

Gary Edmond*

WHAT LAWYERS SHOULD KNOW ABOUT THE FORENSIC ‘SCIENCES’

ABSTRACT

Recent and authoritative reports from the US, Canada and the United Kingdom question many types of forensic science and medicine evidence used routinely in criminal proceedings across the common law world. This article reviews recent reports produced by the National Academy of Sciences (US), the National Institute of Standards & Technology and National Institute of Justice (US), Lord Campbell in Scotland, and Justice Goudge in Ontario, in order to assess their implications for expert evidence and legal practice in Australia. The article suggests that Australian legal institutions have not performed well in response to forensic science and medicine evidence and remain largely oblivious to serious epistemic infirmities.

I THE WORLD OF (FORENSIC SCIENCES) TURNED UPSIDE DOWN

The bottom line is simple: In a number of forensic science disciplines, forensic science professionals have yet to establish either the validity of their approach or the accuracy of their conclusions, and the courts have been utterly ineffective in addressing this problem. For a variety of reasons — including the rules governing the admissibility of forensic evidence, the applicable standards governing appellate review of trial court decisions, the limitations of the adversary process, and the common lack of scientific expertise among judges and lawyers who

* Professor, Australian Research Council Future Fellow, and Director, Expertise, Evidence & Law Program, School of Law, The University of New South Wales, Sydney 2052, Australia (email: g.edmond@unsw.edu.au); Professor (fractional), School of Law, Northumbria University, England; and Member, Advocacy and Justice Unit, The University of Adelaide. This research was supported by the ARC (FT0992041, LP100200142 and LP120100063). An earlier version was presented to the South Australian judicial education seminar in March 2013 and to the Judicial College of Victoria in June 2013. Thanks to Emma Cunliffe, David Hamer, Kent Roach, Mehera San Roque, Andy Roberts and Andrew Ligertwood for comments and suggestions. The author was an adviser to the Inquiry into Pediatric Forensic Pathology in Ontario, 2007–08 and is a member of the Standards Australia committee for Forensic Science. Accepted 26 April 2013.

must try to comprehend and evaluate forensic evidence — the legal system is ill-equipped to correct the problems of the forensic science community.¹

Until recently, courts in most jurisdictions admitted almost all of the forensic science and medicine evidence adduced by the State. There were few challenges to this evidence and, as it turns out, remarkably few effective challenges. The liberal admission and use of incriminating ‘expert’ opinions seems to have been based on long-standing trust invested by judges (and jurors) in forensic practitioners,² a gradual weakening of exclusionary rules, and a related confidence in the accusatorial trial’s mechanisms for identifying and conveying evidentiary weakness.³ However, trials and appeals have not performed well. They have not held forensic analysts to account and have not, systematically or otherwise, exposed and addressed pervasive and profound epistemic problems plaguing the contemporary forensic sciences. Trial and appellate judges, and presumably jurors, seem to have accepted that forensic analysts were generally proffering opinions derived from scientifically predicated techniques, where highly qualified analysts had rigorously followed research-based protocols. In many cases, and perhaps most, they were wrong to do so. It is significant, as this essay explains, that no court in any jurisdiction independently arrived at the conclusions produced by the Committee on Identifying the Needs of the Forensic Science Community assembled by the National Academy of Sciences (United States) reproduced above.

This essay reviews recent inquiries into the forensic sciences and medicine across the Anglophone world.⁴ Spanning several advanced common law jurisdictions and ranging from general reviews of the forensic sciences to more focused inquiries into paediatric forensic pathology and even the controversy created by a questioned latent fingerprint attribution, the inquiries are remarkably consistent in their findings

¹ Committee on Identifying the Needs of the Forensic Science Community, National Research Council, *Strengthening Forensic Science in the United States: A Path Forward* (National Academies Press, 2009) 53 (‘NRC Report’).

² The word ‘expert’ is italicised or featured with scare quotes because in many circumstances we do not know if the relevant individuals actually possess expertise or special abilities.

³ The weakening of exclusionary rules, often associated with free proof (ie the admission of all logically relevant evidence), is conventionally associated with Jeremy Bentham (1748–1832), and more recently the work of Sir Rupert Cross and Larry Laudan. See, eg, Larry Laudan, *Truth, Error, and Criminal Law: An Essay in Legal Epistemology* (Cambridge University Press, 2005).

⁴ While the reports are drawn exclusively from common law jurisdictions, many of the epistemic problems and issues apply in other systems, including ‘inquisitorial’ systems — where there are often few possibilities to realistically challenge the evidence produced by a court-appointed expert. See Gary Edmond and Joelle Vuille, ‘Comparing the Use of Forensic Science Evidence in Australia, Switzerland and the United States: Transcending the Adversarial/Non-Adversarial Dichotomy’ (2014) 54 *Jurimetrics Journal* 221.

and recommendations.⁵ Each inquiry identifies serious and pervasive problems with forensic science and medicine evidence, the practices behind them, as well as organisational structures and management. Simultaneously, they explain or imply that lawyers and judges have not appreciated the magnitude or prevalence of these issues and so have been correspondingly ineffective in their practice. Having reviewed reports of inquiries from the US, Scotland and Canada, Part III of the essay turns to consider some of the unsettling implications for both accusatorial trial systems in general and Australian legal practice in particular. In doing so, the essay confronts and rejects the contention that the forensic sciences in Australia should be considered exceptional.

The title of this essay uses the term ‘forensic sciences’ rather than the singular ‘science’. The distinction is significant. It is important to recognise variation across the forensic sciences.⁶ The many problems and issues confronting the forensic sciences and medicine are not evenly distributed. Complicating things considerably, there is no simple correlation between longevity or legal acceptance of techniques and derivative opinions and their validity and reliability. Some of the oldest techniques (eg latent fingerprint comparison) never became scientific. Some of the most recent techniques are among the most reliable and some of the concerns raised in the reports infect even techniques built on contemporary biology (eg DNA profiling), toxicology and chemistry (eg CG-MS). For even techniques that are demonstrably reliable are often applied and practiced in ways that are insufficiently attentive to the limits of validation and risks posed by human factors — particularly contextual bias and the cross-contamination of evidence. Forensic science and medicine evidence also, and inescapably, depends on collection practices and chains of custody.⁷ Consequently, the various reports and their recommendations offer a wake-up call of general application. They raise profound questions and difficulties concerning the continuing use of forensic science and medicine evidence, as the many problems, uncertainties and threats are addressed, and in some cases the necessary research is retrospectively attempted.

What we should do now, and how we should respond to decades of convictions based on questionable techniques and exaggerated incriminating opinions (mis)portrayed as ‘expert’, ‘scientific’ and/or ‘reliable’, is far from obvious. The question of whether lawyers and judges are capable of domesticating the forensic sciences and what they should do given the inadequacies of historical performances is again, unclear. Bumbling along, as we appear to have done for more than a century, seems not merely misguided but, in the wake of the collective weight of these and other reports,

⁵ The ‘voice’ in Part II reflects the tone of the reports. The author’s voice appears in the residual sections.

⁶ These are discussed in Part II and Part III, especially Part III(A).

⁷ The meaning ascribed to artefacts (as evidence) also depends on stories and assumptions, sometimes as mundane as who owned and wore an article of clothing or the cultures of sharing clothing amongst family and friends. See, eg, Michael Lynch et al, *Truth Machine: The Contentious History of DNA Finger-Printing* (University of Chicago Press, 2008).



undesirable.⁸ Simultaneously, it threatens the legitimacy of traditional criminal justice institutions.

For those who wonder whether these recent reports exaggerate the apparently parlous condition of many areas of forensic science and medicine, I direct their attention to the various authors and committees.⁹ Most of these inquiries and reports had eminent lawyers and judges centrally involved (eg Judge Edwards as Co-Chair, Lord Campbell as Commissioner and Justice Goudge as Commissioner).¹⁰ The reference to judicial participants is not merely an appeal to legal authority, but rather an admonition to

⁸ See, eg, Manitoba, Commission of Inquiry into Certain Aspects of the Trial and Conviction of James Driskell, *Report of the Commission of Inquiry into Certain Aspects of the Trial and Conviction of James Driskell* (Queen’s Printer, 2007) (Chair: Patrick LeSage); Ministry of the Attorney General, Ontario, *Report of the Kaufman Commission on Proceedings Involving Guy Paul Morin* (Queen’s Printer, 1998) (Chair: Fred Kaufman). Going slightly further back we might consider: Sir John May, *Final Report: Return to an Address of the Honourable the House of Commons dated 30 June 1994 for a Report into the Circumstances Surrounding the Conviction Arising Out of the Bomb Attacks in Guildford and Woolwich in 1974* (London: HMSO 1994); Viscount Runciman, *Royal Commission on Criminal Justice*, (London: HMSO, 1993); Paul Roberts, Chirs Wilmore and Gywnn Davis, *The Role of Forensic Science Evidence in Criminal Proceedings: Royal Commission on Criminal Justice Research Study No 11* (London: HMSO, 1993); Trevor Rees Morling, *Report of the Commissioner the Hon Mr Justice TR Morling: Royal Commission of Inquiry into the Chamberlain Convictions* (Government Printer, 1987); Carl R Shannon, *Royal Commission of Inquiry in Respect to the Case of Edward Charles Splatt* (Government Printer, Adelaide, 1984). On ‘bumbling along’, Cf Charles Lindblom, ‘The Science of Muddling Through’ (1959) 19 *Public Administration Review* 79.

⁹ The reader might also be interested in some of the preceding research and reviews see, eg, D Michael Risinger, Mark P Denbeaux and Michael J Saks, ‘Exorcism of Ignorance as a Proxy for Rational Knowledge: The Lesson of Handwriting “Expertise”’ (1989) 137 *University of Pennsylvania Law Review* 731; Michael D Risinger et al, ‘The *Daubert/Kumho* Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion’ (2002) 90 *California Law Review* 1; Michael Saks and Jonathan J Koehler, ‘The Coming Paradigm Shift in Forensic Identification Science’ (2005) 309 *Science* 892; Margaret Berger, ‘What Has a Decade of *Daubert* Wrought?’ (2005) 95 *American Journal of Public Health* S59; Margaret Berger, ‘Expert Testimony in Criminal Proceedings: Questions *Daubert* Does Not Answer’ (2003) 33 *Seton Hall Law Review* 1125; David L Faigman, *Legal Alchemy: The Use and Misuse of Science in the Law* (WH Freeman, 1999); Erica Beecher-Monas, *Evaluating Scientific Evidence: An Interdisciplinary Framework for Intellectual Due Process* (Cambridge University Press, 2007); Kelly Pyrek, *Forensic Science under Siege: The Challenges of Forensic Laboratories and the Medico-Legal Investigation System* (Academic Press, 2007).

¹⁰ See also Ian Binnie, ‘Science in the Courtroom: The Mouse that Roared’ (2008) 27(2) *Advocates Society Journal* 11; Lord Leveson, ‘Expert Evidence in Criminal Courts — the Problem’ (Lecture to the Forensic Science Society, UK, November 2010); Judge Nancy Gertner, ‘The Need for a Research Culture in the Forensic Sciences’ (2011) 58 *UCLA Law Review* 789, 798; Lord Thomas, Chief Justice of England, ‘Expert Evidence and the Future of Forensic Science in Criminal Trials’ (Criminal Bar Association Kalisher Lecture, 14 October 2014).



consider the findings of senior lawyers and, very importantly, others who have devoted considerable attention to the issues. On this point, it is informative to reproduce the reflections of Judge Harry T Edwards, Senior Circuit Judge and Chief Judge Emeritus for the United States Court of Appeals for the District of Columbia Circuit. Writing about his experiences chairing the multidisciplinary committee responsible for the epigraph at the beginning of this essay, Edwards explained:

I started this project with no preconceived views about the forensic science community. Rather, I simply assumed, as I suspect many of my judicial colleagues do, that forensic science disciplines typically are well-grounded in scientific methodology and that crime laboratories and forensic science practitioners follow proven practices that ensure the validity and reliability of forensic evidence offered in court. I was surprisingly mistaken ...¹¹

Given the breadth and depth of the reports, in combination with the range and number of eminent individuals involved in producing them, it seems just as inappropriate to adhere steadfastly to preconceived views as it is to dismiss their findings and recommendations pre-emptively.

For those who might contend that this essay exaggerates the findings and recommendations, I have intentionally included a considerable number of extracts and references, such that many of the findings, criticisms and recommendations are expressed in the words of actual commissioners and committees. I would, nevertheless, encourage any would-be sceptic to finger through the reports themselves.

A Introduction to the Reports

This essay reviews four reports focused on the forensic sciences and/or forensic medicine, namely: *Strengthening Forensic Science in the United States: A Path Forward* ('NRC Report'), *The Scottish Fingerprint Inquiry Report* ('SFI Report'), *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach* ('NIST Report') and the *Inquiry into Pediatric Forensic Pathology in Ontario* ('Goudge Report').¹² Two of the reports emerged from the US, one from Scotland and the fourth from Canada. The essay endeavours to present the reader

¹¹ Harry T Edwards, 'Solving the Problems That Plague the Forensic Science Community' (2009) 50 *Jurimetrics Journal* 5, 7. See also Harry T Edwards, 'The National Academy of Sciences Report on Forensic Sciences: What It Means for the Bench and Bar' (2010) 51 *Jurimetrics Journal* 1.

¹² Committee on Identifying the Needs of the Forensic Science Community, National Research Council, *Strengthening Forensic Science in the United States: A Path Forward* (National Academies Press, 2009) ('NRC Report'); Lord Campbell, *The Fingerprint Inquiry Report* (APS Group Scotland, 2011) ('SFI Report'); Expert Working Group on Human Factors in Latent Print Analysis, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach* (US Department of Commerce, National Institute of Standards and Technology, 2012) ('NIST Report'); Stephen T Goudge, *Inquiry into Pediatric Forensic Pathology* (Queen's Printer, 2008) ('Goudge Report').

with an overview of the four reports in order to convey the remarkable consistency in both the depth and types of problems they describe, as well as the recommendations advanced by diffuse committees and commissioners from quite different common law jurisdictions. What follows focuses primarily on the findings and recommendations pertaining to scientific and medical evidence.

The inquiries and reports are all recent, published between 2008 and 2012 respectively. They arose from a range of mistakes, published scholarly concerns and from the impact of new technologies. The emergence of DNA profiling, with its rigorous mainstream scientific foundation and serendipitous value for criminal justice systems, threw an unexpected and, as it turns out, somewhat unwelcome light on many aspects of practice in the forensic sciences and medicine. In combination with the work of Innocence Projects, and in particular, the refinement of DNA profiling (through reforms to the collection, handling, processing, and the application of population genetics and statistics), these developments produced greater awareness of the frailties of many forms of forensic science. Simultaneously, they illuminated the performance of expert witnesses and prosecutors — and indirectly juries and judges — as convictions predicated upon speculative, exaggerated and mistaken forensic science and medicine evidence were painstakingly overturned.¹³ Most of the reports discussed in this essay emerged in response to problems (or mistakes) associated with latent fingerprint evidence.¹⁴ The NRC and NIST Reports were, in part, responses to the FBI’s mistaken attribution of a latent print recovered from a fragment of a train bomb in Madrid to Brandon Mayfield.¹⁵ Lord Campbell’s inquiry into fingerprint evidence in the McKie case followed protracted controversy surrounding an attribution by a Scottish fingerprint bureau. The fourth, the Goudge Inquiry, flowed from a series of mistakes and wrongful convictions associated with flawed forensic pathology evidence in Ontario.

This essay is written for a legal audience, primarily lawyers and judges exposed to forensic science and medicine evidence in the course of their practice. It aims to encourage prosecutors to reconsider their professional obligations and performances as ‘ministers of justice’, to embolden defence lawyers to challenge techniques and opinions that have not been evaluated (even if they have been uncritically accepted for decades) and to pay close attention to analytical processes and reports, and for

¹³ Barry Scheck, Peter Neufeld and Jim Dwyer, *Actual Innocence* (New American Library, 2003); Brandon L Garrett, *Convicting the Innocent: Where Criminal Prosecutions Go Wrong* (Harvard University Press, 2011).

¹⁴ The focus on fingerprints is revealing because fingerprint comparison is one of the oldest of the institutionalised forensic sciences, remains in widespread use, has an iconic status, and is still presented (and understood) as basically infallible evidence of identity. See Federal Bureau of Investigation, *The Science of Fingerprints: Classification and Uses* (Department Of Justice, 1984). Three of the reports directly challenge this status. See generally Simon A Cole, *Suspect Identities: A History of Fingerprinting and Criminal Identification* (Harvard University Press, 2001).

¹⁵ See generally Office of the Inspector General, US Department of Justice, *A Review of the FBI’s Handling of the Brandon Mayfield Case* (2006).

judges to begin to refine their admissibility jurisprudence and temper their, apparently misguided, confidence in the protections afforded by trial safeguards.¹⁶ For regardless of what the institutionalised forensic sciences and medicine do in response to these reviews and recommendations, lawyers and judges would seem to be *on notice*.¹⁷

II INSIGHTS FROM THE UNITED STATES, SCOTLAND AND CANADA

A Strengthening Forensic Science in the United States (2009)

The forensic science system, encompassing both research and practice, has serious problems that can only be addressed by a national commitment to overhaul the current structure that supports the forensic science community in this country. This can only be done with effective leadership at the highest levels of both federal and state governments, pursuant to national standards, and with a significant infusion of federal funds.¹⁸

The most important, most authoritative and most disruptive of the recent inquiries and reports is a broad ranging independent review undertaken in the United States.¹⁹ In 2006, Congress authorised the National Academy of Sciences ('NAS') to undertake a review of the forensic sciences.²⁰ The NAS assembled a multidisciplinary committee from its National Research Council ('NRC') — the Committee on Identifying the Needs of the Forensic Science Community ('the NRC Committee') — composed of statisticians, a judge (Edwards), law professors, a chemist, forensic scientists, an engineer, biologists, computer scientists and a medical examiner.²¹ The Committee's report, *Strengthening Forensic Science in the United States: A Path*

¹⁶ It provides something of an evidentiary background to the arguments advanced in Gary Edmond, 'Specialised Knowledge, the Exclusionary Discretions and Reliability: Reassessing Incriminating Expert Opinion Evidence' (2008) 31 *University of New South Wales Law Journal* 1; Gary Edmond, 'The Admissibility of Forensic Science and Medicine Evidence under the Uniform Evidence Law' (2014) 38(3) *Criminal Law Journal* 136; Gary Edmond, '(Ad)Ministering Justice: Expert Opinion Evidence and the Professional Responsibilities of Prosecutors' (2013) 36 *University of New South Wales Law Journal* 921.

¹⁷ In some ways the situation is reminiscent of the constructive knowledge (and responsibility) imputed in tobacco litigation. For an overview, see Robert N Proctor, *Golden Holocaust: Origins of the Cigarette Catastrophe and the Case for Abolition* (University of California Press, 2012).

¹⁸ NRC Report, above n 12, xx.

¹⁹ Interestingly, the National Institute of Justice ('NIJ') undertook an inquiry a decade before the NRC review. See NIJ, US Department of Justice, *Forensic Sciences: Review of Status and Needs* (Gaithersburg, Maryland, February 1999). The earlier report acknowledges some of the limitations but represents insiders' perspectives and is far less critical.

²⁰ *Science, State, Justice, Commerce and Related Agencies Appropriations Act of 2006*, Pub Law No 109-108, 119 Stat 2290 (2005).

²¹ NRC Report, above n 12, v.

Forward (known variously as the NAS or NRC Report) followed a multi-year inquiry involving consultations and submissions.²²

The report begins by introducing the reader to the fact that that ‘[t]he term “forensic science” encompasses a broad range of disciplines’ exhibiting ‘wide variability with regard to techniques, methodologies, reliability, level of error, research, general acceptability, and published material’.²³ They range from laboratory-based disciplines to those based on the interpretation of patterns. Consequently, the term applies to ‘a broad array of activities, with the recognition that some of these activities might not have a well-developed research base, are not informed by scientific knowledge, or are not developed within the culture of science.’²⁴

The NRC Committee found the existing system of forensic sciences to be fragmented and inconsistent. Having acknowledged variation, along with the fact that many large state and federal laboratories were better resourced and staffed than the myriad of smaller laboratories and police departments, the NRC Committee found the general absence of ‘meaningful’ standards to be disconcerting:

Often there are no standard protocols governing forensic practice in a given discipline. And, even when protocols are in place (eg SWG standards), they often are vague and not enforced in any meaningful way. In short, the quality of forensic practice in most disciplines varies greatly because of the absence of adequate training and continuing education, rigorous mandatory certification and accreditation programs, adherence to robust performance standards, and effective oversight. These shortcomings obviously pose a continuing and serious threat to the quality and credibility of forensic science practice.²⁵

The problems, however, ran deeper. The absence of standards and, in many cases, constraints on practice followed the widespread paucity of research and a research culture:

²² Ibid 37, 40–3, 44–8. According to the NRC Report, the inquiry was commissioned in the aftermath of the FBI’s mistaken attribution of a fingerprint on a bomb in Madrid, coincident with ‘the growing number of exonerations resulting from DNA analysis’, the ‘greater expectations for precise forensic science evidence raised by DNA testing [forcing] new scrutiny on other forensic techniques’, chronic underfunding and case backlogs, emerging criticisms from beyond the forensic science communities, disquieting episodes of hubris and fraud and even the impact of media and CSI.

²³ Ibid 38.

²⁴ Ibid 39, 182, 188. On this last point see Jennifer L Mnookin et al, ‘The Need for a Research Culture in the Forensic Sciences’ (2011) 58 *UCLA Law Review* 725; David A Harris, *Failed Evidence: Why Law Enforcement Resists Science* (New York University Press, 2012).

²⁵ NRC Report, above n 12, 6 (citations omitted). ‘SWG’ is the acronym for Scientific Working Group.

Little rigorous systematic research has been done to validate the basic premises and techniques in a number of forensic science disciplines. The committee sees no evident reason why conducting such research is not feasible ...²⁶

The lack of systematic research was, and remains, a serious problem. It is problematic because many of the forensic science disciplines proffer interpretations that are not supported by research. Prominent examples are comparison and pattern recognition techniques, often associated with identification. Many types of ‘identification’ evidence (eg those relying on images, voices, hair and fibres, tool and bite marks, ballistics, documents, foot, shoe and tyre marks, fingerprints and so on) purport to ‘match’ a trace with a specific source.²⁷ The NRC Committee was sceptical about such claims:

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.* In terms of scientific basis, the analytically based disciplines generally hold a notable edge over disciplines based on expert interpretation. But there are important variations among the disciplines relying on expert interpretation. For example, there are more established protocols and available research for fingerprint analysis than for the analysis of bite marks. ... *The simple reality is that the interpretation of forensic evidence is not always based on scientific studies to determine its validity. This is a serious problem.* Although research has been done in some disciplines, *there is a notable dearth of peer-reviewed, published studies establishing the scientific bases and validity of many forensic methods.*²⁸

The NRC Committee found the institutionalised forensic sciences to be under-resourced and to maintain ‘thin ties to an academic research base that could support the forensic science disciplines and fill knowledge gaps.’²⁹ Though, of ‘the various facets of under resourcing, the committee [was] most concerned about the knowledge base’:³⁰

A body of research is required to establish the limits and measures of performance and to address the impact of sources of variability and potential bias.

²⁶ Ibid 189.

²⁷ Ibid 183.

²⁸ Ibid 7–8, 87, 100, 128 (emphasis added). See also 42: ‘The fact is that many forensic tests — such as those used to infer the source of toolmarks or bite marks — have never been exposed to stringent scientific scrutiny.

²⁹ Ibid 15.

³⁰ Ibid. On the need for collaboration with ‘broader science and engineering communities’, see 189.



Such research is sorely needed, but it seems to be lacking in most of the forensic disciplines that rely on subjective assessments of matching characteristics.³¹

When it came to legal reliance on forensic science evidence, the NRC Committee’s assessment was stark:

The law’s greatest dilemma in its heavy reliance on forensic evidence, however, concerns the question of whether — and to what extent — there is *science* in any given forensic science discipline.

Two very important questions should underlie the law’s admission of and reliance upon forensic evidence in criminal trials: (1) the extent to which a particular forensic discipline is founded on a reliable scientific methodology that gives it the capacity to accurately analyze evidence and report findings and (2) the extent to which practitioners in a particular forensic discipline rely on human interpretation that could be tainted by error, the threat of bias, or the absence of sound operational procedures and robust performance standards. These questions are significant. *Thus, it matters a great deal whether an expert is qualified to testify about forensic evidence and whether the evidence is sufficiently reliable to merit a fact finder’s reliance on the truth that it purports to support. Unfortunately, these important questions do not always produce satisfactory answers in judicial decisions pertaining to the admissibility of forensic science evidence proffered in criminal trials.*³²

During the course of its deliberations the NRC Committee came to the conclusion that ‘truly meaningful advances will not come without significant concomitant leadership from the federal government.’³³ Recognising that the ‘forensic science enterprise lacks the necessary governance structure to pull itself up from its current weaknesses’ it recommended establishing a new national institute to begin the ‘substantial improvement’ required to place forensic science and medicine on a firm scientific foundation.³⁴ The NRC Committee advocated strong and independent leadership and governance, closely connected with the ‘Nation’s scientific research base’.³⁵ For the Committee it was vital that any ‘entity that is established to govern the forensic science community cannot be principally beholden to law enforcement’, rather ‘it must be equally available to law enforcement officers, prosecutors, *and* defendants in the criminal justice system.’³⁶ This led to the first of the NRC Committee’s 13 recommendations:

³¹ Ibid 8.

³² Ibid 9, 43, 87, 111 (emphasis added).

³³ Ibid 16.

³⁴ Ibid 16, 37.

³⁵ Ibid 16.

³⁶ Ibid 17.



To promote the development of forensic science into a mature field of multidisciplinary research and practice, founded on the systematic collection and analysis of relevant data, Congress should establish and appropriate funds for an independent federal entity, the National Institute of Forensic Science (NIFS).³⁷

The NRC Committee proposed the following as ‘minimum criteria’ for the national institute:

- It must have a culture that is strongly rooted in science, with strong ties to the national research and teaching communities, including federal laboratories.
- It must have strong ties to state and local forensic entities as well as to the professional organizations within the forensic science community.
- It must not be in any way committed to the existing system, but should be informed by its experiences.
- It must not be part of a law enforcement agency.
- It must have the funding, independence, and sufficient prominence to raise the profile of the forensic science disciplines and push effectively for improvements.
- It must be led by persons who are skilled and experienced in developing and executing national strategies and plans for standard setting; managing accreditation and testing processes; and developing and implementing rulemaking, oversight, and sanctioning processes.³⁸

Accordingly, NIFS would be obliged to address deficiencies and ‘focus on’:

- a) establishing and enforcing best practices for forensic science professionals and laboratories;
- b) establishing standards for the mandatory accreditation of forensic science laboratories and the mandatory certification of forensic scientists and medical examiners/forensic pathologists — and identifying the entity/entities that will develop and implement accreditation and certification;
- c) promoting scholarly, competitive peer-reviewed research and technical development in the forensic science disciplines and forensic medicine;
- d) developing a strategy to improve forensic science research and educational programs, including forensic pathology;
- e) establishing a strategy, based on accurate data on the forensic science community, for the efficient allocation of available funds to give strong support to forensic methodologies and practices in addition to DNA analysis;
- f) funding state and local forensic science agencies, independent research projects, and educational programs as recommended in this report, with conditions that aim to advance the credibility and reliability of the forensic science disciplines;

³⁷ Ibid 19, 78 [Recommendation 1].

³⁸ Ibid 18–19.

- g) overseeing education standards and the accreditation of forensic science programs in colleges and universities;
- h) developing programs to improve understanding of the forensic science disciplines and their limitations within legal systems; and
- i) assessing the development and introduction of new technologies in forensic investigations, including a comparison of new technologies with former ones.³⁹

NIFS would provide an important ‘interface between the forensic science and medical examiner communities and basic sciences’.⁴⁰

In advocating a national institute the NRC Committee was conscious of the historical proximity of forensic science and medicine to law enforcement. It insists on distance and autonomy because the ‘best science is conducted in a scientific setting as opposed to a law enforcement setting.’⁴¹ Recommendation 4 places firm and conspicuous emphasis on independence:

To improve the scientific bases of forensic science examinations and to maximize independence from or autonomy within the law enforcement community, Congress should authorize and appropriate incentive funds to the National Institute of Forensic Science (NIFS) for allocation to state and local jurisdictions for the purpose of removing all public forensic laboratories and facilities from the administrative control of law enforcement agencies or prosecutors’ offices.⁴²

The NRC Committee characterises the identification of errors, their elimination and the estimation of remaining errors as ‘a key task’ for those designing validation studies ‘as well as for the analyst applying a scientific method to conduct a particular analysis’.⁴³

All results for every forensic science method should indicate the uncertainty in the measurements that are made, and studies must be conducted that enable the estimation of those values. ... the accuracy of forensic methods resulting in classification or individualization conclusions needs to be evaluated in well-designed and rigorously conducted studies. The level of accuracy of an analysis is likely to be a key determinant of its ultimate probative value.⁴⁴

The ‘existence of several types of potential error makes it absolutely critical ... to be explicit and precise in the particular rate or rates referenced in a specific setting.’⁴⁵

³⁹ Ibid 19–20.

⁴⁰ Ibid 189.

⁴¹ Ibid 23.

⁴² Ibid 24 [Recommendation 4].

⁴³ Ibid 116.

⁴⁴ Ibid 184, 122.

⁴⁵ Ibid 122.

Identifying and understanding error were linked to the ‘self-correcting nature of science’; particularly the need to be ‘as cautious as possible before asserting a new “truth”’ and advocating its adoption among the forensic sciences.⁴⁶

The NRC Committee was also concerned about the variation and inconsistency of ‘terminology used in reporting and testifying about the results of forensic science investigations’, insisting that it ‘must be standardized’.⁴⁷

[M]any terms are used by forensic examiners in reports and in court testimony to describe findings, conclusions, and the degrees of association between evidentiary material (eg hairs, fingerprints, fibers) and particular people or objects. Such terms include but are not limited to ‘match,’ ‘consistent with,’ ‘identical,’ ‘similar in all respects tested,’ and ‘cannot be excluded as the source of.’ The use of such terms can have a profound effect on how the trier of fact in a criminal or civil matter perceives and evaluates evidence. Yet the forensic science disciplines have not reached agreement or consensus on the precise meaning of any of these terms. ... This imprecision in vocabulary stems in part from the paucity of research in forensic science and the corresponding limitations in interpreting the results of forensic analyses.⁴⁸

Problems with terminology were linked to the lack of research, vague standards and deficient reporting. In response the NRC Committee recommended:

The National Institute of Forensic Science (NIFS), after reviewing established standards such as ISO 17025, and in consultation with its advisory board, should establish standard terminology to be used in reporting on and testifying about the results of forensic science investigations. Similarly, it should establish model laboratory reports for different forensic science disciplines and specify the minimum information that should be included. As part of the accreditation and certification processes, laboratories and forensic scientists should be required to utilize model laboratory reports when summarizing the results of their analyses.⁴⁹

Because limitations with techniques and opinions were rarely disclosed and generally inconspicuous in reports and testimony, the NRC Committee insists that expert reports must be ‘complete and thorough’, containing ‘at minimum, “methods and materials,” “procedures,” “results,” “conclusions,” and, as appropriate, sources and magnitudes of uncertainty in the procedures and conclusions (eg levels of confidence).’⁵⁰ Reports should provide sufficient detail ‘to enable a peer or other courtroom participant to understand and, if needed, question the sampling scheme, process(es) of analysis,

⁴⁶ Ibid 125.

⁴⁷ Ibid 21.

⁴⁸ Ibid 185–6.

⁴⁹ Ibid 22 [Recommendation 2]. ISO 17025 is a generic standard for the testing and calibration of laboratories.

⁵⁰ Ibid 21, 186.

or interpretation.’⁵¹ Reports and ‘courtroom testimony stemming from them, must include clear characterizations of the limitations of the analyses, including associated probabilities where possible.’⁵²

The NRC Report draws conspicuous and unprecedented attention to the need for research into different types of bias and the potential for bias to affect the work of forensic scientists (and others in the criminal justice system).⁵³ Noting that in other areas of scientific and medical research and practice — such as biomedical clinical trials of treatment protocols and drugs — elaborate procedures are employed to minimise the threats posed by conscious and unconscious biases. The NRC Report refers to a number of sources of bias threatening the forensic sciences, including: the discounting of base rates; framing and suggestion; institutional pressures and urgency; seeing ‘patterns that do not actually exist’, and anchoring.⁵⁴ In implicit juxtaposition to practice across the forensic sciences, the NRC Report notes that ‘these sources of bias are well known in science, and a large amount of effort has been devoted to understanding and mitigating them.’⁵⁵ The NRC Report dismisses the contention that such biases can be overcome through character or experience. ‘Such biases are not’, as the NRC Committee explained, ‘the result of character flaws; instead they are common features of decision-making, and they cannot be willed away.’⁵⁶

The NRC Committee observed that the forensic sciences could benefit from the large body of research on human performance (and human factors) in diagnostic medicine and cognitive science.⁵⁷ It simultaneously insisted on the need for forensic science disciplines ‘to develop rigorous protocols for performing subjective interpretations, and ... equally rigorous research and evaluation programs.’⁵⁸ Recommendation 5 embodies the NRC Committee’s response to the threats posed by bias:

⁵¹ Ibid 135.

⁵² Ibid 186, 21–2. In contrast, they noted that:

Failure to acknowledge uncertainty in findings is common: Many examiners claim in testimony that others in their field would come to the exact same conclusions about the evidence they have analyzed. Assertions of a ‘100 percent match’ contradict the findings of proficiency tests that find substantial rates of erroneous results in some disciplines (ie voice identification, bite mark analysis)

at 47.

⁵³ See Gary Edmond et al, ‘Contextual Bias and Cross-Contamination in the Forensic Sciences: The Corrosive Implications for Investigations, Plea Bargains, Trials and Appeals’ (2015) 14 *Law, Probability and Risk* 1.

⁵⁴ NRC Report, above n 12, 122–4.

⁵⁵ Ibid 124.

⁵⁶ Ibid 122. See also Risinger et al, ‘The *Daubert/Kumho* Implications of Observer Effects in Forensic Science’, above n 9.

⁵⁷ NRC Report, above n 12, 8.

⁵⁸ Ibid 188.

The National Institute of Forensic Science (NIFS) should encourage research programs on human observer bias and sources of human error in forensic examinations. Such programs might include studies to determine the effects of contextual bias in forensic practice (eg studies to determine whether and to what extent the results of forensic analyses are influenced by knowledge regarding the background of the suspect and the investigator's theory of the case). In addition, research on sources of human error should be closely linked with research conducted to quantify and characterize the amount of error. Based on the results of these studies, and in consultation with its advisory board, NIFS should develop standard operating procedures (that will lay the foundation for model protocols) to minimize, to the greatest extent reasonably possible, potential bias and sources of human error in forensic practice. *These standard operating procedures should apply to all forensic analyses that may be used in litigation.*⁵⁹

Validation studies, in conjunction with attention to error, uncertainty and threats from bias, would enable multidisciplinary groups to develop standards and protocols derived from relevant research. These are embodied in Recommendation 6:

To facilitate the work of the National Institute of Forensic Science (NIFS), Congress should authorize and appropriate funds to NIFS to work with the National Institute of Standards and Technology (NIST), in conjunction with government laboratories, universities, and private laboratories, and in consultation with Scientific Working Groups, to develop tools for advancing measurement, validation, reliability, information sharing, and proficiency testing in forensic science and to establish protocols for forensic examinations, methods, and practices. Standards should reflect best practices and serve as accreditation tools for laboratories and as guides for the education, training, and certification of professionals.⁶⁰

Using examples from medical research and practice, the NRC Committee drew attention to the need for 'quality control, assurance, and improvement'.⁶¹ The NRC Report explains the need for 'systematic and routine feedback' to facilitate 'continuous improvement'.⁶² These needs are encapsulated in Recommendation 8:

Forensic laboratories should establish routine quality assurance and quality control procedures to ensure the accuracy of forensic analyses and the work of forensic practitioners. Quality control procedures should be designed to identify mistakes, fraud, and bias; confirm the continued validity and reliability of standard operating procedures and protocols; ensure that best practices

⁵⁹ Ibid 24 (emphasis added).

⁶⁰ Ibid 24–5. On the value of standards, see 194, 201–6; on proficiency testing and certification, see 206–10.

⁶¹ Ibid 25.

⁶² Ibid.



are being followed; and correct procedures and protocols that are found to need improvement.⁶³

According to the NRC Committee, the quality assurance regime should apply to both laboratories (accreditation) and individuals (certification):

Laboratory accreditation and individual certification of forensic science professionals should be mandatory, and all forensic science professionals should have access to a certification process. ... No person (public or private) should be allowed to practice in a forensic science discipline or testify as a forensic science professional without certification. Certification requirements should include, at a minimum, written examinations, supervised practice, proficiency testing, continuing education, recertification procedures, adherence to a code of ethics, and effective disciplinary procedures. All laboratories and facilities (public or private) should be accredited, and all forensic science professionals should be certified, when eligible, within a time period established by NIFS.⁶⁴

Having criticised the guild-like systems operating in many areas of forensic science, the NRC Committee explained that ‘training should move beyond apprentice-like transmittal of practices to education based on scientifically valid principles.’⁶⁵ The NRC Committee was convinced that these are ‘best learned through formal education. Apprenticeship has a secondary role; under no circumstances can it supplant the need for the scientific basis of education and of the practice of forensic science.’⁶⁶ The NRC Committee was also concerned about the absence of ‘formal and systematically applied standards or standardization requirements for forensic science education programs, making the quality and relevance of existing programs uncertain.’⁶⁷ In response, Recommendation 10 places emphasis on attracting and funding graduate students in fields critical to the forensic sciences and to design programs that ‘cut across organizational, programmatic and disciplinary boundaries.’⁶⁸ Undertaking ‘research and exposure to research’ are characterised as ‘a critical component of an appropriate forensic science education.’⁶⁹ The need for formal education was explicitly extended to other legal actors, through support for ‘legal education programs for law students, practitioners, and judges.’⁷⁰ For

⁶³ Ibid 26.

⁶⁴ Ibid 25, 195–200. On ethics, see 212–4.

⁶⁵ Ibid 26–7, 217.

⁶⁶ Ibid 238. See also Michael J Saks and David L Faigman, ‘Failed Forensics: How Forensic Science Lost Its Way and How It Might Yet Find It’ (2008) 4 *Annual Review of Law & Social Science* 149.

⁶⁷ NRC Report, above n 12, 237.

⁶⁸ Ibid.

⁶⁹ Ibid 230.

⁷⁰ Ibid 28, 234–7.



lawyers and judges often have insufficient training and background in scientific methodology, and they often fail to fully comprehend the approaches employed by different forensic science disciplines and the degree of reliability of forensic science evidence that is offered in trial.⁷¹

Moving from the general to the particular, the Report reviews a range of specific forensic science disciplines and practices. These include those dealing with biological evidence, controlled substances, friction ridge analysis, shoeprints and tyre tracks, tool mark and firearm identification, hair and fibre evidence, document examination, paint and coatings evidence, explosives evidence and fire debris, odontology, blood pattern analysis, digital and multimedia analysis. As we have seen, the Committee was generally critical of pattern recognition and comparison techniques, finding they frequently lack appropriate research, do not employ validated techniques or meaningful standards, do not provide any indication of limitations and errors, expose the analysts to information that has the potential to bias them, and that results tend to be expressed in forms that are often experientially-based or intuitive rather than from the statistical analysis of relevant data.⁷²

The NRC Committee's reaction to friction ridge analysis (ie fingerprints, palm prints and sole prints) conspicuously influenced two of the reports considered below — the NIST and SFI Reports. In order to provide the reader with a flavour of the tone of the more narrowly focused sections, the following extract outlines the NRC Committee's primary concerns with friction ridge comparisons, particularly latent fingerprints.⁷³ As the 'Summary Assessment' explains, current practices tend to elide limitations and overstate the value of the evidence:

Historically, friction ridge analysis has served as a valuable tool, both to identify the guilty and to exclude the innocent. Because of the amount of detail available in friction ridges, it seems plausible that a careful comparison of two impressions can accurately discern whether or not they had a common source. *Although there is limited information about the accuracy and reliability of friction ridge analyses, claims that these analyses have zero error rates are not scientifically plausible.*

ACE-V provides a broadly stated framework for conducting friction ridge analyses.⁷⁴ 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 repeatability and transparency; and does not guarantee that two analysts following it will obtain the same results. For these reasons, merely following the

⁷¹ Ibid 238.

⁷² Ibid ch 5.

⁷³ Although 'latent' means invisible, it is ordinarily used to refer to chance or accidental impressions (often left at a crime scene or on an object). 'Patent' prints are those that are obvious to the human eye, and it is often used to refer to reference prints.

⁷⁴ ACE-V is the dominant friction ridge 'method'. The acronym stands for Analysis, Comparison, Evaluation, and Verification. See the discussion in Part II(B).

steps of ACE-V does not imply that one is proceeding in a scientific manner or producing reliable results. A recent paper by Haber and Haber presents a thorough analysis of the ACE-V method and its scientific validity. Their conclusion is unambiguous: ‘We have reviewed available scientific evidence of the validity of the ACE-V method and found none.’

...

Error rate is a much more difficult challenge. 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 (ie 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 (eg errors in executing the process steps, as well as errors in human judgment).

Some scientific evidence supports the presumption that friction ridge patterns are unique to each person and persist unchanged throughout a lifetime. Uniqueness and persistence are necessary conditions for friction ridge identification 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.⁷⁵

The NRC Committee approvingly quoted a prominent legal scholar:

At present, fingerprint examiners typically testify in the language of absolute certainty. ... Given the general lack of validity testing for fingerprinting; the relative dearth of difficult proficiency tests; the lack of a statistically valid model

⁷⁵ NRC Report, above n 12, 142–5 (emphasis added) (citations omitted), citing Lyn Haber and Ralph N Haber, ‘Scientific Validation of Fingerprint Evidence under *Daubert*’ (2008) 7 *Law, Probability and Risk* 87. On the issue of uniqueness, see NRC Report, above n 12, 188–9. See also Michael Saks and Jonathan J Koehler, ‘The Individualization Fallacy in Forensic Science’ (2008) 61 *Vanderbilt Law Review* 199; Simon A Cole, ‘Forensics without Uniqueness, Conclusions without Individualization: The New Epistemology of Forensic Identification’ (2009) 8 *Law, Probability and Risk* 233; Jonathan J Koehler and Michael J Saks, ‘Individualization Claims in Forensic Science: Still Unwarranted’ (2010) 75 *Brooklyn Law Review* 1187.

of fingerprinting; and the lack of validated standards for declaring a match, such claims of absolute, certain confidence in identification are unjustified ... Therefore ... fingerprint identification experts should exhibit a greater degree of epistemological humility. Claims of 'absolute' and 'positive' identification should be replaced by more modest claims about the meaning and significance of a 'match.'⁷⁶

During the course of the inquiry the Committee recognised that many of the problems confronting the forensic sciences also applied to the 'medicolegal death investigation system'.⁷⁷ The NRC Committee was, once again, confronted and concerned by tremendous variation in qualifications, standards, and practices as well as the limited funding available to the medical examiner community and its counterparts.⁷⁸ There were, for example, 'no mandated national qualifications or certifications for death investigators.'⁷⁹ In addition to the need for more resources, through Recommendation 11, the NRC Committee advocated: converting the remaining coroners to medical examiner systems (dominated by qualified physicians, preferably pathologists); making 'death investigation ... clearly independent of law enforcement';⁸⁰ establishing a scientific working group ('SWG') for forensic pathology and medicolegal death investigation to 'develop and promote standards for "best practices, administration, staffing, education, training, and continuing education"'; the accreditation of medical examiner offices 'pursuant to NIFS-endorsed standards'; and all autopsies performed by 'board certified forensic pathologist[s]'.⁸¹ The Report identified a conspicuous need for basic and translational pathology research.⁸² Anticipating resistance, the Committee indicated that 'federal incentives' will be required to implement changes in each state.⁸³

Overall, the NRC Report identifies and explains the pressing need for reform. The proposed national institute was a response to the poor performance and 'modest leadership' provided by the National Institute of Justice and the FBI Laboratory.⁸⁴

⁷⁶ NRC Report, above n 12, 142, 184, quoting Jennifer Mnookin, 'The Validity of Latent Fingerprint Identification: Confessions of a Fingerprinting Moderate' (2008) 7 *Law, Probability and Risk* 127. More generally, see Cole, *Suspect Identities*, above n 14.

⁷⁷ NRC Report, above n 12, 265. The Report also recognises the important roles played by the forensic sciences in 'natural and human-made mass disasters' and in 'the gathering of effective and timely intelligence and investigative information on terrorists and terrorist groups.' See 33, 265, 279–85: Recommendation 13.

⁷⁸ Ibid 29–30.

⁷⁹ Ibid 264.

⁸⁰ Ibid 252.

⁸¹ Ibid 29–30, 251–2. These structural recommendations are perhaps the least applicable of the various recommendations to Australia, and exemplify some important differences in medico-legal death investigation between the jurisdictions.

⁸² Ibid 265.

⁸³ Ibid.

⁸⁴ Ibid 78–9.



Neither ‘entity has recognized, let alone articulated, a need for change or a vision for achieving it.’⁸⁵ In addition, the NRC Report suggests that laboratories and their budgets should be ‘independent of or autonomous within law enforcement agencies’.⁸⁶

The forensic science community needs strong governance to adopt and promote an aggressive, long-term agenda to help strengthen forensic science. Governance must be strong enough — and independent enough — to identify the limitations of forensic science methodologies and must be well connected with the Nation’s scientific research base in order to affect meaningful advances in forensic science practices. The governance structure must be able to create appropriate incentives for jurisdictions to adopt and adhere to best practices and promulgate the necessary sanctions to discourage bad practices. It must have influence with educators in order to effect improvements to forensic science education. It must be able to identify standards and enforce them.⁸⁷

The NRC Report was not only concerned about the forensic sciences and the institutions responsible for their administration. The NRC Committee recommended a new national institute because lawyers and judges were considered incapable of providing the necessary supervision. In contrast to the usual legal valorisation of adversarial mechanisms and trial safeguards, the NRC Committee drew attention to the poor performance of courts and expressed doubts about the ability of legal institutions to substantially reform practice:

The report finds that the existing legal regime — including the rules governing the admissibility of forensic evidence, the applicable standards governing appellate review of trial court decisions, the limitations of the adversary process, and judges and lawyers who often lack the scientific expertise necessary to comprehend and evaluate forensic evidence — is inadequate to the task of curing the documented ills of the forensic science disciplines. This matters a great deal, because ‘forensic science is but the handmaiden of the legal system.’ ... there are serious issues regarding the capacity and quality of the current forensic science system; yet, the courts continue to rely on forensic evidence without fully understanding and addressing the limitations of different forensic science disciplines.⁸⁸

Interestingly, the proposed NIFS has yet to be established and the NRC Report has not exerted a great deal of direct influence in US courts.⁸⁹ It has, however, been influential on forensic science and medicine communities, their conferences, their publications and their attempts to begin to modify practices, at least behind the scenes. One of the

⁸⁵ Ibid 16, 184.

⁸⁶ Ibid 184.

⁸⁷ Ibid 79.

⁸⁸ Ibid 85, 12, 53, 96, 109, 110.

⁸⁹ While there have been quite a few references, including US Supreme Court decisions such as *Melendez-Diaz v Massachusetts*, 129 S Ct 2527 (2009), as yet there is little evidence of more stringency in admissibility decision-making. In February 2013, the US Government announced that it was establishing a Commission, under the auspices



challenges with such a critical report is how investigators, analysts and courts should respond as practices, techniques and expressions need to be studied and repaired, and in some cases jettisoned, ‘on the run’.

*B The Report of The Expert Working Group on Human Factors in
Latent Print Analysis (2012)*

Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach (‘NIST Report’)⁹⁰ seems to have been a response to both the NRC inquiry (and Mayfield), background issues pertaining to fingerprints following experimental research led by Itiel Dror and Brad Ulery, and continuing critical commentary by scholars such as Simon Cole, Jennifer Mnookin and Ralph and Lyn Haber.⁹¹ As the title suggests, this report is focused on the role of ‘human factors’ in latent fingerprint comparison.⁹² It was prepared by a large multidisciplinary committee — the Expert Working Group on Human Factors (the Expert Working Group or (‘Working Group’)) — where latent print specialists were well represented. The two and a half year review and report were jointly sponsored by two US federal entities, the National Institute of Standards and Technology (‘NIST’) and the National Institute of Justice (‘NIJ’).

‘Human factors’ concern ‘the interaction between humans and products, decisions, procedures, workspaces, and the overall environment encountered as work and in daily living’.⁹³ They are present ‘in any experience- and judgment-based analytical process such as latent print examination.’⁹⁴ Attention to human factors research is intended ‘to enhance quality and productivity in friction ridge examinations and to reduce the likelihood and consequences of human error at various stages in the

of the NIST to begin the arduous work of developing standards and reforming the forensic sciences. Some committees have now been established and held preliminary meetings.

⁹⁰ NIST Report, above n 12.

⁹¹ Cole, *Suspect Identities*, above n 14; Itiel Dror, David Charlton and Alisa E Peron, ‘Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications’ (2006) 156 *Forensic Science International* 74; Jennifer Mnookin, ‘The Courts, the NAS, and the Future of Forensic Science’ (2010) 75 *Brooklyn Law Review* 1209; Lyn Haber and Ralph N Haber, ‘Scientific Validation of Fingerprint Evidence under *Daubert*’, above n 75; Brad Ulery et al, ‘Accuracy and Reliability of Forensic Latent Fingerprint Decisions’ (2011) 108 *Proceedings of the National Academy of Sciences of the United States of America* 7733. See also Jason Tangen, Matthew Thompson and Duncan McCarthy, ‘Identifying Fingerprint Expertise’ (2011) 22 *Psychological Science* 995.

⁹² NIST Report, above n 12, vii, viii. The Working Group was composed of experts from ‘forensic disciplines, statisticians, psychologists, engineers, other scientific experts, legal scholars, and representatives of professional organizations.’ The group reached substantial agreement on most, though not all, issues.

⁹³ NIST Report, above n 12, vi.

⁹⁴ *Ibid.*



interpretation of evidence.⁹⁵ That is, to establish strategies and a culture capable of recognising and responding to human factors and the risks they pose to latent print evidence.⁹⁶

The NIST Report begins with a critical review of the dominant approach to latent print comparison known by the four phases of Analysis, Comparison, Evaluation, and Verification or ACE-V:⁹⁷

Although ACE-V is a systematic process, meaning that the examination proceeds in an orderly and logical fashion, this does not, by itself, demonstrate that the results are accurate and reproducible. In 2009, a committee of the National Research Council (NRC) stated that ACE-V is ‘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. ... Merely following the steps of ACE-V does not imply that one is proceeding in a scientific manner or producing reliable results.’ Additional study is required to ascertain precisely how well examiners using the process perform under either controlled conditions or in casework ...

Although many in the latent print community describe the ACE-V process as a scientific method, the issue is not the label that can or should be attached to the process with respect to human factors. ACE-V is a systematic, skill-based, and widely used process for determining whether two impressions have a common origin. ACE-V designates a logical sequence for a complex process of judgment, but ACE-V itself does not provide substantive guidance about standards to be applied within this sequence. Therefore, even though two examiners might both assert (correctly) that they are using ACE-V, they may be employing different cognitive processes. Those differences create opportunities for human factors to come into play.⁹⁸

⁹⁵ Ibid vi–vii.

⁹⁶ Ibid 172. Human factors are prominent in other areas of human endeavour, such as in medical treatment, the operation of nuclear facilities and air traffic control. Attention to them is intended to improve decision-making and system performance especially where mistakes are potentially catastrophic. See Mark Sanders and Ernest McCormick, *Human Factors in Engineering and Design* (McGraw-Hill Companies, 7th ed, 1993).

⁹⁷ NIST Report, above n 12, 1. On ‘ACE-V’, see above n 74.

⁹⁸ NIST Report, above n 12, 9, 123–4 (citations omitted).

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

at 39.



Of particular concern, given the dearth of detailed standards, was the manner in which fingerprint examiners report and testify about their conclusions, especially the identification of a putative source to the exclusion of all others. According to the Working Group ‘this claim is needlessly strong’:

a fingerprint identification was traditionally considered an ‘individualization,’ meaning that the latent print was considered identified to one finger of a specific individual as opposed to every other potential source in the universe. However, the recent attention focused on this issue reveals that this definition needlessly claims too much, is not adequately established by fundamental research, and is impossible to validate solely on the basis of experience. Nor does fingerprint evidence have objective standards or a well-validated statistical model that can provide an objective measure of the strength of the fingerprint evidence in a given instance. Therefore, examiners should not claim to be able to exclude every other finger in the world as a potential source. Rather, an identification decision suggests a substantial enough similarity that the examiner believes that the two impressions originated from a common source. But whether any other finger in the world might also be able to leave an impression with a comparable amount of similarity is not fully known, and the examiner’s testimony should not suggest otherwise. Regardless of the specific words used to describe an identification, examiners should refrain from claiming that an identification means that they have excluded all other individuals in the world.⁹⁹

In the absence of statistical data on the prevalence of print features, the move between declaring a match and equating that with positive identification becomes, in effect, ‘a leap of faith’:

Because not enough is known about rarity, and even that which is known is not necessarily part of an examiner’s formal training, this aspect of interpretation is often an implicit judgment based on the examiner’s experience.¹⁰⁰

Concerns about ‘over-expressing the evidence’ led to Recommendation 3.7 — the first in the ‘Summary of Recommendations’:¹⁰¹

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.¹⁰²

⁹⁹ Ibid 72, 197. Interestingly, even some of the technical advisory groups have begun to respond to this criticism: at 72. See also Simon A Cole, ‘Individualization is Dead, Long Live Individualization! Reforms of Reporting Practices for Fingerprint Analysis in the United States’ (2014) 13 *Law, Probability and Risk* 117.

¹⁰⁰ NIST Report, above n 12, 48, 63.

¹⁰¹ Ibid 72.

¹⁰² Ibid 72. See also NIST Report at 77: ‘examiners should qualify their conclusions instead of stating an exclusion of identification in absolute terms.’

This bears re-reading, as the Working Group recommended against the use of latent fingerprint evidence to positively identify or individualise.¹⁰³ Notwithstanding this disruptive recommendation, the Working Group was agnostic about the form in which conclusions should be expressed and the circumstances in which ‘a qualified rather than an absolute conclusion is warranted’ (Recommendation 3.8).¹⁰⁴ Aware of ongoing statistical investigation, the Working Group suggested that probabilities and likelihood ratios will eventually provide more appropriate means of expressing results — that is, more empirically robust and qualified conclusions. Moreover, recourse to data sets was perceived as likely to provide examiners with more options than the traditional, and rather clunky, categories of ‘identification, exclusion, or inconclusive’.¹⁰⁵

Notwithstanding the ‘increasing number of research projects’ the Working Group insisted that ‘additional research should be undertaken’.¹⁰⁶ There are, as the Working Group noted, ‘many questions ... that presently lack definitive answers.’¹⁰⁷ They identified ‘a critical need for a focused program of research into the interpretive process that is at the heart of ACE-V’.¹⁰⁸ This would involve ‘standardized methods for feature selection’ and a better understanding of ‘the link between variations in feature selection and the ultimate decision’.¹⁰⁹ Significantly:

The Working Group found no research that explicitly addresses utility or sufficiency in the context of latent print analysis. This is unsurprising, for a critical piece for any such research — the definition and validation of a metric for assessing utility — is missing. ... Opening the box to study the process of judgment in every phase of ACE-V would provide the empirical foundation from which to develop best practices for each part of the process. As a result, the Working Group recommends:

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

¹⁰³ In addition, the existence of large national databases led the Working Group to recommend that examiners should always consider ‘the possibility and dangers of incidental similarity’ between prints: Ibid 199 [Recommendation 3.6].

¹⁰⁴ Ibid 198. Recommendation 3.8 recognises that in some circumstances a positive identification could be made, but such a finding is not appropriate in the vast majority of circumstances.

¹⁰⁵ Ibid 198, 86, 96.

¹⁰⁶ Ibid 203.

¹⁰⁷ Ibid.

¹⁰⁸ Ibid. Findings have shown wide variation in feature selection.

¹⁰⁹ Ibid 204.

- c. Identify key factors related to variations in performance of latent print examiners during the interpretation process.¹¹⁰

The current use of ‘subjective, experienced-based judgments of the probability’ to weigh the significance of individual features, along with research suggesting inconsistency between examiners, led the Working Group to advocate the assembly of data sets (Recommendation 4.2), statistical research and algorithms for some parts of the process (Recommendation 3.9).¹¹¹ The move to statistically-based forms of interpretation would also require pedagogical reform:

Because statistical information plays a fundamental role in weighting latent print feature evidence, training should include the best available empirical information and should educate examiners about probabilistic reasoning in using that information. (Recommendation 3.5)¹¹²

The latent print examiner community should expand the training of examiners in elementary probability theory to enable examiners to properly utilize the output of probabilistic models. (Recommendation 4.3)¹¹³

Having recommended against source attributions ‘to the exclusion of all others’, the Working Group endorsed the NRC Report’s concern about the lack of ‘agreement ... on the precise meaning’ of terms such as “‘match,” “consistent with,” “identical,” “similar in all respects tested,” and “cannot be excluded as the source of””, describing it as a “‘critical” problem’.¹¹⁴ Rather than insist ‘on the use of any single set of terms’, the Working Group sought to consider a range of opinions including probabilistic approaches (eg random match probabilities and likelihood ratios), and different kinds of individualisation — ie global and specific.¹¹⁵ Statistical analysis of data could support probabilistic models ‘of various kinds’ and provide ‘a firmer foundation for and opportunities to improve the judgments of examiners’ and ‘supply objective probabilities that would be useful to judges and juries’.¹¹⁶ Without directly advocating a particular approach (because of the inability to achieve consensus within the Working Group), the NIST Report appears most sympathetic to the likelihood ratio approach, where the evidence is presented in terms favouring ‘one

¹¹⁰ Ibid [Recommendation 3.9].

¹¹¹ Ibid. Automated fingerprint identification system (AFIS) technologies could be enhanced, particularly through greater interoperability and compatibility with tools and software programs (Recommendation 4.1). These are systems used by fingerprint examiners to assist with the location of potentially matching prints. See also NRC Report, above n 12, 276 and Recommendation 12.

¹¹² NIST Report, above n 12, 202.

¹¹³ Ibid.

¹¹⁴ Ibid 128; NRC Report, above n 12, 185.

¹¹⁵ NIST Report, above n 12, 128–9.

¹¹⁶ Ibid 205.

proposition (identity of sources) versus another (non-identity of sources).¹¹⁷ For the Working Group, this approach presents ‘a more modest claim than an absolute source attribution ... Furthermore, it would be far less vulnerable to the charge of over-claiming.’¹¹⁸ This assumes, of course, that underlying models are ‘adequately validated and accepted in the scientific community, and the courts must be persuaded that expressing the strength of evidence in the form of a likelihood ratio is not too confusing for juries.’¹¹⁹

The need for qualification in the way conclusions are expressed in reports and courtroom testimony followed from limitations with current practice and their subjective dimensions:

Current methods for making these interpretations are based on professional knowledge and experience rather than on formal decision thresholds or statistical models. The process known as ACE-V organizes the interpretations and decisions of an examiner into a useful and logical sequence, but descriptions of this process do not detail the substantive content of the various steps. Examining latent prints and exemplars necessitates judgment and expertise, which inevitably makes the interpretive process partly subjective. With this subjectivity of interpretation comes the possibility of reduced performance due to a wide range of human factors issues.¹²⁰

The NIST Report devoted considerable attention to issues related to determining whether prints were ‘sufficient’ for analysis and comparison and the kinds of tolerances that the analyst should accept. The Working Group was concerned that these issues had not been subjected to clear and consistent guidelines. Many ‘[d]ecision making processes’, it explained, ‘are neither clearly defined a priori nor made using an objective, validated metric.’¹²¹ Some examiners may be willing to analyse and declare a match while others might deem the same latent print insufficient for analysis. As the Working Group noted: ‘[t]here are no formal thresholds for any sufficiency determinations, and feature selection and weighting are matters of personal judgment’, and ‘none of [the] variabilities — of features across a population of fingers or of repeated impressions left by the same finger — has been

¹¹⁷ Ibid 134, quoting Christophe Champod, ‘Interpretation of Friction Ridge Examinations (Fingerprints)’ in Andre Moenssens and Allan Jamieson (eds), *Wiley Encyclopedia of Forensic Science* (Wiley and Sons, 2009) Vol 3, 1508–11.

¹¹⁸ NIST Report, above n 12, 134, 138.

¹¹⁹ Ibid 135. This approach sits awkwardly with jury capacity, see Kristy Martire et al, ‘The Psychology of Interpreting Expert Evaluative Opinions’ (2013) 45 *Australian Journal of Forensic Sciences* 305; Andrew Ligertwood and Gary Edmond, ‘Expressing Evaluative Forensic Science Opinions in a Court of Law’ (2012) 11 *Law, Probability and Risk* 289.

¹²⁰ NIST Report, above n 12, 198.

¹²¹ Ibid 45.

characterized, quantified, or compared.¹²² They also drew attention to the problem of distortion in prints along with the lack of information about ‘an examiner’s ability to identify types of distortion’.¹²³ The Working Group recommends that:

‘Each agency or forensic service provider should define ‘suitable’ or ‘sufficient’ in its standard operations procedures. Guidelines should be as explicit as possible about what is expected for sufficiency determinations at different stages of the latent print examination process (Recommendation 3.4).’¹²⁴

Subjective assessments, especially in the absence of detailed standards, introduce a range of dangers. Given its orientation, the NIST Report is conspicuously attentive to threats from cognitive bias. For

Observers’ expectations have been shown to influence judgment in a broad range of tasks. Especially when confronted with ambiguous stimuli, people tend to see what they hope or expect to see. ... some information about the origin of a latent print can facilitate accurate results, but other contextual information can produce confirmation bias. Extraneous information can influence people acting in good faith and attempting to be fair interpreters of the evidence.¹²⁵

The NIST Report is not critical, or disparaging, of ‘subjective elements in human interpretation’.¹²⁶ These dimensions, as the Working Group recognises, constitute an essential component of judgment and expertise.¹²⁷

To recognize that latent print examiners are potentially subject to bias is not to single them out but rather to suggest that they are not exempt from those cognitive biases that all interpreters of data and information face.¹²⁸

¹²² Ibid 16, quoting NRC Report, above n 12, 144. The Working Group continued at 69: ‘This is not to suggest that experts are poor at any of these tasks. Rather, it is simply to note the absence of objective criteria.’ There were also questions, as in the SFI Report, about how to handle the re-assessment of parts of a latent print after looking at the exemplar. This is described as ‘recursive practice.’ See NIST Report, above n 12, 42–3, especially 51.

¹²³ NIST Report, above n 12, 66, 50. There is ongoing controversy around what constitutes a complex latent print. Studies by Dror et al found inconsistency between analysts, and even with the same analyst analysing the same print at different times. See Itiel Dror et al, ‘Cognitive Issues in Fingerprint Analysis: Inter- and Intra-Expert Consistency and the Effect of a “Target” Comparison’ (2011) 208 *Forensic Science International* 10.

¹²⁴ NIST Report, above n 12, 199.

¹²⁵ Ibid 10 (citations omitted).

¹²⁶ Ibid 39.

¹²⁷ On expertise generally, see Itiel Dror, ‘The Paradox of Human Expertise: Why Experts Get it Wrong’ in Narinder Kapur (ed) *The Paradoxical Brain* (Cambridge University Press, 2011) 177.

¹²⁸ NIST Report, above n 12, 40.



Introducing human factors into the mix acknowledges the ‘possibility of reduced performance because of human factors issues’¹²⁹ and is intended to remove the culture of blame by ‘systematically design[ing] safety into the process’.¹³⁰ Rather than simply blame individuals for errors and mistakes, a human factors approach is more holistic — extending attention to ‘the design, working conditions, or management culture of the total system’.¹³¹ The Working Group was desirous that the forensic sciences should emulate other scientific and medical practices involving the identification, reduction and quantification of error:¹³²

A basic tenet of experimental science is that ‘errors and uncertainties exist that must be reduced by improved experimental techniques and repeated measurements, and those errors remaining must always be estimated to establish the validity of our results.’ What applies to physics and chemistry applies to all of forensic science: ‘A key task ... for the analyst applying a scientific method is to conduct a particular analysis to identify as many sources of error as possible, to control or eliminate as many as possible, and to estimate the magnitude of remaining errors so that the conclusions drawn from the study are valid.’ In other words, errors should, to the extent possible, be identified and quantified.¹³³

There is, in consequence, a need to define error precisely, collect data about performance and error, use well-defined proficiency tests to provide feedback and exclude unnecessary contextual information.¹³⁴ That is, to generate ‘system-wide’ reform (Recommendation 7.1).¹³⁵

The Working Group placed conspicuous emphasis upon the importance of learning from mistakes: ‘Numerous studies have found that without quick and accurate feedback on correct and incorrect judgments, experience does not enhance expertise and that experts routinely overestimate their skills.’¹³⁶ They also noted that forensic analysts ‘do not routinely receive ... prompt and frequent feedback.’¹³⁷

¹²⁹ Ibid 39.

¹³⁰ Ibid 21, 39.

¹³¹ Ibid 140.

¹³² Drawing on the National Academy of Sciences, Institute of Medicine, Committee on Quality of Health Care in America report: *To Err Is Human: Building A Safer Health System* (National Academies Press, 1999).

¹³³ NIST Report, above n 12, 21, quoting NRC Report, above n 12, 111.

¹³⁴ NIST Report, above n 12, 24. The Working Group seem relieved that the latent print community had, in recent years, begun to ‘acknowledge that errors do occur and furthermore that claims of zero error rate in the discipline are not *scientifically* plausible’ at 32 (emphasis in original).

¹³⁵ Ibid 203.

¹³⁶ Ibid 23 (citations omitted).

¹³⁷ Ibid (citations omitted).



Conceptual limitations and threats from subjectivity, led the Working Group to attempt to increase the ‘transparency of the process and to insulate the examiner from extraneous influences.’¹³⁸ Research drawn from psychological studies led to proposals for blinding examiners to information and offering multiple exemplars for comparison:

Procedures should be implemented to protect examiners from exposure to extraneous (domain-irrelevant) information in a case. (Recommendation 3.3).¹³⁹

For cognitive biases can be subtle and are ‘usually unknown to the observer.’¹⁴⁰ Vulnerability and transparency could be better managed through enhanced documentation:

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 3.1).

Modifications to the results of any stage of latent print analysis (eg feature selection, utility assessment, discrepancy interpretation) after seeing a known exemplar should be viewed with caution. Such modifications should be specifically documented as having occurred after comparison has begun. (Recommendation 3.2).¹⁴¹

The Report was also concerned with the communication of results to ‘lay consumers’:

Developing and implementing procedures and practices that encourage experts to communicate their findings accurately and fairly to lawyers, judges, and juries and to detect and correct errors in this process is a crucial component of a system that reduces the opportunities for errors in the production and presentation of courtroom fingerprint evidence.¹⁴²

¹³⁸ Ibid 198.

¹³⁹ Ibid.

¹⁴⁰ Ibid 12. On sequential unmasking, see NIST Report at 52 and see also Dan Krane et al, ‘Sequential Unmasking: A Means of Minimizing Observer Effects in Forensic DNA Interpretation’ (2008) 53 *Journal of Forensic Science* 1006.

¹⁴¹ NIST Report, above n 12, 198. The FBI has recently adopted a process described as linear ACE-V. This involves using the steps associated with ACE-V, though in a way where exposure to information and documentation of impressions guide progression through the process. This is a generally positive development but does not address validation and reliability directly.

¹⁴² Ibid 113, 199, see also [Recommendation 5.1] at 94.

According to Recommendation 6.1:

The trial preparation process should address the presentation of technical information in lay terms, the organization of the direct examination, possible cross-examination, and the possible use of visual aids.¹⁴³

To increase the likelihood that the evidence ‘will be used properly, fully, and fairly’ by ‘lay consumers’, Recommendation 5.2 lists ‘the minimum information that should appear in the summary report’:¹⁴⁴

A report should:

- a) Identify the latent print examiner(s);
- b) Describe the items submitted to the examiner(s);
- c) List the procedures used by the examiner to develop, visualize, or enhance the friction ridge impressions;
- d) List all comparisons conducted;
- e) State all conclusions with the method used to reach them;
- f) Note any important limitations to the conclusions;
- g) Indicate whether a verification was made and whether there was any conflict of opinion among examiners prior to the reported conclusions;
- h) Note (or refer to external documentation of) any information about the case that the examiner(s) received;¹⁴⁵
- i) Note the existence of additional documentation; and
- j) Define important technical terms, either explicitly or by reference to an authoritative, readily available source.¹⁴⁶

The report should provide sufficient information to enable a reviewing expert to verify the examiner’s assessment.¹⁴⁷ The ‘level of detail is sufficient to inform investigators, prosecutors, defense counsel, other experts, and judges or juries as to how the analysis was conducted and what conclusions were reached.’¹⁴⁸

Enhanced transparency and tempering the expression of conclusions are characterised as consistent with the obligations of latent print examiners ‘to their profession and to the court ... complying [with] legal demands such as disclosure, and being truthful.’¹⁴⁹ To assist in these ends:

¹⁴³ Ibid 115.

¹⁴⁴ Ibid 90, 95.

¹⁴⁵ Ibid 100.

¹⁴⁶ Ibid.

¹⁴⁷ Ibid 90.

¹⁴⁸ Ibid 101.

¹⁴⁹ Ibid 117, quoting Bruce Budowle et al, ‘A Perspective on Errors, Bias, and Interpretation in the Forensic Sciences and Direction for Continuing Advancement’ (2009) 54 *Journal of Forensic Sciences* 795, 803. On disclosure, see also NIST Report, above n 12, 91.

Forensic service providers should adopt codes of ethics that require testifying in a nonpartisan manner; answering questions from both the prosecution and the defense directly, accurately, and fully; and providing appropriate scientific information before, during, and after trial.¹⁵⁰

The Working Group explains ‘the importance of “staying within the bounds or limits of what the science can provide” and providing appropriate scientific information’.¹⁵¹ This, again, is indexed to Recommendation 3.7, above.¹⁵² One of the implications, following from ‘absolute certainty [being] unattainable in science’ is that ‘the witness must be prepared to acknowledge and discuss the possibility that an opinion is not correct.’¹⁵³

Recognising that it will be difficult to estimate errors, both false positives and false negatives, the NIST Report discusses the virtues of using: non-blind proficiency tests; realistic, blind proficiency tests; verifications; random audits of case reports; and controlled experiments. All have limitations, but all have the potential to inform our understanding and improve the performance of latent print examiners and the use of their interpretations.¹⁵⁴

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 6.3).¹⁵⁵

The point is to refine practices and develop means of expressing fingerprint evidence in ways that reflect limitations and the real possibility of error.¹⁵⁶

¹⁵⁰ NIST Report, above n 12, 200 [Recommendation 6.2].

¹⁵¹ Ibid.

¹⁵² Ibid.

¹⁵³ Ibid 201. Cf Simon A Cole, ‘Who Speaks for Science? A Response to the National Academy of Sciences Report on Forensic Science’ (2010) 9 *Law, Probability and Risk* 25.

¹⁵⁴ NIST Report, above n 12, 33.

¹⁵⁵ Ibid 201, 74. It is unlikely that there will be a single error rate. See Jonathan Koehler, ‘Fingerprint Error Rates and Proficiency Tests: What They Are and Why They Matter’ (2008) 59 *Hastings Law Journal* 1077; Jonathan Koehler, ‘Proficiency Tests to Estimate Error Rates in the Forensic Sciences’ (2013) 12 *Law, Probability and Risk* 89 and Gary Edmond, Jason M Tangen and Matthew B Thompson, ‘A Guide to Interpreting Forensic Testimony: Scientific Approaches to Fingerprint Evidence’ (2014) 13 *Law, Probability and Risk* 1.

¹⁵⁶ See Tangen, Thompson and McCarthy, ‘Identifying Fingerprint Expertise’, above n 91; Brad Ulery et al, ‘Accuracy and Reliability of Forensic Latent Fingerprint Decisions’, above n 91.

The NIST Report also advocated the need for quality assurance mechanisms, including: ‘requirements and guidelines for reporting, documentation, and testimony’ and sampling for compliance (Recommendation 6.4).¹⁵⁷ Enhanced supervision and management were characterised as ‘essential to risk reduction and quality assurance and control. Effective management requires good information about the incidence and sources of errors.’¹⁵⁸ Collecting this information would require developing a culture where ‘openness about errors’ is ‘not necessarily a path to punitive sanctions but rather is part of an effective system to detect deviations from desired practices and incorrect judgments’ (Recommendation 9.1).¹⁵⁹ This means that errors should be identified and tracked (Recommendation 9.2) and procedures should be implemented for prevention and resolution of disputes (Recommendation 9.4).¹⁶⁰ In addition, there is a need for forensic science providers to be accredited, and for the latent print community to ‘develop and implement a comprehensive testing program that includes competency testing, certification testing and proficiency testing’ (Recommendations 9.3 and 9.4) and for the certification and continuing education of analysts (Recommendation 8.7).¹⁶¹

The NIST Report decries the existing ‘heterogeneity in curricula, instructors, pedagogy, documentation, and mentorships’, and laments the ‘few enforceable standards’.¹⁶² In addition to improved training, the Working Group recommended continuing education and research into the education and aptitudes suited to training a latent examiner (Recommendation 8.1).¹⁶³ The Working Group also recommended mentor programs and accreditation of institutional training programs (Recommendation 8.8).¹⁶⁴ Training should be supported by a multidisciplinary expert group responsible for developing ‘a latent print educational textbook, practical exercises, and assessment tests’ (Recommendation 8.4), a clearinghouse of materials and publications (Recommendation 8.6), and funding to support training programs (Recommendation 8.2).¹⁶⁵

The Working Group’s recommendations are consistent with the thrust of the NRC Report. Nevertheless, when it came to detailed recommendations the Working Group

¹⁵⁷ NIST Report, above n 12, 138–9, 201.

¹⁵⁸ Ibid 201.

¹⁵⁹ Ibid.

¹⁶⁰ Ibid.

¹⁶¹ All this should be supplemented with ‘medical surveillance’, such as annual vision testing (Recommendation 7.2). Performance would also be improved by directing more attention to the physical environment, particularly the hardware and software, routinely used by examiners (Recommendation 7.1).

¹⁶² NIST Report, above n 12, 165.

¹⁶³ Ibid 209.

¹⁶⁴ Ibid 210.

¹⁶⁵ Ibid 203. Notably, authors of the textbook, ‘should not be confined to experts in latent print examination, but should include experts on cognitive issues, statistics, and forensic science’: at 167–8.

was undecided on a range of issues (notably the expression of opinions, above). The lack of relevant research meant that for some members of the Working Group specific recommendations were considered inappropriate at this stage.¹⁶⁶ They were conscious of the difficulty of proposing ‘modifications to an interpretive system that is largely non-formalized.’¹⁶⁷

Given the insufficient research base, the Working Group placed emphasis on the need for transparency through documentation:

Documentation serves to maximize the transparency of the interpretative process and to provide a record that can be useful for many purposes, including reports and testimony, future research and evaluation, and quality assurance.¹⁶⁸

Documentation should strive to capture dimensions of practice such as the enhancement of images of prints, the examiner’s impressions at particular stages of analysis, processing and disagreement between examiners.¹⁶⁹

Similarly, the lack of research on the precise effects of bias led the Working Group to adopt a precautionary approach:

given the decades-long research into the significant effects of cognitive bias in other domains, it seems wise to minimize the potential for such biases in latent print interpretation, even in the absence of definitive research results for latent print analysis.¹⁷⁰

Any exposure to domain irrelevant information should be minimized and documented because:

the possibility of biases influencing the decision making process of examiners cannot be dismissed, a report should reveal the context of the examination by describing or referring the reader to the information about the case that an examiner received.¹⁷¹

Overall, the Working Group endorsed the conclusion of the NRC Report, insisting that: ‘fingerprint experts should exhibit a greater degree of epistemological humility.

¹⁶⁶ Ibid 49, 53, 74, 78.

¹⁶⁷ Ibid 63, 8, 19.

¹⁶⁸ Ibid 41, 49, 53, 66, 90, 94. See also Recommendation 3.1 at 42.

¹⁶⁹ Ibid 81–2, 185. See also SWGIT (Scientific Working Group for Imaging Technology) and the need to validate enhancement technologies.

¹⁷⁰ Ibid 43, 44.

¹⁷¹ Ibid 97 quoting SWGFAST 2009. ‘SWGFAST’ is the Scientific Working Group for Friction Ridge Analysis, Study and Technology.



Claims of “absolute” and “positive” identification should be replaced by more modest claims about the meaning and significance of a “match.”¹⁷²

Notwithstanding its focus on fingerprints, there are few doubts about the general application of the NIST Report, its findings and recommendations to many other forensic science disciplines and techniques:

Although this report explicitly addresses only the procedures for performing a latent fingerprint examination and communicating the results, much of the analysis and many of the recommendations are applicable to other forensic science disciplines. Issues of cognitive bias, standardization of procedures, documentation of examinations, working conditions, error detection and correction, and accuracy in testimony — among many others — cut across the forensic sciences.¹⁷³

C The Scottish Fingerprint Inquiry (2011)

The Scottish Fingerprint Inquiry and Report (‘SFI Report’) represent the culmination of more than a decade of controversy surrounding the (mis)attribution of a latent print in a murder investigation. In 1997, as part of the routine examination of a murder scene in Kilmarnock, latent prints were collected and checked against those of investigating police officers.¹⁷⁴ One of the latent fingerprints recovered from an internal doorframe, adjacent to the body of Marion Ross, was attributed by examiners at the Scottish Criminal Records Office (SCRO) to Detective Shirley McKie. McKie had been involved in the investigation but insisted that she had never entered the property and could not, therefore, have been responsible for the print.¹⁷⁵ The eventual prosecution of David Asbury for the murder of Ross relied upon incriminating fingerprint evidence. McKie’s insistence that she had not entered the building introduced complications with the fingerprint evidence. McKie was called as a witness and denied that she had entered the premises. Her unwillingness to accept that it was her print and that she must, therefore, have improperly entered a crime scene, created difficulties for prosecutors and led to McKie’s subsequent prosecution for perjury. With the aid of foreign fingerprint experts testifying on her behalf, McKie successfully resisted that prosecution. The controversy arising from the ‘identification’ and McKie’s unsuccessful prosecution led to several internal and public reviews.¹⁷⁶ The Scottish Fingerprint Inquiry was the most recent and comprehensive.

¹⁷² NIST Report, above n 12, 130.

¹⁷³ *Ibid* ix.

¹⁷⁴ This was part of a routine process to exclude any marks left inadvertently by investigators thereby enabling the focus to be placed on prints that might be relevant to the investigation of the crime.

¹⁷⁵ There was some initial consideration of transfer but this possibility was eventually rejected.

¹⁷⁶ Technically, according to prevailing bureau practice, this was an elimination rather than an identification. Different levels of confidence and review were applied to identifications as opposed to eliminations.



The Commissioner, Sir Anthony (now Lord) Campbell, concluded that a mistake had been made in the attribution of the print to McKie.¹⁷⁷ Through the course of his review Campbell identified many problems and made numerous recommendations to reform the organisation, practice and presentation of fingerprint evidence in Scotland. The final report of the Scottish Fingerprint Inquiry runs to almost 800 pages. Like the Goudge Inquiry — Part II(D), below — it was a full-time public inquiry. Rather than attempt to reproduce the case-specific issues and the intricacies of Scottish practice, in this context it seems appropriate to focus on the major recommendations pertaining to latent fingerprint evidence. To a considerable degree, these reinforce points advanced in the NRC and NIST Reports and, once again, are directly relevant to other forensic sciences, particularly those involving interpretation and comparison.

The SFI Report concluded that there were many problems with the organisation of the Scottish Fingerprint bureau as well as the practices of examiners.¹⁷⁸ Despite these problems and routine overreaching in expert reports, the inquiry heard testimony about how it was relatively rare for a fingerprint examiner to appear in court and almost unheard of for there to be any kind of challenge to identifications.¹⁷⁹ By an Act of Parliament, fingerprint examiners with sufficient experience were effectively ‘authorised’ and excused from having to appear where their evidence was uncontested.¹⁸⁰ This expedient mechanism merely reinforces unquestioned faith in the reliability (or infallibility) of fingerprint comparison and individualisations.¹⁸¹ Confidence can also be observed in the willingness to prosecute McKie for perjury on the basis of a recovered latent fingerprint even though she was a serving police officer, denied entering the premises, and the police officers accompanying McKie at the scene corroborated her version of events.¹⁸²

Of the 86 recommendations made in the SFI Report the following were listed as the ten ‘key recommendations’:

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. [Recommendation 1]

¹⁷⁷ Interestingly, at some level the controversy continues, even though those alleging a misattribution seem to have prevailed. This is interesting because it demonstrates that in the absence of knowledge of the correct answer (ground truth) it can be difficult to know who is right. This is, in part, why less categorical forms of practice and expression have been recommended.

¹⁷⁸ For a short summary, see Simon A Cole and Andrew Roberts, ‘Certainty, Individualisation, and the Subjective Nature of Expert Fingerprint Evidence’ [2012] *Criminal Law Review* 824.

¹⁷⁹ SFI Report, above n 12, 193–4, 237.

¹⁸⁰ *Criminal Procedure (Scotland) Act 1995* (Scot) ss 280, 281.

¹⁸¹ SFI Report, above n 12, 559.

¹⁸² Compare Frank Vincent, *Report: Inquiry Into the Circumstances That Led to the Conviction of Mr Farah Abdulkadir Jama* (Victorian Government Printer, May 2010).

2. 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 3]
3. Examiners should receive training which emphasises that their findings are based on 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’. [Recommendation 2]
4. Differences of opinion between examiners should not be referred to as ‘disputes’. [Recommendation 4]
5. The SPSA’s Standard Operating Procedures should set out in detail the ACE-V process that is to be followed. [Recommendation 20]¹⁸³
6. Features on which examiners rely should be demonstrable to a lay person with normal eyesight as observable in the mark. [Recommendation 9]
7. Explanations for any differences between a mark and a print require to be cogent if a finding of identification is to be made. [Recommendation 11]
8. A finding of identification should not be made if there is an unexplained difference between a mark and a print. [Recommendation 12]
9. The SPSA should develop a process to ensure that complex marks are treated differently. The examination should be undertaken by three suitably qualified examiners who reach their conclusions independently and make notes at each stage of their examination. The substantive basis for the examiners’ conclusions should be reviewed. The reasons why they have reached their respective conclusions should be explored and recorded, even where they agree that an identification can be made. [Recommendation 42]
10. An emphasis needs to be placed on the importance not only of learning and practising the methodology of fingerprint work, but also of engaging with members of the academic community working in the field. [Recommendation 16]¹⁸⁴

The Commissioner, like the Expert Working Group, recommended that fingerprint examiners should not purport to possess an infallible method and should generally not make positive identifications without some qualification (Recommendations 1 and 3, above).

Like the NRC and NIST Reports, the SFI Report documents inadequacies in the research base. Specific areas warranting attention include:

- (i) the frequency of particular characteristics or combinations of characteristics in fingerprints;
- (ii) the use of data as to the frequency of particular characteristics or combinations of characteristics as a means of assisting examiners in their work;
- (iii) the weight to be given to third level detail, and as to its reliability;

¹⁸³ ‘SPSA’ is the acronym for the Scottish Police Services Authority.

¹⁸⁴ SFI Report, above n 12, 740 (citations omitted).

- (iv) distortion and the effect of movement;
- (v) which marks ought to be assessed as complex;
- (vi) the specific factors that may cause variations among examiners; and
- (vii) contextual bias.¹⁸⁵

Recommendation 83 extends this to include ‘probabilistic analysis’.¹⁸⁶ As part of the need to develop the research base, there was a conspicuous need for fingerprint examiners to engage with mainstream research scientists:

Engaging with the academic community

Recommendation 16 (Para 35.135) [Number 10, above]

...

Recommendation 18 (Para 35.142) The SPSA, in conjunction with members of the academic community as appropriate, should design a practical system for examiners to assess and evaluate (a) tolerances and (b) any reverse reasoning.¹⁸⁷

The Report also draws explicit attention to the serious threat posed by human factors, in the guise of contextual bias:

Fingerprint methodology

Recommendation 6 (Para 35.137) The SPSA should review its procedures to reduce the risk of contextual bias.

Recommendation 7 (Para 35.138) The SPSA should ensure that examiners are trained to be conscious of the risk of contextual bias.

Recommendation 8 (Para 35.139) The SPSA should consider what limited information is required from the police or other sources for fingerprint examiners to carry out their work, only such information should be provided to examiners, and the information provided should be recorded.¹⁸⁸

The SFI Report makes many recommendations to improve practice, including: the need to ‘set out in detail the ACE-V process to be followed’ (Recommendation 20);¹⁸⁹ documenting and attaching less weight to ‘characteristics first found at the comparison stage’ (Recommendation 26);¹⁹⁰ emphasising the need to focus on

¹⁸⁵ Ibid 752 [Recommendation 82]. See also SFI Report at 432, 702, 734.

¹⁸⁶ Ibid 752.

¹⁸⁷ Ibid 742 (emphasis in original).

¹⁸⁸ Ibid 741. See also SFI Report at 743 [Recommendation 24].

¹⁸⁹ Ibid 743.

¹⁹⁰ Ibid. That is, once they have had an opportunity to look at both prints side-by-side.

‘tolerances, the quality of similarities, the nature of differences, any explanations for differences, the extent to which reverse reasoning may have been employed and the sufficiency of matching characteristics’ during ‘the evaluation stage’; and blinding reviewers to the reasoning during the verification stage.¹⁹¹ Attention was also directed to improving the quality of images, particularly digital images and documenting changes to digital images (Recommendations 37–40).¹⁹²

The Report placed conspicuous emphasis on the subjective (ie non-certain and interpretive) nature of fingerprint comparison (Recommendation 1).¹⁹³ In addition to the need for procedures to manage disagreement between examiners (Recommendations 34 and 36), it stressed the need for practitioners to:

conduct their individual ACE comparisons conscious of the fact that they are working in a field where there is no certainty and where there is scope for differences of opinion. When it comes to verification, examiners should be encouraged to be open and to adopt a challenging attitude to the opinions of other examiners, irrespective of seniority. Standard Operating Procedures should emphasise that the fact that one examiner reaches the opposite conclusion from another, or entertains any doubt, does not necessarily cast any aspersion on the competence of either examiner.¹⁹⁴

The SFI Report placed emphasis on the need for improved ‘record-keeping and note-taking’ (Recommendations 44–52).¹⁹⁵ Recommendations 54–6 address the need to provide information to prosecutors, and Recommendations 60–3 address disclosure and the provision of access to the defence.¹⁹⁶ Sensitive to the circumstances of its origin, the Report was concerned that those ‘identified’ (and their legal representatives) should have access to all images of prints, not only those relied upon by the state’s examiners.

The SFI Report also insisted on the need for training, improved performance management (Recommendations 70–4) and the certification and authorisation of examiners (Recommendations 76–80).¹⁹⁷ Notwithstanding the need for examiners to be authorised to prepare reports and testify, the Report was open to the possibility of allowing those who were not authorised under Scottish legislation to act as expert

¹⁹¹ Ibid 654–5 [36.113].

¹⁹² Ibid 745.

¹⁹³ Ibid 741.

¹⁹⁴ Ibid 655 [36.118], 744 [Recommendation 33].

¹⁹⁵ Ibid 746–7.

¹⁹⁶ Ibid 747–9.

¹⁹⁷ Ibid 750–1.

witnesses (under common law principles), so that the defence, in particular, might have access to potentially critical perspectives and insights.¹⁹⁸

In terms of the provision of evidence, Recommendation 59 lists the factors that should be included in an examiner's report.¹⁹⁹ Recommendations 64 and 65 insist on the need to 'pay particular attention to ensuring that fingerprint evidence is presented to the court in such manner as to be readily understood by the judge and jury' and 'exploring' the use of technology to assist with the presentation.²⁰⁰

In addition, there was a perceived need for both examiners and prosecutors to maintain 'up-to-date knowledge' of cases and developments in Anglophone jurisdictions, particularly where 'courts, inquiries or other investigating bodies have made significant criticism of existing fingerprint practice' (Recommendation 85).²⁰¹

D *The Inquiry Into Pediatric Forensic Pathology In Ontario (2008)*

The Goudge Report followed a public inquiry into forensic medicine, specifically forensic pathology, in the wake of a series of wrongful convictions in Ontario based on mistaken or misleading expert evidence.²⁰² Conducted by Justice Stephen Goudge, a judge of the Court of Appeal for Ontario, the Inquiry was wide-ranging

¹⁹⁸ See the discussion in Michael Lynch and Simon Cole, 'Science and Technology Studies on Trial: Dilemmas of Expertise' (2005) 35 *Social Studies of Science* 269; Simon A Cole, 'A Cautionary Tale about Cautionary Tales about Intervention' (2009) 16 *Organization* 121.

¹⁹⁹ SFI Report, above n 12, 679–80, 748, Recommendation 59:

Each examiner's separate opinion should cover: (i) the images of the mark and also the specific print used in the comparison; (ii) the examiner's opinion about the quality of the mark; (iii) if the examiner considers the mark to be complex; (iv) whether third level detail is relied upon and the fact that such detail still requires to be supported by further research that has been validated; (v) identifying any differences between mark and print; (vi) a summary of the reasons why any differences between mark and print have been discounted and whether the examiner relies on objective studies and evidence to account for such differences or on common sense and experience; (vii) the characteristics relied on in making the identification, the number of such characteristics, and the classification of such characteristics, (e.g. ridge ending, bifurcation); (viii) a marked up image of the mark and print with a legend specifying the type of the ridge detail (including any third level detail) relied upon and the associated ridge counts; (ix) the opinion of the examiner; (x) any consultation with another examiner during the ACE-V process, including any facilitated discussion or panel review; and (xi) the fact that any novel method such as probabilistic analysis has been used or relied on.

²⁰⁰ *Ibid* 680, 749.

²⁰¹ *Ibid* 734, 752. This is a thinly-veiled reference to the NRC Report.

²⁰² The Goudge Inquiry was one of a series of inquiries following wrongful convictions in Canada. See *Re Truscott 2007 ONCA 575* and the reports discussed above at n 8.

in its review of forensic pathology.²⁰³ Goudge, as Commissioner of the Inquiry, made 169 recommendations pertaining to the organisation, practice and presentation of forensic pathology evidence in Ontario. This brief review focuses primarily on findings and recommendations directed to the practice of forensic pathology and the presentation of medical evidence in inquiries and legal proceedings.

The Goudge Inquiry was established in response to a number of wrongful convictions based on flawed paediatric forensic pathology evidence. Although one pathologist (Dr Charles Smith) was responsible for the Inquiry, having testified beyond his abilities and exaggerated his evidence — in service to the prosecution — leading to several wrongful convictions, Goudge found that many of the problems with forensic pathology were systemic.²⁰⁴ The major thrust behind Goudge’s recommendations is the need to professionalise, or for greater professionalisation of, forensic pathology. In terms of the practice of forensic pathologists and their production of expert opinions about cause of death there was a conspicuous need for more attention to scientific research, enhanced communications, greater transparency and improved documentation.

Central to Goudge’s recommendations for improvements in practice was the need for better training, certification and accreditation (Recommendations 2, 3, 7).²⁰⁵ This would involve both quality assurance and control (Recommendation 51) and improved recruitment and continuing education for pathologists (Recommendation 60).²⁰⁶ Goudge suggested that there was a need for improved standards and an institutional culture where pathologists were open to constructive criticism. The need for improvement ranges from dramatic structural reform to the organisation and governance of forensic pathology in Ontario (eg Recommendations 101–28) to the practice of forensic pathologists.²⁰⁷ In addition to accreditation and certification, reforms would involve a greater emphasis on teamwork (Recommendation 82) and developing guidelines ‘to assist forensic pathologists in adhering to best practice at or surrounding the autopsy’ (Recommendation 83).²⁰⁸

²⁰³ The author was an adviser and appeared before the Inquiry. See Gary Edmond, ‘Pathological Science? Demonstrable Reliability and Expert Pathology Evidence’ in Kent Roach (ed), *Pediatric Forensic Pathology and the Justice System* (Queen’s Printer for Ontario, 2008) 96. On the Canadian position, more generally, see Gary Edmond and Kent Roach, ‘A Contextual Approach to the Admissibility of the State’s Forensic Science and Medical Evidence’ (2011) 61 *University of Toronto Law Journal* 343. The Commission also consulted the Director of the Victorian Institute of Forensic Medicine, Professor Stephen Cordner.

²⁰⁴ See also William C Thompson, ‘Beyond Bad Apples: Analyzing the Role of Forensic Science in Wrongful Convictions’ (2008) 37 *Southwestern Law Review* 1027.

²⁰⁵ Goudge Report, above n 12, 295–6, 301–2.

²⁰⁶ *Ibid* 353–4, 361–2.

²⁰⁷ *Ibid* 436–69.

²⁰⁸ *Ibid* 402–5.

On the subject of bias, the main issues for forensic pathologists are characterised as whether they should attend the death scene and whether details of investigations (eg social-psychological information) should be disclosed to them, whether by investigating police or others.²⁰⁹ In response, Goudge tends to defer to the anticipated professionalism of forensic pathologists and comes out against withholding information.²¹⁰ Recognising differences of opinion over the value of blinding pathologists to information, and noting that Dr Smith had often relied heavily on social-psychological factors in drawing unfounded conclusions, Goudge nevertheless concluded that it would be inappropriate to blind investigating pathologists to relevant information, and indeed recommended that they should in many cases attend scenes.²¹¹ Recommendation 73(e) does, however, advocate developing protocols explaining ‘the types of information that should and should not be provided to the forensic pathologist.’²¹² These were consolidated in Recommendation 75, which stresses the need for forensic pathologists to ‘remain vigilant against confirmation bias or being affected by extraneous considerations.’²¹³ In terms of reaching a conclusion, ‘circumstantial evidence or non-pathology information’ ‘should never’ bear ‘the entire burden of support for an opinion’ (Recommendation 93).²¹⁴ Similarly, if a pathologist relies ‘in whole or part, on consultation with other experts’ that information should be documented and disclosed (Recommendation 94).²¹⁵

Generally, there was a need to disclose and explain the basis for opinions and their underlying facts:

Recommendation 95

- a) The articulation of the basis for the forensic pathologist’s opinion in a completely transparent way is at the cornerstone of evidence-based pathology.
- b) Forensic pathology opinions, whether given in writing or in oral communication, should articulate both the pathology facts found and the reasoning process followed, leading to the opinions expressed.²¹⁶

²⁰⁹ Where, for example, a child dies and the parents are on welfare and had a history of mistreatment and/or neglect.

²¹⁰ This may be a necessary consequence of the manner in which medico-legal death investigation is undertaken, but Goudge does not exhibit the same degree of concern about bias as the subsequent inquiries and reports. A reader might obtain the impression that the Commissioner thought that bias and other human factors could be overcome by being aware of dangers.

²¹¹ Goudge Report, above n 12, 377. Compare S Betts, *A Critical Analysis of Medical Opinion Evidence in Child Homicide Cases* (PhD Thesis, The University of New South Wales, 2013).

²¹² Goudge Report, above n 12, 384.

²¹³ Ibid 390.

²¹⁴ Ibid 422.

²¹⁵ Ibid 423.

²¹⁶ Ibid 427.

This becomes even more conspicuous in Recommendation 97, describing a code of practice for forensic pathologists that would include ‘principles that should guide them as they write their reports and the information that should be contained in them.’²¹⁷ This ‘should include at least the following’:

- a) the principles set out in Recommendation 84 [discussed below];
- b) guidance on the content of their autopsy and consultation reports (particularly where they may be used by the justice system), including
 - i) the subjects mandated by the *Code of Practice and Performance Standards for Forensic Pathologists* in England and Wales;
 - ii) details of each expert’s academic and professional qualifications, experience, and accreditation relevant to the opinions expressed in the report, as well as the range and extent of this expertise and any limitations on it;
 - iii) the levels of confidence or certainty with which the opinions are expressed;
 - iv) any alternative explanations that are raised by the pathology or by the reported history associated with the deceased’s death, with an analysis of why these alternative explanations can or cannot be ruled out;
 - v) what the pathologist has to say that is relevant to the live or pertinent issues in the case and why;
 - vi) any area of controversy that may be relevant to their opinions, placing their opinions in that context;
 - vii) any limits of the science relevant to the particular opinions;
 - viii) the extent to which circumstantial or non-pathology information has been used or relied on, and its impact on the reasoning and opinions;
 - ix) any other expert opinions relied upon;
 - x) the pathology facts found and the reasoning process that was followed, leading to the opinions expressed; and
 - xi) a glossary of medical terms, if helpful, to assist in communicating opinions in plain language to lay readers.
- c) guidance on
 - i) language to be used or avoided, and the dangers associated with the use of particular terms;
 - ii) how best to think about and articulate levels of confidence or certainty;
 - iii) the need to avoid the formulation or articulation of opinions in terms of proof beyond a reasonable doubt;
 - iv) the need to avoid default diagnoses;
 - v) the importance of recognizing and identifying for others the limits of their own expertise and of avoiding the expression of opinions that fall outside that expertise; and

²¹⁷ Ibid 429.

- vi) the cautions that should surround the use of circumstantial evidence or non-pathology evidence.²¹⁸

As this and other recommendations make clear, Goudge placed particular emphasis on the need for improved communication, transparency and documentation. As part of the general need for transparency, there is an explicit preference for the dissemination of information to be by way of writing and, if verbal, to be documented (Recommendations 74 and 76).²¹⁹ The need for documentation also applies to the record of analysis and autopsies along with appropriate retention of organs, samples and exhibits (Recommendation 77).²²⁰ There was also a perceived need for more timeliness in the provision of forensic pathology reports (Recommendation 80), for pathologists to exercise caution in the provision of preliminary opinions (Recommendation 78) and to reduce preliminary opinions to writing to avoid their potential to be misunderstood and to mislead (Recommendation 79).²²¹

The Goudge Report dedicated considerable attention to problems in the interaction between forensic pathology and the criminal justice system. Because many aspects of ‘a pathologist’s opinion may cause misunderstanding’, Recommendation 84 describes ‘general principles’ informing ‘the way pathology opinions are communicated’:

To reduce the risk of their being misunderstood, the most important parts of a forensic pathologist’s opinion should be expressed in writing at the earliest opportunity.

The ability of the various consumers of a forensic pathologist’s opinion — including peer reviewers, coroners, and stakeholders in the criminal justice system or child protection proceedings — to understand, evaluate, and potentially challenge the opinion requires that it be fully transparent. It should clearly state not just the opinion but the facts on which the opinion is based, the reasoning used to reach it, the limitations of the opinion, and the strength or degree of confidence the pathologist has in the opinion expressed.

Although some of the consumers of a forensic pathologist’s opinion are experts, such as peer reviewers, many are lay persons who have little or no understanding of technical language. It is essential that the pathologist’s opinion be understood by all the users. It must therefore be communicated in language that is not only accurate but also clear, plain, and unambiguous.

In expressing their opinions, forensic pathologists should adopt an evidence-based approach. Such an approach requires that the emphasis be placed on empirical evidence, and its scope and limits, as established in large measure by

²¹⁸ Ibid 429–30.

²¹⁹ Ibid 386–90, 391.

²²⁰ Ibid 392.

²²¹ Ibid 392–401.

the peer-reviewed medical literature and other reliable sources. This approach places less emphasis on authoritative claims based on personal experience, which can seldom be quantified or independently validated.²²²

During the course of the Inquiry:

it became apparent that there is no common understanding of how forensic pathologists think about their level of confidence or certainty in their opinions; how they articulate this level, if at all, when communicating their opinions; and how they might strive to sharpen their perception and articulation of the level of certainty in their views. Misunderstanding can arise in a number of ways. Of greatest concern is the possibility that the criminal justice system, in its search for certainty, will interpret a pathology opinion as reflecting a higher level of confidence than the expert intended.²²³

This led to Recommendation 86 (and 99), where it is suggested that forensic pathologists should report as accurately as possible pending the results of multidisciplinary work on a ‘common language to describe what forensic pathologists have to say’.²²⁴ It is also recommended (87 and 88) that forensic pathologists ‘should not formulate or articulate their opinions’ in terms of ‘proof beyond a reasonable doubt’ and should not change their level of confidence ‘depending on the forum in which the opinions are expressed’.²²⁵

Goudge emphasised the importance of disclosing the limits of opinions and the existence of differences of opinion and controversy. Recommendation 91 states:

- a) Forensic pathologists should clearly communicate, where applicable, areas of controversy that may be relevant to their opinions and place their opinions in that context.
- b) They should also clearly communicate, where applicable, the limits of the science relevant to the particular opinions they express.
- c) They should remain mindful of both the limits and the controversies surrounding forensic pathology as they form their opinions and as they analyze the level of confidence they have in those opinions.
- d) These obligations extend to the content of post-mortem or consultation reports, to verbal communications, and to testimony.²²⁶

The need for positive disclosure extends to ‘a positive obligation to recognize and identify for others the limits of their expertise. Forensic pathologists should avoid expressing opinions that fall outside that expertise. When invited to provide such opinions, they should make the limits of their expertise clear and decline to do so’

²²² Ibid 408.

²²³ Ibid 410–11.

²²⁴ Ibid 413, 435.

²²⁵ Ibid 414.

²²⁶ Ibid 419.

(Recommendation 92).²²⁷ Furthermore, forensic pathologists should not engage in ‘default diagnoses’ (Recommendation 89) and should outline, in reports, ‘the alternative or potential diagnoses that may arise in a case’ (Recommendation 90).²²⁸

Because the Inquiry emerged out of wrongful convictions, the Commissioner devoted time and resources to the ‘Role of the Court’ in an attempt to ‘assist in making the courts less vulnerable to unreliable expert evidence’.²²⁹ In this vein, Goudge explains that:

The justice system should place a premium on the reliability of expert evidence if it is to maximize the contribution of that evidence to the truth-seeking function and be faithful to the fundamental fairness required of the criminal process.²³⁰

The Commissioner stressed that the ‘judge must bear the heavy burden of being the ultimate gatekeeper in protecting the system from unreliable expert evidence.’²³¹ Among the main recommendations is the need for the court to ‘clearly define the subject area of the witness’s expertise and vigorously confine the witness’s testimony to it’ (Recommendation 129), and for judges to be vigilant ‘gatekeepers’ attentive to the reliability of expert evidence regardless of whether techniques are classified as novel (Recommendations 130, 131 and 133).²³² To assist in this capacity, the National Judicial Institute (of Canada) ‘should consider programs for judges on threshold reliability and the scientific method in the context of determining the admissibility of expert scientific evidence’ (Recommendation 134).²³³

Goudge also recommends both a Code of Practice for pathologists (Recommendation 98) and a Code of Conduct for Expert Witnesses providing:

that experts have a duty to assist the court ... and that this duty overrides any obligation to the person from who they received instruction or payment [Recommendation 136].²³⁴

The Recommendations for the trial, particularly the need for better funding for the defence follow recognition that trials and appeals had not uncovered systemic problems with the delivery of pathology evidence in prosecutions arising from paediatric injury and death.

²²⁷ Ibid 420.

²²⁸ Ibid 417.

²²⁹ Ibid 470, 471.

²³⁰ Ibid 484. See especially Recommendations 130 and 131, at 487, 496.

²³¹ Ibid 470.

²³² Ibid 487.

²³³ Ibid 502.

²³⁴ Ibid 505. These are consistent with codes already in place in England and Wales and many Australian jurisdictions.

E Thematic Summary of Recommendations

The table below provides a useful review of recommendations on some of the key issues. The table reinforces how many of the findings and recommendations are shared across the reports and would seem to be widely applicable to many types of forensic science and medicine in use across common law jurisdictions.

RECOMMENDATIONS	NRC REPORT	NIST REPORT	SFI REPORT	GOUDGE REPORT
Need for research, validity and reliability	Rec. 1(c), 1(i), 3	Rec. 3.7, 3.9, 4.1, 4.2	Rec. 81–4	Rec. 4, 69
Identification and reduction of error	5	3.6, 6.3, 9.1, 9.2	67, 69	
Identification and reduction of bias	5	3.3, 7.1	6–8, 18, 21, 22, 24, 29–32, 42	70, 75, 110
Current expression problematic	2	3.7	1, 2, 3	84, 89
Develop probabilistic approach	3	3.5, 4.2, 4.3	83	
Documentation, transparency and disclosure		3.1, 3.2, 6.2, 6.4	5, 17, 26, 39, 40, 44–53, 56–9	74–6, 79, 93–5
More training, formal qualifications and continuing education	1(d), 1(g)	4.3, 8.3, 8.4, 8.5, 8.7	17, 70–2, 85, 86	6, 7, 60, 88, 135
Comprehensible to end users	1(h)	5.1, 6.1	9, 64, 65	84–7, 90, 91, 96, 99, 100
Code of ethics/impartiality	9	6.2		97, 98, 136
Research on expression	3	3.8	82	84
More detailed reports	2	5.2	57, 59	
Accreditation, certification and proficiency	1(g), 6, 7, 8	6.4, 8.7, 8.8, 9.3, 9.4, 9.5	76–80	2, 32, 42–6, 50–3, 57, 59
More funding required	1, 10, 11, 12, 13	4.4, 8.2		8–11
Need to engage with scholars	6	8.4	16–18	3, 5, 33
Detailed standards and protocols	1(a), 1(b), 1, 6	3.4	10–12, 19, 20, 23, 25, 33–6	34, 52, 54, 68, 69
Legal responses	1(h)	6.1	60–3	101–28, 129–40
Need for independence	4			
Reliability for admissibility	Yes	Yes		Yes — ‘threshold reliability’
Total recommendations	13	34	86	169

Table 1: A Summary of The Main Findings and Recommendations

III IMPLICATIONS FOR AUSTRALIA (AND BEYOND)

The third part of this essay is a commentary. Before embarking on this analysis I want to make or reiterate four important points. First, many forensic science and medical techniques are insufficiently reliable. That is, we do not know if they work, how well or in what conditions. They are insufficiently reliable because appropriate scientific research has not been performed.²³⁵ Consequently, the validity and reliability of many techniques in routine use is simply unknown and standards and accreditation are not predicated upon the results of research.²³⁶ Secondly, the reports stress the importance of transparency. They place considerable emphasis on the need for enhanced documentation and disclosure to lawyers, courts and other experts. Transparency is important generally, though especially important where there are serious limitations, such as little underlying research or analysts being exposed to threats from cognitive bias. Thirdly, longstanding (and continuing) use, previous legal recognition, and the experience of examiners and judges, are not presented in the reports as grounds for maintaining confidence in particular techniques, interpretations or practices. The fourth point is that the kinds of issues raised in these reports emerge relatively rarely in trials and appeals in Australia and do not feature prominently in the relevant jurisprudence.²³⁷

In the remainder of this essay, it is my intention to briefly review some of the findings, recommendations and insights with an eye to their implications for forensic science and medicine evidence and criminal proceedings in Australia. Many raise unsettling questions about the performance of lawyers and judges interpreting and applying rules as well as the effectiveness of adversarial mechanisms.

A Australian Exceptionalism?

Critically important to this discussion is the relevance of these findings and recommendations to forensic science and medicine in Australia. To the question ‘Do they apply to Australia?’, the answer is an almost unqualified ‘Yes’. Australia is not exceptional in its production, presentation and treatment of forensic science and medicine evidence.

²³⁵ Appellate courts in England and Wales have merely required that, if requested, analysts should disclose the fact that there has been no validation study and there is no database behind claims about frequency. Even though the seriousness, and implications, of these oversights are rarely developed. See, for example, *R v Atkins* [2009] EWCA Crim 1876 (30 July 2009); discussed in Gary Edmond et al, ‘*Atkins v The Emperor: The “Cautious” Use of Unreliable “Expert” Opinion*’ (2010) 14 *The International Journal of Evidence & Proof* 146.

²³⁶ NRC Report, above n 12, 135. This extends to techniques that are otherwise reliable, see William C Thompson, ‘Painting the Target around the Matching Profile: The Texas Sharpshooter Fallacy in Forensic DNA Interpretation’ (2009) 8 *Law, Probability and Risk* 257.

²³⁷ Andrew Ligertwood and Gary Edmond, *Australian Evidence* (LexisNexis, 5th ed, 2010).

While there are significant differences, particularly around admissibility jurisprudence, rules and procedures, traditions of practice and even levels of accreditation and certification, most of the problems with forensic sciences in North America and the United Kingdom have direct implications for Australia.²³⁸ This is readily discernible in relation to the inadequate research base, the forms of expression (used in reports and testimony), the failure to disclose error rates, the reluctance (or inability) to introduce probabilistic forms of expression (whether frequentist or likelihood ratios) based on data, and the unwillingness to shield analysts from domain irrelevant information. Two examples are illuminating. Taking the research base, as an obvious and decisive example (that trumps accreditation and standards), the NRC Report explained that many of the techniques, especially those that were not derived from developments in mainstream biology and chemistry, do not have an adequate research base.²³⁹ For, where there is no research base in the United States (or England or Canada or Germany or Japan) we can confidently assume that none exists. The globalised nature of the modern sciences and international collaboration in law enforcement and counter-terrorism efforts means that missing research is missing everywhere. Australia is not in possession of a clandestine research repository. The fact that Australian laboratories might be better accredited, to varying degrees have standards and protocols, and better-trained personnel, cannot overcome the failure to have undertaken validation studies, confirm that techniques and practices are efficacious, or the need to attend to uncertainty and error.²⁴⁰

A second, more specific example concerns latent fingerprint evidence. Australian fingerprint examiners continue to rely on ACE-V as a comprehensive and adequate description of their scientific credentials and continue to express the results of their comparisons in positive terms — equating a match with the identification of an individual to the exclusion of all other persons. While latent fingerprint evidence will usually have considerable probative value, Australian examiners, like their British and North American counterparts, tend to be indifferent or unresponsive to what have become notorious issues in the practice and reporting of latent fingerprint evidence. Australian examiners are not shielded from gratuitous (ie domain irrelevant) information and routinely provide positive identification evidence.²⁴¹ These practices are inconsistent with the recommendations of the three reports that considered latent fingerprint evidence.

²³⁸ Australia has been a leader in the accreditation of laboratories. However, accreditation depends on the validation of techniques. Having accredited laboratories using techniques that have not been validated and with unknown levels of error is likely to disguise real risks and limitations as analysts appeal to formal processes that may be devoid of content. Accreditation without validation is a triumph of form over substance.

²³⁹ It trumps accreditation and standards because these should be predicated upon research and validation studies for particular practices.

²⁴⁰ Accreditation and standards mean that a process is in place and that analysts are doing the same things, not that the processes or practices actually work.

²⁴¹ See Edmond, Tangen and Thompson, 'A Guide to Interpreting Forensic Testimony', above n 155.

While the reports have undoubted relevance and, in many cases, application to Australia, it is also appropriate to recognise the variegated nature of the forensic sciences and their institutions both within and across Australian jurisdictions.²⁴² Unremarkably, there are anomalies and counter examples. The Victorian Institute of Forensic Medicine, for example, has been a leader in the development of standards and the provision of qualified evidence in reports and testimony. Several of the Institute's most senior forensic pathologists have, in the face of controversy and/or inadequate research foundations, been willing to acknowledge uncertainty and have manifested a creditable reluctance to express strong opinions under pressure from police and prosecutors.²⁴³ Similarly, document examiners in the Victoria Police Forensic Services Department have endeavoured to respond to perceived weaknesses and threats from bias by restricting access to domain irrelevant information and developing probabilistic approaches to their analyses.²⁴⁴ Australia's version of a NIFS is simultaneously endeavouring to improve the standard and performance of the forensic sciences, but in terms of resourcing, independence from law enforcement, scientific leadership and authority to impose standards, is but a pale shadow of the institute proposed in the NRC Report.²⁴⁵ The basic need for more research, for techniques, processes and the expression of opinions to be derived from research (and data), for attention to human factors, and for greater transparency about processes and limitations, seems — notwithstanding a few conspicuous exceptions — to be as applicable to Australia as elsewhere.

Those who would contend that the convergent findings and recommendations from other advanced common law jurisdictions do not apply to Australia have responsibility for making that case persuasively.²⁴⁶ In anticipation, it is important to emphasise that accreditation, a process to develop generic standards for the forensic sciences and Australia already having a National Institute of Forensic Sciences²⁴⁷ are not credible responses to the lack of research, a history of non-disclosure (see Part III(B))

²⁴² See Andrew Ligertwood, 'Forensic Science Expressions and Legal Proof' (2013) 45 *Australian Journal of Forensic Sciences* 263.

²⁴³ See *R v Matthey* (2007) 177 A Crim R 470; *Gilham v The Queen* [2012] NSWCCA 131 (25 June 2012). Professor Cordner's 'modest' approach to the interpretation of the injuries in *Gilham* led two senior prosecutors to characterise him as an unreliable witness. See Justice McClellan's criticisms of these characterisations in *Gilham* at [351]ff, [383]ff and [396]–[412].

²⁴⁴ Interestingly, Victoria seems to be a something of an Australian leader in these areas.

²⁴⁵ See, for example, Frank Vincent, *Independent Review of the National Institute of Forensic Science* (July 2014).

²⁴⁶ These issues are so serious that the burden cannot rest with critics.

²⁴⁷ Australia's NIFS is not the kind of independent, science-driven institution recommended in the NRC Report. Although it was originally more independent, NIFS is effectively captured by law enforcement: funded by state and federal police and dominated by individuals who are or were police and/or forensic scientists. The Australian NIFS reports to the Australia New Zealand Policing Advisory Agency (ANZPAA) board. See Alastair Ross, 'Forensic Science in Australia — Can We Learn from International Reports?' (2011) 43 *Australian Journal of Forensic Sciences* 135.



below) and continuing indifference to notorious threats from contextual and other forms of bias.²⁴⁸ Similarly, the fact that we have been doing things for a long time and have not (yet) exposed as many wrongful convictions as any other jurisdiction is no answer to substantial criticisms — see Part III(J) below.²⁴⁹

The findings and recommendations in the reports make clear that many of our trials and appeals have been substantially unfair. Forensic science and medicine evidence has been routinely misrepresented in reports and testimony (regardless of whether the analysts and lawyers knew) and there have been relatively few substantial challenges to forensic science evidence that was insufficiently reliable. Fingerprint evidence continues to be opaque; expressed in absolute terms — linking a latent print to the accused as opposed to every other person who has ever lived. Incriminating opinions derived from images, voices, bite marks, tools and firearms, documents, fires and explosives, shoe and tire marks, injuries to children (and so on) are routinely expressed in terms that are not consistent with available research (ie *what is known*). Even DNA evidence, the poster-child for reliable forensic science techniques, has very real risks of error.²⁵⁰ We have already encountered mistakes in the collection and continuity of DNA evidence, and increasing sensitivity, mixed samples, along with new techniques and applications, create difficult interpretive issues that are yet to be studied or credibly addressed.²⁵¹ Moreover, Australia has few coherent

²⁴⁸ The author is the only active member of the Standards Australia committee on the reference for the forensic sciences who is not a current or former member of a police force or forensic science institution. Revealingly, there are no independent scientists, no cognitive scientists, and no other lawyers involved in the drafting and revision processes.

²⁴⁹ We do not have independent mechanisms such as Criminal Cases Review Commissions or well-functioning innocence projects in Australia.

²⁵⁰ NRC Report, above n 12, 101, 106. ‘The goal is not to hold other disciplines to DNA’s high standards in all respects; after all, it is unlikely that most other current forensic methods will ever produce evidence as discriminating as DNA. However ... the least that the courts should insist upon from any forensic discipline is certainty that practitioners in the field adhere to enforceable standards, ensuring that any and all scientific testimony or evidence admitted is not only relevant, but reliable.’ See also David H Kaye, *The Double Helix and the Law of Evidence* (Harvard University Press, 2010) and Lynch et al, *Truth Machine*, above n 7. Interestingly, the NIST Report cites studies that suggest error rates for DNA testing in European laboratories have ‘stabilized in the range of 0.4% to 0.7%.’ S Rand et al, ‘The GEDNAP (German DNA Profiling Group) Blind Trial Concept Part II: Trends and Developments’ (2004) 118 *International Journal of Legal Medicine* 83.

²⁵¹ See, for example, Vincent, *Report: Inquiry Into the Circumstances That Led to the Conviction of Mr Farah Abdulkadir Jama*, above n 182. More generally, there are dangers where DNA analysts interpret results, especially mixed samples and LCN profiles after they have been exposed to the profile of the suspect. Analysts should be required to *identify and document* what are peaks, stutters and other anomalies before being exposed to the profile of the suspect. There are, with almost all techniques, problems in moving from the data to the interpretation and in conveying the opinion/conclusion to the decision-maker in ways that are simultaneously consistent with what is known, comprehensible and meaningful in terms of criminal proof. See William C



review mechanisms beyond appeals. Unlike England, Wales and Scotland and many jurisdictions in the United States, our post-conviction review processes are generally narrow and incumbents incredulous in disposition.²⁵²

B *When are the Forensic Analysts Planning to Come Clean?*

While the various findings and recommendations may come as something of a surprise to many legal practitioners and judges, they are by now familiar to forensic analysts. Most forensic analysts have some familiarity with these reports and recommendations. All Australian fingerprint examiners, by way of example, are aware of the controversy surrounding Mayfield and McKie and the subsequent reports.²⁵³ Conferences and working groups have devoted time and discussion to their implications. Indeed, the initial shock and denial has largely passed, as many analysts and institutions are now struggling to come to terms with the need for research and reform. However, as the reports explain, many analysts do not have the time, the necessary skills and/or resources to address fundamental epistemic problems. Most forensic analysts, including latent fingerprint analysts, are not trained in research methods, statistics or cognitive science. Consequently, they are not in a good position to respond to the various recommendations unilaterally.²⁵⁴

In general, we are yet to see responses to the criticisms of current practices, particularly in relation to research, revised terminologies and disclosure. Thus far, and notwithstanding the chorus of criticisms that both pre-date and post-date the reports, forensic analysts relying upon techniques that have never been validated, are yet to disclose these oversights and the corrosive implications for their 'evidence'. They have not volunteered that extant processes do not shield against notorious cognitive biases, they have not explained that many opinions are little more than speculation and that many traditional practices are without empirical foundation. They have not conceded that in many areas they have no idea about their ability or error rate.²⁵⁵ As a

Thompson, 'What Role Should Investigative Facts Play in the Evaluation of Scientific Evidence?' (2011) 43 *Australian Journal of Forensic Sciences* 123 and *Tuite v The Queen* [2015] VSCA 148 (12 June 2015).

²⁵² David Hamer, 'Wrongful Convictions, Appeals, and the Finality Principle: The Need for a Criminal Cases Review Commission' (2014) 37 *University of New South Wales Law Journal* 270.

²⁵³ They are also aware of the work of critics such as Professor Simon Cole, Professor Jennifer Mnookin and others. However, they do not refer to notorious criticisms in reports or testimony.

²⁵⁴ Research scientists have begun to develop statistical models and undertake appropriate validation studies so that we have a clearer idea about abilities, errors and the value of matches. Ironically, some of the pioneers of fingerprint comparison, such as Sir Francis Galton (1822–1911) and Henry Faulds (1843–1930), were sophisticated mathematicians aware of many statistical and practical-epistemic problems.

²⁵⁵ It may be that many techniques will turn out to be quite reliable. The point is that we should not take this on trust when ability and accuracy can be measured and techniques can be improved.

community, forensic scientists have been recalcitrant, sometimes duplicitous, in their failure to proactively concede notorious epistemic constraints and bring them to the attention of users, whether lawyers, judges or jurors.

The continuing silence, especially from leaders and managers, along with conscious omissions from expert reports and trial testimony, is nothing short of scandalous. Forensic scientists have a duty to the court and a commitment to truth. They must respond to the criticisms and professional obligations and disclose epistemic limitations.

C ‘Reliability’ in Admissibility and Discretionary Decision-Making

The widespread failure of forensic scientists to unilaterally disclose limitations — notwithstanding the ‘oath’, admissibility standards, procedural rules, codes of conduct and practice directions (see Part III(G), below) — means that the burden falls on the lawyers and the trial. In practice the burden tends to fall almost entirely on the defence. This sits awkwardly with the presumption of innocence (and the burden of proof) and, as we have seen, is a responsibility that the system is not, at present, capable of bearing.²⁵⁶ Nevertheless, the findings and recommendations in these reports reinforce the need for attention to the value of forensic science and medicine evidence and active judicial gatekeeping at the admissibility stage. Admissibility is important because the trial has not proven to be consistently effective in exposing or addressing weaknesses — see Part III(D) below. Incriminating expert opinion evidence should be demonstrably reliable. Unreliable and insufficiently reliable techniques and opinions should be excluded.

It is essential for lawyers and judges to direct their attention to the reliability of incriminating expert evidence.²⁵⁷ This applies across the board and is not restricted to novel techniques.²⁵⁸ Expert opinions derived through techniques that have not been evaluated, or were derived through processes where the analyst was unnecessarily exposed to gratuitous information, or are not expressed in terms that have an appropriate foundation in research, have no place in a rational system of justice. They are not susceptible to rational evaluation by laypersons either individually or as part of a case.²⁵⁹ In their admissibility jurisprudence around both the exception for opinions

²⁵⁶ Inattention to the reliability of incriminating expert opinion evidence, and presenting incriminating opinions without appropriate qualifications and caution threatens to subvert criminal proof.

²⁵⁷ This can be quite a complicated undertaking and it may be that models of reliability must respond, to some extent, to the particular type of expertise. See Edmond, ‘Pathological Science?’, above n 203.

²⁵⁸ See *R v Trochym* [2007] 1 SCR 239. The recent decision of the Supreme Court of Victoria Court of Appeal in *Dupas v The Queen* (2012) 218 A Crim R 507 represents a positive development, but there needs to be a positive obligation on the state rather than an enhanced opportunity for the defence to challenge.

²⁵⁹ Recall, NRC Report, above n 12, 184: ‘The level of accuracy of an analysis is likely to be a key determinant of its ultimate probative value.’ See also Gary Edmond, ‘Forensic Science Evidence and the Conditions for Rational (Jury) Evaluation’ (2015) 39 *Melbourne University Law Review* (forthcoming).

based on ‘specialised knowledge’ and (mandatory and) discretionary exclusions, trial and appellate judges should direct their attention to the reliability of techniques and derivative opinions and the very real dangers that follow from unreliability and inattention to appropriate practices.

While reliability ought to be a prerequisite for admissibility, it is not my intention to suggest that a formal reliability threshold will operate as some kind of panacea.²⁶⁰ After all, many jurisdictions with an explicit reliability standard admitted, and continue to admit, incriminating opinions that are insufficiently reliable — either unreliable or of unknown reliability. Nevertheless, apprised of some of the problems and recommendations, lawyers and judges are now in a better position *and obliged* to respond to the reliability of expert evidence. Lawyers and trial judges should be willing and able to ask to see research and to ask about procedures, standards, rates of error and limitations. They should expect to see reference to relevant published studies. In many cases, inattention to these sorts of issues threatens both the fairness of the trial and the safety of any conviction. Actual reliability would seem to be a condition precedent to admission. Inattention or insufficient attention to reliability has led to the current state of affairs where unreliable and speculative opinions are sometimes treated as scientifically predicated and reliable, occasionally infallible, evidence of guilt.

Courts have been too accommodating in their responses to the state’s incriminating expert evidence. They have ‘certified’ techniques and experts prematurely; thereby allowing untested and therefore speculative forms of evidence into trials, and required the defence to somehow identify, explain and successfully convey limitations at the accused’s peril. Symbolically, as well as practically, it seems that requiring ‘experts’ called by the state to demonstrate actual abilities and readily concede limitations and uncertainties is consistent with the basic goals of the accusatorial trial and criminal proof. Inattention to reliability places decision-makers in an impossible position and subverts the goal of doing justice in the pursuit of truth.²⁶¹

D Trial Safeguards are Weak and Have Been Ineffective Against Expert Evidence

Attention to reliability, along with the willingness to exclude relevant though insufficiently reliable incriminating expert evidence, is vitally important because adversarial mechanisms and trial safeguards have not proven effective.²⁶² Individually

²⁶⁰ See Gary Edmond et al, ‘Admissibility Compared: The Reception of Incriminating Expert Opinion (ie Forensic Science) Evidence in Four Adversarial Jurisdictions’ (2013) 3 *University of Denver Criminal Law Review* 31.

²⁶¹ H L Ho, *A Philosophy of Evidence Law: Justice in the Search for Truth* (Oxford University Press, 2008).

²⁶² Here it is important to acknowledge that serious (albeit usually unsuccessful) challenges occasionally occur in the US where coordinated innocence projects and attentive scholars often cooperate in an attempt to expose problems and limitations through strategic litigation. Challenges that extend beyond the superficial are highly irregular in the US and Australia, but they occur more regularly in the US just because of the sheer volume of cases and levels of specialisation.

and in combination, prosecutorial restraint (as a ‘minister of justice’), admissibility standards and judicial discretions, cross-examination, defence (ie rebuttal) experts, restrictions on expression (eg similarity evidence rather than positive identifications), directions and warnings, the onerous burden of proof, and even appeals, have not afforded credible protection to those confronted with unreliable opinions or expert evidence that was not appropriately constrained.²⁶³ Significantly, they did not expose the problems identified in the various reports nor reveal their prevalence across the forensic sciences. Standard judicial directions and warnings, in particular, do not refer to, nor have they substantially engaged with, the many issues raised in the reports. It is perhaps illuminating that, years after the publication of highly critical reports, legal safeguards are yet to expose serious (and now notorious) epistemic frailties.

Traditional legal practices and protections did not persuade judges to alter their historically accommodating response to incriminating expert evidence. They did not effectively convey the range and depth of problems, nor the inability of many disciplines and practitioners to fully gauge their significance, let alone address them. The focus on the individual case and the rather myopic manner in which cases are tried and appealed, seem to have made it difficult for trial and appellate judges to appreciate (or respond to) some of the systemic dimensions at play across a wide range of techniques and practices. There has, for example, been no systematic attempt to consider the way human factor issues might impact across all interpretive processes, or the way that different types of identification and comparison evidence might be expressed in court. In consequence, legal responses to different types of expert evidence tend to be largely oblivious to structural similarities (and dangers) and the responses to different techniques (eg using DNA profiling, bite marks, CCTV images and voice recordings to assist with identification) tend to be not merely unprincipled, but epistemologically incoherent.²⁶⁴

Trial safeguards and protections (and human rights instruments) can, in some circumstances, afford very effective means of identifying and presenting evidentiary weaknesses to the tribunal of fact. On most occasions they do not. In practice, trial safeguards and commitment to a fair trial often have more of a discursive or rhetorical flavour than a substantial one. Historically, trial and appellate judges have placed great store in the effectiveness of admissibility rules, the power of cross-examination, their own directions and instructions to the jury, along with the jury’s ‘common sense’. Notwithstanding this seemingly inexhaustible faith, none of these and other protections consistently, nor effectively, exposed the profound problems with many types of forensic science and medicine.

In a recent review of expert opinion evidence in criminal proceedings in England and Wales, the Law Commission expressed doubts about widespread legal confidence in traditional trial safeguards:

²⁶³ Gary Edmond and Mehera San Roque, ‘The Cool Crucible: Forensic Science and the Frailty of the Criminal Trial’ (2012) 24 *Current Issues in Criminal Justice* 51.

²⁶⁴ See Edmond et al, ‘Admissibility Compared’, above n 260.

Cross-examination, the adduction of contrary expert evidence and judicial guidance at the end of the trial are currently assumed to provide sufficient safeguards in relation to expert evidence ... However, ... it is doubtful whether these are valid assumptions.²⁶⁵

Similar concerns were expressed, more starkly, by Judge Edwards in the aftermath of the NRC Report:

Unfortunately, the adversarial approach to the submission of evidence in court is not well suited to the task of finding ‘scientific truth.’ The judicial system is encumbered by, among other things, judges, lawyers, and jurors who generally lack the scientific expertise necessary to comprehend and evaluate forensic evidence in an informed manner; defense attorneys who often do not have the resources to challenge prosecutors’ forensic experts; trial judges (sitting alone) who must decide evidentiary issues without the benefit of judicial colleagues and often with little time for extensive research and reflection; and very limited appellate review of trial court rulings admitting disputed forensic evidence. Furthermore, the judicial system embodies a case-by-case adjudicatory approach that is not well suited to address the systematic problems in many of the various forensic science disciplines.²⁶⁶

Trials and appeals have not identified profound and widespread problems with the forensic sciences and medicine. Trial and appellate judges have not provided relevant or practical leadership and guidance in response to systemic problems with forensic science and medicine evidence. They have preferred to admit expert evidence and, when challenged, exhibited a tendency to valorise longstanding legal techniques, such as cross-examination and judicial directions, and privilege the experience of state-employed analysts — see Part III(I), below.

²⁶⁵ Law Commission of England and Wales, *Expert Evidence in Criminal Proceedings in England and Wales*, 34 Law Commission Report No 325 (HMSO, 2011) 1.20, 1.24. Compare the US Supreme Court majority’s confidence in these mechanism in *Daubert v Merrell Dow Pharmaceuticals Inc* 113 S Ct 2786, 2798 (1993) (*‘Daubert’*) (italics added): ‘Vigorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking *shaky* but admissible evidence’. Simultaneously, the Supreme Court dismissed concerns about ‘the capabilities of the jury and of the adversary system generally’ as ‘overly pessimistic’.

²⁶⁶ Edwards, ‘Solving the Problems’, above n 11, 19 (emphasis added). Testifying before the Goudge Inquiry, Justice LeSage stated (at 501):

I must say it came as somewhat a shock to me, having spent 40 years in the justice system, to hear some of the scientific experts speak of the uncertainty and lack of clarity in areas of science that I had always thought of as much more certain than they really are. And I felt very guilty that I had not better educated myself on these areas long before.

E Bias and Cross-Contamination

One of the central and recurring themes in all of the reviews was the issue of bias. This is revealing because the danger posed by contextual bias has played almost no role in the discussion of forensic science and medicine evidence in trials and appeals. Significantly, the inquiries and reports were not especially interested in bias in the sense of overt partisanship (of the kind that tends to animate judges and law reform).²⁶⁷ Rather, they focused attention on cognitive influences that tend to operate below the threshold of consciousness and are, therefore, difficult and sometimes impossible for analysts to overcome.

All of the reports recognise that exposure to gratuitous information — along with other influences and environmental factors — has a well-established tendency to subtly contaminate human cognition. The NRC Report, the NIST Report and the SFI Report each insist on the need for further research in this area and, in the absence of research, precautionary steps to prevent cognitive processes and institutional pressures improperly contaminating the interpretation and evaluation of evidence. In practice, this requires managing access to information about a case or investigation and may require shielding analysts from some kinds of information (blinding), and/or gradually exposing them to information as they document their initial impressions (sequential unmasking). It may also require the redesign of processes and computer programs (eg AFIS) that currently provide analysts with access to more information than is necessary.

Remarkably, these sorts of issues have rarely surfaced in trials and appeals.²⁶⁸ This is remarkable because threats to cognitive processes have a demonstrated tendency to change how humans interpret evidence. Experiments have demonstrated that analysts may change their opinions when exposed to information that is not relevant to their analysis. In one series of studies fingerprint examiners (and DNA analysts), who did not know they were part of an experiment, were exposed to case information unrelated to their analyses. They were asked to re-assess samples they had previously matched (several years earlier) while under the impression that they had not seen the particular latent prints before. As part of the study the analysts were provided with information about the investigation, which implied that the prints did not match. The results indicate that a large proportion (up to 80 per cent in one study) produced interpretations, on the central issue of whether two prints or profiles matched, that

²⁶⁷ English and Australian judges, in contrast, have been concerned with adversarial partisanship, especially in civil proceedings. See Lord Woolf, *Access to Justice* (HMSO, London, 1996); see also Ian Freckelton, Prasuna Reddy and Hugh Selby, *Australian Judicial Perspectives: An Empirical Study* (Melbourne, Australian Institute of Judicial Administration, 1999); New South Wales Law Reform Commission, *Expert Witnesses*, Report No 109 (2005). See also Peter Huber, *Galileo's Revenge: Junk Science in the Courtroom* (Basic Books, 1991).

²⁶⁸ Contrast Shannon, *Royal Commission of Inquiry in Respect to the Case of Edward Charles Splatt*, above n 8.

were inconsistent with previous interpretations of the same material by the same examiner.²⁶⁹

Threats from bias are real and, significantly, cannot be easily compensated for in the course of accusatorial proceedings. Once analysts are exposed to (cognitive) contamination, the threat cannot be corrected by informing the jury about potential dangers. What, after all, should a judge tell a jury? That the results of the fingerprint comparison are probably correct, but the analyst was unnecessarily exposed to the kind of information that has been shown to make them produce mistakes. How does a jury make sense of such evidence in the particular case? Obviously, the appropriate course of action is to shield analysts from domain-irrelevant information unless there is experimental evidence that exposure is unlikely to influence the analysis.

At this point it is worth stating emphatically that long experience and familiarity with the dangers of cognitive bias do not enable analysts to resist its pernicious effects. This is not an issue of morality or character but rather the manner in which humans process information. The only effective means of avoiding the danger of contamination is through shielding.

Another troubling issue arising out of inattention to bias concerns the way in which expert opinions are presented at trial and the danger of double counting when it comes to evaluating the evidence. At present, in almost all cases, forensic analysts are exposed to a great deal of (no doubt titillating) information that is not relevant to their practice.²⁷⁰ Nevertheless, at trial, forensic science and medicine evidence is routinely presented as *independent* support (or corroboration) for the case against the accused. In reality, because of the danger posed by contamination, we do not know if such interpretive evidence is actually independent. If an interpretation (ie opinion) was shaped or strengthened, for example, by other information (such as knowledge that the suspect had made certain admissions, or the detectives indicated that the suspect had antecedents) then claims about the independence of the expert evidence are misleading. The forensic science evidence is no longer independent of other strands of information — regardless of whether that information is reliable or admissible.²⁷¹ While the evidence may still have probative value it has less probative value than opinions that were derived in ignorance of such information.²⁷² Once

²⁶⁹ Dror, Charlton and Peron, 'Contextual Information' above n 91; Itiel E Dror and Greg Hampikian, 'Subjectivity and Bias in Forensic DNA Mixture Interpretation' (2011) 51 *Science and Justice* 204.

²⁷⁰ One of the difficulties is sometimes determining what is relevant. This, as Goudge's recommendations imply, is usually more of an issue for forensic pathologists than fingerprint analysts. See also Emma Cunliffe, 'Judging Fast and Slow: Using Decision-making Theory to Explore Judicial Fact Determination' (2014) 18 *The International Journal of Evidence & Proof* 139.

²⁷¹ It is not only reliable knowledge that has the power to influence. Innuendo, unfounded beliefs and commitments, institutional pressures and expectations, can all unwittingly shape cognitive processing.

²⁷² See Edmond et al, 'Contextual Bias and Cross-Contamination in the Forensic Sciences', above n 53.

contaminated the probative value is unknown (reliability is compromised), but the danger of unfair prejudice to the accused through misrepresentation, over-valuation, the danger of mistakes and the impossibility of exploring the issue in relation to any specific analysis and opinion will often be extreme.²⁷³

Where, as in most cases, there is no need to expose the analyst to background information, this information should not be conveyed and should not be accessible to the analyst.²⁷⁴ Shielding the analyst has the benefit of avoiding threats to cognition, strengthening their evidence (by making it genuinely independent), and enabling analysts to provide independent feedback on, rather than endorsement of, investigative suspicions and leads. Warning fact-finders about forensic science evidence represented as independent corroboration not actually being independent is, in contrast, not an effective means of addressing very serious dangers that could, in the vast majority of cases, be avoided relatively easily. In practice it places an obligation on the trial judge and the defence lawyer to explain the dangers from bias — even though lawyers and judges had not previously appreciated their significance, and there will be little chance of identifying any effects from bias in the specific case (regardless of whether they were present). It also imposes the dangers created by insensitivity to bias on the accused even though investigators could have employed procedures that avoided the risk.

Moreover, the trial is not necessarily stuck with the contaminated evidence. In many cases the actual interpretation could be re-conducted by a separate analyst (preferably from a different laboratory) shielded from gratuitous information. In the absence of evidence demonstrating that bias is not an issue, trial and appellate judges should be very cautious about admitting contaminated evidence. The issue of bias and the cross-contamination of evidence are difficult to manage at trial. In particular, subtle exposure and contamination are often difficult to trace retrospectively. Because effects tend to operate below the threshold of consciousness they are difficult to explore through cross-examination. Of significance, the passing of gratuitous information is seldom documented, often informal and may even be forgotten. Actual exposure — rather than a witness’s recollection or demonstrable impact on their practice or evidence — ought to guide decisions about admissibility and reliance.²⁷⁵ Responsibility for demonstrating that exposure to domain irrelevant information does not bias practice, or the conditions in which analysts are capable of resisting some kinds of unnecessary contamination, properly resides with the State and its analysts.

²⁷³ Here the only way the issue can be explored by the defence is by introducing general theories of cognitive problems. There is very little chance of identifying contamination and its specific effect(s) post factum.

²⁷⁴ Sometimes the process itself can be suggestive, such as where the analyst is only presented with one comparator. So, for example, where police present images of only one suspect, along with CCTV images of an armed robbery, the process conveys their belief that the suspect is the offender. This is standard practice in forensic image comparison in Australia, England and Wales.

²⁷⁵ Failure to address the issue in reports and documents should be viewed very critically.

F *Jury Not Properly Positioned to Evaluate the Evidence*

Regardless of one's position on the continuing use of jurors, it follows from the failure of forensic scientists to appreciate and disclose limitations, and the failure of lawyers and judges to independently identify and adequately convey (or respond to) them, that juries have not been placed in conditions that are conducive to the rational evaluation of incriminating opinion evidence and proof of guilt more generally.²⁷⁶ When it comes to jury evaluation of expert evidence, and the combination of expert evidence with other forms of evidence, trial and appellate court confidence and deference would seem to be misplaced.²⁷⁷

G *Court-Appointed Experts, Concurrent Evidence and Codes Of Conduct*

Of all the inquiries only Goudge and, to a more limited degree, Campbell devote attention to legal — as opposed to scientific — practice. Revealingly, there is little support for importing putative solutions from non-adversarial systems.²⁷⁸ Indeed, Goudge considered but explicitly rejected them: 'Court-appointed or joint experts are not recommended for cases involving paediatric forensic pathology.'²⁷⁹ Whereas judges frequently express concerns about the partiality or partisanship of expert witnesses, especially expert witnesses called by plaintiffs in civil suits and defendants in criminal proceedings, the discussion in the reports is instead focused on problems — flowing from the lack of research, weak standards and inattention to human factors — *impacting on state-employed forensic analysts*. The emphasis on reliability, research, standards, and transparency is intended to make the occasional incidence of fraud more difficult as well as address systemic problems and risks created by human factors. Significantly, risks are not removed and may not even be addressed through recourse to different, apparently more neutral or independent, experts. For, even genuinely independent experts must employ valid techniques and can have their opinions unwittingly contaminated through exposure to gratuitous information. In the Mayfield case, an independent expert confirmed the mistaken attribution made by three FBI fingerprint examiners.²⁸⁰ All persons, including court-appointed experts (and jurors and judges), are vulnerable to cognitive bias and error.

²⁷⁶ For an orthodox expression, see Ronald Allen and Joe Miller, 'The Common Law Theory of Experts: Deference or Education' (1993) 87 *Northwestern University Law Review* 1131. See also Edmond, 'Forensic Science Evidence and the Conditions for Rational (Jury) Evaluation', above n 259.

²⁷⁷ Here, claims about judge and jury agreement might be rendered problematic to the extent that both groups have relied on quite simplistic impressions of expertise and expert evidence.

²⁷⁸ Edmond and Vuille, 'Comparing the Use of Forensic Science Evidence in Australia, Switzerland and the United States', above n 4.

²⁷⁹ Goudge Report, above n 12, 506–7 [Recommendation 137]. Goudge accepted that 'people such as Professor Sir Roy Meadow, whose testimony was later discredited in England, or Dr Smith might have been precisely the type of well-known "experts" appointed by the court or chosen by the parties as joint experts.'

²⁸⁰ The independent reviewer was not blinded to the conclusions of the other examiners.

Only the Goudge Inquiry considered receiving evidence concurrently. Goudge recommended more pre-trial meetings between experts, and between experts and lawyers. He also proposed trialling concurrent evidence procedures. Accepting that pre-trial meetings and concurrent evidence procedures may have some potential to enhance efficiency and improve communication and comprehension, it is important to recognise that achieving consensus among experts brought together (by the parties) on a concurrent panel is not a substitute for evidence of validity and reliability.²⁸¹ Three fingerprint examiners agreeing in a concurrent session does not somehow overcome the failure to validate techniques or determine the level of error. We should not forget that several examiners agreed with the result in each of the McKie and Mayfield misattributions.

Codes of conduct and professional ethics (and concurrent evidence) have some obvious value and potential, but they have not made a conspicuous difference to the performance of expert witnesses in criminal proceedings. Breaches are rarely disciplined and do not necessarily lead to the exclusion of evidence.²⁸² As with rules of admissibility, codes have been interpreted and applied loosely. They provide a basic normative framework where breaches are unlikely to lead to the exclusion of opinion evidence or even sanctions. This includes flagrant breaches by expert witnesses called by the prosecutor in criminal proceedings.²⁸³

H Trials (and Appeals) Are Not a Credible Test of Forensic Science Techniques

Given the preceding discussion, particularly express concerns about the historical weakness of legal proceedings; it probably goes without saying that the trial does not constitute a credible assessment of a technique.²⁸⁴ Accepting that occasionally a well-resourced defendant might be able to use safeguards to challenge evidence and generate doubts, in the vast majority of cases (ie quotidian trials) weaknesses and limitations are not identified or explored in detail during the course of proceedings.²⁸⁵ The fact that an expert or a particular technique has been previously admitted does not tell us anything about validity or reliability. Earlier decisions, sometimes

²⁸¹ In practice the defence may not have access to an expert and in many areas the state has an effective monopoly on current practitioners.

²⁸² This is also true for breaches of prosecutorial duties, they are hardly ever pursued. See Edmond, ‘(Ad)Ministering Justice’, above n 16.

²⁸³ See discussion in *Wood v The Queen* (2012) 84 NSWLR 581, 618 [723]ff.

²⁸⁴ NRC Report, above n 12, 42. There is a common belief among those in the forensic science and law enforcement communities that: ‘ability to withstand cross-examination in court when giving testimony related to these tests was sufficient to demonstrate the tests’ reliability’. For a notorious US example, see *United States v Haavard* 117 F Supp 2d 848 (SD Ind 2000).

²⁸⁵ Compare Sheila Jasanoff, *Science at the Bar: Law, Science and Technology* (Harvard University Press, 1995) and Gary Edmond, ‘Science in Court: Negotiating the Meaning of a “Scientific” Experiment During a Murder Trial and Some Limits to Legal Deconstruction for the Public Understanding of Law and Science’ (1998) 20 *Sydney Law Review* 361.

used as ‘precedent’, may have been decided under different admissibility rules and almost never involve serious review of the technique(s). In consequence, earlier jurisprudence and previous admissibility decisions might not be particularly useful guides for contemporary practice.²⁸⁶ Moreover, using prior decisions and convictions as admissibility heuristics is unlikely to prevent evidence ‘creep’ as experts extend the scope of claims derived from untested techniques once admitted; constrained only by the incredulity of technically-limited lawyers, judges and jurors.

The evaluation of techniques and proficiency of individuals should be carefully undertaken separate from legal proceedings; preferably independently and in circumstances where the correct answer (ie ‘ground truth’) is known.²⁸⁷ Formal evaluation should produce an indication of validity and reliability. Such studies help us to understand the scope of application, the limits and errors, as well as the kinds of things that are legitimate to say when reporting the results from a particular technique.²⁸⁸ Notably, trials and appeals provide none of these.²⁸⁹ Similarly, trials and appeals cannot sensibly address threats from human factors, other than to bluntly recognise their possibility though without providing a mechanism to gauge their impact (or substantially address the risks created) in the instant case.

I Experience and Peer Review: Necessary but Seldom Sufficient

The various reports recognise, both implicitly and explicitly, that the experience of an analyst is not capable of validating a technique. Experience alone does not provide a safeguard in relation to forensic science and medicine.²⁹⁰ Experience is necessary though almost never sufficient to ground admissibility.²⁹¹ Experience comes into its own through the application of validated techniques and methods.²⁹²

²⁸⁶ See, eg, Simon Cole, ‘Grandfathering Evidence: Fingerprint Admissibility Rulings from *Jennings* to *Llera Plaza* and Back Again’ (2004) 41 *American Criminal Law Review* 1189.

²⁸⁷ Although, studies that simply compare whether analysts use similar practices or produce similar results (technically reliability, but without validation) can be informative in identifying potential problems.

²⁸⁸ Interestingly, in *Daubert* the US Supreme Court indicated that these could be useful factors to consider; but they have not been taken seriously and applied by lower courts, especially in criminal proceedings.

²⁸⁹ We may be confident about convictions, but they very rarely afford certain evidence of guilt or a particular state of affairs. To the extent that lawyers and judges negotiate expressions of opinions, these tend to be speculative.

²⁹⁰ The reports were concerned about analysts expressing opinions about the significance of matches — particularly derived from latent fingerprints — on the basis of experience as opposed to probabilities derived from statistical analysis of data.

²⁹¹ Particularly where the techniques are, or it is envisaged that they will be, used routinely. Such techniques should be properly evaluated.

²⁹² Cf *R v Weller* [2010] EWCA Crim 1085 (4 March 2010).

Relying on experience tends to lead to bare assertions (so-called ipse dixit) that are difficult to challenge and seemingly persuasive. It tends to limit the potential for challenges to methods, ability and accuracy — and in a way that converts them to attacks on morality, character and credibility.²⁹³ Moreover, because courts have often admitted forensic science and medicine evidence that is basically experiential, they are not usually receptive to subsequent admissibility challenges where it is suggested that the same (or similar) techniques may be speculative or misguided.

Unwittingly, experience has become a problem for courts because of their inclination to use the long experience of forensic analysts as grounds for admission (and reliance). Experience is a convenient (and simple) heuristic that enables judges to defer to the accommodating decisions of earlier courts or the length of time a person (or institution) has been doing something, without ever having to consider validation studies, reliability and limitations. Many rules and statutes (eg Federal Rules of Evidence, r 702,²⁹⁴ Uniform Evidence Law, s 79) include ‘experience’ as a term that facilitates the admission of otherwise inadmissible opinions.²⁹⁵ Recourse to ‘experience’ could be limited, however, by attending to the need for ‘knowledge’ — and in Australia reading ‘reliability’ into ‘specialised knowledge’. Such an approach would require even those presenting opinions derived from their experience to present ‘good grounds’ — that is, demonstrative evidence — for believing that techniques and opinions are sufficiently reliable.²⁹⁶

Perhaps the take away message is to recognise that when it comes to forensic science and medicine — especially in response to forensic science techniques that are, or are likely to be, in routine use — experience (and long use) cannot support the weight of admissibility. Prosecutors and trial judges, as well as defence lawyers, are obliged to direct attention to formal evidence of reliability. A witness should not be able to vouch for her performance on the basis of long experience, previous admission or prior convictions. None of these provide meaningful evidence of ability or accuracy.

None of the reports proposed long experience as a viable exception to the formal evaluation of techniques. The limits of experience reinforce the primacy of reliability as a prerequisite for admissibility.

²⁹³ And, employment history rather than validation studies, error rates and proficiency.

²⁹⁴ *Federal Rules of Evidence* 28 USC r 702 (1975).

²⁹⁵ *Evidence Act 1995* (Cth); *Evidence Act 2011* (ACT); *Evidence Act 1995* (NSW); *Evidence (National Uniform Legislation) Act 2011* (NT); *Evidence Act 2001* (Tas); *Evidence Act 2008* (Vic).

²⁹⁶ In *Daubert*, the US Supreme Court explained that “‘knowledge’ connotes more than subjective belief or unsupported speculation’. The term ‘applies to any body of known facts or to any body of ideas inferred from such facts on good grounds’. Cf *Tuite v The Queen* [2015] VSCA 148 (12 June 2015).

Just as experience has serious limitations, so too peer review does not serve to validate a technique or result and is often quite a weak form of quality assurance.²⁹⁷ Weakness is accentuated where techniques have not been evaluated and where those undertaking review are not blinded to contextual information and the workings and conclusions before undertaking their review. In the circumstances leading to the Scottish Fingerprint Inquiry, quite elaborate processes of review did not identify or address problems and even discrepancies. Similarly, review did not prevent the mistaken identification of Mayfield by the FBI. Indeed, in the Scottish case the fingerprint bureau was organised in such a way that disagreement could be effectively circumvented, and subsequently concealed, by obtaining alternative reviewers and not disclosing discordant opinions.

J Wrongful Convictions and Substantially Unfair Proceedings

On the basis of the reports, it is my contention that a great deal of evidence has been misleadingly presented in investigations, plea negotiations, trials and appeals as reliable forensic science or expressed in terms that do not reflect the actual limitations of the underlying techniques.²⁹⁸ Many trials will have been rendered substantially unfair on that basis, and in some proportion of trials the defendant will have lost the benefit of reasonable doubt because mistakes, techniques not being reliable, exaggeration and misrepresentation, technical limitations, and inattention to cognitive contamination, were not identified or explained. In some trials the incriminating forensic science and medical evidence — derived from techniques that are yet to be evaluated, using processes that were inattentive to cognitive risks — will have simply been wrong.

In making this claim about the near universal misuse of forensic science and medicine evidence, it is not my intention to suggest that all or even most of these convictions are mistaken. In the vast majority of cases we do not know for certain that a particular person is guilty. Even so, many past convictions were compelling without forensic science and medicine evidence, and sometimes weak forensic science and medicine evidence may have contributed to compelling cases. The concern is that people have been convicted in circumstances where real limitations with evidence were not disclosed and where there were real dangers that evidence was cross-contaminated and trial judges, appellate judges and jurors were not genuinely alive to these significant threats to proof.

²⁹⁷ Goudge Report, above n 12, 334. Second opinions are often presented as ‘peer review’. See also SFI Report above n 12; and Gary Edmond, ‘Judging the Scientific and Medical Literature: Some Legal Implications of Changes to Biomedical Research and Publication’ (2008) 28 *Oxford Journal of Legal Studies* 523.

²⁹⁸ See for example, National Research Council, *Forensic Analysis Weighing Bullet Lead Evidence* (National Academies Press, 2004) and National Research Council, *On the Theory and Practice of Voice Identification* (National Academies Press, 1979). Interestingly, the FBI reviewed all the cases where there had been misrepresentations of bullet lead evidence. Contrast the (non)responses to the use of bullet lead, voice spectroscopy, microscopic hair comparison and other problematic forensic science evidence in Australia.

Those who insist on being shown wrongful convictions are, to some extent, asking the wrong question. The more appropriate and fundamental question for a criminal justice system that takes itself and its central precepts seriously is to ask: ‘Has incriminating expert opinion evidence been presented accurately and transparently (and in a manner that is comprehensible to the fact-finder)?’ In a large proportion of cases, both past and present, the answer must be ‘No’. Incriminating expert evidence continues to be presented without conveying *known* limitations, uncertainties and risks.

The fact that we have not uncovered many wrongful convictions, in systems that sometimes maintain strong, if seemingly irrational, commitments to their performance and the effectiveness of protections, might not be considered particularly surprising. Alternatively, it might be that Australian criminal justice systems are better than those in the United States, the United Kingdom and Canada at not convicting the innocent. That, however, needs to be demonstrated rather than asserted. Recent successful appeals in *Morgan*, *Gilham*, *Wood*, *Honeysett*, *Eastman* and *Keogh* all raise damaging questions about forensic science and medicine evidence as well as the ability of trials, and criminal justice personnel, to identify weak performances and suspect expert evidence.²⁹⁹

On the subject of uncovering wrongful convictions, involving the misuse or misrepresentation of forensic science evidence, it is interesting to consider the Expert Working Group’s response:

the chances of uncovering an erroneous identification are remote. Most fingerprint identifications are not challenged in court either because the defendant pled to some other charge or because the defense did not obtain a second opinion. Further, after conviction, the opportunities for innocent persons to obtain new evidence and have their convictions reviewed and overturned are still extremely rare.³⁰⁰

Undoubtedly influenced by the difficulties encountered by US-based innocence projects overturning convictions through exonerating DNA, the Working Group recognized just how difficult it is to excavate mistakes and address errors retrospectively. Such insights have an obvious salience to both the McKie and Mayfield cases, because the controversy that emerged from each of these incidents was highly unusual. It seems very likely that it was the unusual case circumstances, rather than the presence of exceptional errors, that led to exposure and notoriety.

²⁹⁹ *Morgan v The Queen* (2011) 215 A Crim R 33; *Gilham v The Queen* (2012) 224 A Crim R 22; *Wood v The Queen* (2012) 84 NSWLR 581; B Martin, Inquiry into the Conviction of David Harold Eastman for the Murder of Colin Stanley Winchester, *Report of the Board of Inquiry* (2014); *Eastman v DPP (ACT) [No 2]* [2014] ACTSCFC 2 (22 August 2014); *Honeysett v The Queen* (2014) 253 CLR 122; *R v Keogh [No 2]* (2014) 121 SASR 307. See also the review of Kathleen Folbigg’s conviction by Emma Cunliffe in *Murder, Medicine and Motherhood* (Hart Publishing, 2011).

³⁰⁰ NIST Report, above n 12, 33.

Unlike the vast majority of individuals who are charged and convicted in criminal justice systems, McKie and Mayfield had resources, supporters and the benefit of external intervention. As a policewoman and lawyer, respectively, they began with levels of social capital and resources that are atypical. Moreover, McKie's father was a senior police officer and the family eventually obtained the assistance of a non-aligned fingerprint examiner from the US who disagreed with the report from the Scottish bureau. In Mayfield's case, Spanish investigators formally rejected the FBI's attribution. They preferred the 'match' with a known terrorist produced by the Spanish National Police. It was in these circumstances that two positive identifications were contested and eventually rejected. In most criminal investigations there is no alternative authority and no *correct* answer. So, unless the suspect is well resourced or problems that occur 'backstage' are disclosed (or emerge), it can be very difficult to effectively challenge incriminating forensic science evidence — particularly the opinions of very experienced (and ostensibly impartial) analysts from prestigious institutions even when they are exaggerated or mistaken.

K. Judicial Notice, Exogenous Knowledge and Engagement (After Aytugrul)

One of the implications flowing from the non-disclosure by forensic analysts — and the incredibly circuitous and capricious manner in which research and knowledge gets before our courts — is that senior Australian judges have not been systematically exposed and are not, therefore, in a position to respond to the issues raised in the various reports. The upshot is that traditional — that is, deferential and accommodating — approaches toward the institutionalised forensic sciences and medicine have been perpetuated by judges, prosecutors and even defence lawyers notwithstanding unprecedented and unanswered criticism and recommendations for reform. Very few Australian courts have made references to the kinds of issues that need to be addressed. Australian courts are yet to reference any of the reports, despite daily reliance on DNA, latent fingerprint, voice, image, ballistic, tool and bite marks, blood spatter, shoe, foot and tire, and many other forms of incriminating evidence. This suggests that existing legal rules and safeguards might be quite limited in their ability to identify and facilitate systematic responses to serious and pervasive problems.

The conspicuous failure to recognise the deep organisational and epistemic problems with many areas of forensic science and medicine would seem to suggest that lawyers and legal institutions need to develop new means of accessing advice on what is known beyond the courts — particularly what is known about the forensic sciences. The fact that lawyers and, in consequence, judges in Australia seem to be largely oblivious to the issues raised in the reports and have yet to engage with the implications in their jurisprudence, means that there are disconcerting discontinuities and ruptures between legal knowledge and practice and *what is known*.

Common law courts have traditionally celebrated their focus on the individual case and ability to adapt, but confronted with systemic problems and issues beyond the experience of judges this orientation may actually be disabling. Insensitive to problems in a core area of proof, appellate courts and trial judges have tended to rehearse longstanding confidence in trial safeguards and legal platitudes while,

perhaps unknowingly, discounting the ever-growing body of critical evidence.³⁰¹ In relation to forensic science and medicine, there is a conspicuous need to develop better-informed processes and systems that help lawyers, judges and experts respond to the range of issues affecting their ability to understand and regulate expert evidence.

The recent High Court decision in *Aytugrul v The Queen*³⁰² reinforces the orthodox expectation that issues such as juror comprehension and reliability (and even legislative facts) should be raised at trial. It reinforces the reluctance of appellate judges to take judicial notice of, or to even consider, issues not raised at trial.³⁰³ This is unfortunate because defence lawyers are not generally resourced or particularly well positioned to raise complex methodological (or policy) issues at trial. Many defence lawyers seem reluctant to insist on reliability before trial judges operating in interpretive traditions that have been insensitive to the actual abilities and accuracy of *experts*. Whatever the precise reasons, defence lawyers (and prosecutorial obligations) have not exposed fundamental and pervasive problems across the forensic sciences and medicine. It seems that we need to consider reform, such as lower thresholds for judicial notice (though retaining scope for the parties to respond), broader scope for the intervention of *amicus curiae*, and perhaps independent advisory panels able to synthesise relevant research in order to provide authoritative, independent advice to lawyers and judges in areas of controversy.³⁰⁴

Regardless of the precise mechanism(s), lawyers and judges must be more attentive to external developments and much more attentive to the performance of their own institutions. There is conspicuous need for more evidence-based practice rather than adherence to tradition in the face of historical failure. Collective legal experience and institutional myths (however sincerely held), should be tempered with independent evidence.³⁰⁵

³⁰¹ Consider the response by J D Heydon, ‘Evidence of Forensic Scientific Opinion and the Rules for Admissibility’ (2015) 36 *Adelaide Law Review* 101.

³⁰² (2012) 247 CLR 170 (*Aytugrul*).

³⁰³ The recent decision in *Aytugrul* is emblematic in this respect. See also Gary Edmond, ‘Bacon’s Chickens? Re-thinking Law and Science (and Incriminating Expert Opinion Evidence) in Response to Empirical Evidence and Legal Principle’ in Justin T Gleeson and Ruth C A Higgins (eds), *Constituting Law: Legal Argument and Social Values* (Federation Press, 2011) 137.

³⁰⁴ See Gary Edmond, ‘Advice for the Courts? Sufficiently Reliable Assistance with Forensic Science and Medicine (Part 2)’ (2012) 16 *The International Journal of Evidence & Proof* 263.

³⁰⁵ It is no coincidence that in failing to attend to the performance of forensic sciences, legal institutions have not generally been particularly interested in their own performances, other than by the crude metrics of disposition under the guise of efficiency.

IV CONCLUSION: WHERE IGNORANCE ISN'T BLISS, 'TIS FOLLY TO BE UNWISE

On the whole the reports find that many forensic science and medicine techniques are not based on independent scientific research. Many techniques and assumptions have never been evaluated and many practitioners have never had their abilities credibly assessed. Surprisingly few forensic analysts know about, or are trained to deal with, issues and factors that may dramatically influence their performances and the reliability of results. They have been especially inattentive to threats to interpretation and the expression of results.³⁰⁶ Many practices and techniques are not standardised, and/or described in sufficient detail, to produce consistent results between analysts. Standards are often imprecise and not based on research. This has meant that — where they exist — accreditation, certification and regulation are generally weaker than they ought to be. The reports are, in addition, generally critical of the proximity, sometimes dependence, of forensic science and medicine on law enforcement.³⁰⁷ One of the main reasons for this state of affairs is the limited funding available for research. Other problems include the lack of formal qualifications among forensic analysts, inadequate scientific leadership and very limited historical contact between forensic analysts and relevant scholarly communities. A very significant factor is the limited scrutiny provided in legal settings. Accommodating legal institutions have effectively circumvented the need for research and attention to reliability while simultaneously providing access and prematurely (or improperly) conferring socio-epistemic legitimacy.

Australian forensic science and medicine and Australian legal institutions need to respond to the serious problems infecting forensic science and medicine evidence, regardless of what comparator countries do. This will require additional funds and new forms of independent leadership for the forensic sciences, preferably at a national level. Simultaneously, there are serious implications for law and legal practice. It requires senior judges (and possibly legislatures) to modify rules, practices, assumptions and interpretations. For, too much insufficiently reliable forensic science and medicine evidence is relied upon in charge and plea negotiations and admitted in Australian criminal proceedings.

Australian trials and appeals have not identified the profound weaknesses in many forms of forensic science and medicine. The vast majority of cases do not provide an adequate assessment of incriminating opinion evidence. They do not place decision makers in an appropriate position to rule on admissibility or rationally determine weight. As this essay has explained, there are very serious challenges confronting

³⁰⁶ See Dawn McQuiston-Surrett and Michael J Saks, 'The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Factfinders Hear' (2009) 33 *Law and Human Behavior* 436; Martire et al, 'The Psychology of Interpreting Expert Evaluative Opinions', above n 119.

³⁰⁷ The reports do not attempt to demonise individual failings and performances. Several make it clear that problems tend to be systemic — generally not caused by 'bad apples'.



not only the forensic sciences in Australia, but also legal practitioners (particularly prosecutors) and judges. While there have been some early efforts in relation to accreditation, the development of standards and the fostering of research in Australia, these are modest and in many cases accreditation and standards do not have the necessary underlying research to instill confidence.

What should lawyers and judges do in response to continuing proffers of incriminating expert opinion evidence? First, they should be willing to ask questions and exclude evidence. More broadly, in consultation with independent multidisciplinary advisory groups, they should begin to experiment with new procedures that are more conducive to the longstanding goals of doing justice in the pursuit of truth. Whatever happens, it is unlikely that this will be a simple or short-lived process. For, we need to develop legal institutions and identify personnel capable of informing legal practice with the knowledge and skills available in other domains. Consequently, there is a need for ongoing review and empirical assessment of legal practice and processes. There are, as the reports all imply, dangers — both epistemic and socio-political — in allowing courts and their ‘handmaidens’ to continue to operate in ways that are detached from what is known — or understood about knowing — in mainstream scientific communities.³⁰⁸

With limited historical interest in the reliability of forensic science and medicine evidence, Australian courts have gradually and unwittingly placed themselves in a state of epistemic bliss. Perhaps a distant prospect, our hope is that ‘Thought would destroy their paradise.’³⁰⁹

³⁰⁸ NRC Report, above n 12, 52: ‘many forensic fields (eg firearms analysis, latent fingerprint identification) are but handmaidens of the legal system, and they have no significant uses beyond law enforcement.’ The relationship might actually be uxorious.

³⁰⁹ Thomas Gray, *An Ode on a Distant Prospect of Eton College* (London 1747).

