with this proposal.[29] Most prominent are: confirmation about the lack of scientific support for contemporary fingerprint comparison practice and underlying assumptions (NRC Rec. 3, LPEHF 3.9 and SFI 2, respectively); the rejection of claims about an infallible method and a zero error rate (NRC p.142, LPEHF 6.3 and SFI 3); and, concern about equating a declared 'match' with positive identification (NRC 3, LPEHF 3.7 and SFI 3).[30] The reports place considerable emphasis on the need for research, standards derived from research (NRC 1, 7, 8, LPEHF 3.4, 3.6, 3.8, 3.9, 8.1), the need to attend to a range of potential biases, and the possibility of shielding analysts from some kinds of information (NRC 5, LPEHF 3.3 and SFI 6, 7, 8). The reports also recognize the need to present opinions derived from fingerprints in a manner that embodies their value and is simultaneously comprehensible to the tribunal of fact (NRC Rec. 2, LPEHF 5.1, SFI 64). The LPEHF Report (Rec. 4.3, 6.3, 9.1, and 9.2) directs attention to the need for examiners to be familiar with error rates, probabilities and statistics and the SFI (Rec. 82, 83) advocates the development of probabilities.

---

[24] See S. Cole and A. Roberts, 'Certainty, Individualisation, and the Subjective Nature of Expert Fingerprint Evidence' [2012] *Criminal Law Review* 824–849.

[25] SFI, para. 35.132.

[26] SFI, para. 35.133.

[27] SFI, para. 38.77.

[28] SFI, para. 35.137. See also Recommendations 7 and 8 at paras 35.138 and 35.139.

[29] L. Haber and R.N. Haber, *Challenges to Fingerprints: A Guidebook for Prosecution and Defense and Examiners* (Tucson, AZ: Lawyers & Judges Publishing Company, 2009); Haber and Haber, 'Scientific validation of fingerprint evidence under *Daubert*'; M.J. Saks & J. Koehler, 'The coming paradigm shift'; M. Saks and D. Faigman, 'Failed forensics: How forensic science lost its way and how it might yet find it' (2008) 4 *Annual Reviews of Law & Social Science* 149; Cole, 'More than zero'; S.A. Cole, 'Who speaks for science? A response to the National Academy of Sciences report on forensic science'; I. Dror and S. Cole, 'The vision in "blind" justice'; G. Edmond and K. Roach, 'A contextual approach to the admissibility of the state's forensic science and medical evidence' (2011) 61 *University of Toronto Law Journal* 343.

[30] References are to the recommendations in the reports.

For the average reader—whether lawyer, judge or potential juror—all of this might come as something of a surprise. For, notwithstanding long reliance on fingerprint evidence, relatively little is known about the performance of fingerprint examiners or the value of their opinions. Contemporary investigative practices and reporting appear to fall well short of the recommendations and advice outlined in the independent reports. Concerns expressed by the NRC, EWGHF, and Lord Campbell, along with several notorious cases of fingerprint misattribution, raise serious (and as yet unresolved) questions about the forensic use of fingerprint evidence.[31] There is, however, an indisputable need to reform the way fingerprint examiners work as well as the manner in which they express their opinions.

Currently, there is a dearth of research. The necessary studies are often beyond the capabilities and competence of fingerprint examiners (and yet to be undertaken, or completed). Understandably, the training of fingerprint examiners is primarily oriented toward comparing fingerprints. Most do not have the methodological skills, funding, time, infrastructure, or experience with research techniques to mount scientific studies of human performance. Moreover, few fingerprint examiners, lawyers, or judges have the time, resources or expertise to track and evaluate extant studies, inquiries and reports, or respond to research as it emerges.[32] Consequently, changes to practices and reporting will require the ongoing assistance of research scientists.[33] Research into expertise and complex sociotechnical systems is the domain of cognitive science and human factors. Researchers in these areas already have the infrastructure in place to conduct the requisite studies, and are well positioned to work with examiners to strengthen the field.

2.2 *Emerging studies*

Scholarly criticisms and recent inquiries have already spawned a range of studies. The first studies focused on consistency and bias in expert decisions, but it is difficult to glean indicators on human matching performance from them.[34] Most of the research is currently in progress, though three studies have recently been published. In a controlled fingerprint matching experiment, Tangen *et al.* found that examiners incorrectly declared 0.68% of similar non-matching prints as 'matching' (false positive errors)—compared to 55.18% for lay persons—and 7.88% of matching prints as 'non-matching' (false negative errors).[35] In a similar experiment that made use of genuine crime scene prints, where the

---

[31] For example, the case of Brandon Mayfield. See Cole, 'More than zero'.

[32] See J. Mnookin, S.A. Cole, I.E. Dror, B.A.J. Fisher, M. Houck, K. Inman *et al*, 'The need for a research culture in the forensic sciences' (2011) 58 *UCLA Law Review* 725–779 at 725, for discussion ('. . . most practicing forensic scientists in pattern and impression evidence, and in most other forensic disciplines as well, are not actually qualified to pursue the necessary research. Until recently, many laboratories did not necessarily require a college degree or any formal science training. Even those with a BS in forensic science or some other scientific discipline have not typically received significant training in the development of research design'.).

[33] In practice, there will be a need for examiners and scientists to work together in ways that slowly refine practices and develop important research skills among leaders in the various forensic science communities.

[34] K. Wertheim, G. Langenburg and A. Moenssens, 'A report of latent print examiner accuracy during comparison training exercises' (2006) 56 *Journal of Forensic Identification* 55–93; I.E. Dror, C. Champod, G. Langenburg, D. Charlton, H. Hunt and R. Rosenthal, 'Cognitive issues in fingerprint analysis: Inter- and intra-expert consistency and the effect of a "target" comparison' (2011) 208 *Forensic Science International* 10–17; G. Langenberg, 'Performance study of the ACE-V process: A pilot study to measure the accuracy, precision, reproducibility, repeatability, and biasability of conclusions resulting from the ACE-V process' (2009) 59 *Journal of Forensic Identification* 219–257; L. Haber and R.N. Haber, 'Letter to the editor. Re: A report of latent Print examiner accuracy during comparison training exercises' (2006) 56 *Journal of Forensic Identification* 493–499.

[35] J.M. Tangen, M.B. Thompson and D.J. McCarthy, 'Identifying fingerprint expertise' (2011) 22 *Psychological Science* 995–997; see also M.B. Thompson, J.M. Tangen, and D.J. McCarthy, 'Expertise in fingerprint identification' (2013) *Journal of Forensic Sciences*. doi: 10.1111/1556-4029.12203.

ground truth is uncertain, they found that examiners incorrectly declared 1.65% of similar non-matching prints as 'matching' (false positive errors)—compared to 55.73% for lay persons—and 27.81% of matching prints as 'non-matching' (false negative errors).[36] In another controlled finger-print matching experiment, Ulery *et al*. found that examiners incorrectly declared 0.1% of similar non-matching prints as 'matching' (false positive errors) and 7.5% of matching prints as 'non-matching' (false negative errors).[37] These results demonstrate that qualified, court-practicing fingerprint exam-iners were far more accurate (and more conservative) than laypersons, and that the rate of false positive errors (i.e. incorrectly reporting that non-matching fingerprints match) in these experimental matching tasks was around 1% and the rate of false negative errors (i.e. incorrectly reporting that matching fingerprints do not match) ranged from 8% to 28%. For criminal justice systems that have routinely relied upon fingerprint evidence for convictions and pleas, these preliminary results should come as a great, if necessarily partial, relief. They suggest that fingerprint examiners have genuine expertise in *discriminating between prints that match and those that do not*.

In conjunction with the findings and recommendations in the various reports, these studies provide a platform upon which to begin reforming the way opinions about fingerprints are represented and used in legal settings. Our proposed guide to interpreting forensic science testimony begins to address some of the conspicuous deficiencies in contemporary fingerprint practice, especially in the reporting and explanation of results. This proposal attempts to take seriously some of the de-stabilizing epistemic and organizational problems raised in scholarly critiques and the recent authoritative and independent inquiries and reports.

We aim, with this proposal, to enhance legal performance by directing attention toward actual abilities, based on emerging evidence. It is clear that fingerprint identification cannot be regarded as an infallible 'methodology' that is detached from human judgement.[38] Given the long history of claims about uniqueness, individualization, and a disembodied identification processes, examiners and their institutions should now begin to replace traditional practices and reporting with evidence-based claims that reflect actual capabilities.[39] Regardless of what forensic scientists do, criminal courts have a principled obligation to truth and justice.[40] Courts, particularly those jurisdictions with a reliability-based admissibility standard, have an obligation to require forensic scientists to present their evidence in ways that embody actual capabilities. This requires evaluating reliability and conveying limitations clearly to the tribunal of fact.

---

[36] M.B. Thompson, J.M. Tangen, and D.J. McCarthy, 'Human matching performance of genuine crime scene latent finger-prints' (2013) *Law and Human Behavior*. doi: 10.1037/lhb0000051

[37] B.T. Ulery, R.A. Hicklin, J. Buscaglia and M.A. Roberts, 'Accuracy and reliability of forensic latent fingerprint decisions' (2011) 108 *Proceedings of the National Academy of Sciences of the United States of America* 7733–7738. See also B.T. Ulery, R.A. Hicklin, J. Buscaglia and M.A. Roberts, 'Repeatability and reproducibility of decisions by latent fingerprint examiners' (2012) 7(3) e32800 *PLoS ONE*. doi:10.1371/journal.pone.0032800.t007.

[38] Cole, 'More Than Zero', ('There is no methodology without a practitioner, any more than there is automobile without a driver, and claiming to have an error rate without the practitioner is akin to calculating the crash rate of an automobile, provided it is not driven.'); see also J.M. Tangen, 'Identification personified' (2013) *Australian Journal of Forensic Sciences*. doi:10.1080/00450618.2013.782339

[39] Unremarkably, perhaps, professional forensic bodies are beginning to revise their practices and standards and contemplat-ing reporting error rates (e.g. SWGFAST and IAI). See R. Garrett, 'Memorandum from the President of the International Association for Identification' (19 February 2009) International Association for Identification, 2009; Scientific Working Group On Friction Ridge Analysis Study And Technology (SWGFAST), 'Standard for the definition and measurement of rates of errors and non-consensus decisions in friction ridge examination (latent/tenprint)' (16 September 2011) Ver. 1.1.

[40] H. L. Ho, *A Philosophy of Evidence Law: Justice in the Search for Truth* (Oxford: Oxford University Press, 2008).

In addition, we take seriously concerns, such as those recently voiced by the Law Commission of England and Wales, about the criminal trial and its limitations with expert opinion evidence.[41] The historically accommodating response to fingerprint evidence, the few substantial challenges, and the vanishingly small number of appellate decisions suggest a legal reluctance (or inability) to appreciate the significance of problems with fingerprint evidence.[42]

Through the provision of a guide, it is our intention to integrate some of the recommendations and emerging research to produce a serviceable tool to assist the legal regulation and use of fingerprint evidence. The Guide is intended to help with the expression and interpretation of opinions about fingerprints by locating them within the appropriate research matrix. We envisage that a version of the Guide would be appended to expert reports prepared by fingerprint examiners, although we also envisage an updated Guide available through a publicly accessible repository.[43] The Guide represents a pragmatic attempt to acknowledge and explain actual abilities as well as non-trivial limitations with fingerprint evidence. It is intended to recognize the existence of genuine expertise in comparison work, expose the weak decision-making framework and problem of extrapolation (i.e. the 'leap' from match to identification), as well as address the historical reluctance to make appropriate concessions in reports and testimony.

3. Insights from medicine: the diagnostic model

In modern diagnostic medicine, the accuracy of a test is inferred from controlled experiments. In home pregnancy testing, for example, a pregnancy test produces a result that reads 'pregnant' or 'not pregnant'—based on the level of human chorionic gonadotropin (hCG) in urine used as a marker for pregnancy—which may or may not agree with the true state of the world. The validity, reliability, and accuracy of the test come from the aggregation of many controlled experiments. So, in a particular case (i.e. when a woman takes a pregnancy test) we can use this aggregated information to infer something about her true state.

Who compared, Fig. 1A depicts the results from an experiment by Tomlinson *et al.*[44] comparing the accuracy of six home pregnancy tests available over-the-counter. The numbers in Fig. 1A represent groups of women who were pregnant or not and who took the Answer™ home pregnancy test,[45] which either resulted in a reading of 'pregnant' or 'not pregnant'. One hundred and twenty pregnant women were given the Answer™ home pregnancy test, 98 of the tests correctly read 'pregnant' and 22 incorrectly read 'not pregnant'. Similarly, 120 women who were not pregnant were given the Answer™ home pregnancy test, 2 of the tests incorrectly read 'pregnant' and 118 correctly read 'not pregnant'.

[41] Law Commission, *Expert Evidence in Criminal Proceedings in England and Wales* (London: The Stationery Office, 2011), para. 1.20; G. Edmond, 'Is reliability sufficient? The law commission and expert evidence in international and interdisciplinary perspective' (2012) 16 *International Journal of Evidence & Proof* 30. See also NRC Report, 85.

[42] Three exceptions, of sorts, include: *United States v Llera Plaza*, 179 F Supp 2d 492, 517 (ED Pa, 2002); *R* v. *Smith* [2011] EWCA CRIM 1296 and Order on Defendant's Motion in Limine, *State* v. *Borrego*, Nos. F12-101 & F12-7083, at 16 (Fla. Cir. Ct. Oct. 25, 2012). Most of the sophisticated challenges have taken place in United States courts, see D.H. Kaye, D.E. Bernstein and J.L. Mnookin, *The New Wigmore: A treatise on Evidence – Expert Evidence*, 2nd ed. (New York, NY: Aspen Publishers, 2011).

[43] Expert reports should contain more detail than at present, see LPEHF Report Recommendation 5.2, and B. Found and G. Edmond, 'Reporting on the comparison and interpretation of pattern evidence: recommendations for forensic specialists' (2012) 44 *Australian Journal of Forensic Sciences* 193–196, as an indication of the kinds of information and considerations that ought to be included.

[44] C. Tomlinson, J. Marshall, and J.E. Ellis, 'Comparison of accuracy and certainty of results of six home pregnancy tests available over-the-counter' (2008) 24(6) *Current Medical Research and Opinion* 1645–1649.

[45] Church & Dwight, Princeton, NJ, USA.

GARY EDMOND *ET AL.*

A	Test said "pregnant"	Test said "not pregnant"		B	Expert said "match"	Expert said "no match"	
Pregnant	98	22	120	Same source	409	35	444
Not pregnant	2	118	120	Different source	3	441	444
	100	140			412	476	

Fɪɢ. 1. Pregnancy test results from Tomlinson *et al.* (2008) and expert fingerprint matching results from Tangen *et al.* (2011).

If a woman purchases an Answer^TM home pregnancy test from the chemist, and tests herself, what could she conclude on the basis of the experiment by Tomlinson and colleagues? If the home pregnancy test read 'pregnant' in this particular case, whether the woman is in fact pregnant (like the 98 for whom the test produced the correct reading) or whether she is not (like the 2 for whom the test produced the incorrect reading), we do not know. Similarly, if the home pregnancy test read 'not pregnant' in this particular case, whether the woman is in fact pregnant (like the 22 for whom the test produced the incorrect reading) or whether she is not (like the 118 for whom the test produced the correct reading), we do not know. This uncertainty does not render the woman helpless; rather, the information could inform her interpretation of the test result (and the question of pregnancy).

We can apply the same diagnostic model to the experiment by Tangen *et al.* on the matching performance of court practicing fingerprint examiners. The results from this experiment are depicted in Fig. 1B. A group of 37 qualified fingerprint examiners examined 444 pairs of fingerprints from the same person, 409 were correctly declared as a 'match', and 35 were incorrectly declared 'no match'. Similarly, the examiners examined 444 pairs of fingerprints from different people, 3 were incorrectly declared as a 'match', and 441 were correctly declared 'no match'.

If a juror hears an examiner give an opinion about whether two prints match (or not) in a criminal case, what could the juror conclude on the basis of the experiment by Tangen and colleagues? If the examiner declared a 'match' we do not know in this particular instance whether the prints are from the same source (like the 409 for which a 'match' opinion was correct) or whether the prints are not from the same source (like the 3 for which a 'match' opinion was incorrect). Similarly, if the examiner said "no match" we do not know whether the prints are from the same source (like the 35 for which a 'no match' opinion was incorrect) or whether the prints are not from the same source (like the 441 for which a 'no match' opinion was correct). This uncertainty does not render the juror helpless; rather, the information could inform his interpretation of the examiner's opinion (and the question of source).

We suggest that an indication of performance (and error) in previous situations, (reasonably) similar to the particular analysis, provides potentially valuable information to those obliged to evaluate fingerprint testimony. This 'statistical base rate' is general information. The juror can reason with this information to infer something about the particular case—to deduce the particular from the general. The juror can also use information about the particular case ('causal base rates') to temper these judgements, if they think the information is relevant. Judgements can be anchored to a plausible base rate and tuned by reasoning, informally, about the information specific to the particular case.[46]

Broadly, the diagnostic model is an approach that offers information about similar situations in order to help decision-makers reason about the present case. The goal for a diagnostic model applied to forensic testimony is to give information to the legal participants to assist their decision-making

[46] D. Kahneman, *Thinking, Fast and Slow* (New York, NY: Farrar, Straus and Girou, 2011).

around admissibility, challenges to evidence, instructions and warnings, and for the jury around the value of evidence and the ultimate conclusion. The goal is to provide information in a way that will help the jury to weigh the evidence, evaluate the arguments, and to judge the degree of belief warranted by the information presented.[47] Often this will involve information about general, or indicative, error rates and practical limitations. As in the diagnostic model, much of the information (i.e. scientific evidence) that can be presented will be based on general data from previous studies (i.e. from beyond the instant case) and the legal participants and fact-finder must reason and make inferences from the general to the particular case.[48]

Little is currently known about the types and forms of information that will assist triers of fact to make optimal decisions. The Guide is presented as a pragmatic intervention: an evolving compromise that endeavours to provide a diagnostic-style framework to improve forensic reasoning. An in-depth treatment of the diagnostic model applied to forensic testimony is forthcoming, but our goal, in the first instance, is to help to ensure that expert evidence is presented in ways that are scientifically tenable. This involves embodying its known value and disclosing limitations in ways that help triers of fact to make sensible decisions about the forensic science evidence in a particular case.

4. A Guide to forensic testimony: fingerprints

This section provides an example of what a guide for fingerprint evidence proffered for identification might look like at this stage. That is, the kind of information or caveats that ought to be included with the fingerprint examiner's report and testimony. It is an intentionally short document that places emphasis on brevity, comprehensibility and the goal of capturing both the considerable evidentiary potential as well as known limitations. Requiring ongoing revision—at least until there are sufficient studies to support a stable consensus—this preliminary version is based on the few scientific studies that have assessed the performance of fingerprint examiners in circumstances where conditions were deliberately controlled. In the remainder of this article we endeavour to unpack the Guide and some of the implications of the recent reports and emerging studies in ways that are sensitive to the criminal justice milieu, and especially the criminal trial.

A Guide to Forensic Testimony: Fingerprints

A decision about whether two fingerprints match or not is based on the judgment of a human examiner, not a computer.

There are several documented cases where an examiner has incorrectly said that two prints 'match' when they actually came from two different people. Laboratory-based experiments suggest that errors of this sort happen infrequently (around 1% of the time).

In practice, however, it is unknown how often examiners say that two fingerprints match when they actually come from two different people.

Without specific evidence, it cannot be known whether an error has occurred in a particular case.

For further information see www.InterpretingForensicTestimony.com

---

[47] S. Pinker, *How the Mind Works* (New York, NY: W. W. Norton & Company, 1997).

[48] D. L. Faigman, J. Monahan and C. Slobogin, 'Group to Individual (G2i) Inference in Scientific Expert Testimony' (July 26, 2013). Available at SSRN: http://ssrn.com/abstract=2298909.

GARY EDMOND *ET AL.*

5. Specific issues arising from the Guide

This proposal is intended to begin the process of practically reforming the way fingerprint (and other types of comparison) evidence is presented—in reports and testimony—and evaluated in courts. While the Guide may be controversial—with fingerprint examiners, other forensic scientists and commentators—it will be virtually impossible to generate consensus in this area. Even in the absence of complete consensus, the fundamental nature of problems with the comparison sciences identified by the NRC, the EWGHF and Lord Campbell, in conjunction with the espoused goals of criminal justice (particularly rectitude, the concern with protecting the innocent, and the need for fair criminal proceedings), mean that we should not persist with our current practices. There is a need to develop some empirically based mechanism to constrain the way opinions about fingerprints are expressed and improve the way lawyers, judges and jurors evaluate fingerprint evidence. Current legal practice tends to be either indifferent to criticism, or has attempted to craft responses on the run. These responses are generally defensive (of past legal practice) and not genuinely engaged with the reports, their recommendations and (emerging) empirical evidence—see Section 6.

There are many difficulties associated with the attempt to ascertain the value of fingerprint evidence, both generally and in specific cases. The commitments of our criminal justice systems, in conjunction with authoritative disclosures about fingerprint (and other forensic science) evidence, however, would seem to dictate the need for the state to present its incriminating expert evidence in ways that embody the value of the evidence. Unavoidably, this requires the proactive disclosure of limitations. Forensic science evidence cannot be admitted on the basis that scientifically notorious limitations are a matter for the trial and weight—to be drawn out through cross-examination, rebuttal experts and judicial instructions—should the evidence be contested. Rather, when it comes to comparison techniques in routine use, it would seem incumbent upon the state to support proffers with evidence of ability and accuracy, and to clearly and effectively consider and explain limitations.

In many areas of forensic science and medicine, levels of error are knowable yet unknown even though there are sometimes good reasons to think that practices and interpretations are error prone. Facial and body mapping, gait analysis, voice comparison, bite marks, blood spatter, document comparison, as well as foot and shoe prints, tool marks, ballistics/firearms, soil, and non-DNA hair comparison are all conspicuous examples of forensic techniques that appear to lack the requisite research (or empirical) foundation.[49] In these areas, the accused is typically left to expose limitations with techniques, as well as the performance of investigators, retrospectively rather than require the state to have studied techniques and tested examiners in order to provide some indication of the existence of expertise and the kinds of errors that are associated with analyses and inferences. While the Guide targets fingerprint evidence, we can imagine similar and in some ways generic versions applied to virtually all expert reports and testimony concerned with identification or sourcing.[50]

---

[49] NRC Report at 7-8, 87. Voice and image comparison are conspicuous examples: see G. Edmond, K. Biber, R. Kemp and G. Porter, 'Law's looking glass: Expert identification evidence derived from photographic and video images' (2009) 20 *Current Issues in Criminal Justice* 337–377; G. Edmond, K. Martire and M. San Roque, 'Unsound law: Issues with ("expert") voice comparison evidence' (2011) 35 *Melbourne University Law Review* 52–112.

[50] Here, it is worth noting that even empirically robust techniques, such as DNA profiling, should be included beneath this rubric. While most technical DNA processes used in criminal justice contexts have been validated, some significant parts of the process are vulnerable to interpretive error and could be readily improved through the provision of information about error. The interpretation of mixed DNA samples is a good example. See W.C. Thompson, 'Forensic DNA Evidence: The Myth of Infallibility' in S. Krimsky and J. Gruber (eds.), *Genetic Explanations: Sense and Nonsense* (Cambridge, MA: Harvard University Press, 2013).

As a heuristic, a guide has the added benefit of informing the use of forensic science and medical evidence not only at trial and on appeal, but also during investigations, decisions to prosecute, and in plea negotiations. In this way, the Guide is practically oriented to the needs of the criminal justice system and its various personnel. In relation to criminal prosecutions, the Guide places both the prosecutor and the accused in an equivalent position vis-à-vis the reliability of a technique and the probative value of the evidence. It is, in addition, intended to discipline forensic scientists (in court) such that they restrict their opinions to assertions that can be supported by current scientific research.

5.1   Experimental evidence and the limitations of empirical error estimates

So far, experiments on the matching accuracy of examiners have been tightly controlled and deliberately artificial in order to balance fidelity, generalizability, and control.[51] They were not designed to resemble the everyday operations of a fingerprint bureau. For example, the examiners studied by Tangen and his colleagues were prevented from making 'inconclusive' judgements, they did not have their usual tools available to zoom, rotate, or apply filters to images, there was no peer assessment or 'verification' of the prints, and so on. On the other hand, the examiners were not provided with any contextual information about the demographics of the source individual, the severity or nature of the case, or attributes or conditions of the latent print, which could potentially sway their judgement. Moreover, the prints were from known sources—i.e. there was no uncertainty about ground truth.[52] Consequently, the error rates incorporated into this first iteration of the Guide are based entirely on the performance of examiners within controlled environments, and limited to assessing the ability to discriminate between matching and non-matching prints.

Errors, however, can arise at any stage of the process, from collecting latent prints at the crime scene to providing testimony in court. The frequency, severity, and kind of error may depend on factors such as an examiner's training and experience, the nature of peer review, the quality of the recovered print (or image), the type of surface and method of retrieving or imaging, the efficiency of the search algorithms used to retrieve the corresponding ten-print candidate, as well as a range of formal and informal practices such as exposing analysts to prejudicial, though domain irrelevant, information.[53] Several experiments are currently being conducted on various aspects of an examiner's workflow across different laboratories, and will ultimately measure the performance of entire pre-trial systems. These will be invaluable for refining estimates of error and should be incorporated into subsequent iterations of the Guide.

It might be argued that the available studies are too limited, lacking ecological validity, to begin to reform the presentation of fingerprint evidence. They do not, after all, address the leap of faith inference from matching to identification.[54] Our response is as follows. First, we are drawing upon the recommendations of authoritative reports by well-credentialed independent bodies. Secondly, although

---

[51] See Thompson, Tangen, & McCarthy, 'Expertise in fingerprint identification' for discussion.

[52] This is a much safer basis than the usual claims about accuracy based on convictions and confessions or the lack of disclosed errors (which assumes the criminal justice system will identify them). L. Haber and R.N. Haber, 'Scientific validation of fingerprint evidence under *Daubert*'.

[53] We mean information that is not relevant to the actual comparison exercise, even if it might be highly probative to the actual facts in issue.

[54] The Guide conveys the difficulty examiners have in breaking out of the matching 'loop' to attach evidentiary significance.

GARY EDMOND *ET AL.*

preliminary and likely to be refined, the validity studies by Ulery *et al.* and Tangen *et al.* (and Wertheim *et al.*, Langenberg *et al.*, and Dror *et al.*), are consistent and all we have in the way of scientific studies at this point in time. Thirdly, failing to respond to the recommendations and insights about the performance of fingerprint examiners maintains the status quo and the dangers associated with equating opinions about matches with error-free identification. Finally, only carefully controlled experiments have the precision and capacity to isolate the conditions and causes of mistakes, and contribute to the development of a system that is resilient to error; making it harder for people to do something wrong and easier for them to do it right—see NRC 3, LPEHF 9.1, 9.2.[55]

5.2   *Uniqueness, variability, and error*

> Uniqueness and persistence are necessary conditions for friction ridge identification [i.e., fingerprint comparison] to be feasible, but those conditions do not imply that anyone can reliably discern whether or not two friction ridge impressions were made by the same person. Uniqueness does not guarantee that prints from two different people are always sufficiently different that they cannot be confused, or that two impressions made by the same finger will also be sufficiently similar to be discerned as coming from the same source. The impression left by a given finger will differ every time, because of inevitable variations in pressure, which change the degree of contact between each part of the ridge structure and the impression medium. None of these variabilities—of features across a population of fingers or of repeated impressions left by the same finger—has been characterized, quantified, or compared.[56]

Examiners have previously claimed that fingerprint identification is infallible and that there is a zero error rate for fingerprint comparisons.[57] These assertions are typically justified by reference to the uniqueness of prints and their longstanding use for purposes of identification.[58] Nevertheless, examiners make mistakes. Courts should recognize that errors are not due to people having identical (i.e. non-unique) fingerprints; errors are due to examiners incorrectly matching prints that are not from the same source and failing to match prints that are from the same source.[59]

In many cases, latent prints can be used to assist with identification by means of their classification as a 'match' or 'non-match'. Qualifying the meaning of a match with an indication of error is intended to draw attention to the non-trivial risk of a mistake in the process of classifying two prints as a match or non-match. In the absence of more accurate comparison practices, or information about the value of apparent matches, such an approach endeavours to reinforce the evidentiary value and reliability of

---

[55] E.g. Institute of Medicine, *To Err Is Human: Building A Safer Health System*; D.D. Woods. L. Johannesen, S. Dekker, R. Cook, N. Sarter, *Behind Human Error*, 2nd ed. (Burlington, TV: Ashgate Publishing, Ltd., 2010).

[56] NRC Report, 144.

[57] Cole, 'More than zero'; FBI, *The Science of Fingerprints*.

[58] In recommending a reliability standard for the admission of expert opinion evidence, even the Law Commission of England and Wales simply assumed that fingerprint evidence—was basically incontrovertible evidence of identification. Notably, this was after more than a century of admission, though before the recent reports and the results of the first validation studies. See Law Commission, *Expert Evidence in Criminal Proceedings in England and Wales,* para. 3.65.

[59] Because the 'methodology' of fingerprint identification cannot be detached from human judgement, the all-too-human foibles of distraction, lapses of attention, fatigue, rushes to judgment, less than perfect information, biases and expectations cannot be avoided even by the most diligent professionals. See J.R. Vokey, J.M. Tangen and S.A. Cole, 'On the preliminary psychophysics of fingerprint identification' (2009) 62 *The Quarterly Journal of Experimental Psychology* 1023–1040. Furthermore, mistakes can happen at any point from the way that prints are collected, stored, filed, or retrieved.

latent fingerprint evidence while recognizing real limitations—including uncertainty around the inference to identification (or 'leap'). The expression of an indicative, or general, error rate recognizes that comparison processes are fallible in circumstances where we are not entirely sure what a match actually means.

Opinions about the uniqueness of fingerprint features must be considered against the demonstrated abilities of examiners. For example, there is no evidence that examiners can judge the statistical base rate of particular fingerprint configurations. Even if they could, there is still a degree of human error associated with such judgements. Such abilities and the level of performance should be empirically demonstrated rather than asserted. Common problems with claims of uniqueness persist, in that examiners have to agree with each other (and themselves) on what counts as a feature, acceptable variations of these features must be specified, and there needs to be a database to draw from that is impervious to the noise and ambiguity that is intrinsic to crime scene prints—see LPEHF Rec. 3.6. In the end, any judgement about the relative occurrence of particular features is a human judgement that is unavoidably prone to error. Claims about a 'methodology' that is detached from human judgement must necessarily be 'off limits'.

5.3   *Towards identification: what might a 'match' mean?*

There are two separate issues pertaining to the presentation of fingerprint evidence that have been conflated historically. Through their concentrated focus, the validation studies help to distinguish them. The first involves the ability of examiners to match a pair (or set) of prints. As we have explained, trained and experienced examiners tend to be good at this, but they are not free from interpretative error. Second, is the issue of what a declared match means in terms of identification. We have referred to this as the inferential leap of faith. The second issue is quite complicated, and there is little evidence that examiners have relevant expertise. Historically, fingerprint examiners assumed that all humans have unique fingerprints (and implicitly, that all the prints they were willing to characterize as 'matching' were produced by a single identifiable source) and, in consequence, if they declared a match they had identified a particular individual to the exclusion of all other persons. As the reports explain, these assumptions are neither empirically based nor plausible. While more work needs to be done on fingerprint comparisons and the conditions in which they are made, the main problem at this stage is moving from a match decision to the attribution of significance in terms of identification. Patently, the declaration of a match does not equate with positive identification. However, on average, a declared match will be probative on the question of identity. The dilemma is how can we make sense of the match evidence. That is, how should we evaluate a putative match?

Courts have experimented with several methods of managing this problem—not only in relation to fingerprint evidence. Apart from probabilistic approaches, associated with DNA evidence, two of the more prominent methods are to limit testimony to the mere description of similarities (so-called 'splitting') and the use of verbal scales.[60] The first approach is conspicuous across a range of emerging comparison sciences. Judges, concerned about underlying evidence, particularly the distribution and frequency of features (whether sub-features of fingerprints or facial features in image comparison), have on occasion-restricted analysts to describing similarities and/or differences. Such an approach

---

[60] See, for example, S.A. Cole, 'Splitting hairs? Evaluating "Split Testimony" as an approach to the problem of forensic expert evidence' (2011) 33 *Sydney Law Review* 459–485.

GARY EDMOND *ET AL.*

would enable a fingerprint examiner to report a match, but to say no more. They would be unable to venture an opinion on the significance of similarities between two prints.

For a variety of reasons this is not an appropriate response to latent fingerprint evidence. First, in the absence of information about the distribution and relationship between fingerprint minutiae, and particularly a range of issues related to differences between all (including same source) fingerprints and fragments of fingerprints, the move from similarities to positive identification is problematic.[61] Secondly, the cultural familiarity with fingerprint evidence, and equating a match with positive identification, is such that describing similarities—regardless of the precise nomenclature—will be a de facto identification and understood as positive identification.[62] On policy grounds, thirdly, fingerprint techniques have been around for so long that it seems inappropriate to simply excuse the tardy performance of examiners in this way. Allowing fingerprint examiners to express opinions about similarities does not address the issue of their rate of error in matching, nor moderate the significance of any match relative to identification. It also presents the jury with a series of similarities without providing rational means of attaching significance.[63] In theory, splitting, removes the probative value of the evidence (to a point that threatens its logical relevance), and in practice it continues to imply positive identification.[64]

A second approach, enabling forensic scientists to go beyond merely describing similarities, is the use of a verbal scale to attach evidentiary significance to the match (or similarities). Image comparison evidence, so-called facial mapping, is a good example. In the absence of DNA-style databases or information about the distribution and independence of facial features, image comparison witnesses have been allowed to move from alleged similarities (or matches) between persons in images to opine about their significance in terms of identification. In Australia, though formally restricted to the description of similarities (and, in theory, differences), there has been slippage with image comparison witnesses testifying in terms of 'high level of anatomical similarity'.[65] In England and Wales, where there are no such prohibitions on positive identification, in recent years image (and other types of) comparison witnesses have adopted verbal scales, such as the one reproduced in Fig. 2, to express their incriminating opinions.[66]

Apart from criticizing the use of numbers to rank the verbal equivalents (in *Atkins*), English courts have tended to accept such expressions provided the jury is informed that they are not derived from a database.[67] The problem, of course, is that the move (or leap) from alleged similarities in facial

---

[61] See also Section 5.2.

[62] We cannot, as Wittgenstein recognized, easily change (or prescribe) the way language will be used and understood. This applies to terms used by expert witnesses, lawyers, judges and jurors. See D. McQuiston-Surrett and M.J. Saks, 'Communicating opinion evidence in the forensic identification sciences: Accuracy and impact' (2008) 59 *The Hastings Law Journal* 1159–1159 and, more generally, L. Wittgenstein, *Philosophical Investigations* (Oxford: Basil Blackwell, 1953).

[63] This is the potential role of probabilistic approaches, and part of the value of an error rate. See Koehler, 'Proficiency tests to estimate error rates'.

[64] Indeed, this is often the very conspicuous implication, see G. Edmond, 'Specialised knowledge, the exclusionary discretions and reliability: Reassessing incriminating expert opinion evidence' (2008) 31 *UNSW Law Journal* 1–55. It is useful to contrast approaches to DNA profiles, see National Research Council, Committee on DNA Technology in Forensic Science, *DNA Technology in Forensic Science* (Washington, DC: National Academies Press, 1992) 74. For a clear and sophisticated exposition, see D. Kaye, *DNA and the Law of Evidence* (Cambridge, MA: Harvard University Press, 2011).

[65] See *Morgan* v. *R* [2011] NSWCCA 257.

[66] See Edmond, Biber, Kemp and Porter, 'Law's looking glass'; Edmond, Martire and San Roque, 'Unsound law'; G. Edmond, R. Kemp, G. Porter, D. Hamer, M. Burton, K. Biber and M. San Roque, '*Atkins v The Emperor*: The "Cautious" use of Unreliable "Expert" Opinion' (2010) 14 *International Journal of Evidence & Proof* 146–165.

[67] *R* v. *Atkins and Atkins* [2009] EWAC Crim 1876 (facial mapping). Such tables tend—even if the assumptions are unknown by analysts or not made explicit to the tribunal of fact—to be informed by Bayesian commitments. See also *R* v. *Dlugosz* [2013]

Level	Description
0	Lends no support
1	Lends limited support
2	Lends moderate support
3	Lends support
4	Lends strong support
5	Lends powerful support

FIG. 2. The kind of verbal scale often adopted by forensic scientists. This version is taken from the expert report in *R v Atkins*: an image comparison (or facial mapping) case.

features to an indication of the value of those similarities—such as 'lends strong support'—as evidence of identification does not have empirical grounding. Those comparing faces and fingerprints are not (at this stage) conversant with the frequency and independence of features and sub-features. As with latent fingerprint examiners, we do not know if they possess expertise in assigning a particular level of significance to apparent similarities.[68] In consequence, it is not clear what probative value(s) we should attach. While some formulations, such as those incorporated in the table (above), might, as qualifications, constitute an improvement over positive identification, this 'solution' remains impressionistic, speculative and quite likely misleading. We contend that the use of verbal scales is inappropriate because it relies on the examiner's impression and privileges untested 'experience'. It cannot be readily assessed and it is not easy to explain the methodological frailties at trial—especially when it is the accused challenging the opinion of an experienced analyst.

One possible compromise is to restrict testimony to similarities (or even 'matches') and incorporate information about the error rate in comparisons while endeavouring to explain that we do not have an established means of moving between the declaration of a match and the attribution of evidentiary significance. This would involve reporting a 'match', an error rate with matching (and possibly other parts of the process when empirical evidence emerges), and informing the jury that a match is not the same as identification. That is, we currently do not know what the relationship is between a match decision and positive identification. The tribunal of fact should not be allowed to attach any significance they want in the absence of information about the indicative value of the opinion. This is the basic approach proposed in the first iteration of the Guide. The important point is to introduce the error rate, rather than rely upon restricting opinions to similarities given the pervasive belief that fingerprints are unique and that a declared (or reported) match is the equivalent of positive identification.

5.4 *Probabilistic evidence, likelihood ratios and error*

There has been an enduring push to 'objectify' fingerprint comparisons in a similar way to the analysis and presentation of DNA evidence.[69] The methodology is fairly straightforward. An examiner can

EWCA (Crim) 2 and contrast *R* v. *T* [2010] EWCA Crim 2439 (shoe prints) where emphasis was placed on the need to disclose calculations and assumptions.

[68] Additional problems are created by the low quality of many images, the ability of persons of interest to disguise themselves, as well as the body changing diachronically.

[69] C. Champod and I. Evett, 'A probabilistic approach to fingerprint evidence' (2001) 51 *Journal of Forensic Identification* 101–22; C. Neumann, 'Fingerprints at the crime-scene: Statistically certain, or probable?' (2012) 9 *Significance* 21–25. More generally, see C. Aitken, P. Roberts and G. Jackson, *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses* (London: Royal Statistical Society, 2010).

GARY EDMOND *ET AL.*

locate and connect a set of features in a fingerprint, which form a polygon where the length of each side and the angles within the polygon are defined numerically. The same measurements are calculated for a second fingerprint. Two hypotheses are then compared: (1) that the two prints come from the same source (i.e. the differences between the polygons are due to the partial or degraded nature of the prints); and (2) that the two prints come from different sources (i.e. the similarities and differences between the polygons are coincidental). One score is then calculated for the first hypothesis by computing the extent to which the same polygon can vary based on a range of marks from the same source. A second score is calculated for the second hypothesis by computing the rarity of this particular polygon compared to polygons that are derived from other sources that have been selected at random. The larger the ratio of within- to between-polygon variation, the stronger the evidence for the first hypothesis compared to the second.

There are several problems with this specific methodology as well as the general approach to providing rarity values in testimony. First, the basic units of analysis in these computations are the distances, directions, and angles among particular configurations of basic fingerprint features (e.g. ridge endings and bifurcations). With highly degraded prints—the sort commonly lifted from crime scenes—there will be very limited consistency among examiners in what counts as a landmark, where they are located and how many to include. These will vary greatly depending on a variety of contextual and human factors including expertise, traditions, fatigue, time constraints and so forth. As a result of this variability, the configurations of the polygons may vary markedly between examiners and even for the same examiner in different circumstances. Indeed, the Automated Fingerprint Identification System (AFIS) currently outputs a numerical measure of similarity based on the minutiae, features, directions, and spatial relationships of the fingerprints that the operator submits to the system.[70] The magnitude of these measures varies greatly depending on the number and nature of the landmarks that each examiner provides, and one can easily cherry pick particular parameters to produce a large or small value. The same problem with consistency is almost certainly true for any probabilistic model that involves human judgement, which tends to subvert the purpose of an objective system. Secondly, the score that is derived to test whether the similarities and differences between the polygons are coincidental is based on polygons that have been generated from other sources *selected at random*. The measure of 'coincidence' may therefore be misleading, particularly when an examiner is faced with the most highly similar candidate print that is retrieved from the database of tens of millions of possibilities. An appropriate measure of coincidence here ought to reflect the confusability of highly similar candidates, which is common in practice. The rarity scores generated by these models may well assist examiners in comparing fingerprints by reducing the amount of time it takes to arrive at a decision, or improve the overall accuracy of their judgement (which is testable), but they are of little use on their own. Thirdly, as discussed in Section 2, the formal training of fingerprint examiners tends to be specific to the workplace conditions (e.g. principles and foundations of friction ridge examination, biology and physiology of friction ridge skin, history of fingerprints).[71] Examiners do not have a background or demonstrated expertise in probability theory, statistics, or

---

[70] I.E. Dror and J.L. Mnookin, 'The use of technology in human expert domains: Challenges and risks arising from the use of automated fingerprint identification systems in forensic science' (2010) 9 *Law, Probability and Risk* 47–67.

[71] See 'Standards for Minimum Qualifications and Training to Competency for Friction Ridge Examiners' (2010, Version 1.0), available at <http://www.swgfast.org/documents/qualifications-competency/100310_Qualifications_Training_Competency_FR_1.0.pdf> (17 October 2012). It is curious that in a field based entirely on human judgment, the people who make important decisions about lives and livelihoods are not trained in the very factors that influence these decisions (e.g. heuristics, biases, memory errors, logical fallacies, cognitive illusions, etc.).

mathematics enabling them to testify about the assumptions and calculations behind these probabilistic models.

Even if an examiner was perfectly consistent in his own analyses of fingerprint landmarks on different occasions (and other examiners agreed), if the resulting probabilities were sufficiently conservative and adequately measured the rarity of the prints, and if the examiner fully comprehended the mathematical basis of the model and effectively communicated it to the court, the problem with the general approach to providing rarity values in testimony remains—in that these values sidestep the very real possibility of human error. Recognition of a real error rate provides an important means of qualifying an expert's interpretation.

The inclusion of an indication of error would seem to place expert opinion evidence in an empirically based format that is conducive to the trial framework and comprehension by lawyers, trial judges and juries.[72] Error rates enable opinion evidence to be presented and interpreted in a manner consistent with fundamental concerns about reasonable doubts. The provision of an indication of error helps to ground expert opinions; helping to prevent prosecutors and judges from simply assuming that the match—particularly claims about uniqueness, infallibility or improbably small random match probabilities—is infallible or that interpretation is mechanical, that protocols were followed, that labelling was accurate, that equipment worked, that interpretations were sound, and reviews effective.[73] While, the reported error rate will not necessarily capture error in the particular case, it provides a useful (and arguably necessary) background rate against which claims about matching can be assessed.[74]

Recognizing and incorporating errors into the reporting of results may, with the assistance of further study, enable fingerprint examiners to move beyond 'match', 'non-match' and 'inconclusive'—to the extent that they are willing to concede greater uncertainty and probably greater risks of error—see LPEHF Rec. 3.8. What the criminal justice system, particularly courts, should do in relation to less reliable forms of incriminating opinion warrants detailed consideration against admissibility standards and discretions such as those oriented to the danger of unfair prejudice to the accused.[75]

---

[72] Though see D. McQuiston-Surrett and M.J. Saks, 'The testimony of forensic identification science: What expert witnesses say and what factfinders hear' (2009) 33 *Law & Human Behavior* 436 and D. McQuiston-Surrett and M. Saks, 'Communicating opinion evidence in the forensic identification sciences: accuracy and impact' (2008) 59 *Hastings Law Journal* 1159; K.A. Martire, R.I. Kemp and B.R. Newell (2013). 'The psychology of interpreting expert evaluative opinions'. *Australian Journal of Forensic Sciences*, 1–10. doi:10.1080/00450618.2013.784361.

[73] Not doing the studies provides a means of never having to disclose *known* limitations. Not sponsoring or requiring studies advances the prosecution case. In consequence, we have an approach where the accused bears the risk of errors rather than the state being obliged to concede the real possibility of a range of errors in each case. Rather than concede an average or indicative error rate, the state obtains the benefit of never reporting (i.e. disclosing) error. See A. Ligertwood and G. Edmond, 'Expressing evaluative forensic science opinions in a court of law' (2012) 11 *Law, Probability & Risk* 80–91. See also G. Edmond and A. Roberts, 'Principles of evidence law and their implications for forensic science and medicine' (2011) 33 *Sydney Law Review* 359.

[74] Koehler, 'Fingerprint error rates and proficiency tests' at 1088 ('The industry wide error-rate estimates provide anchors for judgments about the risks of error in individual cases.' The *base rate fallacy* refers to the tendency to believe that general error rates can be ignored when some special information about the case (e.g. the unusual proficiency of the examiner or the 'difficulty' of the specimen) is offered). Historically, and explicitly in DNA appeals, English courts have been unsympathetic to the idea that a risk of error should be incorporated into the expression of the expert's opinion. See discussion in M. Lynch, S. Cole, R. McNally, K. Jordan, *Truth Machine: The Contentious History of DNA Finger Printing* (Chicago: University of Chicago Press, 2008). See also J. Koehler, 'The psychology of numbers in the Courtroom: How to make DNA-match statistics seem impressive or insufficient' (2001) 74 *S California Law Review* 1275 at 1299–1300.

[75] See discussion in Sections 6 and 7.

5.5   *Improving performance and reducing bias*

It is vital that examiners are not begrudging or disingenuous in their recognition and attribution of possible errors—both generally and specifically. The failure to accept or recognize the existence of errors distorts their evidence, and makes it more difficult to improve processes and performance. It also shifts responsibility to the accused to somehow identify errors made in the process of recovery, collection, analysis, interpretation and reporting of opinions retrospectively during an adversarial trial.

   Given the interpretive (or subjective) nature of comparisons, the various reports recommend undertaking research into bias and developing practices that prevent the detrimental effects it may have on analysis—NRC Rec. 5, LPEHF 3.3, SFI 7, 8 and 9. Fingerprint examiners are, in most bureaus, not insulated from case information when undertaking their analyses or even when selecting prints for comparison from fingerprint databases. This means that their comparisons may have been influenced by information that is not relevant to their analysis, such as the criminal record of suspects with prints similar to those recovered or the fact that the accused made a retracted confession or was believed by investigating police to be guilty.[76] Generally, given the absence of empirical evidence, it seems appropriate to withhold such information from analysts, even if only until they have made and recorded their first analysis of the evidence—NRC Rec. 5, SFI 8.[77] Exposure to domain irrelevant information should be documented, and courts should be interested in such exposures when considering both admissibility and weight. It may be that we should consider excluding comparison evidence where the examiner has been unnecessarily exposed to gratuitous information.[78] Matches (and non-matches) declared after exposure to domain irrelevant information are at risk of having been influenced by that information. Where the examiner has been exposed to prejudicial information, any match decision cannot be understood as independent corroboration and should not be presented as such at trial.

5.6   *Individual performance*

The error rates reported in the Guide are based on controlled experiments of the matching performance of qualified fingerprint examiners. Given the dearth of research on the competency of examiners at all stages of fingerprint analysis, or on the factors that influence performance, and in the absence of comprehensive (i.e. industry wide) measures of accuracy, the performance of examiners in these experiments can only be used as a generic or indicative rate for the field. The error rate does not apply to individual examiners or even particular fingerprint bureaus. In terms of individual performances, we can assume that they vary—between examiners, as well as over time and conditions—but we do not know which examiners, bureaus, work contexts, or situations are more or less error prone. This is part of the problem. Most forms of practice do not involve making comparisons in controlled conditions. In principle, it seems useful to have an indication of a general error rate for comparisons

---

[76] See Busey and Dror, 'Special abilities'; Dror *et al.*, 'Contextual information renders experts vulnerable'; Dror and Cole, 'The vision in "blind" justice'; Dror *et al.*, 'When emotions get the better of us'; Dror and Rosenthal, 'Meta-analytically quantifying the reliability and biasability of forensic experts'.

[77] D. Krane *et al.*, 'Sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation' (2008) 53 *Journal of Forensic Science* 1006; W.C. Thompson, 'What role should investigative facts play in the evaluation of scientific evidence?' (2011) 43 *Australian Journal of Forensic Sciences* 123–134.

[78] Exclusion is appropriate because we can always ask another examiner to undertake the same comparison 'blind', thereby removing any risk.

(and other parts of the process).[79] These can be used as a base rate, even if there are arguments about the performance of the individual examiner and the value of the prints in specific cases.[80]

Here, it is important to distinguish the proficiency tests currently used by fingerprint examiners, such as those provided by Collaborative Testing Services, Incorporated. These tests have been trenchantly criticized and are insufficient for measuring accuracy.[81] In order to make general claims about accuracy—beyond those of specific prints at a specific level—different (and randomized) sets of prints for each examiner are needed. Commercial proficiency tests do not adequately address the general issue of expert matching accuracy and were not designed to disentangle the factors that affect matching accuracy.

6. Trial safeguards are not a viable mechanism for managing speculative opinions

Regardless of the admissibility standard, whether focused on 'specialized knowledge' (and reliability), a field of knowledge or assisting the tribunal of fact, the opinions of fingerprint examiners, about whether two prints match or do not match, should ordinarily be admissible because matches appear to be accurate and probative on the question of identity.[82] Admissibility should be subject, however, to the disclosure of limitations (i.e. an error rate and exposure to potentially biasing information and breaches of standards and protocols) and the witnesses restricting their report or testimony to areas in which they possess demonstrable expertise. The state should be obliged to study techniques in widespread use and to proactively concede limitations.[83] Drawing on the various recommendations, we would contend that latent fingerprint evidence should not be admitted at trial unless known limitations and measures of performance are included with the opinion.

The Guide is also motivated by emerging evidence that trial safeguards are not consistently effective in exposing and conveying limitations with incriminating expert opinion evidence.[84] Appellate review has also been ineffective at exposing errors or even overconfidence.[85] Common law courts have been remarkably accommodating toward the opinions of state-employed fingerprint examiners and other forensic scientists.[86] This accommodating posture may have produced complacency and

---

[79] Thompson, Tangen and McCarthy, 'Expertise in fingerprint identification' ('. . . of course, the courts will be concerned with data that will help fact finders make optimal decisions. But we don't demand, for example, individual error rates for a medical doctor or a field-wide error rate in medical diagnosis; we only demand performance measures of the instrument or test on average. To ask for error rates associated with a particular individual on a particular test seems, rightly, inappropriate in medicine. Similarly, focusing on the individual is the wrong level of analysis when attempting to characterize the accuracy of the forensic fingerprint identification system. A broader question concerns the level of analysis that is appropriate for presenting evidence and associated rates of error in court. At the extreme, an examiner could report how accurate they are at matching a whorl type print, lifted from a crime scene, on a wooden surface, using magnetic black powder, in a particular department, in a particular country, on a Tuesday, and so on'.)

[80] Koehler, 'Proficiency tests to estimate error rates in the forensic sciences'.

[81] Thompson, Tangen and McCarthy 'Expertise in fingerprint identification'; Haber and Haber, 'Scientific validation of fingerprint evidence under *Daubert*'; Koehler, 'Fingerprint error rates and proficiency tests'; Vokey, Tangen and Cole, 'On the preliminary psychophysics of fingerprint identification'.

[82] Though on the question of 'specialized knowledge' see also Section 7.

[83] D.S. Medwed, *Prosecution Complex: America's Race to Convict and Its Impact on the Innocent* (New York: NYU Press, 2012).

[84] G. Edmond and M. San Roque, 'The cool crucible: Forensic science and the frailty of the criminal trial' (2012) 24 *Current Issues in Criminal Justice* 51–68. Trial safeguards are more likely to be effective where the witness, prosecutor and judge understand limitations and them pro-actively.

[85] Even U.S. trials and appeals have not been effective at systematically identifying and conveying limitations.

[86] See G. Edmond, S. Cole, E. Cunliffe and A. Roberts, 'Admissibility compared: The reception of incriminating expert opinion (i.e., forensic science) evidence in four adversarial jurisdictions' (2013) 3 *University of Denver Criminal Law Review* 31–109.

GARY EDMOND *ET AL.*

overconfidence (among lawyers, forensic scientists, judges, jurors and the public), and may have contributed to the remarkable paucity of research in many areas of forensic science. While limitations might be explored on the voir dire, or via cross-examination, rebuttal witnesses and judicial directions during the trial, such safeguards have not proved to be effective in exposing limitations or identifying real risks of error. Significantly, problems with fingerprint evidence did not emerge from the crucible of the trial.

Trial counsel and judges should be more attentive to expert evidence (and its bases).[87] The Guide is intended to help in this regard. We should not discount the fact that fingerprint examiners continue to characterize a 'match' as positive identification or something practically indistinguishable. To the extent that qualifications have occasionally and sporadically arisen they are often begrudging and dismissive—e.g. reluctantly conceded as hypothetical (or theoretical) possibilities (LPEHF 6.3)—or transformed into (un)ambiguous similarity evidence.[88] While many examiners are working on im-proving practices and should be cooperating with scientists to change how comparisons are undertaken and interpreted, experience suggests that we cannot rely on the fingerprint community to unilaterally reform their practices. The Guide is intended to help stimulate this transformation while providing a framework that enables legal institutions to utilize fingerprint evidence on a more empirically secure footing.

7. 'Expertise': a practical ability, tacit or explicit knowledge?

An interesting aspect of the emerging research on the performance of fingerprint examiners is that their judgements and decisions may not be (fully) explicable.[89] Many jurisdictions, such as those based on the Federal Rules of Evidence (FRE) in the US and the uniform evidence law (UEL) in Australia, specify the need for knowledge—'scientific, technical and other specialized knowledge' and 'specialized knowledge', respectively—in their admissibility standards for expert opinion evidence. It may be that the abilities or expertise of examiners is not readily reducible to articulable knowledge or a particular method, but attributable to learned abilities to compare prints and recognize *apparent* similarities and differences. Whether this is actually 'knowledge' is an interesting question, though there is little doubt that comparisons by trained and experienced examiners are generally reliable.[90]

Research on medical decision-making has shown that in some circumstances the factors (purportedly) relied upon by experts may not be clearly defined or verbalized, yet performance

---

[87] As Allen and Miller explain, this is required under the orthodox approach to the accusatorial trial: R. Allen and J. Miller, 'The Common law theory of experts: Deference or education' (1993) 87 *Northwestern University Law Review* 1131.

[88] Cole, 'Splitting Hairs?'.

[89] Expertise in a domain does not necessarily include the ability to articulate the basis of that expertise or the reasoning behind judgments and decisions. For example, asking experts to describe what they are doing can hurt performance and, despite the expectations/requirements of courts, experts may not have 'knowledge' or access to the basis of their decisions. See T.D. Wilson, *Strangers to ourselves: Discovering the adaptive unconscious* (Cambridge, MA: Harvard University Press, 2002); R.E. Nisbett and T.D. Wilson, 'Telling more than we can know: Verbal reports on mental processes' (1997) 83 *Psychological Review* 231–259; C.S. Dodson, M.K. Johnson and J.W. Schooler, 'The verbal overshadowing effect: Why descriptions impair face recognition' (2007) 25 *Memory & Cognition* 129–139; M.C. Fox, K.A. Ericsson and R. Best, 'Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods' (2011) 137 *Psychological Bulletin* 316–344; M. de Vries, C.L.M. Witteman, R.W. Holland, A. Dijksterhuis, 'The unconscious thought effect in clinical decision making: An example in diagnosis' (2010) 30 *Medical Decision Making* 578–81; G.R. Norman and K.W. Eva, 'Diagnostic error and clinical reasoning' (2010) 44 *Medical Education* 94–100.

[90] See also M. Polanyi, *The Tacit Dimension* (New York: Anchor Books, 1967); H. Collins, *Changing Order: Replication and Induction in Scientific Practice* (Beverly Hills, CA: Sage, 1985). H. Collins, *Tacit and Explicit Knowledge* (Chicago: University of Chicago Press, 2010).

may still be accurate.[91] That is, the information that examiners think they rely on, or use to rationalize their decision-making (retrospectively), may bear limited resemblance to the information that they actually use. Processing may not, in fact, be entirely conscious. In contradistinction to legal practice and assumptions about expertise (based on 'knowledge'), if comparisons are based, to some non-trivial extent on intuitive or unconscious processes, then it may prove to be very difficult to expose error at trial and the analyst might remain sincerely confident even when mistaken.[92] Conventional trial safeguards, such as cross-examination may, in consequence, be of limited value in exposing limitations or errors in such circumstances—see Section 6.

8. Jury access to images and working notes

When asked to compare and match fingerprints laypersons are quite error prone. Vokey *et al.* demonstrated that novices generally have the ability to match prints, but their performance deteriorates dramatically when it comes to distinguishing highly similar but non-matching prints; such as those obtained through database searches.[93] Tangen *et al.* found that laypersons incorrectly declared 55.18% of these similar non-matching prints as 'matching', compared to 0.68% for examiners.[94] Lay vulnerability to error raises enduring questions about whether the tribunal of fact—jury or trial judge— should be allowed to examine and compare prints.[95] Empirical studies might be interpreted such that they counsel against allowing laypersons to undertake their own comparisons.[96] Jurors and judges will be particularly vulnerable to error, and this will be accentuated by: the lack of feedback; exaggerated confidence in their own abilities; and exposure to (potentially biasing) contextual information. This ought to lead judges to exclude images of prints, working notes and mark-ups unless the interests of justice demand admission or the defence seeks their introduction. Lay impressions, about whether two prints match or do not match, are likely to be error prone and should not generally be encouraged. Given the levels of error, there are real dangers in allowing the jury to undertake its own interpretation of contested prints and admission of such images raises the real, and perhaps insurmountable, risk that this will occur.

---

[91] For example, dermatologists, radiologists, and other medical professionals become highly skilled in discriminating between complex and highly variable visual patterns that are difficult to articulate to diagnose various diseases or conditions: L.R. Brooks, G.R Norman and S.W. Allen, 'Role of specific similarity in a medical diagnostic task' (1991) 120 *Journal of Experimental Psychology: General* 278.

[92] K. Krug, 'The relationship between confidence and accuracy: Current thoughts of the literature and a new area of research' (2007) 3 *Applied Psychology in Criminal Justice* 7–41, for a review. The feeling of confidence that we experience is not a reliable guide to the validity of the information and can be due to a variety of factors that may be unrelated to the judgment in question. For example, presenting a name repeatedly on one occasion produces a feeling of familiarity on another occasion. This illusion of "pastness" makes us feel like we are remembering and is accompanied by false ratings of confidence. Repetition is just one way of inducing false confidence. Easy-to-process material—through visual ease, linguistic ease (e.g. rhyme), or even from being in a good mood—can result in overconfidence. See also D. Kahneman, *Thinking, Fast and Slow* (Farrar, Straus and Girou, 2011); A.L. Alter and D.M. Oppenheimer, 'Uniting the tribes of fluency to form a metacognitive nation' (2009) 13 *Personality and Social Psychology Review* 219–235.

[93] Vokey, Tangen and Cole, 'On the preliminary psychophysics of fingerprint identification'.

[94] Tangen, Thompson and McCarthy, 'Identifying fingerprint expertise'.

[95] Generally, fingerprint analysts will only testify when they declare a match.

[96] Juries may perform better than jurors, but how much better is uncertain and many of the risks are likely to impact on the group as well as individuals. There is no evidence that judicial performance is superior to the performance of jurors.

9. Reports rather than testimony?

Because of the many, very real complications with fingerprint evidence and the danger of opinions being presented in terms that exceed what can be empirically sustained, a slightly more radical response might be to limit the presentation of fingerprint evidence to a short documentary form embodying the kinds of issues identified in the Guide.[97] That is, in most cases the examiner should not testify in person and the prints and working notes should not be adduced or admitted. These documents should, of course, be disclosed to the defence. Such an approach would not only help to prevent testimonial misrepresentations, but it would also save time and resources. Again, where the defence requires attendance, or where attendance is in the interests of justice, the examiner should testify in person.[98]

10. Conclusion

The proposed Guide represents a pragmatic attempt to address criticisms of fingerprint methodology and the way most comparisons are currently reported and explained. It is intended to begin to address and recognize the existence of genuine expertise in undertaking comparison work, the weak underlying decision-making framework, and the historical reluctance among examiners to make appropriate concessions in reports and testimony. If this approach is conceived as excessively empirical (or requires too much evidentiary support), we would remind the reader that for a century courts have allowed techniques to be misrepresented even though they could have been studied and improved. Along with the National Research Council of the United States National Academy of Sciences, the National Institute of Justice and the National Institute of Standards and Technology (US), Lord Campbell (*Scottish Fingerprint Inquiry*), and others, we are supporting the introduction of accountability mechanisms in forensic science reporting and testimony.

At the base of this proposal is a growing chorus of criticism about the kinds of studies and evidence that should underpin the forensic sciences—at least, when they are relied upon in criminal proceedings. We, along with others, believe that courts have an obligation to require evidence of ability—actual expertise and rates of accuracy—particularly about forensic science and medicine techniques in routine use. Empirical studies provide evidence of ability, they enable those evaluating the evidence to have a clearer idea of the value of evidence than is usually provided through cross-examination or judicial cautions, and the studies will often inform the manner in which experts should be allowed to express their opinions as well as the scope of their testimony. All of these can be observed in the recent experiments on the abilities and accuracy of fingerprint examiners.

The Guide, and particularly the focus on error rates, offers a useful means of assessing and regulating a range of comparison practices, especially those where the likelihood of generating useful probability rates seems remote or even unlikely. Even if fingerprint examiners and others develop probabilistic approaches, there will be a need to consider how error rates should be incorporated into the results, as the various reports recommend. For many other types of comparison practices (e.g. images, voices, ballistics, tool marks, bites, footprints, tire prints, pattern marks and so on), error rates will provide important insights into the value of the evidence—by highlighting the abilities of the witnesses—that enable judges (and lawyers) to determine the admissibility of opinions and their

---

[97] See Found and Edmond, 'Reporting on the comparison and interpretation of pattern evidence'.

[98] Though perhaps, without direct/examination-in-chief. See generally *Melendez-Diaz* v. *Massachusetts*, 129 S. Ct. 2527 (2009).

weight should they be sufficiently reliable for admission. Preliminary studies suggest that not all comparison and identification techniques will be as accurate as fingerprint.

We accept that, notwithstanding its empirical sensitivities, this is a pragmatic response or compromise. We also accept that others may have alternative, perhaps better, ideas about how we can respond to the frailties of latent fingerprint evidence. Our proposal is based on what we currently know, empirically, about latent fingerprint evidence in combination with the realization that investigators and courts are unlikely, and perhaps unable, to respond unilaterally to latent fingerprint evidence. No doubt many examiners, and others, will be concerned about even these modest, empirically inflected impositions. While acknowledging that the Guide and the imposition of an indicative error rate represents a considerable departure from historical assumptions and practices, there is no reason to continue to admit and accept opinions about latent fingerprints in the conventional accommodating manner. Our proposal accepts that latent fingerprint evidence is potentially powerful evidence of identity, and we believe that the Guide provides a compromise that reflects current knowledge and abilities, as well as what we now know about the limitations of fingerprint evidence (as well as the limitations of trials and appeals).

The legal system should not avert its eyes to mainstream scientific consensus around frailties and errors in the way comparison sciences are practiced and reported. Disregarding scientific consensus threatens the legitimacy of legal institutions and undermines their ability to deliver accurate verdicts and, simultaneously, justice.[99]

Funding

[99] Edmond and Roach, 'A contextual approach'.