

THE 'SCIENCE' OF MISCARRIAGES OF JUSTICE

GARY EDMOND*

I FORENSICS ON TRIAL

Many, perhaps most, wrongful convictions involve forensic scientific and medical evidence that was exaggerated, misleading or simply mistaken.¹ Many ordinary criminal proceedings are substantially unfair, some are no doubt miscarriages of justice, because forensic science and medicine evidence that is not known to be reliable is admitted and relied upon. For a variety of reasons serious limitations with expert evidence seem to be infrequently identified, explained or conveyed. Those responsible for admissibility, evaluating the evidence, or reviewing the evidence and the safety of the conviction on appeal, are not necessarily apprised of serious limitations and methodological heterodoxy. This is an appalling vista.² This article endeavours to provide some explanation for why so much unreliable, speculative and weak expert evidence is relied upon in criminal proceedings and how we might refine our practices to reduce the volume of such evidence and, simultaneously, the number of miscarriages of justice. This article suggests that among the most useful reforms that could be imposed is the imposition of a reliability standard to regulate the admission and enhance the evaluation of incriminating expert opinion evidence.³

This article draws on the experience of past miscarriages of justice, royal commissions and criminal case review commissions, the experience of innocence projects, as well as extensive commentary on forensic science and medicine

* Professor, Australian Research Council Future Fellow, and Director, Expertise, Evidence and Law Program, Faculty of Law, The University of New South Wales, Sydney 2052, Australia, email: g.edmond@unsw.edu.au; Professor, School of Law, Northumbria University; and Visiting Professor, Peter Wall Institute of Advanced Studies, University of British Columbia. This research was supported by the Australian Research Council (FT0992041 and LP120100063). Thanks to Emma Cunliffe, Mehera San Roque, David Hamer, Andrew Ligertwood and referees for their comments.

1 See, eg, *Morgan v The Queen* (2011) 215 A Crim R 33 ('Morgan'); *Wood v The Queen* [2012] NSWCCA 21 ('Wood'); *Gilham v The Queen* (2012) 224 A Crim R 22 ('Gilham').

2 Cf *McIlkenny v Chief Constable of the West Midlands* [1980] 1 QB 283.

3 For reasons that will become obvious, this article focuses exclusively on expert evidence adduced by prosecutors (ie, on behalf of the Crown or state). When speaking about reliability this essay is referring to the trustworthiness of evidence. 'Reliability' also possesses a technical, or scientific, meaning that is far more constrained.

evidence developed over several decades.⁴ It also draws on my own experience observing trials and appeals, reading expert reports, transcripts and judgments across several domestic and foreign jurisdictions, as well as my participation on a national standards committee.⁵ While these insights are important, they supplement a confronting recent development: reviews and reports by peak scientific and technical bodies. Over the last five years there have been several important inquiries into the forensic sciences. The resulting reports characterise 'law's greatest dilemma in its heavy reliance on forensic evidence' as 'the question of whether – and to what extent – there is *science* in any given forensic science discipline.'⁶ Here, in the context of a discussion of forensic science evidence and miscarriages of justice, it is my intention to raise two closely related issues that are central to the question of the need for a reliability-based admissibility standard.

The first is that many of the forensic sciences do not appear to be demonstrably reliable. That is, we do not know whether many of the techniques used in criminal investigations and relied upon in criminal prosecutions actually work (and, compounding things, some of the reliable techniques are not always performed in ways that are conducive to reliable outcomes). These disturbing findings are briefly discussed in Part II. Secondly, Australian lawyers and judges did not, in some kind of endogenous manner, recognise or develop their procedures and jurisprudence in ways that identified or provided credible responses to the concerns expressed in the recent reports. To put this more

-
- 4 United Kingdom, Report into the Circumstances Surrounding the Conviction Arising out of the Bomb Attacks in Guildford and Woolwich in 1974, *Final Report* (1994); United Kingdom, Royal Commission on Criminal Justice, *Report* (1993); Northern Territory, Royal Commission of Inquiry into the Chamberlain Convictions, *Report of the Commissioner* (1987); Manitoba, Commission of Inquiry into Certain Aspects of the Trial and Conviction of James Driskell, *Report* (2007); Ontario, Commission on Proceedings Involving Guy Paul Morin, *Report* (1998); *R v Truscott* (2007) 50 CR (6th) 1 (Ontario Court of Appeal). For critical commentary, see Clive Walker and Keir Starmer (eds), *Miscarriages of Justice: A Review of Justice in Error* (Oxford University Press, 1999); Richard Nobles and David Schiff, *Understanding Miscarriages of Justice: Law, the Media and the Inevitability of a Crisis* (Oxford University Press, 2000); Michael Naughton, *Rethinking Miscarriages of Justice: Beyond the Tip of the Iceberg* (Palgrave Macmillan, 2012); Brandon L Garrett, *Convicting the Innocent: Where Criminal Prosecutions Go Wrong* (Harvard University Press, 2011); Michael J Saks and Jonathan J Koehler, 'The Coming Paradigm Shift in Forensic Identification Science' (2005) 309 *Science* 892; Margaret A Berger, 'What Has a Decade of *Daubert* Wrought?' (2005) 95 *American Journal of Public Health* 59; Peter J Neufeld, 'The (Near) Irrelevance of *Daubert* to Criminal Justice and Some Suggestions for Reform' (2005) 95 *American Journal of Public Health* 107; Erica Beecher-Monas, *Evaluating Scientific Evidence: An Interdisciplinary Framework for Intellectual Due Process* (Cambridge University Press, 2007).
 - 5 See, eg, Gary Edmond et al, 'Admissibility Compared: The Reception of Incriminating Expert Evidence (ie, Forensic Science) In Four Adversarial Jurisdictions' (2013) 3 *University of Denver Criminal Law Review* 31; 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) forthcoming *Jurimetrics Journal*. The author is a member of the Forensic Analysis CH-041 Committee of Standards Australia.
 - 6 Committee on Identifying the Needs of the Forensic Science Community, 'Strengthening Forensic Science in the United States: A Path Forward' (Report, National Academy of Sciences, 2009) 9 (emphasis in original) ('*NAS Report*').

starkly, our trial system – with its oft-valourised values, rules, safeguards and experienced personnel – neither recognised nor exposed fundamental and widespread problems with many techniques and a great deal of forensic science evidence. Rather, persisting with longstanding and antiquated approaches to expertise and fairness, our courts have tended to discount or overlook serious problems. Trial judges have been overly accommodating in the admission of incriminating expert evidence and appellate courts have, in general, been insensitive to problems and dangers with a great deal of contemporary forensic science and forensic medicine. The failure to recognise endemic limitations and react systematically has, ironically, contributed to the continuing, and perhaps expanding, use of unreliable, weak and speculative incriminating expert evidence. Simultaneously, it has rendered courts oblivious to the depth and prevalence of evidentiary problems, the erosion of the standard of proof and, quite likely, the prevalence of miscarriages of justice.

II WHAT DO WE NOW KNOW ABOUT THE FORENSIC SCIENCES?

To support the need for a reliability standard, this article draws on three recent reports that provide unprecedented insight into the condition of the forensic sciences. The most important, most wide-ranging and most critical of these reports was based on an inquiry by a committee of the United States (‘US’) National Academy of Sciences (‘NAS’). Over two years, this formidable multidisciplinary committee received submissions, held hearings and made inquiries. The result of that process was the *NAS Report*.⁷ The *Report* captures the committee’s surprise at what it uncovered: many of the forensic sciences appear to lack scientific foundations. The *Report* is, in response, highly critical of a great deal of contemporary forensic science (and medicine) evidence and practice. What follows draws heavily on this seminal report, particularly the findings and recommendations – see Boxes 1 and 2.

Box 1: *NAS Report* – Findings

Validation of techniques

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. ... there is a notable dearth of peer-reviewed, published studies establishing the scientific bases and validity of many forensic methods.⁸

Little rigorous systematic research has been done to validate the basic premises and techniques ... The committee sees no evident reason why conducting such research is not feasible ...⁹

7 For a detailed review of this and other recent reports, see Gary Edmond, ‘What Lawyers Should Know about the Forensic “Sciences”’ (2014) forthcoming *Adelaide Law Review*.

8 *NAS Report*, above n 6, 8.

9 *Ibid* 189.

Standards

Often there are no standard protocols governing forensic practice in a given discipline. And, even when protocols are in place ... they often are vague and not enforced in any meaningful way. ... These shortcomings obviously pose a continuing and serious threat to the quality and credibility of forensic science practice.¹⁰

Accuracy and uncertainty

Few forensic science methods have developed adequate measures of the accuracy of inferences made by forensic scientists. 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.¹¹

Terminology and expression

[M]any terms are used by forensic examiners in reports and in court testimony ... Such terms include ... 'match', 'consistent with', 'identical', 'similar in all respects tested', and 'cannot be excluded as the source of.' ... 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 ... and the corresponding limitations in interpreting the results of forensic analyses.¹²

Contextual bias (and human factors)

Some initial and striking research has uncovered the effects of some [cognitive and contextual] biases in forensic science procedures, but much more must be done to understand the sources of bias and to develop countermeasures. ... The forensic science disciplines are just beginning to become aware of contextual bias and the dangers it poses. The traps created by such biases can be very subtle, and typically one is not aware that his or her judgment is being affected.¹³

Following the inauguration of the NAS inquiry, the US National Institute of Standards and Technology ('NIST') and the National Institute of Justice ('NIJ') jointly sponsored a review of latent fingerprint evidence, particularly the role of 'human factors'.¹⁴ A large multidisciplinary committee, the Expert Working Group on Human Factors produced a report that identified many areas of practice that require attention and reform. At about the same time that the NIST/NIJ review was established in the US, a senior Scottish judge was conducting a public inquiry into fingerprint evidence in Scotland.¹⁵ Endorsing and affirming many of the findings and recommendations made by the NAS committee, these reports were critical of the way a potentially valuable technique had been processed and reported in ways that were (and remain) insensitive to limitations and inattentive to a range of substantial risks.

10 Ibid 6.

11 Ibid 184.

12 Ibid 185–6.

13 Ibid 184–5.

14 Expert Working Group on Human Factors in Latent Print Analysis, 'Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach' (Report, US National Institute of Standards and Technology and US National Institute of Justice, 17 February 2012) ('*NIST/NIJ Report*').

15 Anthony Campbell, 'The Fingerprint Inquiry Report' (Report, The Fingerprint Inquiry Scotland, December 2011) vol 1 ('*Scottish Fingerprint Inquiry Report*').

What did the reviews uncover that so unsettled the scientists, engineers, statisticians and judges involved?¹⁶ In summary, the reports concluded that many forensic science techniques in routine use had not been validated. This means that we do not know whether the techniques actually work and cannot be confident about the scope of any claimed abilities.¹⁷ Moreover, standards are often vague, unenforced or not derived from empirical studies.¹⁸ In addition, relatively few forensic scientists are shielded from domain irrelevant information. This is significant because several prominent experimental studies have confirmed the corrosive impact of contextual influences (eg, suggestion) on human practices, particularly interpretation – see ‘Contextual bias’ in Boxes 1, 2, and 3.¹⁹

The implication from the NAS and other reports is that many forensic science and medicine techniques are of unknown value. In many cases we do not know if analysts can actually do the things that courts routinely enable them to testify about. And, if they can do them, we do not have a clear idea about how accurate they are, or the limits of any abilities. This led to a series of substantial recommendations for reform – see Box 2.

Box 2: NAS Report – Recommendations

Recommendation 1: 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).²⁰

Recommendation 2: 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.

-
- 16 See Harry T Edwards, ‘Solving the Problems that Plague the Forensic Science Community’ (2009) 50 *Jurimetrics* 5; Harry T Edwards, ‘The National Academy of Sciences Report on Forensic Sciences: What It Means for the Bench and Bar’ (2010) 51 *Jurimetrics* 1.
 - 17 It is important to acknowledge that many of the techniques are probably reliable. Unfortunately, in the absence of studies we do not know how probative (or error prone) the techniques actually are. And, surprisingly few techniques are designed or conducted in ways that reduce threats from contextual bias and cross-contamination.
 - 18 Standards have not, for example, led latent fingerprint examiners to agree on whether a latent print was suitable for comparison or whether particular minutiae were similar, or how much similarity was sufficient for a ‘match’: see Box 3; 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.
 - 19 Itiel E Dror, David Charlton and Ailsa E Péron, ‘Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications’ (2006) 156 *Forensic Science International* 74. See generally D Michael 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.
 - 20 *NAS Report*, above n 6, 19. The National Institute recommended in the *NAS Report* bears limited resemblance to the Australian NIFS. The Australian NIFS is a very small institution that is funded by state and territory police commissioners and has very few research scientists among the small number of staff.

Similarly, it should establish model laboratory reports for different forensic science disciplines and specify the minimum information that should be included.²¹

Recommendation 3: Research is needed to address issues of accuracy, reliability, and validity in the forensic science disciplines. The National Institute of Forensic Science (NIFS) should competitively fund peer-reviewed research in the following areas:

- (a) Studies establishing the scientific bases demonstrating the validity of forensic methods.
- (b) The development and establishment of quantifiable measures of the reliability and accuracy of forensic analyses. ... The research by which measures of reliability and accuracy are determined should be peer reviewed and published in respected scientific journals.
- (c) The development of quantifiable measures of uncertainty in the conclusions of forensic analyses.²²

Recommendation 4: 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.²³

Recommendation 5: [E]ncourage research programs on human observer bias and sources of human error in forensic examinations.²⁴

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.²⁵

Recommendation 7: Laboratory accreditation and individual certification of forensic science professionals should be mandatory, and all forensic science professionals should have access to a certification process.²⁶

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.²⁷

Recommendation 9: The National Institute of Forensic Science (NIFS), in consultation with its advisory board, should establish a national code of ethics for all forensic science disciplines and encourage individual societies to incorporate this national code as part of their professional code of ethics.²⁸

The reports suggest that forensic science evidence adduced by the state is often presented such that the claimed probative value extends beyond what is

21 Ibid 22.

22 Ibid 22–3.

23 Ibid 24.

24 Ibid.

25 Ibid 24–5.

26 Ibid 25.

27 Ibid 26.

28 Ibid.

known and, therefore, appropriate. Two hard cases exemplify this tendency. The first involves latent fingerprint evidence and the second concerns DNA profiling.

In the wake of notorious misattributions, the *NIST/NIJ* and the *Scottish Fingerprint Inquiry Reports* identified numerous aspects of the practices around latent fingerprint comparison and reporting that required research and/or reform. This is interesting, because the first independent validation studies (reported in 2011) confirmed that latent fingerprint examiners possess genuine expertise, when it comes to comparing prints.²⁹ Nevertheless, both reports concluded that fingerprint examiners continue to make stronger claims than their techniques and abilities can sustain – see Box 3. All of the reports caution against positively identifying persons on the basis of latent fingerprint comparisons, insist that examiners should not suggest that their method is infallible, and should respond to the danger of cognitive bias. It is unclear how the abilities of latent print examiners relate to the question of identity, and both fingerprint reports stress the need to attend to standards guiding sufficiency assessments, matches, and how these should be expressed in expert reports and testimony.³⁰ The fingerprint reports also suggest that latent fingerprint evidence should be probabilistic rather than categorical (ie, match, non-match and inconclusive).

Box 3: Findings and Recommendations on Latent Fingerprint Evidence

NAS Report

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 of fingerprinting and the lack of validated standards for declaring a match, such claims of absolute, certain confidence in identification are unjustified. ... Claims of 'absolute' and 'positive' identification should be replaced by more modest claims about the meaning and significance of a 'match'.³¹

Scottish Fingerprint Inquiry Report

Recommendation 2: Examiners should receive training which emphasises that their findings are based on their personal opinion; and that this opinion is influenced by the quality of the materials that are examined, their ability to observe detail in mark and print reliably, the subjective interpretation of observed characteristics, the cogency of explanations for any differences and the subjective view of 'sufficiency'.³²

-
- 29 The first validation studies for latent fingerprint comparisons were only published in 2011, more than a century after fingerprint evidence was accepted in many common law courts. See Jason M Tangen, Matthew B Thompson and Duncan J McCarthy, 'Identifying Fingerprint Expertise' (2011) 22 *Psychological Science* 995; Bradford T 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.
- 30 That is, conditions in which a print is deemed adequate for analysis/comparison and how much detail is required to support a 'match' (as opposed to exclusion), as well as what a 'match' means in terms of identification.
- 31 *NAS Report*, above n 6, 106 fn 79, quoting Jennifer L Mnookin, 'The Validity of Latent Fingerprint Identification: Confessions of a Fingerprinting Moderate' (2008) 7 *Law, Probability and Risk* 127, 139.
- 32 *Scottish Fingerprint Inquiry Report*, above n 15, 741.

Recommendation 3: Examiners should discontinue reporting conclusions on identification or exclusion with a claim to 100% certainty or on any other basis suggesting that fingerprint evidence is infallible.³³

Recommendation 6: The Scottish Police Services Authority should review its procedures to reduce the risk of contextual bias.³⁴

NIST/NIJ Report

Recommendation 3.3: Procedures should be implemented to protect examiners from exposure to extraneous (domain-irrelevant) information in a case.³⁵

Recommendation 3.4: Each agency or forensic service provider should define 'suitable' or 'sufficient' in its standard operations procedures. These guidelines should be as explicit as possible about what is expected for sufficiency determinations at different stages of the latent print examination process.³⁶

Recommendation 3.7: Because empirical evidence and statistical reasoning do not support a source attribution to the exclusion of all other individuals in the world, latent print examiners should not report or testify, directly or by implication, to a source attribution to the exclusion of all others in the world.³⁷

Recommendation 3.9: The federal government should support a research program that aims to:

- (a) Develop measures and metrics relevant to the analysis of latent prints;
- (b) Use such metrics to assess the reproducibility, reliability, and validity of various interpretive stages of latent print analysis; and
- (c) Identify key factors related to variations in performance of latent print examiners during the interpretation process.³⁸

Recommendation 6.3: A testifying expert should be familiar with the literature related to error rates. A testifying expert should be prepared to describe the steps taken in the examination process to reduce the risk of observational and judgmental error. The expert should not state that errors are inherently impossible or that a method inherently has a zero error rate.³⁹

Notwithstanding the reports and their recommendations, latent fingerprint examiners in the US, Scotland and Australia continue to report their conclusions in terms that are logically indefensible and beyond what their techniques can sustain. That is, they purport to make positive identifications (or individualisations) – excluding every other person, whether they state it or not, on the planet. The recommendations confirm that fingerprint evidence is presented in terms that are not empirically warranted. Latent fingerprint evidence is, therefore, systematically exaggerated in reports and courtroom testimony. Moreover, risks are not taken seriously enough. Latent fingerprint examiners in

33 Ibid.

34 Ibid.

35 *NIST/NIJ Report*, above n 14, 44.

36 Ibid 54.

37 Ibid 72.

38 Ibid 76.

39 Ibid 127.

Australia are routinely, though unnecessarily, exposed to domain irrelevant information. And, they never incorporate real risks of error into their reports and testimony.⁴⁰

Even the value of DNA evidence is routinely exaggerated. While analysts sometimes round down the ‘denominator’ so that the random match probability or likelihood ratio is purportedly ‘conservative’ (very often just higher than the number of people currently alive on earth – eg, 1×10^{10}), complex and difficult interpretative choices are not always conceded or explained.⁴¹ More problematically, the real risk of error (particularly human error in collection, continuity and interpretation) tends to be systematically omitted from reports and testimony.⁴² Consequently, the defence must somehow identify errors and limitations, in retrospect and in circumstances where they were not present during the collection, transportation, storage and handling, analysis, preparation, testing, interpretation and report writing.⁴³ After the fact, the accused is left to somehow identify errors that were not picked up by investigators and analysts.⁴⁴

These examples indicate how techniques that are basically reliable (ie, have been validated, such as DNA analysis and, more recently, latent fingerprint comparison) may encounter problems. There are even more serious problems where the value of techniques and abilities are uncertain. Such problems appear to be widespread, and presumably run much deeper among the many forensic science and medicine techniques that have not been independently evaluated – eg, ballistics and tool marks, shoe, tyre and foot prints, image, voice, bite mark and gait comparison, and so on. The state should not obtain an evidentiary or procedural benefit from the failure of its forensic analysts to evaluate techniques or shield themselves from notorious risks. Such oversights should not be left to ‘weight’. The failure to report in empirically predicated terms, and the failure to unilaterally provide information on uncertainty and error, distort the meaning of forensic science evidence and subvert rational fact-finding. The failure to reform practice and reporting also reveals a remarkable intransigence in the face of a chorus of criticisms and recommendations by authoritative scientific bodies.

40 See Gary Edmond, Matthew B Thompson and Jason M Tangen, ‘A Guide to Interpreting Forensic Testimony: Scientific Approaches to Fingerprint Evidence’ (2014) 13 *Law, Probability and Risk* 1.

41 Notably, problems created by mixed or very small samples. See Itiel E Dror and Greg Hampikian, ‘Subjectivity and Bias in Forensic DNA Mixture Interpretation’ (2011) 51 *Science and Justice* 204; Jonathan J Koehler, ‘The Psychology of Numbers in the Courtroom: How to Make DNA-Match Statistics Seem Impressive or Insufficient’ (2001) 74 *Southern California Law Review* 1275. See also David H Kaye, *The Double Helix and the Law of Evidence* (Harvard University Press, 2010); Michael Lynch et al, *Truth Machine: The Contentious History of DNA Fingerprinting* (University of Chicago Press, 2009).

42 On error in DNA profiling (in Germany), see S Rand et al, ‘The GEDNAP Blind Trial Concept Part II: Trends and Developments’ (2004) 118 *International Journal of Legal Medicine* 83.

43 The case of Farah Jama indicates how difficult this can be. See Victoria, Inquiry into the Circumstances That Led to the Conviction of Mr Farah Abdulkadir Jama, *Report* (2010).

44 The fact that responses to serious dangers and risks of error are not necessarily embodied in standard operating procedures, protocols and standards makes it more difficult for the defence to persuade courts and fact-finders of their significance.

III CHANGING HEURISTICS: WHAT SHOULD AN ADMISSIBILITY STANDARD LOOK LIKE?

Before explaining what an appropriate admissibility standard should look like, it is useful to contrast conventional Australian admissibility criteria with factors that are prominent in recent reviews of the forensic sciences by scientific organisations and commissioners.

At common law and under the uniform evidence law ('UEL'), when considering the admissibility of expert evidence Australian judges tend to focus on the existence of a 'field' and whether the witness is an 'expert' (usually determined by training, study and/or experience).⁴⁵ There is, in addition, an expectation that the evidence will assist the jury, although there is no formal requirement that the expert evidence must be beyond 'common knowledge'.⁴⁶ At common law and under the UEL, the expert's opinion should be based on identified facts and assumptions, preferably on admissible evidence.⁴⁷ Some judges also require the basis of the opinion, or reasoning, to be transparent.⁴⁸ Ordinarily, the reliability of expert opinion evidence is considered to be an issue of weight for the tribunal of fact.⁴⁹

The text of section 79(1) of the UEL requires that the opinion should be based on 'specialised knowledge' based on 'training, study or experience'. In practice, prosecutors and judges have demonstrated little interest in developing the meaning of 'specialised knowledge'.⁵⁰ Drawing on their common law experience they have substituted the easier, though frequently deceptive, question of whether the witness has formal training, study or experience that is, or appears, relevant. In consequence, few criminal decisions extend beyond a superficial parsing of 'training, study or experience'.

45 See, eg, *Clark v Ryan* (1960) 103 CLR 486; *R v Bonython* (1984) 38 SASR 45. Most Australian jurisdictions, namely the Commonwealth (ie, federal courts), New South Wales, the Australian Capital Territory, Tasmania, Victoria and the Northern Territory, now largely use UEL: see *Evidence Act 1995* (Cth); *Evidence Act 1995* (NSW); *Evidence Act 2011* (ACT); *Evidence Act 2001* (Tas); *Evidence Act 2008* (Vic); *Evidence Act 2013* (NT). The other jurisdictions, South Australia, Western Australia and Queensland, use common law and domestic statutes.

46 See, eg, UEL s 80.

47 See *Makita (Australia) Pty Ltd v Sprowles* (2001) 52 NSWLR 705 ('*Makita*'); *Dasreef Pty Ltd v Hawchar* (2011) 243 CLR 588 ('*Dasreef*').

48 *HG v The Queen* (1999) 197 CLR 414 ('*HG*'); *Dasreef* (2011) 243 CLR 588.

49 Recently, in *Dupas v The Queen* (2012) 218 A Crim R 507, the Court of Appeal in Victoria suggested that judges ought to consider the reliability of expert evidence when undertaking the balancing exercise mandated by s 137 of the UEL. This approach is not consistent with practice in NSW: *R v XY* [2013] NSWCCA 121. See Gary Edmond et al, '*Christie*, Section 137 and Forensic Science Evidence (after *Dupas v The Queen* and *R v XY*)' (2014) forthcoming *Monash Law Review*. The Canadians consider reliability when evaluating probative value: see, eg, *R v Abbey* (2009) 97 OR (3d) 330, [87] (Doherty JA) (Ontario Court of Appeal); *R v Aitken* (2012) 92 CR (6th) 384, [78] (Hall JA) (British Columbia Court of Appeal) ('*Aitken*').

50 The main exceptions are *Makita* (2001) 52 NSWLR 705, 743–4 [85] (Heydon JA) and *R v Tang* (2006) 65 NSWLR 681 ('*Tang*'). 'Scientific knowledge' is often overlooked or, somewhat tautologically, equated with 'training, study or experience'.

In the leading decision in New South Wales, the Court of Criminal Appeal explained that '[i]n the immediate context of "specialised knowledge" ... the word "knowledge" connotes more than subjective belief or unsupported speculation.'⁵¹ At the same time, and somewhat inconsistently, the Court disavowed the need for attention to reliability: the focus of attention must be on the meaning of the statutory phrase 'specialised knowledge', rather than by invoking extraneous ideas such as 'reliability'.⁵² This approach displaced New South Wales from the international convergence around reliability standards, persuasive US Supreme Court jurisprudence on the meaning of the word 'knowledge' (see Part IV), and the more recent advice of the NAS and NIST.

At this point, it is useful to contrast our legal categories with criteria characterised by the NAS and others as fundamental.

Table 1: Criteria for Assessing (the Admission of) Expert Evidence

Australian Courts	NAS
[Legally recognised] 'field'	Independent (pre-litigation) validation studies
[Apparent] expertise	Error rates and uncertainties disclosed
Practical experience [untested]	Rigorous proficiency testing
Recognition by other courts	Supported by peer reviewed publication (persuasive to multidisciplinary audiences) ⁵³
Previous involvement in investigations and proceedings	Probabilistic expression of results
[Apparent] impartiality	Processes to remove contextual bias and errors created by human factors (and the need to document exposure to domain irrelevant information)
Science/non-science	Independence from law enforcement
'[S]pecialised knowledge'	Detailed and transparent reports
Novelty/longevity of practice	Standards developed and applied
Assist the jury	Standardised expressions (preferably probabilistic expressions derived from formal studies)
General acceptance [in a particular 'field']	Full disclosure

The differences between these columns are stark and revealing. Generally, the factors identified by the NAS are readily identifiable. Most have an independent existence beyond the criminal justice system. We can, for example, determine whether validation studies have been conducted. Validation

51 *Tang* (2006) 65 NSWLR 681, 712 [138] (Spigelman CJ), quoting *Daubert v Merrell Dow Pharmaceuticals Inc*, 509 US 579, 590 (Blackmun J) (1993).

52 *Tang* (2006) 65 NSWLR 681, 712 [137] (Spigelman CJ).

53 Revealingly, there is no extensive collection of studies, such as the Cochrane Collaboration database of clinical trials and epidemiological studies developed for medical practice and research, for the forensic sciences. This is, in part, a result of limited research funding and the non-scientific origins of many forensic 'science' disciplines.

studies are important because they guarantee that the technique actually does what it is supposed to and in what conditions. They also enable us to construct error rates that provide an indication of how well the technique works. In most cases we can assess the proficiency of analysts using validated techniques. We can determine what information is required for analysis and ascertain whether the analyst was unnecessarily exposed to domain irrelevant information. We can examine expert reports to determine whether there is sufficient detail to enable an independent scientist (let alone a defence lawyer) to understand the process and review the findings.⁵⁴ We can determine whether conclusions and expressions are based on empirical studies or are intuitive or speculative.

By comparison, the legal categories (in the left column of Table 1) tend to be more opaque and reveal little about the value of techniques and actual abilities of analysts – eg, the existence of a 'field', the untested experience of a forensic analyst, and previous appearances in courts. They do not, for example, provide evidence of actual expertise. Similarly, the existence of a 'field' is declaratory and superficial. Does it really matter if there is a 'field'? The demonstrated abilities of analysts supported through validation studies tend to be far more important and far more elucidatory. Moreover, categories such as a 'field' (and 'general acceptance') are rarely considered in any detail. When it comes to 'facial mapping' or 'forensic gait analysis', for example, courts do not undertake a survey and lawyers do not provide the materials for them to do so. Rather, judges make a determination (really a declaration) about whether some set of practices constitutes a 'field', usually based on whether the analyst has qualifications in an apparently related domain such as anatomy or podiatry.⁵⁵ 'Facial mapping', as just one example, does not appear to exist beyond the courtroom.⁵⁶

There is no substantial jurisprudence on the meaning of 'field' and frustratingly little on 'specialised knowledge'. Formal qualifications, like previous experience doing something, do not necessarily provide credible evidence of expertise. They do not always engage with abilities, accuracy, standards or the limits of techniques and expertise. Continuing with the previous example, formal training and experience as an anatomist or podiatrist does not obviously enable the analyst to reliably identify features or make comparisons from CCTV and other surveillance images. These 'skills', substantially different to the everyday practices and experience of anatomists and podiatrists, should be formally evaluated.⁵⁷ Characterising disparate sets of assumptions, practices and techniques as a 'field' circumvents inquiry into the fundamental questions of ability and accuracy. That is, does the analyst possess

54 See, eg, *Pan Pharmaceuticals Ltd (in liq) v Selim* [2008] FCA 416.

55 See, eg, Gary Edmond et al, 'Law's Looking Glass: Expert Identification Evidence Derived from Photographic and Video Images' (2009) 20 *Current Issues in Criminal Justice* 337.

56 Cf *R v Dastagir* (2013) 224 A Crim R 570; *Tang* (2006) 65 NSWLR 681; *Morgan* (2011) 215 A Crim R 33; *Honeysett v The Queen* [2013] NSWCCA 135; *Murdoch v The Queen* (2007) 167 A Crim R 329.

57 Even if they have experience in investigations, unless we assess performance in controlled conditions (against 'ground truth') we do not know their abilities or the value of any experience.

relevant expertise? Whereas the criteria recommended by the NAS and other attentive groups focus attention on independent evidence of expertise, legal decision-making creates (or reifies) some of the very categories (eg, 'field' and previous admission) used to support admissibility.

On reflection, superficial attention to the existence of a 'field' has been a particularly unfortunate criterion, particularly in relation to emerging techniques. It perpetuates a credibility cycle, where earlier legal recognition in Australia (or elsewhere) confirms the existence of the 'field'. Discredited techniques, such as voice spectroscopy, bullet lead analysis and the use of bite marks as identification evidence all have limited, if any, probative value.⁵⁸ Yet, courts in the US and Australia admitted and relied upon such untested techniques, in many instances for decades, before those beyond the courts persuaded investigative institutions (and occasionally courts) to modify their practices. Revealingly, in these examples, courts drew upon linguistics, chemistry and dentistry as 'fields' capable of supporting practices that had not been validated and were found to be error prone.

It is no coincidence that many forensic analysts invoke legal recognition, previous involvement in investigations and earlier appearances in courts as support for their *expertise*. They also place great store in the fact that they have been doing things for years or perhaps decades even though many techniques and capabilities have never been formally evaluated. The lack of formal evaluation means that opinions are routinely expressed in terms that were described by the NAS and NIST as inappropriate or 'implausible', and dangers created through gratuitous exposure to domain irrelevant information, institutional arrangements and speculative procedures are largely ignored (by those ignorant of, or indifferent to, the very real risks). These are all problems that have been facilitated by legal (mis)understandings and premature legal recognition.

Inattention to reliability is all the more surprising because 'specialised knowledge' (from section 79(1) of the UEL) might – and arguably should – be read in a way that requires lawyers and judges to attend to the kinds of criteria proposed by NAS. 'Knowledge' was read in this way by the US Supreme Court in *Daubert v Merrell Dow Pharmaceuticals Inc*:

The adjective 'scientific' implies a grounding in the methods and procedures of science. Similarly, the word '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 or accepted as truths on good grounds.' Webster's Third New International Dictionary 1252 (1986).⁵⁹

And in *Kumho Tire Co Ltd v Carmichael*:

In *Daubert*, the Court specified that it is the Rule's word 'knowledge,' not the

58 See, eg, US National Research Council, *On the Theory and Practice of Voice Identification* (NAS, 1979); Erica Beecher-Monas, 'Reality Bites: The Illusion of Science in Bite-Mark Evidence' (2009) 30 *Cardozo Law Review* 1369.

59 509 US 579, 590 (Blackmun J) (1993) ('*Daubert*'). Ironically, this definition was endorsed in *Tang* (2006) 65 NSWLR 681, 713 [139] (Spigelman CJ).

words (like 'scientific') that modify that word, that 'establishes a standard of evidentiary reliability.' Hence, as a matter of language, the Rule applies its reliability standard to all 'scientific,' 'technical,' or 'other specialized' matters within its scope.⁶⁰

Australian courts have not read the need for reliability into 'specialised knowledge'. Rather, judges in UEL jurisdictions have tended to persist with their common law categories and effectively ignored 'specialised knowledge'. Inattention to 'knowledge' means that most prosecutors and judges proceed as if section 79(1) merely requires an opinion to be based on 'training, study or experience'. The few courts that have gone further have suggested that section 79 requires attention to the form in which the expert evidence is presented in order to enable evaluation.⁶¹ If we consider Table 1, the *NAS Report* implies that for forensic science evidence to count as 'specialised knowledge' it requires attention to the features in the right-hand column rather than the left. It is important to recognise that the NAS did not regard the existence of 'fields', 'experience', or legal recognition as particularly informative. On partisanship, the NAS proposed removing forensic science from the control (and funding) of law enforcement – Recommendations 1 and 4 in Box 2.

What should we require from forensic scientists?⁶² According to the various reports and extensive scholarly commentary, Box 4 describes useful factors for determining the admissibility and weight of forensic science and medicine evidence.

Box 4: More Meaningful Admissibility Criteria for the Forensic Sciences

For techniques in regular use, especially those techniques purporting to provide 'identification' or comparison evidence, the NAS and other reports suggest the need to direct attention to:

Validation. Validation indicates whether a technique does what it purports to and in what conditions. There should be independent validation or rigorous proficiency studies published in a publicly accessible peer reviewed repository. Any validation studies should be relevant to the application used in the specific case.

Error rate. An indication of limitations or uncertainties and indicative error rates derived from validation and/or rigorous proficiency studies. This should be reported with the opinion and will assist those obliged to evaluate the evidence to determine its probative value.

Standards. A description/reference to the relevant standards (or protocol) and whether they were applied.

Expertise (or proficiency). Evidence that the analyst is qualified/certified and experienced with the validated technique. Ideally, this evidence should be based on rigorous, independent proficiency studies.

Contextual bias. The analyst should not be exposed to domain irrelevant information or placed in conditions

60 526 US 137, 147 (Breyer J) (1999) ('*Kumho*').

61 See *HG* (1999) 197 CLR 414; *Dasreef* (2011) 243 CLR 588.

62 For a more detailed account, 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) 190.

that might adversely impact on the analysis. Information provided to the analyst should be controlled and documented.

Expression of results. The opinion should be expressed in terms that are empirically-predicated (ie, supported by the validation studies).

Report. The report should be transparent and provide sufficient detail to enable an independent scientist to understand what was done. It should include references to supporting and critical literatures.

Here I am describing what we should expect in relation to forensic science and medical evidence derived from techniques in routine use or likely to be used regularly. This includes DNA profiling; latent fingerprint evidence; ballistics; image comparison (eg, face and/or body mapping and gait analysis); voice comparison; document; bite mark and handwriting evidence, and so on. On the basis of the NAS and other reports, our courts should be focused on whether a specific technique has been validated and whether the individual(s) are proficient and experienced (eg, formally certified) with the particular technique and operating in an accredited laboratory.⁶³ Ordinarily, validation studies should be independent assessments. The technique and the analysts should be assessed in conditions where the correct answer (ie, ground truth) is known. Standards should be derived from validation and other studies. Courts should be attentive to the existence and application of standards. Courts should be keenly interested in the limits of techniques as well as error rates. Not only should an indicative error rate be included in the expert report and testimony, but appellate courts should offer guidance on the kinds of error rates that the legal system is willing to tolerate. Generally, techniques with high error rates should not be admitted. Would we, for example, want to rely on techniques that had an error rate of 20 or 40 per cent (ie, were wrong one or two times out of five)? Results should be reported in terms that reflect the known capabilities and limits of the technique (and analyst). There should be a strong preference for forms of expression that are probabilistic – derived from validation studies and/or databases – rather than impressionistic or speculative.⁶⁴

Courts must be extremely cautious about allowing forensic analysts to express opinions solely or primarily on the basis of their experience, even if their evidence is tempered (as in *Tang*).⁶⁵ In most cases experience should be derived using techniques that are demonstrably reliable. The value of experience obtained

63 Bryan Found and Gary Edmond, 'Reporting on the Comparison and Interpretation of Pattern Evidence: Recommendations for Forensic Specialists' (2012) 44 *Australian Journal of Forensic Sciences* 193.

64 At this stage verbal formulations seem problematic and large numbers appear to create difficulties. See Kristy A Martire, Richard I Kemp and Ben R Newell, 'The Psychology of Interpreting Expert Evaluative Opinions' (2013) 45 *Australian Journal of Forensic Sciences* 305. Of interest, Justice McClellan's approach, if not his unilateral research, in *Aytugrul v The Queen* (2010) 205 A Crim R 157, 173–6 [86]–[95], appears sound. Cf *Aytugrul v The Queen* (2012) 247 CLR 170.

65 See, eg, Daniel Kahneman, Paul Slovic and Amos Tversky (eds), *Judgment under Uncertainty: Heuristics and Biases* (Cambridge University Press, 1982); Daniel Kahneman and Gary Klein, 'Conditions for Intuitive Expertise: A Failure to Disagree' (2009) 64 *American Psychologist* 515.

using techniques that have not been validated is uncertain. Real risks of error and real limitations may not be known, disclosed or considered. Such risks may be far from trivial or theoretical. Moreover, they may not be obvious and may even appear counterintuitive. Courts should also require documentation about what the analyst was told and in most cases be willing to exclude expert evidence where the analyst was unnecessarily exposed to suggestive information. Expert reports should include much more information about what was done, the information disclosed to the analyst (formally or informally), the interpretive processes, the basis for the particular interpretation and expression as well as limits of the process, relevant literatures and critical perspectives. Experts should not come to court without disclosing the existence of criticisms of their current practices, especially those produced by peak scientific organisations such as the NAS. Such non-disclosure by prosecutors and experts breaches respective duties to the court and the profession.⁶⁶ Without appropriate disclosure there is a much greater likelihood that incriminating expert evidence will be over-valued.

Obviously, there is a need for flexibility with admissibility standards and criteria. In introducing scope for discretion five issues deserve emphasis. First, judges should be cautious about admitting incriminating expert evidence where the underlying technique cannot be formally evaluated. Most forensic techniques, particularly those associated with linking a trace to a particular source or individual, are – as the *NAS Report* makes clear (see Box 1 'Validation of techniques') – susceptible to testing.⁶⁷ The fact that formal evaluation has not been undertaken should sound alarms. It should not be excused in order to expedite the admission of untested (and therefore speculative) techniques into criminal proceedings even if the opinions are superficially plausible and/or qualified – ie, read down.⁶⁸

Secondly, recognition of the need for flexibility is not intended to encourage the creation of exceptions and inconsistencies that circumvent the primary need for formal evaluation. Rather, trial and appellate judges should carefully consider appropriate reliability criteria (always sensitive to the way the evidence will be used in the case).⁶⁹ Most types of genuine expertise (including non-scientific forms of expertise) have a range of methodological canons, standards and accepted limitations. These should guide the admission, scope and use of incriminating expert evidence. Where formal validation is not possible, alternative criteria should be identified and applied.⁷⁰ Characterising *expert* evidence as non-scientific, technical or specialised does not provide a warrant to

66 Gary Edmond, '(Ad)ministering Justice: Expert Evidence and the Professional Responsibilities of Prosecutors' (2013) 36 *University of New South Wales Law Journal* 921, 931–7.

67 See, eg, *R v Li* (2003) 140 A Crim R 288; *R v Atkins* [2010] 1 Cr App R 8 ('*Atkins*'); *Aitken* (2012) 92 CR (6th) 384 (British Columbia Court of Appeal).

68 *Tang* (2006) 65 NSWLR 681. *Contra* Simon A Cole, 'Splitting Hairs? Evaluating "Split Testimony" as an Approach to the Problem of Forensic Expert Evidence' (2011) 33 *Sydney Law Review* 459.

69 In *Kumho*, Scalia J insisted that 'choosing the manner of testing expert reliability – is not discretion to abandon the gatekeeping function': 526 US 137, 158–9 (1999).

70 See Edmond, 'Pathological Science?', above n 62.

ignore the need for validation or other indicia of reliability. Judges should be reluctant to admit forensic science evidence, adduced by the prosecutor, that depends exclusively or primarily on the untested experience and abilities of forensic scientists or other individuals – whether investigators or highly credentialed academics.⁷¹ They should be extremely anxious about *ipse dixit*.

Thirdly, as a general rule, the admissibility of expert evidence should be considered independently from the other evidence. It does not matter if there is compelling evidence or the case otherwise appears strong. If a technique – especially a technique in routine use – has not been formally evaluated then in most cases the evidence should be excluded until it has been.

Fourthly, acknowledging the failure to have undertaken appropriate studies should not provide a back door to the courtroom.⁷² Being aware of oversights and limitations – ie, knowing what should have been done – does not provide means to rationally evaluate techniques and derivative opinions that have not been properly evaluated.

Finally, judges should direct attention to the information provided to the analyst, especially if the information is not required for their analysis. Certain kinds of information have a demonstrated tendency to mislead and so, as a general rule, exposure to gratuitous information should lead to the exclusion of derivative opinions. Another analyst can, after all, be invited to undertake the same analysis in conditions that are not suggestive or otherwise vulnerable to cognitive contamination.

Whether NAS-style admissibility criteria should be applied symmetrically – to evidence adduced by the defendant – remains an open question. Principle might be invoked to suggest that the accused should be entitled to adduce expert evidence derived from techniques that are less well established.⁷³ Regardless, the imposition of a reliability standard that is symmetrical between the parties would, if enforced, reduce substantial unfairness and the risk of miscarriages of justice. The defence would no longer be obliged to persuade the jury – against appearances, legal institutional endorsement and at considerable risk to the innocent accused – that many forensic science techniques in routine use are of unknown value, though likely to be less probative than is suggested by the prosecutor and forensic analyst at trial.

Any admissibility standard needs to attend to the value of the expert evidence as well as the effectiveness of the trial and appellate processes at handling (incriminating) expert evidence – see Part V. In terms of the effectiveness of trial and appellate processes, we need to consider the likelihood that limitations will be *exposed, explained and understood* by judges and the tribunal of fact. Evidence must be presented in a manner that is conducive to comprehension. It must be very likely that the expert evidence and its limitations will be

71 See Gary Edmond and Mehera San Roque, 'Quasi-Justice: Ad Hoc Expertise and Identification Evidence' (2009) 33 *Criminal Law Journal* 8.

72 See, eg, *Atkins* [2010] 1 Cr App R 8.

73 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, 406–8.

understood, otherwise the system cannot claim to be rational or fair. As well as probative value, attention should be directed at the ability of the defence to effectively counter expert evidence; the not unrelated issues of the extent to which the evidence might descend into a protracted contest where the jury will be obliged to try to resolve fundamental disagreements (eg, around the reality of expertise); as well as the cost and fairness to the accused, the jury and the public of such 'battles'.⁷⁴ The value of transparent practices predicated upon validated techniques should be self-evident.

In the absence of information about the value of techniques and derivative opinions, lay jurors and judges have been required to determine the weight of expert evidence on the basis of whatever transpires at trial. Allowing jurors to assign probative value to forensic science evidence in the absence of formal assessment is not a rational approach to expert evidence and proof. We have required lay individuals to determine the value of expert evidence without providing the kind of information necessary to make sense of it. In most cases, jurors are left to assess the value of incriminating expert evidence on the basis of the performance of the 'expert' witness at trial: how they perform in credibility-oriented cross-examination, where emphasis is placed on experience and previous performances rather than demonstrable expertise.⁷⁵

Judges should be very reluctant to allow criminal trials to descend into arguments about whether fields exist, how scientific they are, whether practices are testable and whether untested techniques might work and how well. These issues should be determined prior to trials and, if necessary, explored on the voir dire. Insufficiently reliable expert evidence should be excluded. Trials, even trials that pay close attention to NAS (and *Daubert*) style criteria, are not capable of validating a technique. Similarly, disclosing the failure to formally evaluate a technique does not provide grounds for rational evaluation, just as 'peer review' does not make a difference where the validity of the technique remains uncertain.⁷⁶

Many institutional insights and advantages are lost through legal (and especially judicial) inattention to the reliability of forensic science techniques. During investigations, false leads are as likely to be confirmed as contradicted when investigators place (excessive) reliance on expert opinions derived from unreliable, weak or speculative techniques. In plea negotiations, trials and

74 These concerns are not new: see Learned Hand, 'Historical and Practical Considerations Regarding Expert Testimony' (1901) 15 *Harvard Law Review* 40.

75 While sociological and psychological studies suggest that in some circumstances motivated citizens can perform reasonably well in their assessment of scientific and technical evidence, trials do not generally provide the sorts of conditions that are conducive to lay evaluation. See generally Alan Irwin and Brian Wynne (eds), *Misunderstanding Science? The Public Reconstruction of Science and Technology* (Cambridge University Press, 1996) and the journal *Public Understanding of Science*.

76 In cases in Australia, in the absence of formal validation some analysts have appealed to 'peer review' by colleagues, friends and family members. In Canada, one analyst recently made reference to self-verification. See *R v Alrekabi* (2007) 4 DCLR (NSW) 292; *Aitken* (2012) 92 CR (6th) 384 (British Columbia Court of Appeal).

appeals, directing insufficient attention to the reliability of techniques means that unreliable, weak and speculative opinions are frequently treated as valuable and corroborated (or corroborating). In many cases trials are substantially unfair because prosecutors adduce and represent forensic science and medicine evidence as more probative than it is known to be. In some cases the absence of formal evaluation means that traditional claims by forensic analysts are perpetuated in courts and believed by prosecutors, judges and defence lawyers (and perhaps jurors) even though they are wrong. Important legal principles and aspirations – around the primacy of truth and justice – have been compromised by inattention to the reliability of forensic science and medicine at the point of admission.

No Australian court requires that incriminating opinions, represented by the prosecution as expert, be demonstrably reliable.⁷⁷ No Australian court has imposed a reliability threshold in its admissibility jurisprudence. The kinds of criteria advanced by peak scientific and technical advisory bodies do not govern admissibility and all too often do not form part of the evidentiary record in criminal proceedings. This is a deplorable state of affairs. Criminal justice practice is displaced from its own aspirations and is not consistent with the independent advice of peak science.

Australian approaches to forensic science and medicine are a century out of date.⁷⁸ Our admissibility jurisprudence directs investigators to focus on secondary and tertiary features of evidence rather than the core issues. Instead of asking can the expert do what she claims and how do we know, attention is directed to whether there is a ‘field’ or another court has recognised the species of expertise and this (kind of) witness. This orientation is particularly unfortunate because many of the forensic ‘sciences’ emerged without sensitivity to scientific processes, values and insights.⁷⁹

IV ‘RELIABILITY’ IN OTHER COMMON LAW SYSTEMS

Other common law systems have already adopted reliability-based admissibility standards. The US Supreme Court, for example, called for ‘relevance and reliability’ when determining the admissibility of ‘scientific

77 There are occasional cases such as *Lewis v The Queen* (1987) 88 FLR 104 where, in the aftermath of *Chamberlain* case, the Court was concerned about the reliability of bite mark evidence provided by one of the British experts involved in the *Chamberlain* conviction.

78 Arrangements between legal institutions and forensic analysts have a curious similarity with the relations of gentlemanly trust and vouching described by Shapin and Schaffer in their study of the Boyle-Hobbes controversy in the 17th century: see Steven Shapin and Simon Schaffer, *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life* (Princeton University Press, 1985).

79 See Jennifer L Mnookin et al, ‘The Need for a Research Culture in the Forensic Sciences’ (2011) 58 *UCLA Law Review* 725.

knowledge' from rule 702 of the US *Federal Rules of Evidence* ('FRE')⁸⁰ in its *Daubert* decision:

But, in order to qualify as 'scientific knowledge,' an inference or assertion must be derived by the scientific method. Proposed testimony must be supported by appropriate validation – ie, 'good grounds,' based on what is known. In short, the requirement that an expert's testimony pertain to 'scientific knowledge' establishes a standard of evidentiary reliability.⁸¹

The Supreme Court proposed four (or five) criteria to assist trial judges with admissibility determinations. The Court explained:

[1] Ordinarily, a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact will be whether it can be (and has been) tested.

...

[2] Another pertinent consideration is whether the theory or technique has been subjected to peer review and publication. Publication (which is but one element of peer review) is not a *sine qua non* of admissibility; it does not necessarily correlate with reliability ... But submission to the scrutiny of the scientific community is a component of 'good science,' in part because it increases the likelihood that substantive flaws in methodology will be detected.

...

[3a] Additionally, in the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error, see, eg, *United States v Smith*, 869 F 2d 348, 353–4 (CA7 1989) (surveying studies of the error rate of spectrographic voice identification technique), and [3b] the existence and maintenance of standards controlling the technique's operation ...

[4] Finally, 'general acceptance' can yet have a bearing on the inquiry. A 'reliability assessment does not require, although it does permit, explicit identification of a relevant scientific community and an express determination of a particular degree of acceptance within that community.' ... Widespread acceptance can be an important factor in ruling particular evidence admissible ...⁸²

In *Kumho*, a subsequent appeal considering whether the *Daubert* criteria applied to 'technical and other specialised (ie, non-scientific) knowledge', also from rule 702 of the *FRE*, the Court affirmed the need to flexibly apply these and other criteria to guarantee reliability. In 2000, the text of rule 702 was redrafted to accommodate the *Daubert* and *Kumho* jurisprudence. It now states:

A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:

- (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
- (b) the testimony is based on sufficient facts or data;

80 Prior to an amendment in December 2000, rule 702 stated, '[i]f scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise.' The Committee on the Judiciary, The United States House of Representatives, *Federal Rules of Evidence* (2010) 14.

81 *Daubert*, 509 US 579, 590 (Blackmun J) (1993). *Contra Tang* (2006) 65 NSWLR 681.

82 *Daubert*, 509 US 579, 593–4 (Blackmun J) (1993).

- (c) the testimony is the product of reliable principles and methods; and
- (d) the expert has reliably applied the principles and methods to the facts of the case.⁸³

The Canadian Supreme Court followed the US lead when the common law categories derived from *R v Mohan*⁸⁴ were supplemented with the *Daubert* criteria in *R v J-LJ* and *R v DD*.⁸⁵ The need for attention to ‘reliability’ was reiterated in *R v Trochym*,⁸⁶ where the majority explained that problems with expert evidence should not be left for weight and concern with reliability should not be limited to novel techniques. The need for attention to the reliability of expert evidence was directly linked to the Canadian experience with wrongful convictions. Justice Deschamps explained:

In recent years, a number of public inquiries have highlighted the importance of safeguarding the criminal justice system – and protecting the accused who are tried under it – from the possibility of wrongful conviction ... ‘[t]he names of Marshall, Milgaard, Morin, Sophonow and Parsons signal prudence and caution in a murder case’ ... In the case at bar, we consider once again the need to carefully scrutinize evidence presented against an accused for reliability and prejudicial effect, and to ensure the basic fairness of the criminal process.⁸⁷

In *R v Abbey*, the Court of Appeal for Ontario explained the need to flexibly apply criteria so that non-scientific forms of expert evidence satisfy ‘threshold reliability’ before being admitted in criminal proceedings.⁸⁸

Finally, a recent report by the Law Commission of England and Wales concluded that the admission of expert evidence in English criminal proceedings was too liberal. The Commission expressed doubts about the effectiveness of trial safeguards (see also Part V):

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, by revealing to the jury factors adversely affecting reliability and weight. However, as we explained in our consultation paper, and repeat below, it is doubtful whether these are valid assumptions. A more credible assumption, at least in relation to complex scientific or technical fields, is that juries will often defer to the expert providing the opinion. If such an expert’s opinion evidence is unreliable, the dangers associated with deference are obvious, particularly if the opinion forms a critical link in the prosecution’s case.⁸⁹

83 *Federal Rules of Evidence*, Pub L No 93–595, §1, 88 Stat 1926 (1975).

84 *R v Mohan* [1994] 2 SCR 9 required: (a) relevance; (b) necessity in assisting jury; (c) no exclusionary rule; (d) properly qualified expert.

85 *R v DD* [2000] 2 SCR 275; *R v J-LJ* [2000] 2 SCR 600. The ‘admissibility of expert evidence’, Binnie J explained, ‘should be scrutinized at the time it is proffered, and not allowed too easy an entry on the basis that all of the frailties could go at the end of the day to weight rather than admissibility.’: *R v J-LJ* [2000] 2 SCR 600, 613 [28].

86 [2007] 1 SCR 239.

87 *R v Trochym* [2007] 1 SCR 239, 249 [1].

88 *R v Abbey* (2009) 97 OR (3d) 330. See also Stephen Goudge, Ontario, Inquiry into Pediatric Forensic Pathology in Ontario, *Report* (2008). The author was an adviser to the Goudge Inquiry.

89 United Kingdom Law Commission, *Expert Evidence in Criminal Proceedings in England and Wales*, Report No 325 (2011) 5 [1.20].

In response, the Commission produced a draft bill proposing a 'sufficiently reliable' admissibility standard and outlining the kinds of criteria that ought to guide the trial judge's assessment of reliability – see Box 5.

Box 5: Criminal Evidence (Experts) Bill (UK) clause 4

Reliability: meaning

- (1) Expert opinion evidence is sufficiently reliable to be admitted if –
 - (a) the opinion is soundly based, and
 - (b) the strength of the opinion is warranted having regard to the grounds on which it is based.
- (2) Any of the following, in particular, could provide a reason for determining that expert opinion evidence is not sufficiently reliable –
 - (a) the opinion is based on a hypothesis which has not been subjected to sufficient scrutiny (including, where appropriate, experimental or other testing), or which has failed to stand up to scrutiny;
 - (b) the opinion is based on an unjustifiable assumption;
 - (c) the opinion is based on flawed data;
 - (d) the opinion relies on an examination, technique, method or process which was not properly carried out or applied, or was not appropriate for use in the particular case;
 - (e) the opinion relies on an inference or conclusion which has not been properly reached.
- (3) When assessing the reliability of expert opinion evidence, the court must have regard to –
 - (a) such of the generic factors set out in Part 1 of the Schedule as appear to the court to be relevant;
 - (b) if any factors have been specified in an order made under Part 2 of the Schedule in relation to a particular field, such of those factors as appear to the court to be relevant;
 - (c) anything else which appears to the court to be relevant.⁹⁰

The influence of *Daubert* and the revised rule 702 on the Law Commission of England and Wales should be obvious.

Unfortunately, the Ministry of Justice considered that a new admissibility standard might create additional costs and so recommended that judges should, instead, be 'provided, at the initial stage, with more information about the expert evidence it is proposed to adduce.'⁹¹

⁹⁰ Ibid 148.

⁹¹ Ministry of Justice, Government of the United Kingdom, 'The Government's Response to the Law Commission Report: "Expert Evidence in Criminal Proceedings in England and Wales"' (Law Com Paper No 325, 21 November 2013) 4 [4]. How this will enable judges to make informed decisions, especially given the persistence of conventional (ie, common law) admissibility jurisprudence in England, is unclear. See Gary Edmond, 'Is Reliability Sufficient? The Law Commission and Expert Evidence in International and Interdisciplinary Perspective (Part 1)' (2012) 16 *International Journal of Evidence & Proof* 30.

V TRIAL ‘SAFEGUARDS’ AND THE PRIMACY OF ADMISSIBILITY

The emergence of the reports, ‘out of the blue’, after more than a hundred years of criminalistics and forensic science, suggests that there might be serious problems with the performance of trials and appeals (and the largely unexamined reliance on forensic science and medicine in plea and charge negotiations). Collectively, prosecutorial obligations, the provision of defence lawyers, admissibility standards, judicial discretions, scope for cross-examination, the occasional provision of defence (usually ‘rebuttal’) experts, warnings and directions, a high standard of proof and appellate review have not prevented miscarriages of justice or, more conspicuously, the continuing acceptance and reliance upon incriminating expert opinions of unknown value.

Attention to the reliability of incriminating expert evidence at the admissibility stage is vitally important because trial safeguards afford inconsistent protection against the many dangers with expert evidence and the influence of expert witnesses – including those using unreliable techniques and proffering misleading or mistaken opinions.⁹² Prosecutors do not generally focus on ‘reliability’ in detail when considering whether to adduce expert evidence. And, when they do adduce expert evidence, the disclosure of serious limitations is usually partial at best.⁹³ Cross-examination tends to be of limited value where defence lawyers are not resourced or prepared, sufficiently proficient or confident in their ability to persuade trial judges and jurors about the significance of dangers. In too many cases defence lawyers do not fully appreciate the significance of limitations and dangers. The defence calls rebuttal experts relatively infrequently.⁹⁴ Despite appearances, there is little evidence that these witnesses effectively identify and explain limitations or repair oversights even in defective expert evidence adduced by the state. Regardless, in most cases the forensic analyst’s response to cross-examination marks the limits of the evidence. Overall, the system appears to be symmetrical and therefore fair, but these appearances do not reflect quotidian realities.⁹⁵

Such observations might seem counterintuitive, but they make more sense when explained in a little detail. Expert witnesses called by the prosecutor are normally state-employed forensic analysts. Their involvement in the case is

92 Richard Kemp, Stephanie Heidecker and Nicola Johnston, ‘Identification of Suspects from Video: Facial Mapping Experts and the Impact of Their Evidence’ (Paper presented at the European Association of Psychology and Law, Maastricht, 2–5 July 2008).

93 This may be caused by ignorance or because prosecutors place exaggerated confidence in admissibility standards and trial safeguards. Nevertheless, institutional credibility and reward structures for prosecutors might not encourage them to challenge widespread commitments and the practical advantages conferred by weak admissibility standards.

94 At the recent NSW Fingerprint Conference on 5–6 November 2013 at HMAS Penguin, a NSW Supreme Court judge with 22 years experience as a barrister and two years sitting as a judge reported that he had never encountered a rebuttal expert in a case involving fingerprint evidence.

95 See, eg, 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.

routine and they usually proffer opinions that are consistent with guilt. They are not paid for their involvement in the particular case. They appear to have no obvious pecuniary interest in the outcome and are portrayed as impartial.⁹⁶ On average they are experienced, often through long 'apprenticeships' and participation in numerous investigations, and usually allude to their previous experience in their expert reports and in qualification during examination-in-chief. Serious limitations with their evidence (such as those identified by the NAS) will, in many cases, not be disclosed or explained. Even when limitations – including fundamental oversights and serious methodological flaws – are identified, it does not follow that they will be understood and credited by the tribunal of fact. Once admitted, incriminating expert opinions will often be refracted through the other incriminating evidence and the substantial experience of the analyst. No matter how weak, speculative or unreliable, they will be considered as part of the overall case against the accused and may be seen to corroborate other strands of evidence, even when the expert evidence was contaminated by the other evidence and is not, therefore, independent.⁹⁷ Again, such limitations are almost never identified and explained.⁹⁸

Rebuttal experts, in contrast, do not enter proceedings with the imprimatur of the state and the prosecutor. Not only are they paid for their appearance, often they are not experienced forensic analysts. Moreover, they generally do not provide positive evidence of innocence. Rather, they tend to criticise the assumptions, methods and interpretations of the highly experienced state-employed forensic analysts. Unlike the state's forensic analysts, rebuttal experts are usually drawn from different disciplines and often in possession of little, if any, forensic experience. Rebuttal experts are often academics and occasionally retired forensic analysts. They, along with their methodological criticisms, are easy to portray as abstract, theoretical and even partisan. These academicians are often unfavourably (and unfairly) juxtaposed to the experienced, and implicitly proven, forensic analysts called by the prosecutor.⁹⁹ Their contributions are often *made to appear* irrelevant to the exigencies of serious criminal investigations. In addition, the defence does not always propose an alternative narrative (beyond doubt or investigative mistakes) to bolster methodological criticisms and concerns raised by rebuttal experts.¹⁰⁰ Consequently, rebuttal expert evidence is

96 Although, there may well be institutional incentives based around remuneration and promotion. The NSW Police Force, for example, has just granted a substantial loading to analysts qualified to appear as expert witnesses.

97 Dan Simon, 'Pseudo-Corroborator' (Paper presented at the Second Annual Workshop on Law, Science, and Evidence, UCLA Law School, 5 June 2013); Gary Edmond et al, 'Contextual Bias and Cross-contamination in the Forensic Sciences: The Corrosive Implications for Investigations, Plea Bargains, Trials and Appeals' (2014) (unpublished, copy on file with author).

98 See Edmond et al, 'Contextual Bias and Cross-contamination', above n 97.

99 Notwithstanding appearances and legal representations, surprisingly often it is the defence expert who advances the more scientifically orthodox perspectives and insights.

100 While there is no formal obligation to do so as the state carries the burden of proof, failure to do so may place the accused at a distinct disadvantage.

often vulnerable to representation as implausible, partial, and inconsistent with the other strands of evidence assembled against the accused.

Directions and warnings are frequently invoked, and sometimes celebrated, as a means of managing unreliable, weak and speculative forms of expert evidence and any dangers that might arise from admission. In reality, directions and warnings appear to be weak and ineffective.¹⁰¹ Not only do studies suggest that they do not work, at least as intended, attention to the content of warnings suggest that they are often pale approximations of what is known (beyond the courts) about many forms of technical practice and the risks created by expert evidence. Most warnings and directions in relation to expert evidence are superficial. They do not, for example, raise or engage with the kinds of issues raised by the NAS and other attentive scientific and technical organisations – see Box 4 and Table 1.¹⁰² They do not provide useful insights or serviceable criteria for accurate decision-making. In some cases they are misguided, likely to mislead or focus unwanted attention on issues. Sometimes they are simply wrong. It is highly unlikely that directions and warnings about incriminating expert evidence correct or substantially address reliability issues.

Consequently, the tribunal of fact, whether a jury or trial judge (and courts of criminal appeal), is in the invidious position of having to assess incriminating expert evidence in circumstances where they are unlikely to appreciate, and unlikely to be told in detail and impartially, about serious methodological oversights, actual abilities, error rates and limitations, general indifference to contextual bias, the lack of standards, and what is actually known about the issue beyond the courtroom. Decision-makers (jurors, judges as well as prosecutors, defence lawyers and even investigators) are expected to make important decisions, that include the assessment of forensic science evidence, in circumstances where the most appropriate evaluative criteria are neither explained nor provided. Complicating decision-making, other incriminating evidence will tend to make forensic science techniques appear reliable even though the opinions of the forensic scientists might have been influenced through unnecessary exposure to this very evidence (and vice versa).¹⁰³ Evidentiary problems may be compounded where evidence is unwittingly counted twice (or more) and what appears to be independent corroboration is, in reality, pseudo-corroboration.¹⁰⁴

A little appreciated insight is the incredible difficulty of addressing, let alone correcting for, the admission of unreliable, weak and speculative expert evidence. Attention to the reliability of techniques and derivative opinions during

101 See Supreme Court of Victoria, ‘Simplification of Jury Directions Report’ (Report to the Jury Directions Advisory Group, August 2012); Jacqueline Horan, *Juries in the 21st Century* (Federation Press, 2012).

102 See, eg, Gary Edmond, Kristy Martire and Mehera San Roque, ‘Unsound Law: Issues with (“Expert”) Voice Comparison Evidence’ (2011) 35 *Melbourne University Law Review* 52.

103 A forensic analyst tasked with an interpretive exercise might know, eg, that the suspect made admissions and has a prior criminal record. In addition, the results of forensic analyses are often fed back to other witnesses, such as eyewitnesses and this has the undesirable effect of strengthening their confidence and belief in ways that have no relationship to accuracy.

104 Simon, above n 97.

admissibility decision-making is fundamental because unreliable, weak and speculative expert evidence is likely to contaminate the trial. It is very likely to be misrepresented, likely to be misunderstood and likely to be unfairly prejudicial to the accused.¹⁰⁵ In most appeals there tends to be considerable deference to jury fact-finding and no means of actually determining what the jury did in relation to the (expert) evidence. If the findings and recommendations of the NAS are correct, then most Australian trials have not and do not expose the kinds of limits and issues necessary to evaluate incriminating expert evidence. Judicial claims about trial safeguards are by and large an article of legal faith or, perhaps more appropriately, dogma.

The inability to consistently focus on the reliability of forensic science evidence means that there are probably many more miscarriages of justice and wrongful convictions in New South Wales (and Australia) than are exposed.¹⁰⁶ We should not forget that the judges who continue to use misguided admissibility criteria and champion ineffective safeguards decide whether there are problems with procedures and proof warranting review, a retrial or acquittal. The NAS and other reviews provide powerful evidence that our courts have been mistaken in both the choice of heuristics and their confidence. In many proceedings resulting in convictions our courts admitted incriminating opinions of unknown value. They allowed these to be presented as 'expert' even though there was no independent evidence of expertise. They also treated cross-contaminated evidence (ie, opinions contaminated by other evidence) as independent support for guilt.

The NAS and other reports provide the courts with an opportunity to break out of their internally referential (vicious) circle.¹⁰⁷ Australian courts should recognise that they are currently engaged in a highly misadvised activity, are contributing to the weakness of forensic science and medicine, are perpetuating rules that blind judges (and others) to real risks from unreliable, weak and speculative forensic science, and are failing to engage with scientific knowledge and authoritative advice (from beyond the criminal justice system and insular forensic 'science' communities).

105 Inattention to NAS-style criteria means that real dangers of unfair prejudice to the accused (UEL ss 135, 137) are systematically overlooked. The tribunal of fact is not, by way of example, provided with appropriate means of rationally evaluating the opinions of those characterised by prosecutors as 'experts'. Moreover, considerable emphasis tends to be placed on issues such as experience and confidence.

106 Obviously not every case with weak, speculative or unreliable expert evidence will be a wrongful conviction, but misrepresentation of expert evidence and heavy reliance on such evidence will have made quite a few convictions substantially unfair.

107 The critical perspectives find support in the work of criminal case review commissions, the work of innocence projects and the growing volume of critical scholarship. See, eg, 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; Mike Redmayne, *Expert Evidence and Criminal Justice* (Oxford University Press, 2001); Garrett, above n 4; Saks and Koehler, above n 4.

VI CONCLUSION: MORE RELIABLE JURISPRUDENCE AND APPROPRIATE HEURISTICS

The huge yellow somethings went unnoticed at Goonhilly, they passed over Cape Canaveral without a blip, Woomera and Jodrell Bank looked straight through them – which was a pity because it was exactly the sort of thing they'd been looking for all these years.¹⁰⁸

Australian admissibility standards, jurisprudence and legal practice bear a pitiful resemblance to the failure of terrestrial tracking stations to detect the spaceships sent to destroy Earth parodied by Douglas Adams. In application, exceptions to the opinion rule for those with 'specialised knowledge' have 'looked straight through' the failure to evaluate techniques, apply empirically-derived standards, report limitations, express opinions in justifiable terms and use processes that limited the threats from notorious cognitive dangers. In the cases where some of these kinds of issues were raised, they were generally left to weight – for a jury to evaluate in the midst of an adversarial criminal proceeding often without meaningful guidance about the significance of failures and oversights.

We should refine our admissibility standards for criminal proceedings to incorporate a formal reliability standard and the kinds of indicia described by the NAS and others. Regardless of what judges and forensic analysts have done or might contend, most forensic science techniques are amenable to some form of validation or proficiency assessment. We should expect to see and ask for independent studies of the technique and in many cases evidence of personal proficiency using validated techniques.¹⁰⁹ We should be concerned with protocols and standards and be alarmed by unnecessary exposure to domain irrelevant information. Failure to test techniques or avoid threats from contextual bias should suggest exclusion. For it is difficult, perhaps impossible, to rationally assess subjective opinions when we are not presented with evidence about the validity and reliability of the underlying techniques. Where such issues are not clearly raised and conveyed through prosecution disclosure, cross-examination (and rebuttal expert witnesses) or judicial instruction, the threat to the rational evaluation of the evidence is considerable. It is important to reiterate that recourse to experience, recognition of putative fields, and apparent necessity do not overcome fundamental epistemic problems.

In criminal proceedings our admissibility standards should stipulate the need for reliability. In UEL jurisdictions this could be read into the much-elided need for 'specialised knowledge' in section 79(1). Regardless of how it is done there is a need for change. There are serious and continuing dangers in persisting with practices that are insufficiently sensitive to the probative value of incriminating expert evidence. In most cases, especially where the opinion involves comparison

108 Douglas Adams, *The Hitchhiker's Guide to the Galaxy* (1979) chapter 3.

109 See, eg, Jonathan J Koehler, 'Fingerprint Error Rates and Proficiency Tests: What They Are and Why They Matter' (2008) 59 *Hastings Law Journal* 1077; Jonathan J Koehler, 'Proficiency Tests to Estimate Error Rates in the Forensic Sciences' (2013) 12 *Law, Probability and Risk* 89.

or 'identification' evidence (eg, from latent fingerprints; tyre; shoe and foot prints; images and voice recordings; documents and handwriting; hairs; ballistics; tool marks; blood spatter; bite marks and so on) we should expect to see the criteria set out in the *NAS Report* (and listed in right-hand column of Table 1) *satisfied*. Expert reports (or certificates) should explicitly address them.¹¹⁰ Where these criteria are not satisfied, certainly in cases involving 'identification' and comparison evidence, courts should not admit the incriminating expert evidence. For, as we have seen, admission lends premature legitimacy, misleads courts and fact-finders, discourages fundamental research, and contributes to the incidence of miscarriages of justice. Reliance on the wrong heuristics for determining both admissibility and weight has contributed not only to miscarriages of justice but also to the difficulty of retrospectively exposing them. Inappropriate heuristics thwart the primary goal of the accusatorial trial – namely 'doing justice in the pursuit of truth'.¹¹¹

Some of our admissibility rules and safeguards might be made to work. This, however, would require changes in the way prosecutors, trial and appellate judges, and defence lawyers approach the state's expert evidence in criminal proceedings.¹¹² Prosecutors should be much more cautious about the expert evidence they adduce. They should seek information from forensic analysts, more detailed reports and proactively jettison insufficiently reliable expert evidence. Where they adduce expert evidence they should disclose and explain limitations.¹¹³ Defence lawyers should challenge the admissibility of expert evidence – though, not necessarily with rebuttal witnesses – and provide trial and appellate judges with a record and materials enabling them to assess the reliability of the technique and opinion.¹¹⁴ Judges in particular should be familiar with recent critiques of the forensic sciences and understand why the concerns expressed in a growing number of independent reports require active responses.¹¹⁵ To successfully resist insufficiently reliable expert evidence and its insidious effects on the rational administration of justice there is a need for much greater awareness of problems with the forensic sciences and for more active

110 Found and Edmond, above n 63. See also the more limited expectations from the influential *Ikarian Reefer* litigation: *National Justice Compania Naviera SA v Prudential Assurance Co Ltd* [No 1] [1993] 2 Lloyd's Rep 68.

111 Hock Lai Ho, *A Philosophy of Evidence Law: Justice in the Search for Truth* (Oxford University Press, 2008) 54.

112 Defence lawyers need to be more proactive, but most of the changes need to come initially from prosecutors and judges.

113 See the discussion based on prosecutorial obligations in Edmond, '(Ad)ministering Justice', above n 66.

114 Perversely, engaging rebuttal experts may make expert witnesses called by the Crown – whose expertise and techniques have not been established experimentally – appear to be part of an established field and may make the trial appear fair – via the symmetry afforded to the defence – even though it does not address fundamental epistemic issues. Taking advice from those with methodological and technical insights might be more useful for understanding reliability and challenging admissibility than actually calling witnesses. This does, however, convert the reliability question into a tactical issue for the defence lawyer and accused. See, eg, *R v Dastagir* (2013) 224 A Crim R 570; *Morgan* (2011) 215 A Crim R 33; *Honeysett v The Queen* [2013] NSWCCA 135.

115 See, eg, *R v Borneyk* [2013] BCSC 1927 (British Columbia Supreme Court).

judicial gatekeeping. Simultaneously, courts (and this means judges) must be much more attentive to the performance of legal institutions, particularly around the use of expert evidence in charge negotiations, trials and appeals. Trial and appellate judges should not invoke platitudes about the effectiveness of trial safeguards, the increasing sophistication of the modern jury, and the collective wisdom and experience of judges in response to sustained and substantiated criticism. We should be slow to overlook the fact that the collective experience of the judges and their close proximity to forensic science and medicine evidence for more than a century did not lead them to recognise the many problems recently revealed through authoritative non-legal reviews.¹¹⁶

In this vein, we might wonder: where are the High Court decisions on validation and acceptable error rates in ‘identification’ (ie, pattern recognition and comparison) sciences? Where are the decisions on contextual bias and the need to exclude incriminating opinions derived in conditions where the analyst was unnecessarily exposed to domain irrelevant information? Where are the decisions rejecting the continued use of face and body mapping evidence or the use of police officers *as ad hoc experts* in voice identification cases where there have been no studies of actual abilities? Where, most importantly, are the decisions requiring trial and appellate courts to consider the reliability of incriminating expert evidence – whether scientific, non-scientific, novel or longstanding? It is disconcerting, but perhaps not surprising, to find that there are no references to the *NAS Report* and other inquiries in reported Australian decisions.

While forensic scientists, prosecutors and judges might argue that foreign reports and insights have limited application to Australia, and particularly Australian courts, two issues warrant specific consideration.¹¹⁷ First, if Australian forensic scientists had undertaken (or had access to) validation studies the *NAS* and *NIST/NIJ Reports* would have cited them. The vast majority of our state-employed forensic analysts are not in compliance with the *NAS* recommendations in their practice and reporting.¹¹⁸ Secondly, who should bear the risk that techniques have not been formally evaluated, along with the misplaced confidence in our current practices? In a system that purports to cherish liberty and uphold the rights of the innocent, it might be considered curious that unanswered critiques of forensic science and medicine (and by implication traditional legal practice) by some of the world’s pre-eminent scientific institutions should be treated with suspicion, or ignored, while legal experience and impressions are privileged. Such an orientation puts non-systematic legal experience above scientific advice, and makes those accused of the most serious crimes bear the risk of incriminating expert evidence being exaggerated or mistaken and the trial safeguards less effective than credible

116 There are no judgments that capture the tone or depth of problems identified in the recent reports.

117 See Edmond, ‘What Lawyers Should Know’, above n 7.

118 Most expert reports (or certificates) do not comply with domestic rules of court. See, eg, standard NSW (or Victorian) latent fingerprint reports and contrast them with domestic codes of conduct for expert witnesses outlined in *Uniform Civil Procedure Rules 2005* (NSW) sch 7.

safeguards ought to be. A long history of miscarriages of justice (eg, *Chamberlain*, *Button*, *Mickelberg*, *Mallard*, and *Stafford*),¹¹⁹ and continuing problems (exemplified in *Morgan*, *Gilham* and *Wood* as well as the cases of *Velevski*, *Folbigg* and *Neill-Fraser*),¹²⁰ are the legacy of inattention to the reliability of incriminating expert evidence and the limited effectiveness of existing rules, jurisprudence and safeguards.¹²¹

Notwithstanding the critical tone of this article, we should not overlook the fact that forensic analysts are overwhelmingly conscientious and the forensic sciences have, in general, been a positive force. They have contributed to the resolution of innumerable crimes and the conviction of innumerable criminals. They have also assisted in removing innocent suspects from suspicion. They have not, however, been as positive a force as they could have been and could yet be. The failure to formally evaluate techniques has resulted in inflated beliefs about the abilities of analysts and exaggerated forms of expression and confidence. In many cases we have no idea about the value of the incriminating expert evidence. Instead of independent assistance, in a surprisingly large number of cases, analysts confirmed the impressions of investigators using techniques that had never been formally evaluated. Opportunities for independent insights were lost as police and forensic analysts (often themselves police) worked cooperatively sharing information and beliefs. Moreover, analysts, many of who were not scientists, relied upon convictions as evidence of their abilities. Most never conceived that convictions did not provide credible feedback on techniques and abilities or lead to improvement in performance.

Those accused of serious criminal acts should not be responsible for exposing fundamental limitations with expert evidence relied upon by the state. They should not bear the risk that trial safeguards might work and methodological frailties – that passed unappreciated or undervalued for decades – will somehow be understood by contemporary jurors and judges. The accused should not be obliged to confront prosecution cases that are ‘propped’ or ‘bloated’ with weak, speculative or unreliable evidence, portrayed as the reliable conclusions of highly experienced experts, independently corroborating other strands of incriminating evidence. The state rather than the accused should bear the many risks associated with the use of insufficiently reliable forensic science techniques and derivative opinions. Insufficiently reliable forensic science evidence should not be adduced or admitted. Such responses will not only improve the forensic sciences but

119 *Chamberlain v The Queen* [No 2] (1984) 153 CLR 521; *Button v The Queen* (2002) 25 WAR 382; *Mickelberg v The Queen* (2004) 29 WAR 13; *Mallard v The Queen* (2005) 224 CLR 125; *R v Stafford* [2009] QCA 407.

120 *Morgan* (2011) 215 A Crim R 33; *Gilham* (2012) 224 A Crim R 22; *Wood* [2012] NSWCCA 21; *Velevski v The Queen* (2002) 187 ALR 233; *Folbigg v R* [2007] NSWCCA 371; *Neill-Fraser v Tasmania* [2012] TASCCA 2.

121 See generally Emma Cunliffe, *Murder, Medicine and Motherhood* (Hart Publishing, 2011); Bibi Sangha, Kent Roach and Robert Moles, *Forensic Investigations and Miscarriages of Justice – The Rhetoric Meets the Reality* (Irwin Law, 2010).

simultaneously will contribute to a reduction in the incidence of miscarriages of justice in ways that are consistent with fundamental legal values and aspirations.