Fingerprint identification: advances since the 2009 National Research Council report

Christophe Champod

School of Criminal Justice, Faculty of Law, Criminal Justice and Public Administration, University of Lausanne, Lausanne, Switzerland

This paper will discuss the major developments in the area of fingerprint identification that followed the publication of the National Research Council (NRC, of the US National Academies of Sciences) report in 2009 entitled: Strengthening Forensic Science in the United States: A Path Forward. The report portrayed an image of a field of expertise used for decades without the necessary scientific research-based underpinning. The advances since the report and the needs in selected areas of fingerprinting will be detailed. It includes the measurement of the accuracy, reliability, repeatability and reproducibility of the conclusions offered by fingerprint experts. The paper will also pay attention to the development of statistical models allowing assessment of fingerprint comparisons. As a corollary of these developments, the next challenge is to reconcile a traditional practice dominated by deterministic conclusions with the probabilistic logic of any statistical model. There is a call for greater candour and fingerprint experts will need to communicate differently on the strengths and limitations of their findings. Their testimony will have to go beyond the blunt assertion of the uniqueness of fingerprints or the opinion delivered ispe dixit.

1. The National Research Council report

In 2009, the US National Research Council (NRC) of the National Academies of Sciences issued a report—referred to hereinafter as the NRC report—that reviewed the state of affairs in the USA of all forensic areas except DNA [1].

After hearing representatives of the profession (notably members of the FBI and SWGFAST) in relation to fingerprints, the report pointed out the following critical elements (condensed from pp. 5.7 to 5.12) that will be the focus of the present contribution:

— After an analysis of the examination steps taken by fingerprint experts, known as ACE-V, the report acknowledges that the whole assessment is entirely left to the appreciation of the expert.
— The threshold used by experts to evaluate their findings and notably to conclude an identification are kept deliberately subjective, so that the examiners can consider in their assessment all features that the mark and the print may display. Training and experience is the mechanism whereby experts acquire a sense of rarity of the fingerprint features and typically how much friction ridge detail could be common to two prints from different sources. The fingerprint community has eschewed numerical scores and corresponding thresholds, because those developed to date have been based only on minutiae, not on the whole range of features of the friction ridge skin. Contrary to DNA, the features used by the experts cannot be defined a priori and depend on the quality of the mark that may be partial, smudged, moved, distorted or difficult to interpret when visualized on an interfering background. It led to a lack of population statistics, although more would be feasible. The area of statistical modelling of fingerprint features is ripe for additional research.
— The results may not be reproducible from expert to expert. Contextual bias has been shown to have an impact on experts’ decisions.
— The reporting of results generally espouses the agreed three possible options: identification, exclusion or inconclusive. Experts are actively discouraged by their professional bodies from expressing themselves in probabilistic terms. When an identification conclusion is reached, the term conveys the notion that the mark and print could not possibly come from two different individuals. Conversely, a conclusion of exclusion makes it impossible for that individual to be the source of the mark.

— The reliability of the ACE-V process remains unknown and could be improved if specific measurement criteria were defined. Those criteria become increasingly important as the quality of the marks reduces. Experts testify in a language of absolute certainty and the report, quoting Mwookin [2], invites experts to adopt more modest claims about the meaning and significance of their findings.

The NRC report drew a rather critical, and for many, unexpected picture of the fingerprint area. It received mixed reactions and commentaries that I shall summarize below.

2. Responses to the National Research Council report

(a) Reactions from practitioners, organizations and scholars

Prestigious scientific journals such as Nature [3] or Science [4] took immediate notice of the publication and depicted an alarming state of affairs (and still are [5]). Giannelli [6] detailed the range of reactions and strategies deployed to cope with the NRC report, ranging from attacking the messenger, blunt resistance, lack of scientific coverage to change to a critic of the Judiciary. I shall restrict my discussion here to the reactions dealing specifically with the strength to be attached to fingerprint evidence without seeking exhaustiveness.

The International Association for Identification (the IAI, the largest professional body dealing with fingerprint identification) issued an immediate reaction.7 The IAI aimed at putting its community at rest by referring to a specific sentence of the NRC report, p. 5-12: ‘it seems plausible that a careful comparison of two impressions can accurately discern whether or not they had a common source’. Without denying the need for further research and policy, the IAI stressed upon the fact that fingerprint evidence shall be considered as reliable when delivered by trained and competent examiners following accepted practices and standards. The IAI furthermore issued a memo to its members, stating its support to many of the report’s recommendations, cautioning against asserting ‘100% infallibility (zero error rate) and advising its members not to state ‘their conclusions in absolute terms when dealing with population issues’.8

The Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST, a US working group dedicated to fingerprint matters) also issued a detailed response.9 Largely in agreement with the report, SWGFAST drew attention to a few additional pieces of scientific research that address some of the expressed concerns and that were either not fully referred to by the report or not published yet. Regarding ‘individualization’ the position statement reads: ‘History, practice, and research have shown that fingerprints can, with a very high degree of certainty, exclude incorrect sources and associate the correct individual to an unknown impression’. And further in its response ‘SWGFAST acknowledges that errors do occur and furthermore that claims of zero error rate in the discipline are not scientifically plausible’. The European Network of Forensic Science Institutes has reacted through its European fingerprint working group (ENFSI–EPWG) [7]. After recalling the different initiatives of the ENFSI–EPWG as well as Interpol that address some of the points made in the report, it was agreed on the need for devoting careful attention to the way fingerprint evidence is reported because (p. 678): ‘such evidence should not be considered as absolute or factually sufficient in itself to exclude every donor on the planet’. In the scholarly literature, critical attention was also given to how fingerprint evidence ought to be conveyed in court. Mwookin [8] insisted on the limitations of fingerprint research at the time of the NRC report and rightly clarified the key issue (p. 1240): ‘The problem with fingerprint evidence is not that it completely lacks probative power, but rather that research on the domain has not yet established the appropriate limits to its probative power, or shown how that value varies depending on its quality or its quantity of information’. Mwookin is calling for a stricter regime to testify to individualization. The question of uniqueness and individualization further attracted numerous commentaries [9–15], as did the question of testifying in court [8,16]. What is clear from the post NRC report scholarly literature is that the days where invoking ‘uniqueness’ as the main (if not the only) supporting argument for an individualization conclusion are over.

SWGFAST responded ultimately to this systematic challenge of traditional reporting practice by a major change of the definition of individualization in its documents. Between 2002 and 2011, it went through the following wording (a more detailed analysis is provided in Cole [14]):

2002 ‘Individualization occurs when a latent print examiner, trained to competency, determines that two friction ridge impressions originated from the same source, to the exclusion of all others’.10

2011 ‘Individualization is the decision by an examiner that there are sufficient features in agreement to conclude that two areas of friction ridge impressions originated from the same source. Individualization of an impression to one source is the decision that the likelihood the impression was made by another (different) source is so remote that it is considered as a practical impossibility’.11

The definition went from ‘excluding all others’ to, now, a decision. Cole [14] highlighted the positive side of what he described as downgrading the claimed strength of the report, in that ‘decision’ sounds weaker than ‘conclusion’ or ‘determination’. For Cole, this change is an additional attempt to preserve the status quo. In my view, it is an improvement on transparency, because, indirectly, it affirms for the first time that probabilities come into play in any fingerprint conclusion. Cole is making a further and very valid point: fingerprint practitioners overall have not grasped the subtle nuance between the above definitions and changes. Some experts have rightly identified the issue associated with an absolute claim made in reference to the Earth population (see two cases quoted in [14]: State v. Hull12 and State v. Doe13). They invoke a local identification (i.e. within a restricted pool of suspects) instead of the individualization to the exclusion of all others on the
Earth (a distinction between local identification and individualization was introduced by Kaye [9,17]). The experts who testified in these cases are well versed in the statistical and decisional aspects that underpin the process and tried to find a way to communicate effectively. But they are the exception more than the rule. It means that fingerprint examiners today have adopted a new language without the necessary educational and scientific change that comes with it. Although training of examiners is a key component of many recommendations made after the NRC report (and similarly after the NIST report [18] or the Fingerprint Inquiry report [19]), we haven’t seen such development yet. It is clear that there is a need to foster a culture of research in all identification areas, including fingerprints [20]. But, first and foremost, what is currently lacking is a proper scientific education in forensic science [21].

(b) Follow-up documents and reviews

A couple of papers [22,23] were published in an attempt to defend the discipline from the critical scrutiny of the NRC report. Overall, they have been written with the implicit objective to maintain the status quo and to simply adapt the reporting practice moving from an absolute to an opinion of certainty. The reader is left then with a vast set of references without the detailed analysis of the consequence of the exposed body of research.14

In 2011, SWGFAST published The fingerprint sourcebook [24] covering many subjects in the fingerprint individualization process. The sourcebook was not a response to the NRC report but was developed in the wake of persistent criticisms and the lack of an authoritative reference work. SWGFAST also submitted a detailed bibliography15 in response to a defined set of questions put forward by the Inter-agency Working Group for Research, Development, Testing and Evaluation (IWG RDT&E—under the White House Office of Science and Technology Policy—OSTP, committee on Science, subcommittee on Forensic Science) set up following the NRC report. In 2013, under the Department of Justice, the National Commission on Forensic Science (NCFS)16 has been established in partnership with the NIST (National Institute of Standard and Technology). This new organ is not alien to the NRC report that was recommending the establishment of a National Institute of Forensic Science to regulate and harmonize forensic science in the US. One of its objectives is to provide recommendations and advice for strengthening the validity and reliability of the forensic sciences. The response from the NCFS on the annotated bibliographies (IWG RDT&E initiative) prepared by the working groups is rather harsh,17 in particular the lack of relevance to support the foundations of the discipline and a lack of rigorous peer-review processes.

An extensive report has been written by a group of experts convened under the NIST/NIJ (National Institute of Standard and Technology/National Institute of Justice) and addresses human factors in latent print analysis [18] (hereinafter referred to as the NIST report). It not only analyses current practices and their contribution to errors but also investigates how to reduce error and how to implement these solutions practically. Some critical recommendations echo the NRC report:

Recommendation 3.1: A report and contemporaneous supporting notes or materials should document the examination to make the interpretive process as transparent as possible. Although the degree of detail may vary depending on the perceived complexity of the comparison, documentation should, at a minimum, be sufficient to permit another examiner to assess the accuracy and validity of the initial examiner’s assessment of the evidence.

Recommendation 3.3: Procedures should be implemented to protect examiners from exposure to extraneous (domain-relevant) 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.

In Scotland, after years of dispute, the final report [19] of the Fingerprint Inquiry (hereinafter referred to as the Fingerprint Inquiry report)—a judicial inquiry devoted to the mis-identification of both Shirley McKie and David Asbury—gives a set of 86 recommendations. Among them, some are crosscutting with the recommendations of the NRC report and the NIST report:

Recommendation 1: Fingerprint evidence should be recognized as opinion evidence, not fact, and those involved in the criminal justice system need to assess it as such on its merits.

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.

Recommendation 9: Features on which examiners rely should be demonstrable to a lay person with normal eye sight as observable in the mark.

Recommendation 53: Subject to any requirement under ISO 17025 and recommendations 50 and 51, note-taking as to the detail found on analysis and the process of comparison, though not mandatory, should become the general practice for all fingerprint comparison work.

Recommendation 66: Before a finding of ‘unable to exclude’ is led in evidence, careful consideration will require to be given to (a) the types of mark for which such a finding is meaningful and (b) the proper interpretation of the finding. An examiner led in evidence to support such a finding will require to give a careful explanation of its limitations.

3. Scientific developments since the National Research Council report

(a) Bias

The issue of bias and lack of reproducibility among experts was highly ranked as a research priority by the NRC report. At that time, the report was informed by studies [25–27] mainly done in the aftermath of the 2004 Mayfield case. Indeed in the investigation that followed that case of mis-identification [28,29], confirmation bias was raised as a significant cause of the error.

Post NRC research on bias led to the same observations: bias exists and may impact decision making in fingerprint examination. For example, the decision as to whether or not the mark is suitable for further comparison (and, ultimately,
identification) can be subject to biasing influences. The knowledge of a previous decision by another expert may, in some situations, influence the decision made [30]. The presence of the potential source comparison print significantly alters the number of characteristics annotated [31]. Study of the verification stage showed that examiners can also be influenced by contextual bias information (e.g. the conclusion of another expert), but rather unexpectedly as it was shown to impact more on the false negatives (false exclusions) than on the false positives (false associations) [32]. During the comparison phase, experts having access to consensus information from other fingerprint experts demonstrated more consistency and accuracy in minutiae selection. They also showed higher accuracy, sensitivity, and specificity in the decisions reported [33]. When practice is analysed from a very operational perspective measuring success rates and efficiency [34], there is no clear evidence that bias (due to the knowledge of contextual element of the case at hand) has a large scale and systematic adverse effect.

(b) Accuracy, reliability and reproducibility of fingerprint experts

The accuracy, reliability, repeatability and reproducibility of fingerprint decisions have been tested in the so-called ‘black box’ study [35,36]. It details, for a corpus of experts, the percentages of decisions of identification and exclusion that indeed correspond to ground truth. A very low rate of false positives was observed (0.1%). Among the marks determined as of value for identification, examiners were unanimous on 48% of mated pairs and on 33% of non-mated pairs (on average, a pair was examined by 23 examiners). This demonstrates a certain lack of consensus. This (lack of) reproducibility was then compared with the repeatability (intra-examiner). Here, 89.1% of individualization decisions and 90.1% of exclusion decisions were repeated. Most changes of opinion were towards inconclusive decisions. However, none of the four false positive errors included in this study were repeated [36]. The observed lack of repeatability and reproducibility increases with the difficulty or complexity of the mark as judged by the examiners.

A similar trend is observed in a recent study where challenging cases were submitted to more than 150 fingerprint experts [37]. On one case of a close non-match (a case selected through the use of an AFIS system to maximize the chance of finding potential adventitious associations), 11 identifications were reported for a total of 124 rendered conclusions. It is important to keep in mind that false exclusions are more prevalent than false identifications. Ulery et al. [35] reported 7.5% of wrong exclusions for 0.1% of wrong association. Tangen et al. [38] obtained, respectively, 7.88% and 0.68%. Finally, Neumann et al. [37], on challenging cases, obtained 4.92% and 0.67%, respectively.

How value judgements are carried out has been studied in the so-called ‘white-box’ studies. These use the quantity of annotated features as well as the quality map and other features, and how these elements link to the value judgement [37,39,40]. Results show that minutiae count is the best predictor of value judgement. However, substantial variations of both annotations and conclusions among examiners have been observed in all studies. When focusing only on minutiae, other studies led to the same conclusion that there is quite an important range of variation between experts [41,42]. This observation is not new: Evett & Williams [43] observed it in the context of the 16-point standard review, but over the years no efforts have been put into managing such variability. Again standardization and training will be a key element for improving the process and providing greater transparency in casework.

(c) Quality measures

Hicklin et al. [44] study the process of analysis including local and overall assessment of clarity. This endeavour is based on a survey of examiners, where the participants were asked to assess the local quality of 70 marks [45]. The standardization of the process is proposed [44], and an interface presented. It paves the way towards automatic methods to efficiently measure the quality (or information content) of marks. Such research opens the possibility of automatically distinguishing between complex marks and non-complex marks as suggested in the Fingerprint Inquiry report. Kellman et al. [46] have also shown the possibility of using metrics characterizing the image of the mark to predict expert performance and the assessment of difficulty in fingerprint comparisons.

(d) Statistical modelling

A few studies came to add to the body of knowledge in relation to minutiae prevalence and distribution [47–49]. Some efforts have recently been put into understanding the contribution of pore structures [50]. But the more significant contributions are in the domain of statistically modelling minutiae variability.

Reviews of the most recent efforts on statistical modelling have been published by Neumann [51], followed by Abraham et al. [52]. It translates a steady trend to invest in research that offers the potential to qualify statistically the probative strength to be assigned to a fingerprint comparison. The purpose here is not to conduct this exhaustive analysis again, but to focus on the main contributions that can have an operational impact.

One model has been published in the Journal of the Royal Statistical Society and represents the culmination of a few years of research conducted by the Forensic Science Service in England and Wales. Neumann et al. [53] presented a very advanced model to compute likelihood ratios (weight-of-evidence to be assigned to a comparison between a mark and a print), integrating variations in annotations and due to distortion. Data concerning the validation of that model are also presented. To date, this work presents the most extensive validation exercise towards a probabilistic system that can be implemented in casework. The operational use of this model has also been studied [54]. The marks considered were those not recovered initially (due to low quality), recovered but considered of insufficient quality for identification in the analysis stage, or marks that were compared with a fingerprint and where the conclusion was inconclusive, all in the normal course of casework. A few additional associations were found by examining a large amount of marks. While a generalized application of the model to all marks does not seem cost-efficient, some contexts where the use of such marks with a probability model is cost-efficient are highlighted.

The second group of models takes advantage of scores obtained from AFIS systems to assign a weight to a given comparison [55,56]. These models have been published and are moving towards operational validation and implementation.
This research progress is not going without healthy scientific debate in the specialized literature [57,58].

(e) Reporting and communication

Found & Edmond [59] articulated, informed by scientists, lawyers and psychologists, the expected contents of a report in the pattern evidence domains. It is followed by a guide for interpreting testimony and for cross-examination [60,61]. It is obvious that the NRC report has changed the attitude of the judiciary who are much more and rightfully inclined to ask for more detailed reports (e.g. a section on the limits and current controversies) and to challenge the fingerprint findings put forward to them. This thus means that experts will have to strengthen their communication skills.

There is a body of work starting to emerge that looks at the readability of reports, the production of standardized report templates and appendices for lay persons, and the adoption of standard terms to communicate the findings (see [62] and associated references).

We still have to see how this will cascade down to the fingerprint practice. An early attempt at conveying fingerprint evidence in probabilistic form has been recently conducted using a mock trial exercise [63]. It emerged that mock jurors (not fingerprint specialists, but people from the general public) overall understood the testimony rather well, and integrated the result into their reasoning.

4. Impact of the National Research Council report in the courtrooms

(a) In the USA

My purpose here is not to carry out a full analysis (interested readers may refer to [64–66]) for the cases in the US post NRC report, but a few highlights below will provide the global tendency. Overall, the NRC report had no impact on trial court acceptance of fingerprint evidence. Only some guidance regarding how fingerprint evidence should be presented has been given by courts. More specifically:

In United States v. Rose,18 the Court ultimately agreed with the prosecution and held that the NRC report identified a need for additional research but did not conclude that fingerprint evidence was unreliable such as to render it inadmissible. Indeed, very quickly, some prosecutors went further in court briefs stating that the NRC Report was not intended to affect admissibility decisions. Judge Edwards (Chairman of the committee who prepared the NRC report) reacted strongly against that view and suggested quite the contrary [67]: judges should account for the NRC report in their decision-making regarding admissibility. Judge Edwards made extensive reference to the order by Judge Gertner.19 She indicated that although the admissibility of this kind of evidence was effectively presumed, largely because of its pedigree—the fact that it had been admitted for decades—admissibility ought not to be presumed but carefully examined in each case in light of the NRC concerns.

In Commonwealth v. Gambora,20 Judge Spina took a strong stance on claim of certainty: ‘Claims of absolute certainty are particularly irresponsible by a science based in large part on human judgment’.

In United States v. Herrera,21 the court went as far as suggesting that fingerprint expert opinion regarding sources is akin to an art expert or similar to eyewitness testimony. The court stated that ‘Matching evidence of the kinds that we’ve just described, including fingerprint evidence, is less rigorous than the kind of scientific matching involved in DNA evidence’; and recognized that ‘evidence doesn’t have to be infallible to be probative’ and declared fingerprint evidence to be admissible.

In United States v. McCluskey,22 the court held that the fingerprint identification testimony, while perhaps not ‘scientific’, is sufficiently reliable to be admitted into evidence at trial, but the expert ‘will not be permitted to testify that any individual is the source of a particular print ‘to the exclusion of all others,’ or that she is ‘100% certain’ about an identification, or any variant thereof. There simply is no evidence in the record to support such a conclusion. To the contrary, the National Research Council, the FBI, and SWGFAST have all recognized the lack of scientific basis for such testimony and have advised against permitting examiners to express opinions to this level of certainty. Such a conclusion lacks a reliable scientific basis.’

These cases attest to the increasing tendency among courts to refrain from accepting fingerprint evidence as facts that can be expressed with 100% certainty or suggesting that the evidence alone is enabling the exclusion of all others in the world except the concerned individual. Legal scholars echoed that call for more humble conclusions [10,11,13].

So did SWGFAST in its 2012 position statement23: ‘SWGFAST recognizes that individualization has been used within the latent print community to mean ‘to the exclusion of all others’. The ability of a latent print examiner to individualize a single latent impression, with the implication that they have definitively excluded all other humans in the world, is not supported by research and was removed from SWGFAST’s definition of individualization’.

(b) In the UK

In R. v. Smith [68], the court of Appeal of England and Wales quashed a conviction and made general observations regarding the provision of fingerprint evidence. The court was astonished by the absence of notes taken at the time of examination, stating that ‘No competent forensic scientist in other areas of forensic science these days would conduct an examination without keeping detailed notes of his examination and the reasons for his conclusions’. In relation to the reports produced, the court’s decision stressed that: The quality of the reports provided by the Nottinghamshire Fingerprint Bureau for the trial reflected standards that existed in other areas of forensic science some years ago and not the vastly improved standards expected in contemporary forensic science. This case is echoing the issues raised in the Finger-print Inquiry in Scotland and has led the UK forensic science regulator to re-think quality standard in this area [69].

5. A way forward?

In my view, there are a few areas of the discipline that require critical attention. I take them in turn and will conclude with some prioritization at the end.
(a) Clarifying the inference process

My position regarding reported conclusions in fingerprint matters was expressed in 2008 already [70]: experts shall abandon the identification/individualization conclusion altogether. Cole followed with the same recommendation in 2009 [71] and some fingerprint experts start to come to the same conclusion [72]. A proper evaluation of the findings calls for an assignment of two probabilities. The ratio between these two probabilities gives all the required information that allows discriminating between the two propositions at hand and the fact finder to take a stand on the case. This approach is what is generally called the Bayesian framework [73]. Nothing prevents its adoption for fingerprint evidence. That vision just acknowledges the fundamental contribution of Biedermann et al. [74,75] on the topic of decision theory. Cole [14] rightly expressed his surprise that SWGFAST was not endorsing decision theory, as in Biedermann et al. [74], to justify its choice for the term ‘decision’ in the definition of identification. This is a mistake; decision theory is the only logical way to understand and justify current practices.

Another, more fundamental point is that SWGFAST appears to misconceive the nature of decision theory as it is endorsed in Biedermann et al. [74]. As Biedermann et al. [76] later emphasized, decision theory is of a normative nature, whereas SWGFAST appears to use the term ‘decision’ in a descriptive sense (i.e. to describe what experts actually do—whereas the normative perspective would imply to take awareness of what, from a logical point of view, ought to be done). To bring it to the point: what SWGFAST does is to ‘construct’ a theory of identification based on what people do. It is like ‘constructing’ a theory for the mathematical operation of addition (e.g. 2 + 2) by asking people (without knowledge of arithmetic, i.e. the theory) what 2 + 2 is and taking their answers (that could be different from 4) as the theory or having a theory that conforms the (logically wrong) empirical result. The proper understanding of the descriptive and normative notions is fundamental and where this is not properly understood it is at the origin of many of the confused discussions in this area.

The NIST report [18, p. 68] decomposed the steps taken by an examiner to reach a decision. The examiner will (1) assign prior probabilities based on a perceived context and a perceived size of the relevant population, (2) assign the weight to be assigned to the findings arising from the fingerprint examination, (3) update the probabilities to get the posterior probabilities and finally (4) make a decision considering the result of (3) and weighted balance between benefits and costs associated with each decision. To the question of whether or not it is appropriate for an expert to take responsibility for that entire decision process, my position remains unchanged [73]; the expert should only devote his or her testimony to the strength to be attached to the forensic findings and that value is best expressed using a likelihood ratio. The questions of the relevant population—which impacts on prior probabilities—and decision thresholds are outside the expert’s province but rightly belong to the fact finder.

The topic of inference is fundamental to the issue. Most of the research in the fingerprint field is empirical, whereas the nature of the problem at first stance is logical, one of the reasonable reasoning in the face of uncertainty. This issue is conceptual, and it needs to be properly understood first before even starting to think about empirical work/research to generate data that could help translating the inferential concept into practice.

(b) Improving transparency

This is a call for greater candour in case files, report writing and the presentation of findings in court: to be able to articulate the weight to be assigned to each feature considered without resorting to comfort statements such as ‘friction ridge skin impressions are unique’,24 ‘all three levels of features have been used in conjunction’, or ‘it is my opinion based on my training and experience’.

Is it just an opinion? Yes, it is, and, at the moment, it does not offer a lot to the fact finder to assess its quality. I agree with Cole [77], there is a risk in ‘opinionization’ because it can range from respectful informed opinions to mere opinions. But more crucially, an opinion of say ‘identification’ between a mark and a print cannot be critically assessed. Indeed there is an impossibility to articulate what comes from the context and the cost–benefit analysis and what comes from the forensic findings, i.e. the comparison itself. I cannot see any other way out than to concentrate the testimony of the expert only on the comparison alone and refrain from any consideration of context or valuation of consequences given by correct and false decisions.

(c) Moving away from considering experts as black boxes

Measuring error rates from experts will provide needed indicators for quality but are not the panacea. The problem is that people (including many scientists) transpose a practice-wide ‘rate’ to any individual examiner in a given case. This is unfair in either sense, unjustifiably penalizing for the good examiners and unjustifiably praising for the bad examiners. The paper by Edmond et al. [60] offers, in my view, no viable solution for this, because it takes a practice-wide ‘rate’ for a constructive piece of information in the individual case. For example, the study by Ulery et al. [35] reported a rate of false positive is 0.1%. I cannot foresee how the fact finder will consider an average error rate when an expert testifies to an opinion of identification. If nothing in the case indicates that the expert deviated from the standard practice, his or her opinion will be simply trusted. It means that that claim that the ‘probability of an error considered as so small as to be dismissed’ will be simply admitted without any further reference to the average value of 0.1%. That process offers no mechanism to effectively measure the actual weight to be attached to the forensic findings. Only structured and systematic research on the features themselves can lead to such a state. That leads me to the next topic.

(d) Introducing statistics and resolving potential conflicting opinions

In my view, statistics are unavoidable for the next generation of fingerprint evidence. And it is not only by analogy to DNA evidence—whose weight is reported most of the time using a probabilistic measure—but also simply because it cannot be avoided scientifically and logically. I foresee the introduction in court of probability-based fingerprint evidence. This is not to say that fingerprint experts will be replaced by a statistical tool. The human will continue to outperform machines for a wide range of tasks such as assessing the features on a mark, judging its level of distortion, putting the elements into its context, communicating the findings and applying critical
thinking. But statistical models will bring assistance in an assessment that is very prone to bias: probability assignment. What is aimed at here is to find an appropriate distribution of tasks between the human and the machine. The call for transparency from the NRC report will not be satisfied merely with the move towards opinions, but also require offering a systematic and case-specific measure of the probability of random association that is at stake. It is the only way to bring the fingerprint area within the ethos of good scientific practice.

Such a transition will come with a few new challenges. The first is on research. At the moment, no published models have gone through an extensive validation and implementation exercise from an operational perspective. It would mean deploying a model in operational practice, defining its scope of operations, monitoring its rates of misleading evidence and defining the standard operating procedures. The second challenge is on the training of experts and the development of mechanisms to resolve conflicting outcomes (i.e. an expert’s opinion and the output of a model). The third challenge is one of communication with the police, the courts and all other stakeholders. The long tradition of usage of fingerprint evidence makes any change difficult and the beneficiaries of such forensic findings would need to be fully informed and educated regarding the changes outlined above.

(e) Being able to rank marks as a function of their level of quality
One of the features of the debate regarding the reliability of fingerprint evidence is that its participants tend to forget that fingerprint marks can display a range of quality, both in terms of extent of the papillary lines shown or their legibility. In a given case, it should be the information content of the mark that should be the main focus of the critical discussion. However, we tend to see discussion in abstracto. One reason for this lies with the difficulty of measuring the information content of a mark (or its level of complexity). Leaving that assessment to the expert is unsatisfactory. I think that the profession needs to develop agreed systematic methods to measure quality. Hicklin et al. [44] have paved the route with the development of a dedicated algorithm to measure the quality of marks. These efforts should be pursued.

(f) Managing the potential risk of bias
The issue of potential bias was raised as one of the top priorities of the NRC report. I expressed elsewhere my concern that there is a risk of putting too much effort on that line of inquiry compared with other research priorities [78]. Dror [79] rightly stressed that an appropriate balance needs to be found between the risks and benefits. Where the balance lies is not fully clear yet but I offer here a few suggestions. The mechanisms allowing us to guard against bias should be proportional to the information content of the marks. In other words, when the mark is rich in information, the risk that bias will significantly impact judgement is limited. The cost of adopting an all-encompassing procedure without any distinction between marks is too high compared with the benefits. This is why the measurement of the quality of marks according to an agreed (i.e. standard) procedure is so critical. Indeed, when the quality is low, all efforts should be put in place in order to mitigate contextual bias and peer-pressure. When the information content of the mark is high, these risks are limited by the quality of the mark itself. Efficient mechanisms range from using an independent panel of experts to enforcing full documentation, established at the time of examination, of the analysis (the mark alone) and the comparison phase with blind verification procedures.

To conclude, I believe that, in the future, fingerprint evidence will not be a matter of individual opinion anymore, but will be constructed through the harmonious play of multiple experts (humans and computer systems), working according to specified procedures, to deliver a consensus probabilistic assignment of the weight to be attached to their forensic findings. They will maintain fully auditable case files established at the time of examination. Such consensus final assignments will best describe the real contribution of the forensic findings compared with the expression of personal, egocentric and generally badly articulated opinions. To get to that stage, I explored above six areas that boil down to two generic priorities. The first is to improve on the conceptual understanding of the nature of the problem at hand. This aspect is not empirical but has to do with the logic of inference and decision. The profession needs to adopt sound principles of logical reasoning and decision-making. It is only after the adoption of an appropriate framework that the second priority comes, namely the operational conditions (statistical tools, quality measures, procedures and documentation, bias minimization) required to support the endeavour.

Competing interests. The author was a member of SWGFAST and is now an invited member of the Friction Ridge subcommittee (part of the Physics and Pattern Evidence Scientific Area Committee) of the Organization for Scientific Area Committees (OSAC). He is a member of the EFPWG (European Fingerprint Working Group) of the ENFSI (European Network of Forensic Science Institutes). He also was a member of the IAI Standardization II committee, of the NIST/NJF working group on human factors in latent print analysis and testified in the context of the Fingerprint Inquiry in Scotland.

Funding. I received no funding for this study.

Acknowledgements. The author thanks Alex Biedermann and Heidi Eldridge as well as the two anonymous referees for their careful review and help in improving the manuscript.

Disclaimer. Any views or opinions presented in this paper are solely those of the author.

Endnotes
1SWGFAST (Scientific Working Group on Friction Ridge Analysis Study and Technology) was a US-based group that provided guidance and elaborated standards in relation to fingerprint evidence. See www.swgfast.org. It has been replaced by the NIST/NJF Friction Ridge subcommittee (part of the Physics and Pattern Evidence Scientific Area Committee) of the Organization for Scientific Area Committees (OSAC).

2The use of the word ‘fingerprints’ is used rather loosely in this paper. Strictly speaking, we should refer to friction ridge skin (covering the inside part of hands and feet) and the associated marks and prints that portions of this skin may deposit. Most non-specialist readers may refer simply to fingerprints when discussing the potential use of the friction ridge skin in the forensic context of the identification of living or dead. For this contribution, it was decided to keep the use of ‘fingerprints’ throughout the text to facilitate communication at the sacrifice of a more rigorous terminology.

3The term ‘expert’ will be used to designate a fingerprint examiner that is asked to carry out fingerprint comparison. That person may also be called in court to testify regarding the identity of source between marks and prints. In some jurisdictions, this practitioner may be legally nominated as an expert or called as an expert witness. In others, he or she may simply be asked to provide to the mandating authority specialist
knowledge that is outside the usual realm of a layperson or a court. In other words, not all fingerprint practitioners will take for themselves the term ‘expert’. The choice of using the term ‘expert’ throughout this text is motivated by the view that most members of the public will refer to a fingerprint specialist as a fingerprint expert, regardless of his or her legal status in the procedure or in court.

4. ‘ACE-V’ is the acronym for Analysis, Comparison, Verification and Verification, the protocol traditionally followed by fingerprint experts to conduct examinations. First unknown marks or prints are analysed to judge whether or not they are suitable for further examination and then compared with the prints of known sources. The output of the comparison leads to an evaluation of the expert to form a conclusion. A second examiner subsequently verifies the conclusions.

5. Fingerprint experts generally define three levels of features. The first level refers to the general flow of the papillary ridges and the general patterns taken by them. The second level accounts for the points of ending or bifurcation of the ridges. Those are called minutiae or Galton details. Others features such as the disruptions of the pattern are associated with scars will count as second-level features. The third level refers to the specific shapes of the edges of the ridges. It also includes the location and forms associated with the pores that are the opening orifices regularly scattered on the top of each ridge, used by the body to secrete the complex products of the eccrine glands.

6. Internationally, there is no common approach regarding the criteria for fingerprint identification. Broadly speaking, the practice divides between countries applying a numerical standard (a fixed number of minutiae in agreement are required to declare an identification), or applying a holistic approach (the assessment is left to the examiner’s judgement based on the whole range of available features). Anglo-American countries are using the holistic approach.


14. It is worth mentioning that the present author was a member of the Standardization II committee that produced the report [22] for the IAI. Apart from the limited move discussed in this paper, one of its key recommendations was to rescind the 1979–1980 IAI resolutions that banned fingerprint experts from expressing conclusions in cases where a ‘categorical’ opinion could not be stated. According to the report, under some defined conditions, experts will be allowed to guide by degree when the case required. It is one essential step in the right direction, recognizing the potential weight to be attached to fingerprint comparisons that would be normally reported as ‘inconclusive’.


24. Even though friction ridge skin impression is unique, because something can only be identical with itself, that is not the issue. In practice, we are facing imperfect reproductions of skin surface marks, that may be of varying quality so that they may be confused with prints from other sources.

References


68. R v Peter Kenneth Smith. 2011 EWCA Crim 1296.


