

FORENSIC SCIENCE EVIDENCE AND THE CONDITIONS FOR RATIONAL (JURY) EVALUATION

GARY EDMOND*

Historically, concerns with jury competence have been assuaged by the celebration of trial safeguards, expressions of confidence in jury abilities, and most recently through initiatives intended to improve the presentation of expert evidence. Whereas trial and appellate judges continue to express confidence in the effectiveness of the adversarial trial and the competence of juries, based almost exclusively on their (individual or institutional) experience, jury researchers have been more attentive to empirical studies of jury performance, particularly how jurors understand complex evidence, probabilities, directions and warnings, and standards of proof. These empirical studies sometimes identify problems with traditional trial practices though often suggest that problems can be improved (or overcome) through more careful presentation. In response, this article contends that legal assumptions and some of the proposals flowing from empirical research are misguided. It will be argued that inattention to the validity and reliability of many forensic science techniques, along with the failure to provide indicative error rates and attend to limitations, proficiency and contextual bias, means that in many cases expert opinion evidence adduced in criminal proceedings is not susceptible to rational evaluation.

CONTENTS

I	Introduction.....	78
II	What Jurors Need: The Basic Thesis (in Its Strong Form).....	80
	A Some Caveats.....	92
III	Enlightenment ‘Experts’: Contrasting Conventional Legal Practices	94

* BA (Hons) (Wollongong), LLB (Hons) (Syd), PhD (Cantab); Professor, Director of Postgraduate Research, and Director of Expertise, Evidence & Law Program, School of Law, The University of New South Wales, Sydney; Research Professor, Northumbria Law School, Northumbria University; member of the Standards Australia Forensic Analysis Committee, Vice President of the Australian Academy of Forensic Sciences; Chair of the Evidence-Based Forensics Initiative. This research was supported by the Australian Research Council (FT0992041 and LP120100063).

A	Inadequate Insights: Legal Criteria and Their Limitations.....	95
1	Conceding Methodological Limitations	95
2	Speculation about the Validity and Reliability of Opinions Derived from Untested Techniques	96
3	Speculation about Error and Uncertainty.....	97
4	Most Legal Criteria and Heuristics.....	98
5	Moderating the Expressions and Terminology Available to the Forensic Analyst	99
6	The Apparent Independence or Impartiality of the Forensic Analyst	100
7	Allowing the Jury to Decide Based on What Transpires at Trial.....	102
8	Deference to the Forensic Analyst	103
B	The Limits of Trial Safeguards in Response to Forensic Science and Medicine Evidence	104
C	The Apparent Strength of the Case and the Unrecognised Problem of ‘Masking’	106
IV	Jury Competence: Jury Research and Legal Attitudes	110
A	Research on Expert Evidence and the Common Law Jury	110
B	Judicial Experience and the Experience of the Law	121
V	Irrationality, Agnosticism and Post-enlightenment Judging.....	125

I INTRODUCTION

Historically, in response to proffers of forensic science and medicine evidence, concerns about the competence of juries have been assuaged by expressions of confidence in jury abilities, the celebration of trial safeguards and most recently through initiatives intended to improve the presentation of expert evidence, notably training, the use of primers, refined terminologies, new display technologies, encouraging questions, and revised directions and warnings.¹ Whereas trial and appellate judges have maintained strong public

¹ Jacqueline Horan, *Juries in the 21st Century* (Federation Press, 2012) chs 2–3; Justice Mark Weinberg, ‘The Criminal Law — A “Mildly Vituperative” Critique’ (2011) 35 *Melbourne University Law Review* 1177; Justice Mark Weinberg et al, ‘Simplification of Jury Directions Project: A Report to the Jury Directions and Advisory Group’ (Supreme Court of Victoria, August 2012); New South Wales Law Reform Commission, *Jury Directions*, Report No 136 (2012). Cf the oft-cited essay by Learned Hand, ‘Historical and Practical Considerations Regarding Expert Testimony’ (1901) 15 *Harvard Law Review* 40, 54.

confidence in the effectiveness of the trial and the competence of juries based largely on their (individual or institutional) experience, experimental psychologists have based their confidence on empirical studies of jury performance, particularly jury verdicts, exit surveys and responses to complex evidence. Where empirical studies identify limitations — often via experimental recreations of adjudicative processes — jury researchers (and others) tend to suggest that these might be substantially mitigated or overcome through more careful presentation of evidence and relatively minor modifications to conventional trial procedures.²

In response, this article explains why pervasive legal confidence, as well as some of the dominant readings of relevant empirical research, seem to be misplaced. It suggests that the reluctance to require techniques (or processes) relied upon by forensic analysts to be formally assessed means that a considerable volume of forensic science and medicine evidence is not conducive to, and cannot be presented in ways that are susceptible to, rational evaluation.³ Liberal approaches to admissibility (specifically current interpretations of uniform *Evidence Acts* ss 55, 56, 79, 135 and 137)⁴ result in the routine admission of *expert* opinions without attending to ‘specialised knowledge’ or the kinds of information and interpretive tools necessary to unpack, understand and evaluate them.⁵ This article explains why the failure to formally test techniques means that in most cases limitations and risks cannot be gauged or

² See, eg, Joel S Cecil, Valerie Hans and Elizabeth C Wiggins, ‘Citizen Comprehension of Difficult Issues: Lessons from Civil Jury Trials’ (1991) 40 *American University Law Review* 727, 764–71. See also the informative study by Jane Goodman-Delahunty and Lindsay Hewson, ‘Improving Jury Understanding and Use of Expert DNA Evidence’ (Technical and Background Paper No 37, Australian Institute of Criminology, 2010).

³ I have used ‘forensic analyst’ rather than ‘forensic scientist’ to accommodate the many forensic analysts who do not possess formal scientific qualifications.

⁴ The *Evidence Acts* comprise seven Australian statutes: *Evidence Act 1995* (Cth); *Evidence Act 2011* (ACT); *Evidence Act 1995* (NSW); *Evidence Act 2004* (Norfolk Island); *Evidence (National Uniform Legislation) Act 2011* (NT); *Evidence Act 2001* (Tas); *Evidence Act 2008* (Vic). Queensland, South Australia and Western Australia have not adopted the uniform legislation.

⁵ Gary Edmond, ‘The Admissibility of Forensic Science and Medicine Evidence under the Uniform Evidence Law’ (2014) 38 *Criminal Law Journal* 136. My earlier work focused more heavily on admissibility and procedural dimensions. This essay questions the possibility of rational analysis. See also Scott Brewer, ‘Scientific Expert Testimony and Intellectual Due Process’ (1998) 107 *Yale Law Journal* 1535. ‘Expert’ is italicised to draw explicit attention to the fact that in many cases we do not know whether the forensic analyst is actually skilled at the specific task. This includes forensic analysts with formal qualifications and experience.

credibly explored via traditional trial mechanisms such as cross-examination (and testimonial concessions), the use of rebuttal witnesses, or careful judicial instruction.⁶ None of these legal safeguards can replace formal scientific evaluation. They do not provide the kinds of information that would enable a decision-maker to assess a technique or ability and rationally assign a probative value to a derivative opinion. This article explains the importance of validating techniques in regular use. Simultaneously, it reinforces the need for more rigorous admissibility gatekeeping by prosecutors and judges. For, unless they are presented in a manner that is susceptible to rational evaluation, the admission of *expert* opinions threatens the overarching objectives of factual rectitude and fairness.

II WHAT JURORS NEED: THE BASIC THESIS (IN ITS STRONG FORM)

There are certain (normative) assumptions, which do not seem to be controversial, governing obligations and the way evidence is adduced and presented in accusatorial trials.⁷ It is, for example, the responsibility of the state, usually through a prosecutor, to persuade the tribunal of fact (or fact-finder) of the accused's guilt.⁸ Requiring the state to prove the case against the accused beyond reasonable doubt encapsulates the burden and standard of proof flowing from the presumption of innocence and the premium placed on not convicting the innocent. Apart from a handful of exceptions, such as where mental incapacity is in issue, the accused has few formal obligations and is not expected to prove his or her innocence.⁹ Because of the way the

⁶ This article is focused on the trial, but similar problems haunt the largely unregulated domain of plea negotiations.

⁷ Paul Roberts and Adrian Zuckerman, *Criminal Evidence* (Oxford University Press, 2nd ed, 2010); H L Ho, *A Philosophy of Evidence Law: Justice in the Search for Truth* (Oxford University Press, 2008); Alex Stein, *Foundations of Evidence Law* (Oxford University Press, 2005); Andrew Ligertwood and Gary Edmond, *Australian Evidence: A Principled Approach to the Common Law and the Uniform Acts* (LexisNexis Butterworths, 5th ed, 2010).

⁸ 'Tribunal of fact', 'fact-finder', and 'trier of fact' are used interchangeably.

⁹ In some instances the burden may shift in relation to a particular issue (eg around mental capacity), though this usually involves either satisfying the fact-finder of some state of affairs on the balance of probabilities or providing an evidentiary basis that requires the state to respond.

burden of proof is allocated the accused does not need to testify or even adduce evidence.

In terms of the evaluation of evidence, there are few rules or prescriptions and even fewer means of imposing them upon the trier of fact. Australian juries decide in camera and do not provide reasons. Their understanding and evaluation of the evidence, along with their reasoning, is largely inscrutable.¹⁰ Nevertheless, the orthodox and longstanding legal commitment underpinning accusatorial prosecutions is to the fact-finder being able to understand all of the evidence presented during the trial. There is no obligation upon the fact-finder to accept it, but the fact-finder should consider all of the admissible evidence.¹¹ In accordance with this commitment, fact-finders must be placed in a position that enables them to understand and evaluate admissible evidence.¹²

This article is concerned with expert opinion evidence adduced by the state and the ability of the jury (and simultaneously lawyers and judges) to understand and evaluate it.¹³ At trial, the fact-finder should be placed in a position to understand and evaluate any opinion presented by a forensic analyst.¹⁴ They should not have to defer to the authority or standing of the *expert* or the plausibility of the claim. Inexorably, this article draws attention to the performance of lawyers and judges, as well as the effectiveness of criminal proceedings, procedures, rules and safeguards in relation to forensic science and medicine evidence. In relation to the focus of this article, the overwhelming majority of forensic science and medicine evidence is prepared

¹⁰ See, eg, *Jury Act 1977* (NSW) s 68A. See also below Part IV.

¹¹ See William Twining, *Theories of Evidence: Bentham and Wigmore* (Weidenfeld and Nicolson, 1985); William Twining, *Rethinking Evidence: Exploratory Essays* (Basil Blackwell, 1990).

¹² See, eg, *Kingswell v The Queen* (1985) 159 CLR 264, 301 (Deane J):

in a legal system where the question of criminal guilt is determined by a jury of ordinary citizens, the participating lawyers are constrained to present the evidence and issues in a manner that can be understood by laymen. The result is that the accused and the public can follow and understand the proceedings.

¹³ See Ronald J Allen and Joseph S Miller, 'The Common Law Theory of Experts: Deference or Education?' (1993) 87 *Northwestern University Law Review* 1131; Gary Edmond and Andrew Roberts, 'Procedural Fairness, the Criminal Trial and Forensic Science and Medicine' (2011) 33 *Sydney Law Review* 359.

¹⁴ *HG v The Queen* (1999) 197 CLR 414, 429 [44] (Gleeson CJ); *Velevski v The Queen* (2002) 187 ALR 233, 275 [182] (Gummow and Callinan JJ); *Makita (Australia) Pty Ltd v Sprowles* (2001) 52 NSWLR 705, 743–4 [85] (Heydon JA); *Hillstead v The Queen* [2005] WASCA 116 (23 June 2005), [48]–[49] (Pullin JA).

and adduced by the state, and the state maintains a special interest in the conduct of criminal proceedings.¹⁵

My thesis builds on these orthodox commitments but introduces a largely unexplored set of issues with radical implications for how we understand the operation of our criminal justice institutions. The basic thesis runs as follows: *specific types of information are required to evaluate (ie 'weigh') most types of forensic science and medicine evidence. In the absence of this information many types of forensic science and medicine evidence are not susceptible to rational evaluation.* To put this in more legally sensitive terms: the attempt to assign a probative value to incriminating opinion evidence is, in the absence of specific kinds of information, unavoidably speculative or impressionistic.¹⁶

Difficulties arise because criminal trials do not usually supply the requisite information or an environment conducive to the rational evaluation of expert opinion evidence.¹⁷ That is, the kinds of information ordinarily required to make sense of opinions — including insights into the conceptual or methodological reasons for specific kinds of testing, the meaning of test results, as well as the way opinions are expressed — are not routinely provided and are rarely presented in a balanced way. In consequence, the fact-finder is expected (and effectively required) to evaluate incriminating opinions and assign a probative value without the kinds of information and assistance that would enable them to do so rationally.

The thesis, and concerns about conspicuous informational voids, emerged from my work on comparison and pattern recognition techniques, sometimes described as the *identification sciences*.¹⁸ These are the staples of the forensic sciences, and include: DNA profiling, drug and chemical assays, comparisons

¹⁵ It is, for example, common for the prosecution to call the only expert evidence at trial: Ian Freckelton et al, 'What Do Real Jurors Think of Expert Evidence?' (Speech delivered at the 11th Annual Australasian Jury Research and Practice Conference: Current Issues in Jury Reform, Research and Policy, The University of Melbourne, 7 November 2014).

¹⁶ These difficulties are not transcended by recourse to other strands of evidence: see below Part III(C).

¹⁷ These difficulties are compounded by the requirement that the presentation and evaluation of the opinion take place in an agonistic (ie an adversarial) environment. This makes admissibility particularly important: see generally Pierre Bourdieu, 'The Force of Law: Toward a Sociology of the Juridical Field' (1987) 38 *Hastings Law Journal* 814.

¹⁸ 'Sciences' is italicised because, as we shall see, many forensic science techniques have not been formally evaluated and do not satisfy the kinds of conditions we might expect from the experimental sciences and biomedicine.

of latent fingerprints, ballistics and tool marks, documents and handwriting, gait, shoe, foot and tyre prints, glass, paint, hair, fibre and bite marks, as well as use of image or voice recordings to assist with identification, some forms of digital evidence, and so forth. Following from the thesis it is my contention that in order to understand and evaluate the probative value of *expert* opinions in these (and many other) domains, quite specific information is required.¹⁹ What follows is a hierarchical list of factors (or criteria) providing the kinds of information that facilitate rational fact-finding.²⁰ The first three are the most important and, to varying degrees, the remaining criteria flow from them.²¹

First, the most fundamental issue in any attempt to evaluate an opinion or result is to know whether the underlying technique or process used to produce it actually works. This applies whether the technique is a complex technical process (such as DNA profiling) or a beguilingly straightforward visual comparison of two images on a screen (for latent fingerprints, ballistics and facial mapping). It also applies to new techniques and those in long use. Determining whether a technique works — ie does what it is supposed to — requires some kind of experimental test or evaluation, usually an *independent validation study*. Validation studies should be conducted in conditions where the ‘ground truth’ is known so that the performance of the technique can be assessed against an answer that is known to be correct.²² Validation studies tell us whether the technique (or process) actually works, how well, and in what conditions.

¹⁹ This is consistent with the best scientific advice: see National Research Council, *Strengthening Forensic Science in the United States* (National Academies Press, 2009) (‘NAS Report’) 7–8, 87:

With the exception of nuclear DNA analysis ... no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source. ... 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.

²⁰ There are other criteria, including some discussed below in Part III(A), but logically we need to know that techniques work and analysts are proficient before we turn to consider claims about partisanship or admission by other courts.

²¹ Gary Edmond, ‘Pathological Science? Demonstrable Reliability and Expert Pathology Evidence’ in Kent Roach (ed), *Pediatric Forensic Pathology and the Justice System* (Queen’s Printer for Ontario, 2008) 96–149.

²² Investigations and legal proceedings do not afford these conditions. Guilt, even when upheld on appeal, is not a ‘correct’ answer and criminal proceedings do not validate techniques.

A recent report on the forensic sciences prepared by the United States research committee of the National Academies of Sciences explained validation (and reliability) in the following terms:

One particular task of science is the validation of new methods to determine their reliability under different conditions and their limitations. Such studies begin with a clear hypothesis (e.g., ‘new method X can reliably associate biological evidence with its source’). An unbiased experiment is designed to provide useful data about the hypothesis. Those data — measurements collected through methodical prescribed observations under well-specified and controlled conditions — are then analyzed to support or refute the hypothesis. The thresholds for supporting or refuting the hypothesis are clearly articulated before the experiment is run. The most important outcomes from such a validation study are (1) information about whether or not the method can discriminate the hypothesis from an alternative, and (2) assessments of the sources of errors and their consequences on the decisions returned by the method.²³

Almost all techniques (or processes), and most expert claims, can and should be subjected to validation or some kind of analogous experimental testing.²⁴ In most cases the results of experimental testing should be published or made available through publicly accessible repositories.²⁵ All expert opinion evidence in uniform *Evidence Acts* jurisdictions should be based on

²³ National Research Council, *NAS Report*, above n 19, 113. Validation and proficiency testing will not prevent genuine contests around continuity, interpretations and significance, and even whether the validation and proficiency tests are adequate. Formal evaluation is, however, likely to limit the breadth of debate and focus attention on factors most relevant to probative value. See also Bruce Budowle et al, ‘A Perspective on Errors, Bias, and Interpretation in the Forensic Sciences and Direction for Continuing Advancement’ (2009) 54 *Journal of Forensic Science* 798, 807.

²⁴ Where appropriate testing or inquiry is not undertaken, problems sometimes emerge subsequently. Consider the failure to direct attention to testing and the reliability of opinions about carbon monoxide and stab wounds in *Gilham v The Queen* [2012] NSWCCA 131 (25 June 2012) and the blood and textile evidence in Justice Trevor Morling’s review of the Chamberlain convictions in Commonwealth, ‘Royal Commission of Inquiry into the Chamberlain Convictions’, *Report* (1987).

²⁵ On the need for public access to the workings of techniques and algorithms, see Lord Chief Justice Thomas, ‘Expert Evidence: The Future of Forensic Science in Criminal Trials’ (Speech delivered at the Criminal Bar Association, Kalisher Lecture, 14 October 2014). Peer review is an important component of both scientific publication and the review of scientific performance. It cannot, however, replace testing. Its value as a check on a process is also limited where review is not blind.

identifiable ‘specialised knowledge’. Validation studies provide the kind of *knowledge* that can support the opinions of those with relevant ‘training, study or experience’.²⁶

Second, as the extract above implies, there is also a need to direct attention to *limitations* and *uncertainties* and the *error rate* associated with the technique.²⁷ These provide insight into limitations, the number of mistakes, and some of the risks associated with the application of the technique (or process).²⁸ Indicative error rates are usually generated through validation studies, and occasionally through system-wide tests.²⁹ Though, uncertainties, limitations and risks might, in addition, be extrapolated from other domains such as biomedicine.³⁰

Third, the last of the major factors is evidence of the analyst’s *proficiency using the validated technique (or process)*. This is to be contrasted with long experience, including experience doing similar things. For, rigorous proficiency testing provides information about the abilities of the forensic analyst. It confirms that the analyst is competent using a particular technique and is an expert in the specific domain.³¹ That is, regardless of qualifications and

²⁶ *Evidence Acts* s 79.

²⁷ In most circumstances testing will produce general information, and this is often the best we can hope for. See David L Faigman, John Monahan and Christopher Slobogin, ‘Group to Individual (G2i) Inference in Scientific Expert Testimony’ (2014) 81 *The University of Chicago Law Review* 417. Medical testing, prognoses, and extrapolations are usually based on the same kind of *general* — or indicative rather than specific — information.

²⁸ All techniques and processes have errors. Errors may occur in the collection, transportation, labelling, storage, description, analysis, interpretation and evaluation of traces, data and evidence. See generally Linda T Kohn, Janet M Corrigan and Molla S Donaldson (eds), *To Err Is Human: Building a Safer Health System* (National Academy Press, 1999).

²⁹ See, eg, 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; Jason M Tangen, Matthew B Thompson and Duncan J McCarthy, ‘Identifying Fingerprint Expertise’ (2011) 22 *Psychological Science* 995. Interestingly, these articles report the results of the first validation studies of latent fingerprint comparison. While the results are generally reassuring, the investigators reported small numbers of false positive and false negative errors.

³⁰ See Expert Working Group on Human Factors in Latent Print Analysis, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach* (US Department of Commerce, National Institute of Standards and Technology, 2012).

³¹ There may be overlap between validation studies and some kinds of proficiency testing: Jonathan J Koehler, ‘Proficiency Tests To Estimate Error Rates in the Forensic Sciences’ (2012) 12 *Law, Probability & Risk* 89.

experience, rigorous proficiency testing tells us whether the forensic analyst performs a task or set of tasks better than non-experts or chance. A significantly enhanced level of performance is precisely what it means to be an expert.³² Evidence that techniques work, and that the forensic analyst performs markedly better than an ordinary person, is what courts should require before opinions are presented to fact-finders.³³

Beyond these fundamental criteria there is a need to attend to *standards*. Are there any? Were they developed through formal testing and were they used in the specific application of the technique in the case? Techniques should be standardised and standards should be based on the results of experimental, usually validation, studies. Standards developed or imposed without formal evaluation may have limited value.³⁴ Similarly, proficiency with a technique that has not been validated, even if the witness purports to have applied 'standards', produces opinions of unknown value.

In many cases there should also be an explanation of *contextual bias* and other threats to cognition and interpretation, along with a description of the efforts to avoid or eliminate them. Notwithstanding their historical interest in partisanship and bias, especially 'adversarial bias' in civil proceedings, Australian judges are yet to respond to the serious threat to forensic science and medicine evidence posed by a range of undesirable practices that remain remarkably common within investigative communities.³⁵ Without wanting to trivialise the dangers created by forensic analysts working intimately with investigators *as a team* or being aligned ideologically with

³² K Anders Ericsson et al (eds), *The Cambridge Handbook of Expertise and Expert Performance* (Cambridge University Press, 2006). Usually expertise is demonstrated through some kind of formal evaluation or test where the analyst demonstrates their proficiency at a specific task. Expertise is not necessarily transferable to different tasks. See also *Velevski v The Queen* (2002) 187 ALR 233, 253 [82] (Gaudron J).

³³ Current interpretations of *Evidence Acts* s 79 do not require this. Indeed, in some cases, trial and appellate judges have been inclined to admit the opinions of forensic analysts rather than leave difficult evidence (such as image comparisons involving poor quality CCTV images) to the jury. Where the task is difficult or error-prone there are problems both in allowing a highly experienced or qualified analyst of unknown ability to express an opinion as well as leaving the task to the jury. See, eg, *R v Tang* (2006) 65 NSWLR 681; *R v Atkins* [2010] 1 Cr App R 8.

³⁴ See the work of Standards Australia, *CH-041 Forensic Analysis*, Standards Development Public Portal <<http://www.sdpp.standards.org.au/ActiveProjects.aspx?CommitteeNumber=CH-041&CommitteeName=Forensic%20Analysis>>.

³⁵ New South Wales Law Reform Commission, *Expert Witnesses*, Report No 109 (2005) 71–6.

the police or prosecutors, these may not be the major threats to cognition and interpretation.³⁶

Several peak scientific organisations have recently expressed grave concerns about the way forensic analysts are routinely provided with information about the case and the accused that is often prejudicial or suggestive but *not required for the analysis*.³⁷ Preliminary studies, along with experiences in other domains of scientific and biomedical endeavour, have demonstrated that even subtle suggestions and environmental influences can change the way analysts select and test samples, interpret data and report experimental results. Specific studies have shown that interpretations of latent fingerprints and DNA electropherograms can be influenced — ie substantially altered and even reversed — by exposing the analysts to gratuitous information or suggestive processes.³⁸ These studies indicate that opinions derived from techniques that are otherwise valid and reliable may be vulnerable to insidious cognitive biases and, moreover, that long experience and even awareness of the dangers posed do not guarantee resilience to such biases.³⁹

The threat posed by contextual bias is particularly invidious because the influence can be subtle. Analysts are not always conscious of their exposure to biasing information or its influence.⁴⁰ The upshot is that exposure and

³⁶ Keith A Findley, 'Innocents at Risk: Adversary Imbalance, Forensic Science and the Search for Truth' (2008) 38 *Seton Hall Law Review* 893; David A Harris, *Failed Evidence: Why Law Enforcement Resists Science* (New York University Press, 2012).

³⁷ National Research Council, *NAS Report*, above n 19, 24 (recommendation 5); National Institute of Standards and Technology, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach* (US Department of Commerce, 2012) ('*NIST Report*') 10, 44 (recommendation 3.3); Sir Anthony Campbell, Submission to the Cabinet Secretary for Justice, Scottish Government, *The Fingerprint Inquiry Report*, December 2011, 741.

³⁸ Itiel E Dror, David Charlton and Ailsa E Péron, 'Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications' (2006) 156 *Forensic Science International* 74; Itiel E Dror and Greg Hampikian, 'Subjectivity and Bias in Forensic DNA Mixture Interpretation' (2011) 51 *Science and Justice* 204.

³⁹ These techniques are demonstrably reliable, but the evidence they produce is not usually presented in courts in ways that facilitate rational evaluation. No error rate is provided and the results are not consistently reported in ways most conducive to lay comprehension. See the recommendations of the National Research Council in *NAS Report*, above n 19; National Institute of Standards and Technology, *NIST Report*, above n 37.

⁴⁰ Sometimes information might be suggestive and even the process of analysis might suggest an answer. In some fingerprint bureaus, for example, fingerprint examiners only review reported 'matches'. This means that the individual in the 'peer review' process knows the

influences might not be documented and may be difficult to explore during the trial. How, for example, do you cross-examine a forensic analyst about unconscious influences?⁴¹ The dangers created by suggestion and other cues and biases are the main reason why most clinical trials are double-blind — neither the doctor nor the patient knows who received the active ingredient and who received the placebo (or comparator). Significantly, there is no evidence that training and experience enable forensic analysts to circumvent the effects of cognitive biases, especially where evidence is ambiguous or interpretation difficult. Experienced medical doctors have proved incapable of avoiding such influences.⁴² The most effective response to threats to cognition seems to be to restrict access to information rather than try to document exposure or somehow tease out subtle psychological influences in their aftermath, sometimes years later. Contextual bias is a real threat to forensic science and medicine evidence but is yet to receive serious consideration by appellate courts.⁴³

Information about the *frequency* of feature(s) in relevant populations and whether they are independent of one another is vital where analysis (or extrapolation from analysis) involves some kind of matching or comparison in an attempt to link a trace to a person or an object.⁴⁴ Where techniques are based on comparisons, it is essential to have an idea, drawn from empirical data, of how frequent features of interest are in specific populations. This

conclusion obtained by an earlier examiner and may even know who the examiner is — with the attendant institutional and personality dynamics.

⁴¹ See Gary Edmond et al, 'How to Cross-Examine Forensic Scientists: A Guide for Lawyers' (2014) 39 *Australian Bar Review* 174.

⁴² I appreciate that some traces and some types of practice seem to be less susceptible to particular context effects and biases. Where the results are unambiguous, such as with a high quality latent print or single DNA profile used to exclude, the chances of context influencing the decision are relatively low. However, as the information becomes less clear (eg the trace or sample is degraded), the risks of influence and error become much more substantial.

⁴³ Cf *Li v The Queen* (2003) 139 A Crim R 281, 291 [71]: drawing upon civil justice authority, Ipp JA indicated that the 'risk of bias (unconscious or otherwise) is no reason not to admit evidence of an expert'.

⁴⁴ These can be complex and controversial. See, eg, National Research Council, *DNA Technology in Forensic Science* (National Academies Press, 1992); National Research Council, *The Evaluation of Forensic DNA Evidence* (National Academies Press, 1996). Studies of juror abilities to recall descriptive information about DNA evidence should not be over-read or freely extended to probabilities: see Valerie P Hans et al, 'Science in the Jury Box: Jurors' Comprehension of Mitochondrial DNA Evidence' (2011) 35 *Law and Human Behavior* 60, 61.

applies to population genetics and DNA databases (for DNA profiling), fingerprints, the soles of shoes, accents and intonations (for voice comparison), face and body features and even tattoos in image comparisons, as well as to posture and movement for forensic gait analysis. In the absence of empirically-based information on the distribution of features, attaching significance to (apparent) similarities or ‘matches’ can be dangerous.⁴⁵ Where there appears to be more than one similar feature we also need to know the degree of in(ter)dependence between them. For, independence of features will often strengthen the significance of actual similarities. This information, along with an awareness of the value of a technique and the level of error (and other uncertainties and limitations), enables opinions to be assessed. In its absence, the appearance of one or more similar features may or may not be particularly probative.

Finally, those asked to evaluate the opinions of forensic analysts need to be confident that the particular *expressions* — ie the form of words — selected to capture results and convey opinions are derived from and supported by independent research.⁴⁶ The way forensic analysts express their opinions (ie conclusion or interpretation) should be based on what validation and proficiency studies can support. The results of validation studies (and information about frequency), as well as error rates, other limitations and uncertainties and proficiency, should structure and constrain the way conclusions are expressed in reports and testimony. Moreover, opinions should be expressed in ways that are both scientifically defensible and simultaneously conducive to comprehension by laypersons.⁴⁷ Managing these needs may be difficult, but however we moderate expressions to ease comprehension at trial, the underlying analytical process must be valid and reliable.⁴⁸

⁴⁵ See the review, and extensive criticism, of similarity evidence advanced during the prosecution of Edward Charles Splatt: South Australia, Royal Commission Concerning the Conviction of Edward Charles Splatt, *Report* (1984) (‘Edward Charles Splatt Report’).

⁴⁶ Problems include undervaluing demonstrably reliable evidence as well as overvaluing evidence of unknown probative value.

⁴⁷ See, eg, Dawn McQuiston-Surrett and Michael J Saks, ‘The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Factfinders Hear’ (2009) 33 *Law and Human Behavior* 436, 448–51. See also National Research Council, *NAS Report*, above n 19.

⁴⁸ Some of the difficulties appear, though without much clarity or resolution, in the following appeals: *Aytugrul v The Queen* (2012) 247 CLR 170 (‘*Aytugrul*’); *R v T* [2010] EWCA Crim 2439 (26 October 2010); *R v Tang* (2006) 65 NSWLR 681. Even if defensible in orthodox legal terms, the High Court decision in *Aytugrul* is practically misguided. Courts and prosecutors

These kinds of information and insights, especially the first three (namely, independent validation studies, error rate and limitations, and proficiency), are vitally important in any attempt to assess the probative value of opinions associated with pattern recognition and comparison techniques and most other kinds of scientific and medical processes. Together, validation studies, error rates and insight into the proficiency (or competence as opposed to bare experience) of the analyst, provide a yardstick against which an opinion can be evaluated.⁴⁹

In principle, given the professional responsibilities and formal obligations of prosecutors (eg bar rules and prosecutorial guidelines) and forensic analysts (eg professional guidelines, codes of conduct and practice directions), information about validity, error rates and limitations, and proficiency should be proactively disclosed and explained during the presentation of the prosecution case.⁵⁰ It is the responsibility of the prosecutor to present expert opinion evidence in a manner that reflects its *known* value.⁵¹ Where techniques have not been formally evaluated or are speculative, the prosecutor should not adduce the opinion. For, it is not an opinion based on ‘specialised knowledge’.⁵²

should not ignore the fact that ordinary persons do not treat different, though mathematically equivalent, formulations in the same way.

⁴⁹ Justice Trevor Morling recognised that the failure to have undertaken appropriate research meant that some of the opinions expressed by highly credentialed and experienced expert witnesses at the Chamberlain trial were ‘of doubtful validity or wrong’: Commonwealth, above n 24, 340.

⁵⁰ Consider the myriad of prosecutorial responsibilities: Legal Services Council, *Legal Profession Uniform Conduct (Barristers) Rules 2015* (at 26 May 2015) rr 83–95; Legal Services Council, *Legal Profession Uniform Law Australian Solicitors’ Conduct Rules 2015* (at 26 May 2015) r 29; Office of the Director of Public Prosecutions (NSW), *Prosecution Guidelines* (1 June 2007). The state, through the prosecutor, should operate as a model litigant. Adherence to codes and positive norms are becoming increasingly important as funding for legal aid and defendants decline. See Gary Edmond, ‘(Ad)Ministering Justice: Expert Opinion Evidence and the Professional Responsibilities of Prosecutors’ (2013) 36 *University of New South Wales Law Journal* 921. On experts see, eg, *Uniform Civil Procedure Rules 2005* (NSW) sch 7; Australian New Zealand Forensic Science Society Inc, *Code of Professional Practice for Members of the ANZFSS* (11 August 2014).

⁵¹ ‘Specialised knowledge’ requires that something is *known*. Significantly, this is the kind of information that provides an indication of the probative value of opinions, allows the trial judge to determine what ‘taken at its highest’ might mean with respect to expert opinion evidence, and also draws attention to some of the risks of unfair prejudice to the defendant: see especially *Evidence Acts* s 137.

⁵² ‘Training, study or experience’ cannot overcome the failure to formally evaluate.

Many years ago, writing for an appeals court in Scotland, Lord President Cooper somewhat presciently drafted the following passage:

Expert witnesses however skilled or eminent can give no more than evidence. They cannot usurp the functions of the jury or Judge sitting as a jury ... Their duty is to furnish the Judge or jury with the necessary scientific criteria for testing the accuracy of their conclusions so as to enable the Judge or jury to form their own independent judgment by the application of these criteria to the facts proved in evidence. The scientific opinion evidence, if intelligible, convincing and tested, becomes a factor (and often an important factor) for consideration along with the whole other evidence in the case, but the decision is for the Judge or jury. In particular the bare *ipse dixit* of a scientist, however eminent, upon the issue in controversy, will normally carry little weight, for it cannot be tested by cross-examination nor independently appraised, and the parties have invoked the decision of a judicial tribunal and not an oracular pronouncement by an expert.⁵³

This extract was reproduced in the High Court of Australia in *Dasreef Pty Ltd v Hawchar* and has been approved by courts of appeal in New South Wales, Victoria and Western Australia.⁵⁴

This article is a modern elaboration of this basic commitment: explaining the need to provide information that enables the fact-finder to evaluate expert opinion evidence and to *exclude the evidence* where this information is not provided.⁵⁵ Unfortunately, many techniques used by forensic analysts and relied upon by prosecutors — including some techniques in regular use — are

⁵³ *Davie v Magistrates of Edinburgh* [1953] SC 34, 40 ('*Davie*'). *Davie* was a civil appeal, and the judgment implies that the admissibility of '*ipse dixit*' was basically harmless. This somewhat complacent approach should not be extended to modern criminal proceedings, especially in jurisdictions requiring opinions to be based on 'specialised knowledge'. See also Edward Charles Splatt Report, above n 45, 186–7; Tony Ward, "A New and More Rigorous Approach" to Expert Evidence in England and Wales? (2015) 19 *International Journal of Evidence & Proof* (forthcoming).

⁵⁴ (2011) 243 CLR 588, 624 [93] (Heydon J). See also *Makita Pty Ltd v Spowles* (2001) 52 NSWLR 705, 729–31 [59]–[63] (Heydon JA); *MA v The Queen* (2013) 40 VR 564, 580 [74] (Osborn JA); *Hillstead v The Queen* [2005] WASC 116 (23 June 2005) [49] (Pullin JA).

⁵⁵ See, eg, *Dasreef Pty Ltd v Hawchar* (2011) 243 CLR 588, 605 [42] (French CJ, Gummow, Hayne, Crennan, Kiefel and Bell JJ): 'A failure to demonstrate that an opinion expressed by a witness is based on the witness's specialised knowledge based on training, study or experience is a matter that goes to the admissibility of the evidence, not its weight'. See also *Re Pan Pharmaceuticals Ltd (in liq)* [2008] FCA 416 (2 April 2008) 10–15, [25]–[37] (Emmett J).

yet to be evaluated and so cannot be expressed in terms that reflect known values or enable rational evaluation.⁵⁶ In the absence of formal evaluation we do not know if techniques and derivative opinions are probative.⁵⁷ Even if we assume that opinions are probative (ie that they could be ‘accepted’ by a jury) as part of our test for relevance we have no way of evaluating them.⁵⁸ Admission in these circumstances invariably introduces a real risk of unfair prejudice to the defendant through misunderstanding, over-reading or unwarranted deference.⁵⁹

A Some Caveats

Before proceeding, a few qualifications and clarifications are useful. In advancing the thesis, it is not my intention to suggest that there is only one way to interpret expert opinion evidence, nor do I intend to suggest that there is only one proper value. Rather, it is my contention that in most circumstances specific types of information need to be provided to, and understood by, a decision-maker before they can *rationally* evaluate — ie attribute probative value or ‘weight’ to — an opinion said to be expert. In most cases, formal evaluation (through independent testing prior to the investigation and trial) will provide a range within which the expert opinion will fall. Of course, where within this range the evidence is to be placed is a matter for the tribunal of fact.⁶⁰ My basic contention is that, absent requisite information, the allocation of a value tends to be capricious (even though the decision-maker may think otherwise). For, there is always a non-trivial danger that the fact-finder will assign a probative value that is higher than the technique (or

⁵⁶ National Research Council, *NAS Report*, above n 19.

⁵⁷ For example, the analysis of bite marks for the purpose of identification (aside from any DNA evidence) represents a technique that notwithstanding long historical use appears to lack probative value.

⁵⁸ *Evidence Acts* ss 55–6.

⁵⁹ There is also a procedural unfairness in having to contest evidence at trial, where the accused is not usually well-resourced and bears the risk of misunderstanding or misuse.

⁶⁰ The results of validation studies provide guidance on how high evidence might be ‘taken at its highest’ when applying *Evidence Acts* s 137. See *R v XY* (2013) 84 NSWLR 363; cf *Dupas v The Queen* (2012) 40 VR 182; *Tuite v The Queen* [2015] VSCA 148 (12 June 2015). See also Gary Edmond et al, ‘Christie, Section 137 and Forensic Science Evidence (After *Dupas v The Queen* and *R v XY*)’ (2014) 40 *Monash University Law Review* 389.

the proficiency of the analyst) can sustain. In the absence of validation and rigorous proficiency testing we cannot discount the gravity of this risk.

The previous paragraph recognises that fact-finders might assign probative value to the opinions of forensic analysts regardless of how the evidence is presented, explained or challenged. The point is not that our fact-finders have been unable to make decisions, rather the problem is that they have been deciding in ways that tend to be inattentive to highly pertinent, and arguably essential, information. The failure to produce and provide the kinds of information that are conducive to rational decision-making not only makes legal proceedings inefficient, cumbersome and inconsistent but it means that they fall short of espoused commitments to fairness and are more likely to result in mistaken verdicts that are difficult to identify and correct.

Some might contend that provided the fact-finder understands the 'bottom line', gist or direction of an *expert's* opinion that is sufficient. Such arguments tend to undervalue the historical commitment to decision-makers actually understanding the evidence (and any limitations or criticisms).⁶¹ Simultaneously, it might encourage the jury to place excessive reliance upon non-scientific evidence and epiphenomena (see below Part III) to mediate the reception and interpretation of opinions presented as scientific or technical.

Finally, in advancing the thesis, it is useful to note that forensic science and forensic medicine evidence can be distinguished from most other types of evidence routinely admitted in criminal proceedings. The state, after all, has far more control over what happens 'backstage.' Significantly, the state is in a position to evaluate techniques, develop empirically-based standards, shield analysts from gratuitous information, enforce rigorous proficiency testing regimes, as well as select the terminology used in reports and testimony.⁶² The

⁶¹ Cf Allen and Miller, above n 13.

⁶² Here, I prefer a functional to an epistemic distinction. This makes sense in criminal proceedings because the state has resources, abilities and obligations to those it accuses and convicts. I disagree with the thesis advanced by Frederick Schauer and Barbara A Spellman, 'Is Expert Evidence Really Different?' (2014) 89 *Notre Dame Law Review* 1 (argued largely in relation to civil justice issues). Most other types of witness come as they are and we have few means beyond plausibility, consistency and demeanour to weigh their testimony. (There are complexities, such as the awareness that certain types of evidence (eg eyewitness evidence and confessions) or classes of witness (eg prison informers) might sometimes have reliability issues and evaluation might be assisted by reference to relevant scientific literatures.) Nevertheless, with forensic science and medicine evidence, especially techniques in routine use, decision-makers can and should have systematic access to information about validity, reliability and proficiency.

fact that these things have not been routinely addressed does not mean that they do not matter or should not be undertaken.⁶³ Rather, these oversights implicate lawyers and judges in the pre-scientific and speculative condition of many techniques and practices used by our forensic analysts and relied upon by prosecutors. Liberal admissibility standards, along with the actual weakness of trial safeguards and prosecutorial obligations (see below Part III(B)), have contributed to the questionable epistemic status of a great deal of forensic science and medicine evidence.

III ENLIGHTENMENT ‘EXPERTS’: CONTRASTING CONVENTIONAL LEGAL PRACTICES

Information conducive to the rational evaluation of ‘opinion based on specialised knowledge’ is sometimes presented in criminal proceedings in Australia. Adequate disclosure and detailed explanation is, however, the exception.⁶⁴ Ordinarily, to the extent that validation, error rates, limitations, and proficiency are raised, they tend to be traversed in rather superficial ways, where it is not obvious that lawyers, judges and fact-finders appreciate their significance or the significance of their absence.

Prosecutors and forensic analysts do not routinely present evidence in ways designed to enhance understanding, let alone facilitate, rational evaluation. Defence lawyers, whether through resourcing constraints, incompetence or for ‘tactical reasons’, irregularly develop oversights, omissions and limitations in ways that are comprehensible to laypersons or would persuade them of their epistemic significance. Judicial instructions and directions rarely capture or convey the significance of omissions and *knowledge gaps*. Judicial instructions might touch upon them but they seldom endorse, or lend legal imprimatur, to the significance of methodological limitations and oversights.⁶⁵ Rather, they tend to leave the significance of these fundamental deficiencies as

⁶³ It is not uncommon for judges to treat the need for validation (or methodological rigour) as an extravagance or even an impossibility. The opposite is usually true: see *Li v The Queen* (2003) 139 A Crim R 281, 288–9 [47]–[56] (Ipp JA); *R v Madigan* [2005] NSWCCA 170 (9 June 2005).

⁶⁴ I can make this claim because the information is either unavailable (recall the conclusion by the National Academy of Sciences in the National Research Council, *NAS Report*, above n 19) or not routinely disclosed in reports or testimony.

⁶⁵ Contrast notorious wrongful convictions where these issues tend to become prominent: see especially Edward Charles Splatt Report, above n 45, 175–7, 182–8.

jury questions — issues of weight for the fact-finder to somehow ‘factor in.’ Similarly, courts of appeal have not taken the failure to validate sufficiently seriously. Trusting in the trial and trial personnel, and in many cases the strength of the overall case (on review), they tend to uphold the admission of opinions based on techniques that are testable but are yet to be independently tested.⁶⁶ Appellate courts display a tendency to defer to the imagined capabilities of laypersons informed by Panglossian impressions of the effectiveness of trial safeguards.

A Inadequate Insights: Legal Criteria and Their Limitations

Lacking the kinds of insights and information that would facilitate rational evaluation of forensic science and medicine evidence, our courts have developed a range of alternative criteria. These include: requiring the forensic analyst to concede that validation testing has not been performed; allowing forensic analysts to offer ‘educated’ guesses about validity, error rates and limitations; imposing constraints on the form of expression the analyst might use; relying on legal (admissibility) criteria such as the existence of a ‘field,’ formal qualifications, experience, previous admission, and apparent impartiality; as well as expecting the defence to expose limitations in conjunction with a tendency to treat defence failures, oversights and omissions as tactical and therefore deliberate.⁶⁷ In sharp contrast to the kinds of indicia outlined above in Part II, these legal criteria (or heuristics) direct attention toward secondary qualities or epiphenomena. The remainder of this Part endeavours to explain some of the problems and limitations with the kinds of criteria and mechanisms ordinarily relied upon to evaluate the opinions of forensic analysts in trials and on appeal.

1 Conceding Methodological Limitations

Disclosing or conceding that formal evaluation has not been undertaken should not provide an admissibility pathway for opinions based on untested techniques.⁶⁸ Significantly, disclosing that appropriate testing has yet to be

⁶⁶ The strength of the overall case is commonly referenced in the application of the proviso.

⁶⁷ See, eg, Australian Law Reform Commission, *Uniform Evidence Law*, No 102 (2006) ch 9.

⁶⁸ See, eg, *R v Tang* (2006) 65 NSWLR 681; *Morgan v The Queen* [2011] NSWCCA 257 (1 December 2011); *Honeysett v The Queen* (2014) 253 CLR 122. The English case of *R v Atkins* [2010] 1 Cr App R 8 is a conspicuous example.

performed does not enable a fact-finder to evaluate the technique. All such concessions do is reveal that scientifically appropriate forms of evaluation have not been conducted. There seems to be a belief that such concessions will enable those lacking methodological training and technical sophistication to appreciate the significance of methodological shortcomings and omissions. However, 'known unknowns' do not place the fact-finder in a position to evaluate opinion evidence. Rather, they provide insight into what needs to be done to produce *knowledge*. The absence of relevant information obliges the fact-finder to rely on criteria and considerations that may not be particularly informative. In the context of the trial, for example, admission might be used by jurors to draw the inference that the technique is basically sound or that other criteria are of more utility than formal evaluation.

2 *Speculation about the Validity and Reliability of Opinions Derived from Untested Techniques*

In the absence of formal evaluation, undertaken in conditions where the correct answer is known, the analyst should not be invited to opine on the probative value of the opinion. In these circumstances claims about validity and reliability, whether explicit or implicit, are unavoidably speculative. The analyst does not know whether the technique works, nor how often, nor how well. This not only applies to untested techniques, but also to novel applications of proven techniques. Independent testing provides insights into validity and reliability. We should not allow forensic analysts to speculate about the value of a technique or express their confidence in a particular opinion in the absence of testing.

Where the analyst is highly trained and experienced, the significance of validation, error rates and limitations, standards, proficiency and contextual bias, may be marginalised or obscured.⁶⁹ Indeed, dangers are acute where the results of testing are not available to mediate the opinions of those presented

⁶⁹ In many trials the forensic analyst's credentials are introduced at length even though it is often not known whether the analyst possesses actual expertise in the domain relevant to the proceedings. Consider the lengthy introduction of anatomists and their credentials in image comparison (ie face and body mapping) cases such as *R v Tang* (2006) 65 NSWLR 681, 684–5 [9], [11], 690–1 [41]–[42] (Spigelman CJ); *Morgan v The Queen* (2011) 215 A Crim R 33, 44 [71], 46 [82] (Hidden J); *Honeysett v The Queen* (2014) 253 CLR 122, 128 [11]–[12] (French CJ, Kiefel, Bell, Gageler and Keane JJ). Cf Alice Towler et al, 'Is Facial Identification Training Effective?' (Paper presented at the Society for Applied Research in Memory and Cognition Conference, Victoria, British Columbia, 27 June 2015).

as experienced. An analyst's experience, confidence and the strength of his or her opinion are not necessarily correlated with ability or accuracy.⁷⁰ Inattention to appropriate criteria introduces the risk that a forensic analyst will (inadvertently) exaggerate the probative value of his or her evidence, and unreasonably reject or trivialise (defence) criticisms and mainstream scientific insights.⁷¹ A significant proportion of our forensic analysts, particularly those in areas such as crash investigation, ballistics and tool marks, document examination, and latent fingerprint, shoe, foot and tyre print comparisons, do not possess formal scientific qualifications. Many more are not conversant with relevant methodological precepts, probabilistic approaches to expression and a range of risks that inform practice in mainstream biomedicine, engineering and science.⁷²

3 *Speculation about Error and Uncertainty*

In the absence of formal evaluation, claims about error rates, accuracy and uncertainty are speculative, and likely to be wrong or misleading. As with speculation about the validity of techniques, forensic analysts should not be allowed to express incriminating opinions at trial on the basis that the defence is entitled to try to obtain concessions about the analyst's impression of errors, uncertainties and other limitations. Making the defence responsible for exposing limitations and oversights will, in many cases, require them to ask a forensic analyst — who has not proactively disclosed or acknowledged limitations in their report and testimony, and might not even

⁷⁰ Obviously, they can be correlated, but those with formal qualifications and experience are not necessarily more skilled than those without formal qualifications or such extensive experience. White et al report that the ability of passport officers to determine whether two portrait photographs are of the same unfamiliar person is unrelated to the duration of employment: David White et al, 'Passport Officers' Errors in Face Matching' (2014) 9 *PLoS ONE* 1, 2.

⁷¹ In some domains forensic analysts are not conversant with methodological canons and limitations that are notorious in mainstream scientific practice. The historical failure to impose procedures to reduce the risks posed by contextual bias represents a conspicuous example of such an oversight.

⁷² See Jennifer L Mnookin et al, 'The Need for a Research Culture in the Forensic Sciences' (2011) 58 *UCLA Law Review* 725. Dangers are acute where the analyst is an ad hoc expert who may be completely displaced from relevant scientific literatures, standard research methods and notorious problems.

understand or accept them — whether such limitations exist and ought to be taken into consideration.⁷³

4 *Most Legal Criteria and Heuristics*

Conventional admissibility criteria and heuristics — such as formal qualifications, a ‘field’, ‘training, study or experience’, prior legal recognition and admission — do not provide direct insight into validity, error rates and limitations, or proficiency.⁷⁴ This observation is both confronting and revealing. Our admissibility criteria do not focus attention on the validity and reliability of expert evidence. They do not require those adducing and relying upon the opinions of forensic analysts — in the vast majority of proceedings, the prosecutor — to demonstrate that the underlying technique is *demonstrably reliable* and that the analyst is proficient in its use. Rather, admissibility is predicated upon superficial or indirect indicia such as formal qualifications in a relevant (or apparently relevant) ‘field’, whether a court has recognised the ‘field’, and whether the forensic analyst has done this sort of thing before or has been previously admitted (in Australia or elsewhere).⁷⁵ Legal criteria are largely focused on the ‘trappings’ of expertise. They reference and privilege past legal practice, antiquated images of expertise and trial safeguards above formal evaluation and evidence of validity and proficiency.⁷⁶ They are relatively easy to satisfy and therefore facilitate the reception of weak, specula-

⁷³ We should not admit the impressions of individual forensic analysts against the advice of peak scientific and technical organisations, such as the National Academy of Sciences and the National Institute of Standards and Technology.

⁷⁴ We could add the desire to ‘assist the jury’ to the list. If the technique is not valid and limitations are not understood, then it is unlikely that derivative opinion evidence will assist the jury.

⁷⁵ Conspicuous examples are face and body mapping. These ‘fields’ do not have an independent existence beyond the courtroom. Consider the discussion in *R v Tang* (2006) 65 NSWLR 681, 696–703 [55]–[81] (Spigelman CJ). See also the legal recognition of forensic gait analysis in Canada in *R v Aitken* [2012] BCJ No 632 (British Columbia Court of Appeal); and in England in *Otway v The Queen* [2011] EWCA Crim 3 (14 January 2011). See, eg, Emma Cunliffe and Gary Edmond, ‘Gaitkeeping in Canada: Mis-steps in Assessing the Reliability of Expert Testimony’ (2014) 92 *Canadian Bar Review* 327.

⁷⁶ In practice, legal expectations have not changed substantially since *Folkes v Chadd* (1782) 3 Dougl 157; 99 ER 589 was decided in the late 18th century.

tive and unreliable opinions.⁷⁷ Legal criteria facilitate the admission of opinions as expert without requiring proof of relevant expertise.

Unfortunately, these kinds of criteria cast an unfortunate pall over the trial as they often loom large, indeed far too large, in the imaginations and understandings of lawyers and judges. Legal criteria are likely to inform and sometimes dominate examination-in-chief and cross-examination, challenges to the opinions of forensic analysts, as well as efforts to repair credibility and probative value in re-examination. They also tend to dominate the representation, and presumably understanding, of forensic science and medicine evidence on appeal. This might not be completely unexpected where the necessary testing has not been performed and the significance of this and other oversights is not always recognised. Attention to legal factors is likely to create confusion and confound good decision-making. Legal factors also tend to waste time and resources as relatively insignificant, secondary considerations are explored at the cost of attention to more fundamental epistemological issues.⁷⁸

5 Moderating the Expressions and Terminology Available to the Forensic Analyst

Several courts have imposed restrictions on the kinds of expressions that particular types of forensic analyst might use as a kind of admissibility compromise.⁷⁹ Though, there is remarkably little coherence across the expressions used for ostensibly similar comparison techniques.⁸⁰ In most

⁷⁷ In many cases, particularly among the historical forensic sciences (such as latent fingerprint, ballistics, handwriting and trace evidence comparisons), practices and 'standards' were developed in response to legal decisions. Courts provided the ultimate forum and the only external check. Many forensic science institutions were (and are) more attentive to what courts permit than orthodox scientific research and methods.

⁷⁸ Rather than subject techniques to validation and proficiency tests, superficial issues are repeatedly contested in trial after trial.

⁷⁹ So for example, (until recently) those comparing images to assist with identity could only describe similarities, whereas latent fingerprint analysts are allowed to equate a 'match' with positive identification — ie individualisation. DNA analysts, in comparison, are required to present their evidence in probabilistic terms. See Michael J Saks and Jonathan J Koehler, 'The Individualization Fallacy in Forensic Science Evidence' (2008) 61 *Vanderbilt Law Review* 199; Simon A Cole, 'Forensics without Uniqueness, Conclusions without Individualization: The New Epistemology of Forensic Identification' (2009) 8 *Law, Probability & Risk* 233.

⁸⁰ Consider, for example, in relation to images: *R v Tang* (2006) 65 NSWLR 681; *Murdoch v The Queen* (2007) 167 A Crim R 329; *Honeysett v The Queen* (2014) 253 CLR 122. In relation to

cases the terms used to express a forensic analyst's opinion are not based on the results of formal testing, do not draw upon empirical evidence of frequencies, and do not incorporate known limitations, uncertainties and the inescapable risk of error.⁸¹ While judicial attempts to moderate the strength of opinions might reduce some of the excesses, they do not address fundamental questions about whether techniques work, how well they work and whether the analyst is proficient. They do not provide information that would enable the fact-finder to evaluate the qualified opinion and may, somewhat perversely, make a forensic analyst, relying upon a technique that is yet to be validated, appear more reasonable than if left to his or her own devices.

6 *The Apparent Independence or Impartiality of the Forensic Analyst*

It is highly desirable to have access to independent expert opinions.⁸² Independence is generally desirable but it cannot overcome the failure to have formally evaluated a technique. Independence cannot make an impression or interpretation, derived from an untested technique, valid.

In thinking about the (mis)use of independence, we should not overlook the structural disadvantages confronting the accused and their defence lawyers at trial.⁸³ Whereas forensic analysts called by the state tend to enter

voices: *R v Leung* (1999) 47 NSWLR 405; *Li v The Queen* (2003) 139 A Crim R 281; *R v Robb* (1991) 93 Cr App R 161. In relation to fingerprints: *Parker v The King* (1912) 14 CLR 681. In relation to shoe prints: *R v T* [2010] EWCA Crim 2439 (26 October 2010). In relation to glass: *R v Abadom* [1983] 1 All ER 364. In relation to DNA: *R v GK* (2001) 53 NSWLR 317; *Aytugrul v The Queen* (2010) 205 A Crim R 157. In relation to gait: *R v Aitken* [2012] BCJ No 632 (British Columbia Court of Appeal); *Otway v The Queen* [2011] EWCA Crim 3 (14 January 2011). In relation to bite marks: *Lewis v The Queen* (1987) 88 FLR 104; *R v Carroll* (1985) 19 A Crim R 10. Appellate courts (in Australia and England) have not recognised that analysts involved in these and other pattern recognition and comparison cases are involved in structurally similar activities. In consequence, the various reporting requirements, and practices appear incoherent from scientific and methodological perspectives.

⁸¹ DNA evidence is an exception to the first two. Most DNA techniques have been validated and rely on the frequencies of alleles derived from population genetics. Though, importantly, DNA analysts do not incorporate error rates into their evidence. See William C Thompson, Franco Taroni and Colin G G Aitken, 'How the Probability of a False Positive Affects the Value of DNA Evidence' (2003) 48 *Journal of Forensic Science* 47; Jonathan J Koehler, 'On Conveying the Probative Value of DNA Evidence: Frequencies, Likelihood Ratios, and Error Rates' (1996) 67 *University of Colorado Law Review* 859.

⁸² National Research Council, *NAS Report*, above n 19, 24 (recommendation 4).

⁸³ See, eg, Shari Seidman Diamond et al, 'Juror Reactions to Attorneys at Trial' (1996) 87 *Journal of Criminal Law & Criminology* 17.

proceedings (portrayed) as disinterested and highly-experienced experts, having few apparent interests in the instant proceedings, expert witnesses called by the defence are relatively easy to portray as interested and partisan. Defence experts are usually paid for the specific appearance and do not obtain the ‘insulation’ available to the state’s forensic analysts from routine employment and experience.⁸⁴ In many cases defence witnesses are not in forensic practice or part of a forensic institution. Most operate primarily as rebuttal witnesses or ‘consultants’. These consultants undertake independent analysis of the specific trace or sample relatively rarely.⁸⁵ Rather than positive evidence of non-guilt, they usually present methodological insights and criticisms directed at the assumptions, practices and the conclusions of the state’s highly-experienced and apparently disinterested forensic analysts. Such rebuttal experts are relatively easy to stereotype as ‘hired guns’ or ivory tower academics whose testimony can be portrayed as motivated or a counsel of perfection.⁸⁶ Compounding their apparent limitations, rebuttal witnesses called by the defence usually only respond to one part of the prosecution case (or story) — the forensic science or medicine evidence — and may be the only witness called by the defendant.⁸⁷ Where the rebuttal expert is espousing mainstream, and highly pertinent, methodological criticisms of the state’s

⁸⁴ Even consultants engaged by the state also seem to benefit from the alignment. In *Wood v The Queen* (2012) 84 NSWLR 581, the bias attributed to a Crown witness by the Court of Criminal Appeal was not developed at trial. Before Associate Professor Cross published his book, limited disclosure made establishing apparent partisanship quite difficult: see Rod Cross, *Evidence For Murder: How Physics Convicted a Killer* (University of New South Wales Press, 2009). Interestingly, on appeal, the Court was uncertain whether serious professional failures, including flagrant noncompliance with the code of conduct for expert witnesses, should lead to the exclusion of opinion evidence: *Wood v The Queen* (2012) 84 NSWLR 581, 619–20 [728]–[731] (McClellan CJ at CL). The ‘Expert Witness Code of Conduct’ is found in *Uniform Civil Procedure Rules* 2005 (NSW) sch 7, and applies to criminal proceedings by virtue of the *Supreme Court Rules* 1970 (NSW) pt 75 r 3J.

⁸⁵ Sometimes testing by the state exhausts the available samples.

⁸⁶ There may be tactical reasons for not calling a rebuttal expert that have nothing to do with the value of their insights. In addition to the way they might be portrayed (discussed in the body of this article), defence counsel may not want the last evidence the jury hears to come from the prosecutor cross-examining the defence expert. In the context of the trial, inattention to validity and reliability can transform basic evidentiary issues into tactical decisions for the defence.

⁸⁷ The importance of stories or narratives has been a recurrent finding in jury research. See generally Nancy Pennington and Reid Hastie, ‘A Cognitive Theory of Juror Decision Making: The Story Model’ (1991) 13 *Cardozo Law Review* 519.

inadequate forensic science evidence, in the midst of adversarial proceedings this may be very difficult for judges and jurors to appreciate. The alignments imposed by our system of investigation and prosecution, along with continuing reliance on misguided legal heuristics, have a tendency to privilege forensic analysts called by the state and in the process obscure the significance of methodological precepts from the mainstream sciences.⁸⁸

7 *Allowing the Jury to Decide Based on What Transpires at Trial*

The jury may be asked to determine guilt without the kinds of information that would enable them to credibly evaluate the expert evidence. Where the jury is not provided with information about validity, limitations, errors and proficiency, assessment of a forensic analyst's impression is unavoidably speculative. In consequence, the jury will be obliged to guess and in the process may rely upon available, though potentially misleading, heuristics.⁸⁹ In many, perhaps most, cases the resources required to rationally evaluate the expert opinion evidence are not provided to them.⁹⁰ In these circumstances there is always a real (ie non-trivial) risk that the forensic analyst cannot do what he or she claims, that he or she will perform at a level significantly below that suggested, or that his or her opinion will have been contaminated by other evidence or the circumstances in which it was obtained. There is always a real risk that the jury will assign a higher probative value, and possibly a much higher probative value, than the technique can support — ie even taken 'at its highest'.⁹¹ Courts should not reward the state's failure to use orthodox scientific methods and testing protocols by allowing the jury to speculate about validity and weight or to assign any weight they deem appropriate. To

⁸⁸ Commitment to the party-led nature of litigation, along with onerous judicial notice doctrines, tends to prevent judicial engagement with mainstream scientific literatures. But see the dissent by McClellan CJ at CL in *Aytugrul v The Queen* (2010) 205 A Crim R 157, 176–7 [101]–[103]. See also Zoe Ratus, 'A Call for Clarity in the Use of Social Science Research in Family Law Decision-Making' (2012) 26 *Australian Journal of Family Law* 81.

⁸⁹ Amos Tversky and Daniel Kahneman, 'Availability: A Heuristic for Judging Frequency and Probability' (1973) 5 *Cognitive Psychology* 207. See also *Davies v The King* (1937) 57 CLR 170, 180 (Latham CJ, Rich, Dixon, Evatt and McTiernan JJ).

⁹⁰ 'The bottom line is simple: In a number of forensic science disciplines, forensic science professionals have yet to establish either the validity of their approach or the accuracy of their conclusions, and the courts have been utterly ineffective in addressing this problem': National Research Council, *NAS Report*, above n 19, 53.

⁹¹ See Edmond et al, 'Christie, Section 137 and Forensic Science Evidence', above n 60.

leave such opinions to the fact-finder always threatens to be unfairly prejudicial to the defendant because validity, reliability, error rate, the value of experience, the issue of contamination and so on, cannot be positively resolved through cross-examination or rebuttal evidence.

Conventional legal approaches to the opinions of forensic analysts oblige fact-finders to rely on their impressions of techniques in conjunction with other considerations such as qualifications and experience, confidence, demeanour, credibility and the analyst's resilience in response to any cross-examination or criticism. Impressions of the analyst's demeanour, credibility and performance, even in conjunction with other considerations such as apparent plausibility, do not provide adequate grounds for determining whether a technique works or whether the analyst is proficient. They do not provide 'good grounds' for ascribing a probative value to the opinion evidence.⁹² Rather, they require fact-finders, to the extent that they engage with the incriminating opinions of forensic analysts, to trust or defer to the expert, to rely on factors that do not directly address the question of expertise (and validity, reliability or proficiency), or to use other strands of prosecution evidence as a 'makeweight' — this last issue is developed in Part III(C).

8 *Deference to the Forensic Analyst*

The risk of deference to the state's forensic analysts is heightened where limitations and uncertainties are not presented to the fact-finder. Earlier, we saw that deference is conventionally seen as undesirable and at odds with our ostensibly rational approach to fact-finding.⁹³ It seems to follow that we should not allow deference — or attribute comprehension to fact-finders — where requisite testing has yet to be performed and limitations with techniques and derivative opinions are not *known* and so cannot be clearly explained.

In sum, the main point is not that conventional legal criteria are without significance. Rather, it is that in general they do not provide direct insight into

⁹² This language is taken from the definition of knowledge used in *Daubert v Merrell Dow Pharmaceuticals Inc*, 509 US 579, 590 (Blackmun J for the Court) (1993). The definition, but not the United States Supreme Court's conspicuous concern with validation and reliability, was endorsed in *R v Tang* (2006) 65 NSWLR 681, 713–14, [139]–[145] (Spigelman CJ); *Honeysett v The Queen* (2014) 253 CLR 122, 131 [23], 132–3 [27] (French CJ, Kiefel, Bell, Gageler and Keane JJ).

⁹³ See, eg, above nn 13–14 and accompanying text.

probative value, the abilities of analysts, or features of the investigative environment that might have compromised the reliability of the opinion evidence. Legal criteria do not provide consistent guidance and are likely to mislead. They might be used to negate or reject specific opinions, such as where an analyst appears excessively partisan, does not possess qualifications or relevant experience, or implodes under skilled cross-examination. Legal criteria are, however, of less utility when it comes to assigning a positive value to an opinion.⁹⁴

Individually and collectively conventional legal insights do not enable us to determine whether a technique works, how well and in what conditions, whether the conclusion is expressed in appropriate terms, and whether the analyst is genuinely expert. These, it might be thought, are serious constraints on decision-making.

B *The Limits of Trial Safeguards in Response to Forensic Science and Medicine Evidence*

More than a century after we began admitting and relying on forensic medicine, toxicology, handwriting comparisons and latent fingerprint evidence, our judges are yet to ask prosecutors and forensic analysts to provide the kinds of information required to make sense of these and other techniques. Australian judges are yet to incorporate the criteria and rigour recommended by mainstream scientists into their ongoing attempts to regulate the admission and use of *expert* opinions in criminal proceedings.⁹⁵ In consequence, legal standards and practices are not merely antiquated but may actually hinder the development of the forensic sciences and the ability of fact-finders to evaluate the opinions of those admitted as experts.⁹⁶

⁹⁴ Consider, for example, the latest offering from the High Court in *Honeysett v The Queen* (2014) 253 CLR 122. The Court did not provide useful criteria for assessing the admissibility or probative value of expert evidence in criminal proceedings. See Gary Edmond, 'A Closer Look at *Honeysett*: Enhancing our Forensic Science and Medicine Jurisprudence' (2015) 17 *Flinders Law Journal* (forthcoming).

⁹⁵ Edmond, 'The Admissibility of Forensic Science and Medicine Evidence', above n 5.

⁹⁶ Because forensic science institutions are very sensitive to legal signals, only exclusion (or perhaps the threat of exclusion) is likely to encourage analysts to undertake appropriate evaluation, standard-setting, proficiency testing and to develop empirically-based reporting practices.

Prosecutors and judges have directed insufficient attention to the symbolic significance of adduction, admission and reliance by the state. The state's reliance on a forensic analyst or technique, especially where the technique is endorsed during examination-in-chief and in summaries, is highly suggestive of validity and reliability. By admitting (or upholding the admission of) the opinions of forensic analysts not *known* to be expert, trial and appellate judges have lent the imprimatur of their courts to techniques of unknown value. In isolation, and without the resources or authority of the state, the defendant is somehow expected to identify actual errors (remote from the crime scene and the laboratory) and expose misrepresentations and exaggerations and persuade the jury of the significance of methodological limitations and omissions.⁹⁷ This might be done, but most defence lawyers are neither sufficiently trained nor resourced for the undertaking and the quotidian adversarial trial is not a particularly conducive forum. The trial is not a level playing field and there is equality of arms between the prosecution and defence in only a tiny proportion of contested cases.⁹⁸ Prosecutors and judges should be much more attentive to the dangers created by the admission of speculative and unreliable opinions.

Adding to the difficulties confronting criminal defendants, the prosecution obtains the significant advantage from the ever-present possibility of deference to the state's forensic analyst and the illegitimate use of other strands of evidence as a 'makeweight' to overcome fundamental frailties with *expert* opinions — see Part III(C). Our current approach to forensic science evidence adduced by the state, given the lack of interest in validity and reliability (and evidence of genuine expertise) for the purposes of determining admissibility, shifts the main evidentiary contest from the *voir dire* to the trial and in effect burdens the defendant. Rather than require the state's *experts* to demonstrate that their techniques work, incriminating opinions are admitted (sometimes without qualification) and the defence assumes responsibility for teasing out

⁹⁷ The defence is often obliged to try to identify (ie piece together) mistakes and problems retrospectively, displaced in time and space from the circumstances in which they occurred. Consider F H R Vincent, 'Inquiry into the Circumstances That Led to the Conviction of Mr Farah Abdulkadir Jama' (Victorian Government Printer, 29 May 2010).

⁹⁸ Conventional commitments may make the trial appear balanced, but if the safeguards do not work well or consistently then equality of arms might be a misnomer or a mere possibility. Defence access to a rebuttal witness may, somewhat perversely, assuage concerns by allowing the contest to go before the jury rather than requiring the trial judge to consider and exclude insufficiently reliable incriminating opinion evidence. See, eg, *R v Jung* [2006] NSWSC 658.

and explaining limitations to the fact-finder during the course of an adversarial proceeding. The fact that judges have not been sufficiently sensitised to the significance of epistemic problems means that they are unlikely to endorse, and in many cases incapable of explaining, serious limitations through the summing up, directions and warnings.⁹⁹

Our current approaches to *expert* opinions, especially opinion evidence adduced by the prosecutor, are not consistent with principle. Unnecessarily, they place onerous burdens on the defendant and the lay fact-finders. Even though in theory the burden of proof rests with the state, we allow the prosecutor to adduce opinions — routinely obtained or presented in ways that are inattentive to notorious risks and authoritative advice. Rather than trust the state and its *experts*, our admissibility rules and practice should only allow those who are known to be experts to assist the jury.

In the absence of formal evaluation, forensic analysts and investigators should not be encouraged to express their opinions. There are real risks of deference, undue influence, subversion of the trial process and fact-finding that are difficult to identify or rectify through our ordinary appeal processes. And, as we shall see, where incriminating opinions are admitted, there are real dangers that the jury will use apparently independent evidence to fill gaps and oversights in the *expert* evidence.

C *The Apparent Strength of the Case and the Unrecognised Problem of 'Masking'*

The thesis casts light on important issues that have been largely overlooked in conventional legal proceedings.¹⁰⁰ Once it is recognised that specific kinds of information are required to evaluate the opinions of those presented as expert witnesses, the roles played by other evidence and the circumstances in which

⁹⁹ Conventional directions and warnings do not seem to be effective, and there are relatively few examples of informed judicial advice on forensic science and medicine evidence. See generally Joel D Lieberman and Bruce D Sales, 'The Effectiveness of Jury Instructions' in Walter E Abbott and John Batt (eds), *A Handbook of Jury Research* (American Law Institute, 1999) 18; James R P Ogloff and V Gordon Rose, 'The Comprehension of Judicial Instructions' in Neil Brewer and Kipling D Williams (eds) *Psychology and Law: An Empirical Perspective* (Guilford Press, 2005). See also the example in A Leo Levin and Robert J Levy, 'Persuading the Jury with Facts Not in Evidence: The Fiction-Science Spectrum' (1956) 105 *University of Pennsylvania Law Review* 139, 168.

¹⁰⁰ See Commonwealth, above n 24, 331.

the opinion was obtained, emerge in ways that have not been considered by trial and appellate courts. In most cases, trial and appellate judges seem to have been reassured (and fact-finders persuaded) by the fact that the incriminating opinion of a forensic analyst appears *consistent with* and *implicitly independent of* other strands of incriminating evidence.¹⁰¹ The apparent strength of the case (or the consistency of the opinion with other inculpatory evidence) is routinely used to support admission, and sometimes to excuse the admission of opinions of unknown probative value. For reasons explained below, relations between different strands of evidence, and the use of other evidence to mediate the reception and use of incriminating opinions, are more complex than is routinely suggested. Our conventional approaches tend to *mask* both the limits of the opinion and the interrelatedness of different strands of evidence (and other information and suspicions informing investigations and analysis) because of the manner in which investigations and forensic analyses are practised and re-presented.

The problem with our approach is that if the technique has not been evaluated, the apparent strength of the case against the accused does not provide a warrant for admitting and relying on opinion evidence of unknown provenance, or treating the state's forensic analysts (and other investigators) as experts. The strength of the case, and the apparent guilt (or conviction) of the accused does not operate as a *de facto* form of validation and should not be used as a makeweight (for admissibility, at least). The techniques and derivative opinions produced by forensic analysts must be able to stand on their own. In terms of assessing the value of techniques and incriminating opinions, it does not matter if the case is otherwise strong or even compelling. Perceived strength should not be used to excuse oversights or as a basis for admitting incriminating opinions based on techniques that have not been formally evaluated.¹⁰²

¹⁰¹ This Part reinforces problems with approaches that accept the ability to ascertain the 'gist' or 'direction' of the opinion evidence. This is a problem for studies that rely on correspondence between the impressions of judges and juries — discussed below in Part IV(A).

¹⁰² This other information cannot directly support validity or reliability. We should not, in addition, overlook the fact that in many cases the expert opinion will have been selected and the evidence adduced because it supported the suspicions of investigators (of which the analyst was apprised) and the case advanced by the prosecutor. See also Emma Cunliffe, 'Judging Fast and Slow: Using Decision-Making Theory to Explore Judicial Fact Determination' (2014) 18 *International Journal of Evidence & Proof* 139.

These difficulties are compounded by the fact that in many investigations the forensic analyst is unnecessarily exposed to other strands of evidence or other prejudicial information (that may or may not be admitted, such as the criminal record of the suspect or the beliefs of an investigating police officer).¹⁰³ In most criminal trials, the dangers posed by contextual bias and cross-contamination are not identified by prosecutors, forensic analysts or defence lawyers and it is unlikely that they are considered by fact-finders. This means that in many trials and appeals, strands of evidence are treated as independent corroboration of guilt when in reality they are not independent and their ability to corroborate might be profoundly compromised. In some cases incriminating opinions will result from suggestion or cognitive contamination rather than the underlying data. The failure to validate and control for contextual bias means that we cannot answer the question of how probative the opinion is, or whether it has been contaminated by other information or the process of its generation.

In practice, inattention to these issues means that opinions are treated as both expert and independent when they may be neither. The failure to consider the validity and reliability of techniques and derivative opinions, or the possession of relevant expertise, as part of the admissibility determination or the application of mandatory and discretionary exclusions means that speculative opinions and impressions are frequently (mis)represented, and presumably treated, as independent and reliable expert evidence. This evidence might be probative, but it might not be. The problem is that we have no means of knowing. Only the knowledge generated through formal evaluation enables us to address our ignorance. In its absence, and in circumstances where analysts are not shielded from suggestive information and processes, we should not defer to purported expertise, treat forensic science evidence as independent, or allow the fact-finder to assign — really guess at — a value.

Analysts using speculative and unreliable techniques will sometimes produce answers that are consistent with factual guilt. Knowledge of other strands of evidence and the suspicions and impressions of investigators will not invariably lead to error, and in many cases will actually render an opinion more probative than it might otherwise appear. In some cases this will be, at

¹⁰³ See Gary Edmond et al, 'Contextual Bias and Cross-contamination in the Forensic Sciences: The Corrosive Implications for Investigations, Plea Bargains, Trials and Appeals' (2014) 13 *Law, Probability & Risk* 1.

least in part, a function of the forensic analyst's familiarity with suggestive information about the suspect (or awareness of the *desired* answer) rather than a reflection of the value of the technique or the analyst's ability. Nevertheless, unless they can demonstrate that exposure to information and suggestive processes do not unnecessarily increase their level of error, forensic analysts should be shielded from gratuitous information and direct communication with investigators.¹⁰⁴ The defence should not bear the responsibility, difficulty, cost and risk of having to retrospectively unpack and successfully convey these dangers to decision-makers when they can and should generally be avoided.¹⁰⁵

In the end, the apparent strength of the case cannot overcome a lack of information about the validity of a technique and the proficiency of the analyst. Relying on the apparent strength of the case tends to elide the methodological failings and the inability to rationally evaluate the opinion. And, to the extent that forensic analysts are influenced by other evidence, or the suspicions of investigators, their opinions are not independent (and vice versa). The apparent strength of the case and inattention to the validity and reliability of techniques have a tendency to mask the limits and uncertainties associated with the opinions of forensic analysts. Yet, they are often used in trials and appeals to suggest that any dangers from unreliability or error are slight or legally insignificant. This Part, in contrast, illustrates why we need to be more careful about the way we combine evidence. By not attending to the probative value of incriminating opinions, which includes the circumstances in which an analyst's opinion was formed, prosecution cases may appear stronger than they actually are.

For admission, expert opinion evidence must stand or fall on its own. Techniques must be valid and reliable and analysts must be proficient in their use. Unless these conditions are satisfied, opinions should not be admitted and combined with other evidence. In effectively overlooking validity, error rates and limitations, proficiency *and independence* we deprive our investiga-

¹⁰⁴ There are ways to facilitate this, such as designating one analyst to liaise with investigators to shield the primary analyst from suggestion and other gratuitous information. See also Dan E Krane et al, 'Sequential Unmasking: A Means of Minimizing Observer Effects in Forensic DNA Interpretation' (2008) 53 *Journal of Forensic Sciences* 1006. The Victorian Police Forensic Science Service has begun to instigate such processes: Office of the Chief Forensic Scientist, 'Message from the Chief Forensic Scientist', *OCFS Annual Summary* (2014) 4–5.

¹⁰⁵ Edmond et al, 'Contextual Bias and Cross-contamination in the Forensic Sciences', above n 103, 18.

tors, prosecutors and fact-finders of independent and reliable expert opinions that might confirm or challenge their suspicions. Inattention to these issues means that in cases where the suspicions of investigators or the incriminating opinion is in fact wrong (or misleading) we have few means of exposing it or gauging its likelihood.

IV JURY COMPETENCE: JURY RESEARCH AND LEGAL ATTITUDES

Perhaps the major potential objection to my thesis, and its implicit threat to criminal trials in their current form, is the contention that jury research and *the experience of the law* confirm that in practice fact-finders perform reasonably well when presented with scientific and technical evidence.¹⁰⁶ For a variety of reasons, several of which are developed below, on closer examination experimental research does not unequivocally support this claim and some of the main justifications appear weaker than proponents have suggested. Significantly, there is no evidence that fact-finders are able to overcome information deficits or untangle different strands of evidence and cross-contamination.

A Research on Expert Evidence and the Common Law Jury

Empirical research indicates that jurors understand the adversary process, that they do not automatically defer to the opinions of experts, and that their verdicts appear to be generally consistent with external criteria of performance.¹⁰⁷

No studies have been conducted that comprehensively monitor *actual* comprehension levels.¹⁰⁸

Some of the most commonly cited evidence in support of the ability of juries to understand scientific and technical evidence — sometimes characterised as *jury competence* — is drawn from observational studies, exit interviews and a series of mock juror experiments. Superficially these studies might appear to

¹⁰⁶ For an overview, see Neil Vidmar et al, 'Amicus Brief: *Kumho Tire v Carmichael*' (2000) 24 *Law and Human Behavior* 387; Neil Vidmar, 'Expert Evidence, the Adversary System, and the Jury' (2005) 95 *American Journal of Public Health* S137; Neil Vidmar and Shari Seidman Diamond, 'Juries and Expert Evidence' (2001) 66 *Brooklyn Law Review* 1121.

¹⁰⁷ Vidmar, 'Expert Evidence, the Adversary System, and the Jury', above n 106, S137.

¹⁰⁸ Horan, above n 1, 71 (emphasis in original).

confirm that juries understand and evaluate forensic science and medicine evidence. In reality, the findings are far more circumscribed than is usually posited. Consider, for example, claims around the significance of correspondence between jury verdicts and judicial impressions of appropriate outcomes.

Numerous investigators and jury proponents have relied on the fact that, when surveyed, judges and jurors tend to agree on the verdict.¹⁰⁹ High levels of agreement are usually taken as a surrogate for the ability of jurors to credibly engage with and even understand scientific and technical evidence. The fact that the rate of disagreement between jury verdicts and judicial ‘verdicts’ does not appear to vary significantly across trials, including those classified by lawyers and judges (and investigators) as complex or involving difficult evidentiary issues, provides reassurance to those who rely on convergence as a sign of comprehension and competence.

On inspection, claims about judge–jury correspondence are not particularly informative and provide very limited insight into their respective abilities to rationally evaluate expert opinion evidence. For a start, analysts and proponents tend to substitute verdict correspondence for understanding and evaluation without necessarily knowing the reasons for the ‘agreement’. We do not know, for example, whether the judge and jury actually agreed about the value or significance of the opinion evidence. This is compounded by the possible influence (whether deliberate or inadvertent) that a judge might exert on the jury and its assessment of the evidence or the verdict during the course of the trial — through responses and demeanour, rulings, directions and the summing up. These kinds of potential influences, which mean that jury verdicts and judicial impressions are not independent, tend to be overlooked because they are difficult and resource-intensive to monitor. Additionally, the fact that in most cases judges have access to inadmissible and usually prejudi-

¹⁰⁹ See, eg, Harry Kalven Jr and Hans Ziesel, *The American Jury* (Little Brown, 1966). The primary authors of the seminal Chicago Jury Project summarised the correspondence in the following terms: ‘The result is a stunning refutation of the hypothesis that the jury does not understand’: at 157. See also Valerie P Hans and Neil Vidmar, ‘*The American Jury at Twenty-Five Years*’ (1991) 16 *Law & Social Inquiry* 323; Paula L Hannaford, Valerie P Hans and G Thomas Munsterman, ‘Permitting Jury Discussions during Trial: Impact of the Arizona Reform’ (2000) 24 *Law and Human Behavior* 359; Larry Heuer and Steven Penrod, ‘Trial Complexity: A Field Investigation of Its Meaning and Its Effects’ (1994) 18 *Law and Human Behavior* 29, 46–9; Theodore Eisenberg et al, ‘Judge–Jury Agreement in Criminal Cases: A Partial Replication of Kalven and Ziesel’s *The American Jury*’ (2005) 2 *Journal of Empirical Legal Studies* 171.

cial information about the defendant that is not properly before the jury, such as previous convictions or an inadmissible confession, means judges and juries are not always making decisions on the basis of the same information.¹¹⁰ Each of these issues threatens the significance we can attach to correspondence and explanations based on capability or competence. However, these methodological problems are compounded, indeed dwarfed, by the concerns at the heart of this article.

Of the cases studied for correspondence, the kinds of information required for rational evaluation were not available in the vast majority.¹¹¹ To make things clear, relevant information was available to neither the judge nor the jury. The lack of relevant information raises fundamental problems with many observational studies, exit surveys and mock juror experiments. Where decision-makers are not provided with the information required to make sense of the opinion evidence, we should not be particularly impressed by correspondence or claims about comprehension. Rather, many decision-makers seem to be oblivious to the significance of omissions and their own ignorance.

What the studies might allow us to infer is that judges and juries are probably using the same kinds of heuristics (and peripheral processing) to assess opinions presented as 'expert' in relation to what transpires at trial in the context of the overall case.¹¹² While judges and jurors may converge on the direction and strength of the evidence (or the verdict), they are probably oblivious to epistemic problems plaguing many forensic sciences.

Another type of evidence used to suggest that jurors are responsible and basically competent in response to proffers of expert opinion evidence is the results of juror exit surveys. These usually involve some kind of post-verdict interview or questionnaire. Although a potentially valuable source of information, there are obvious limitations with ex post facto reconstructions and rationalisations of past behaviour. When conducted in ways that are sensitive

¹¹⁰ In some studies the judge's 'verdict' was obtained via self-reporting, so depending on when the result was recorded it might have been influenced by knowledge of the jury decision.

¹¹¹ It could not have been provided to judges and juries because in most cases it did not exist.

¹¹² This seems consistent with research by Joel Cooper, Elizabeth A Bennett and Holly L Sukel, 'Complex Scientific Testimony: How Do Jurors Make Decisions?' (1996) 20 *Law and Human Behavior* 379; Joel Cooper and Isaac M Neuhaus, 'The "Hired Gun" Effect: Assessing the Effect of Pay, Frequency of Testifying, and Credentials on the Perception of Expert Testimony' (2000) 24 *Law and Human Behavior* 149.

to the issues and evidence presented in the particular trial there is some scope to expose both understanding and ignorance of forensic science and medicine evidence, to the extent that limitations were constructively explored during proceedings or known to the investigator. Typically, post-verdict surveys tend to treat the expert evidence and engagement with it in rather superficial ways. There is an understandable tendency to limit analysis of the expert evidence to the way it was presented at trial or to ask jurors to report on whether they understood the evidence or found it challenging. Because constructive engagement occurs in relatively few trials, unless jurors leave the courtroom aware that they were not provided with information that would have enabled them to evaluate the *expert* opinion evidence, then we cannot be particularly confident about understanding and performance, even where self-reports are generally positive.¹¹³ More robust studies of juries (and the effectiveness of trial mechanisms) might contrast post-verdict jury impressions with what is known about techniques and opinions beyond the courtroom.¹¹⁴

¹¹³ We can be more confident about self-reports that express difficulty or a lack of understanding. A New Zealand study reported five cases among its sample (of 48, though fewer than half the cases in the sample featured expert evidence) where jurors candidly acknowledged that they did not 'grasp the evidence'. The report suggests that some of these cases resulted in perverse, compromised and hung verdicts: see 'New Zealand Law Commission, *Juries in Criminal Trials: Part Two*, Preliminary Discussion Paper No 37 (1999) vol 1, 52–7, vol 2, 19, 25–7. The American Bar Association sponsored a study of four complex cases. The report found that jurors experienced difficulty with some of the evidence, but nevertheless concluded that jurors were not improperly influenced by the expert evidence, recognised points of expert agreement, and were critical of the evidence presented by those who appeared to be 'hired guns': Special Committee on Jury Comprehension (ABA), *Jury Comprehension in Complex Cases* (1989) iii, 29–31, 40, 42. See also Richard Lempert, 'Civil Juries and Complex Cases: Taking Stock after Twelve Years' in Robert E Litan (ed), *Verdict: Assessing the Civil Jury System* (Brookings Institution, 1993) 181, 183–228.

¹¹⁴ See also Molly Selvin and Larry Picus, *The Debate over Jury Performance: Observations from a Recent Asbestos Case* (Rand, 1987); John S DeWitt, James T Richardson and Lyle G Warner, 'Novel Scientific Evidence and Controversial Cases: A Social Psychological Examination' (1997) 21 *Law & Psychology Review* 1. There are, in addition, studies contrasting jury verdicts with the pre-litigation advice of medical doctors advising insurance companies in relation to claims. Investigators reported a statistically significant correspondence between physician assessments of negligence and jury verdicts. See Mark I Taragin et al, 'The Influence of Standard of Care and Severity of Injury on the Resolution of Medical Malpractice Claims' (1992) 117 *Annals of Internal Medicine* 780; Frank A Sloan et al (eds), *Suing for Medical Malpractice* (University of Chicago Press, 1993); Henry S Farber and Michelle J White, 'Medical Malpractice: An Empirical Examination of the Litigation Process' (1991) 22 *Rand Journal of Economics* 199. These studies are not without value, but correspondence cannot be equated

Interestingly, one small study did precisely this. Sanders interviewed jurors after they gave a verdict in favour of a plaintiff in a civil trial where it was alleged that the anti-nausea drug Bendectin, ingested during the first trimester of the mother's pregnancy, had caused birth defects.¹¹⁵ Drawing upon a fairly strong consensus among research scientists, where epidemiological studies and meta-analyses of epidemiological studies are widely accepted as the most robust form of evidence on the issue of causation,¹¹⁶ Sanders' interviews revealed the divergent ways jurors valued the expert evidence presented at trial. In addition to the epidemiological and secular trend evidence,¹¹⁷ this included in vitro and in vivo studies and the opinions of treating clinicians.¹¹⁸ The jurors interviewed in Sanders' study (who seem to have made a factually wrong decision, because it is unlikely that Bendectin was the cause) did not seem to accept (or apparently appreciate) the relative strength of the epidemiological evidence and the significance of the secular trend data. In the aftermath of an emotionally demanding trial around a child plaintiff suffering from serious birth defects, the jurors described expert witness effectiveness in ways that were not consistent with the probative value

with jury understanding. In addition, assigning liability, like guilt, might not always be dependent on, or directly related to, the scientific and medicine evidence.

¹¹⁵ Joseph Sanders, 'Jury Deliberation in a Complex Case: *Havner v Merrell Dow Pharmaceuticals*' (1993) 16 *Justice System Journal* 45.

¹¹⁶ This normative approach is endorsed by influential scientific and technical organisations. The Cochrane Collaboration, the leading repository of medical research, places different types of medical evidence (eg meta-analyses, randomised clinical trials, case-control studies, clinical opinion and other types of studies) in a formal hierarchy — with 'expert opinion' on the lowest rung. See also the formal response to research and data suggested by The Intergovernmental Panel on Climate Change.

¹¹⁷ The number of birth defects did not appear to be correlated with the appearance and withdrawal of Bendectin from the United States market: Sanders, above n 115, 51–4.

¹¹⁸ See also Carl F Cranor, *Toxic Torts: Science, Law, and the Possibility of Justice* (Cambridge University Press, 2006). It may be that the different, and more diffuse, policy goals associated with tort and product liability litigation — such as deterrence and compensation — might warrant more liberal approaches toward admissibility in civil proceedings. Liberal admission of expert opinions might, for example, help to encourage manufacturers to undertake research or avoid exposing customers, workers and others to potentially risky products and environments. See generally Gary Edmond and Mehera San Roque, 'Just(.) Quick and Cheap: Do We Need More Reliable Expert Evidence in Civil Proceedings?' in Michael Legg (ed), *The Future of Dispute Resolution* (LexisNexis Butterworths, 2013) 72.

of the techniques underlying the various opinions.¹¹⁹ The wider history of Bendectin litigation suggests that United States state and federal judges — who sometimes upheld jury verdicts for plaintiffs on issues of general and specific causation — were not always appreciative of the relative merits of the different types of expert evidence relied upon by the parties.¹²⁰

Another strand of inquiry, namely studies of mock jurors, has also been used to support the competence of juries.¹²¹ These studies are often read in ways that suggest jurors do not defer to experts or abandon their decision-making responsibilities when confronted with complex scientific and technical evidence.¹²² Once again, upon closer examination, many of these studies perpetuate the kinds of misguided assumptions criticised in this article, especially those discussed in Part III(A). In studies focused on expert evidence, understanding or critical engagement is frequently credited to jurors who recognise, on the basis of exposure to a limited script (or video or role-play), that one of the mock experts is aligned with a party or possesses an interest in the case. Other forms of creditable engagement include sensitivity to the *expert's* formal qualifications and experience. Superficial engagement and issue spotting are used as proxies for critical engagement, comprehension

¹¹⁹ It may be that the jurors decided on other grounds, such as sympathy for the plaintiff, the availability of insurance or corporate malfeasance, but that hardly assists our concern with jury comprehension of scientific and technical evidence: Shari Seidman Diamond and Neil Vidmar, 'Jury Room Ruminations on Forbidden Topics' (2001) 87 *Virginia Law Review* 1857.

¹²⁰ Joseph Sanders, *Bendectin on Trial: A Study of Mass Tort Litigation* (University of Michigan Press, 1998); Michael D Green, *Bendectin and Birth Defects: The Challenges of Mass Toxic Substances Litigation* (University of Pennsylvania Press, 1996).

¹²¹ There are, in addition, several studies of real jurors engaged in actual deliberations (mainly from Arizona): see, eg, Scott E Sundby, 'The Jury as Critic: An Empirical Look at How Capital Juries Perceive Expert and Law Testimony' (1997) 83 *Virginia Law Review* 1109. These studies do not suggest that jurors have acute methodological sensitivities and are able to overcome informational deficits. Rather, they have tended to focus on civil trials or sentencing in criminal proceedings. The expert evidence in many of the criminal proceedings was psychiatric opinions offered by non-treating physicians in the sentencing phase. There are a variety of reasons why jurors, especially jurors in capital cases, might not be particularly well disposed to such psychiatric evidence.

¹²² Though some of these studies suggest that when confronted with complex evidence, laypersons often rely on 'peripheral cues' and broader contextual factors. See Horan, above n 1, 61; Cooper, Bennett and Sukel, above n 112, 382; Ryan J Winter and Edith Greene, 'Juror Decision-Making' in Francis T Durso (ed), *Handbook of Applied Cognition* (John Wiley & Sons, 2nd ed, 2007) 739.

and competence.¹²³ Evidence of actual understanding and questions around the ability to evaluate expert evidence tend to be elided.¹²⁴ Like exit surveys, that rely on the respondent's recollection or engagement with the evidence presented at trial, these studies tend to assume that superficial engagement is meaningful and sufficient to demonstrate competence or facilitate rational decision-making. Such readings tend to misunderstand both the informational needs and the performance of (mock) jurors. They tend to reproduce, albeit faintly, the unsatisfactory conditions of ordinary trials where jurors are obliged to rely on inappropriate and frequently misleading heuristics or defer to the authority of those presented as experts.

There is, in addition, a smaller set of mock juror studies that provide more direct insights into the ability of respondents to identify methodological limitations and other problems with expert evidence. These studies tend to

¹²³ Daniel W Shuman, Anthony Champagne and Elizabeth Whitaker, 'Assessing the Believability of Expert Witnesses: Science in the Jurybox' (1996) 37 *Jurimetrics Journal* 23, 30; Daniel W Shuman, Elizabeth Whitaker and Anthony Champagne, 'An Empirical Examination of the Use of Expert Witnesses in the Courts — Part II: A Three City Study' (1994) 34 *Jurimetrics Journal* 193. In another study, Shuman and Champagne conducted telephone surveys of citizens who had served as civil jurors: Daniel W Shuman and Anthony Champagne, 'Removing the People from the Legal Process: the Rhetoric and Research on Judicial Selection and Juries' (1997) 3 *Psychology, Public Policy and Law* 242, 255. They concluded that jury performance was rational because jurors relied upon credentials, recollection of the facts of the case, quality of reasoning and perceptions of impartiality and believability (ie credibility) to assess the evidence:

We did not find evidence of a 'white coat syndrome' in which jurors mechanistically deferred to certain experts because of their field of expertise. Instead we found jurors far more skeptical and demanding in their assessments. Jurors made expert-specific decisions based on a sensible set of considerations — the expert's qualifications, reasoning, factual familiarity and impartiality. Our data do not lend support to the critics who paint jurors as gullible, naïve or thoughtless persons who resort to irrational decision-making strategies that they rely on superficial considerations.

Similarly, in a study by Vidmar, the ability to 'articulate the main medical issues in the case and recognize the basic points made by the opposing experts' along with attention to motives of witnesses and 'incompleteness' of the record were considered as evidence of competence: Neil Vidmar, *Medical Malpractice and the American Jury: Confronting the Myths about Jury Incompetence, Deep Pockets, and Outrageous Awards* (University of Michigan Press, 1995) 127–60. See also Sanja Kutniak Ivković and Valerie P Hans, 'Jurors' Evaluation of Expert Testimony: Judging the Messenger and the Message' (2003) 66 *Law & Social Inquiry* 441.

¹²⁴ Because of the artificial nature of the enterprise, in some mock jury studies there is no way to rationally evaluate the opinion evidence and the (legal) heuristics relied upon by the investigator do not enable the underlying research question to be answered. See also Diamond and Vidmar, above n 119, 1859–64.

confirm that respondents are either insensitive or insufficiently sensitive to substantial methodological limitations and oversights, even when they are identified and explained.¹²⁵ Across a range of studies, it is uncommon for mock jurors to unilaterally recognise serious methodological defects and oversights.¹²⁶ Similarly, studies that provide opportunities for mock jurors to moderate their evaluation of evidence in ways that are consistent with a range of normative expectations, especially around responses to probabilistic or statistical evidence, have not produced particularly reassuring results.

To be fair, limitations within empirical studies have not passed without comment. Helpfully, a New Zealand Law Commission discussion paper acknowledged

practical problems in identifying juror incompetence in empirical research. For example, determining whether what appears to be difficulty in comprehension is due to the jurors' inability or due to the failure by counsel to present matters clearly.¹²⁷

¹²⁵ Margaret Bull Kovera and Bradley D McAuliff, 'The Effects of Peer Review and Evidence Quality on Judge Evaluations of Psychological Science: Are Judges Effective Gatekeepers?' (2000) 85 *Journal of Applied Psychology* 574; Margaret Bull Kovera et al, 'Expert Testimony in Child Sexual Abuse Cases: Effects of Expert Evidence Type and Cross-Examination' (1994) 18 *Law and Human Behavior* 653; Margaret Bull Kovera, Bradley D McAuliff and Kellye S Hebert, 'Reasoning about Scientific Evidence: The Effects of Juror Gender and Evidence Quality on Juror Decisions in a Hostile Work Environment Case' (1999) 84 *Journal of Applied Psychology* 362, 366–7, 371–2; Lora M Levett and Margaret Bull Kovera, 'The Effectiveness of Opposing Expert Witnesses for Educating Jurors about Unreliable Expert Evidence' (2008) 32 *Law and Human Behavior* 363; Cooper, Bennett and Sukel, above n 112, 381–2; Julie A Buck, Kamala London and Daniel B Wright, 'Expert Testimony Regarding Child Witnesses: Does It Sensitize Jurors to Forensic Interview Quality?' (2011) 35 *Law and Human Behavior* 152. See also Jason Schklar and Shari Seidman Diamond, 'Juror Reactions to DNA Evidence: Errors and Expectancies' (1999) 23 *Law and Human Behavior* 159.

¹²⁶ See, eg, Brian L Cutler, Hedy R Dexter and Steven D Penrod, 'Expert Testimony and Jury Decision Making: An Empirical Analysis' (1989) 7 *Behavioral Sciences & the Law* 215. This suggests that if issues are not raised and adequately explained during the trial they are unlikely to be considered by fact-finders.

¹²⁷ New Zealand Law Commission, above n 113, vol 1, 52 [220]. See also Richard O Lempert, 'Civil Juries and Complex Cases: Let's Not Rush to Judgement' (1981) 80 *Michigan Law Review* 68; Vidmar, 'Expert Evidence, the Adversary System, and the Jury', above n 106, 142; William W Schwarzer, 'Reforming Jury Trials' (1991) 132 *Federal Rules Decisions* 575, 586–8.

Similarly, Sanders reminds us that it is important

to distinguish mistaken verdicts from irrational ones. ... In many complex cases, any inadequacies in the jury deliberation are masked by the fact that a verdict for either party appears to be reasonable. Absent some insight into the deliberation process (e.g., jury mistakes in answering special instructions or, more reliably, post-verdict interviews with jurors), we will not know about jury shortcomings.¹²⁸

Unfortunately, such qualifications and insights are not routinely incorporated into rather glib summaries.

Judge-jury correspondence and exit surveys conducted on trials where information about the validity and reliability of techniques and proficiency (and other issues) was not presented do not provide grounds for maintaining confidence in jury (or judicial) performances. In such conditions, judge-jury correspondence suggests that by and large these groups tend to be oblivious to the question of whether the necessary information has been provided, and instead rely on the direction of the evidence and a range of epiphenomena associated with *expert* opinions (see above Part III(A)) and the case (see above Part III(C)). Even 'sceptical' judges and juries seem to be incapable of overcoming, and in many cases recognising, the dearth of relevant information.¹²⁹ Moreover, there is little evidence that judges or juries are disaggregating evidence or are aware of the potential for non-independence and cross-contamination. There are no studies, and little evidence, to suggest that judges or juries will recognise or somehow 'allow for' methodological oversights and other problems. The absence of information does not invariably lead juries to ask insightful questions and has not always sufficiently alerted judges to require that pertinent questions be answered.¹³⁰ On reconsideration, judge-jury research and most exit surveys might be read in ways that

¹²⁸ Sanders, above n 115, 46 (citations omitted).

¹²⁹ On 'sceptical' effects, see Shuman and Champagne, above n 123.

¹³⁰ Where juries do ask about relevant issues, legal responses are frequently unsatisfactory because the relevant information is not available. This often leads trial judges to trivialise the issue or allow analysts to speculate in order to provide some 'assistance'. See, eg, *R v Tang* (2006) 65 NSWLR 681, 701-3, [74]-[81], 707-8 [106]-[110] (Spigelman CJ); *Morgan v The Queen* (2011) 215 A Crim R 33, 50-61 [102]-[146] (Hidden J); *Penza v The Queen* [2013] NSWCCA 21 (15 February 2013) 8-10 [17]-[23] (Hoeben JA); *Kosian v The Queen* (2013) 237 A Crim R 156, 168-71 [47]-[59] (Redlich JA).

confirm the poor performance of trial personnel in response to proffers of expert testimony.¹³¹

The studies of most direct application to the contemporary forensic sciences are those concerned with the ability of laypersons to understand probabilistic and statistical evidence.¹³² An extensive body of research using real jurors (and judges), mock jurors as well as experimental subjects in non-legal contexts, confirm the difficulties citizens confront when presented with probabilistic and statistical evidence.¹³³ As forensic science and medical techniques are gradually validated, we will begin to have access to empirically-derived information on the value of techniques, error rates and limitations, proficiency and frequencies. Much of the resulting information will be probabilistic in form. Ironically, the very information required to make rational use of expert opinions will often be available in forms that juries and

¹³¹ This does not mean that judges and juries have produced factually mistaken verdicts, but rather that their verdicts, to the extent that they rely on forensic science and medicine evidence, might not be the result of understanding and rational evaluation.

¹³² See, eg, Hans et al, above n 44, 61; see generally Richard E Nisbett (ed), *Rules for Reasoning* (Lawrence Erlbaum Associates, 1993). Laypersons typically function well in familiar or everyday situations, but often encounter difficulties in new settings or circumstances.

¹³³ Jonathan J Koehler, Audrey Chia and Samuel Lindsey, 'The Random Match Probability in DNA Evidence: Irrelevant and Prejudicial?' (1995) 35 *Jurimetrics* 201; 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; Jonathan J Koehler, 'When Are People Persuaded by DNA Match Statistics?' (2001) 25 *Law and Human Behavior* 493; Schklar and Diamond, above n 125, 178–80; Shari Seidman Diamond and Jonathan D Casper, 'Blindfolding the Jury to Verdict Consequences: Damages, Experts and the Civil Jury' (1992) 26 *Law & Society Review* 513; Suzanne O Kaasa et al, 'Statistical Inference and Forensic Science: Evaluating a Bullet Lead Match' (2007) 31 *Law and Human Behavior* 433; 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; Nicholas Scurich and Richard S John, 'Mock Jurors' Use of Error Rates in DNA Database Trawls' (2013) 37 *Law and Human Behavior* 424, 429; William C Thompson, 'Are Juries Competent to Evaluate Statistical Evidence?' (1989) 52(4) *Law & Contemporary Problems* 9; William C Thompson and Edward L Schumann, 'Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy' (1987) 11 *Law and Human Behavior* 167; David L Faigman and A J Baglioni, 'Bayes' Theorem in the Trial Process: Instructing Jurors on the Value of Statistical Evidence' (1988) 12 *Law and Human Behavior* 1. See generally Gary L Wells, 'Naked Statistical Evidence of Liability: Is Subjective Probability Enough?' (1992) 62 *Journal of Personality and Social Psychology* 739; Daniel Kahneman, *Thinking, Fast and Slow* (Farrar, Straus and Giroux, 2011) pts 2–3.

judges tend to struggle with.¹³⁴ Unfortunately, experimental studies suggest that difficulties with probabilistic and statistical evidence are unlikely to be alleviated through simple solutions such as instructional videos or converting numbers to verbal formulations or tables.

Somewhat more positive in their conclusions, another substantial body of mainly ethnographic and observational research, affirming the ability of laypersons to understand and engage with scientific and technical knowledge in a variety of social, public and professional settings, complicates the foregoing assessment. These studies confirm that laypersons can acquire levels of *interactional* expertise¹³⁵ through activities as diverse as: leisure pursuits (eg bird watching, gardening, chess, and astronomy); health problems (eg HIV or diabetes); work (eg ecology and meteorology through to farming); and even participation in political movements and litigation (eg opposing the extraction of coal seam gas or the farming of genetically modified crops).¹³⁶

Where laypersons have the motivation to acquire understanding they often do so.¹³⁷ Conditions conducive to such engagement, apart from motivation, include considerable time, access to materials and resources, the ability to interact with actual experts, scope to ask questions or receive informal training, and the provision of critical feedback. Few of these conditions are available, or fully available, to jurors and judges. Whereas laypersons, beyond the jury, tend to be motivated to expend time, energy and resources developing or acquiring skills, sometimes over long periods of time, jurors are placed in conditions where all of the information they are said to require is given to them. Their exposure is responsive and largely passive. Indeed, they are actively discouraged from undertaking independent research and are not

¹³⁴ Mistakes by fact-finders and respondents are not all in one direction. Problems with probabilities and statistics include respondents undervaluing the strength of DNA matches.

¹³⁵ Harry Collins and Robert Evans, *Rethinking Expertise* (University of Chicago Press, 2007) 28–35.

¹³⁶ Influential works include Alan Irwin and Brian Wynne (eds), *Misunderstanding Science? The Public Reconstruction of Science and Technology* (Cambridge University Press, 1996); Alan Irwin, *Citizen Science: A Study of People, Expertise and Sustainable Development* (Routledge, 1995); Steven Epstein, *Impure Science: AIDS, Activism and the Politics of Knowledge* (University of California Press, 1996); Mike Michael, *Technoscience and Everyday Life: The Complex Simplicities of the Mundane* (Open University Press, 2006).

¹³⁷ Although in most cases the level of understanding only rises to an ability to comprehend and discuss basic and intermediate issues, not to become a *contributory* expert (ie a research scientist): Collins and Evans, above n 135, 14.

always encouraged to ask questions.¹³⁸ Like most judges, they receive no feedback on their interpretation of the evidence.¹³⁹ Consequently, the diverse assortment of case studies on the public understanding of science and technology have limited application to our extant criminal justice processes.¹⁴⁰

Overall, available research literatures do not support the contention that (notwithstanding the dilatory performance of lawyers) jurors are performing well in response to proffers of expert opinion. Jurors, judges, researchers and experimentalists have been insufficiently sensitive to the information and environments required to understand and evaluate forensic science and medicine evidence.

B *Judicial Experience and the Experience of the Law*

Judicial experience and ‘the experience of the law’ are also positively inclined toward the abilities and performance of juries.¹⁴¹ These perspectives tend to be a mixture of normative commitments, articles of faith, and bare assertions.¹⁴²

¹³⁸ See, eg, *Jury Act 1977* (NSW) s 68C. See also J B Hunter, *Jurors’ Notions of Justice: An Empirical Study of Motivations to Investigate & Obedience to Judicial Directions* (UNSW Jury Study, 2014).

¹³⁹ Judges may very occasionally receive feedback from appellate courts, but this is not based on ground truth and is often part of more complicated institutional and appellate processes. On the importance of critical feedback for learning and expertise, see Kahneman, above n 133, 241–2.

¹⁴⁰ Some of my early work on jury abilities was insufficiently attentive to the criminal justice context. Trial observations have revealed the lack of formal testing of forensic science techniques and the weakness of trial safeguards in identifying and conveying limitations. See Gary Edmond and David Mercer, ‘Scientific Literacy and the Jury: Reconsidering Jury “Competence”’ (1997) 6 *Public Understanding of Science* 329 and the discussion of ‘civic epistemology’ in Sheila Jasanoff, *Science at the Bar: Law, Science and Technology in America* (Harvard University Press, 1995).

¹⁴¹ The most conspicuous Australian exception is Justice Peter McClellan. Justice McClellan questioned whether the failure to provide reasons is consistent with a fair and rational system of criminal justice and expressed concerns about the ability of juries to understand some scientific and technical evidence: Justice Peter McClellan, ‘Juries — Common Sense and the Truth’ (Speech delivered at the New South Wales Crown Prosecutors’ Annual Conference, Sydney, 25 March 2008); Justice Peter McClellan, ‘Looking Inside the Jury Room’ (Speech delivered at the Law Society of New South Wales Young Lawyers Annual Criminal Law Seminar, Sydney, 5 March 2011).

¹⁴² See, eg, *Gilbert v The Queen* (2000) 201 CLR 414, 425 [31] (McHugh J): ‘Put bluntly, unless we act on the assumption that criminal juries act on the evidence and in accordance with the directions of the trial judge, there is no point in having criminal jury trials’; *R v Yuill* (1993)

They lack credible empirical foundations — apart from the fact that jurors turn up, frequently produce verdicts, and in many cases *appear* to take their responsibilities seriously. None of this tells us much about jury capabilities, how jurors evaluate complex evidence, or if they somehow manage to overcome the informational void confronting them in most criminal proceedings involving forensic science and medicine evidence.

One study asked Australian judges about jury comprehension and understanding. The survey included the following questions:

- Q5.1 In the cases over which you have presided which area of expertise do you think has presented the most difficulty for jurors to comprehend?
- Q5.3 Do you think that the jurors have comprehended the expert evidence before your summing up?
- Q5.5 Should matters involving complex and conflicting expert scientific evidence be withdrawn from juries and be determined by judges alone or by some other means?
- Q5.10 From the following list please circle the three factors which you consider to be the most persuasive for jurors when an expert is giving oral evidence: (a) their appearance; (b) clarity of explanation; (c) educational qualifications; (d) prior experience in the field; (e) prior experience as an expert witness; (f) familiarity with the facts; (g) impartiality; (h) publications.

69 A Crim R 450, 453 [6] (Kirby ACJ): ‘Courts will assume that jurors, properly instructed, will accept and conform to the direction of the trial judge to decide the case solely on the evidence placed before them in the court’; *R v Glennon* (1992) 173 CLR 592, 603 (Mason CJ and Toohey J): ‘The law ... proceeds on the footing that the jury, acting in conformity with the instructions given to them by the trial judge, will render a true verdict in accordance with the evidence’. On expert evidence, see *Velevski v The Queen* (2002) 187 ALR 233, 275 [182] (Gummow and Callinan JJ):

Juries are frequently called upon to resolve conflicts between experts. They have done so from the inception of jury trials. Expert evidence does not, as a matter of law, fall into two categories: difficult and sophisticated expert evidence giving rise to conflicts which a jury may not and should not be allowed to resolve; and simple and unsophisticated expert evidence which they can. Nor is it the law, that simply because there is a conflict in respect of difficult and sophisticated expert evidence, even with respect to an important, indeed critical matter, its resolution should for that reason alone be regarded by an appellate court as having been beyond the capacity of the jury to resolve.

Q6.5 Are most experts who give evidence before you representative of the views of their discipline?¹⁴³

Such questions reinforce how misguided this domain can be. Accepting that those completing bounded (eg multiple choice) surveys should not be too heavily reproached for their responses, the answers to such questions could not be considered as anything more than impressions.¹⁴⁴ Regardless of the responses and the interpretations attributed by investigators, the only reasonable response to these questions can be: how would I know?¹⁴⁵ Judges have *no meaningful interactions* with juries and any claims about jury competence and performance are impressionistic; based on demeanour and apparent attentiveness, the occasional jury question and the verdict.¹⁴⁶ Judicial opportunities to observe juries and verdicts provide very constricted insights into the reasoning, understanding and performance of jurors and juries. Surveys of judges may cast a good deal of light on the abilities of investigators, along with some indication of the perceived value of judicial knowledge and legal experience, but they tell us almost nothing about the capabilities of juries.¹⁴⁷

¹⁴³ Ian Freckelton, Prasuna Reddy and Hugh Selby, *Australian Judicial Perspectives on Expert Evidence: An Empirical Study* (Australian Institute of Judicial Administration, 1999) 65–73, 150–3.

¹⁴⁴ *Ibid.* While the authors recognised, perhaps because about a third of the judicial respondents did not answer the questions, that judges have few bases for answering questions about jury responses to expert evidence ‘by reason of the limited feedback which they receive from jurors’, they nevertheless proceeded to discuss the judicial responses in a non-reflexive way: at 65. See generally the discussion in Gary Edmond, ‘Judging Surveys: Experts, Empirical Evidence and Law Reform’ (2005) 33 *Federal Law Review* 95.

¹⁴⁵ Judge Shannon made this clear in the Edward Charles Splatt Report, above n 45, 286:

It is not possible ... to say that one item of evidence was or would have been of little or minimal importance to the jury. ... Once a particular item was admitted into evidence then the weight or significance attached to it by the jury is completely unknown and remains incalculable.

¹⁴⁶ Judges might read the jury literature, but as this article endeavours to explain, this might not be particularly helpful. On jury questions, see above n 130 and accompanying text. On understanding scientific and technical issues, see Harry Collins, *Are We All Scientific Experts Now?* (Polity, 2014).

¹⁴⁷ Conversely, knowing the kinds of evidentiary issues that create difficulties for judges (even from self-reports) is probably far more salient. See the survey by Freckelton, Reddy and Selby, above n 143, which collected quite a bit of information in this domain.

Interestingly, there is not much evidence that judges perform significantly better with expert evidence than juries.¹⁴⁸ There are few reasons to believe that judges are more conversant with scientific and technical evidence or capable of resisting the insidious psychological impact of inadmissible evidence or cross-contamination when combining evidence and evaluating guilt.¹⁴⁹ There is little independent evidence that legal training or experience assists with fact evaluation or logic.¹⁵⁰ To the extent that judges might on occasion perform better with expert evidence than some juries (or mock jurors), that could be a result of tertiary education, participation in judicial education programs featuring expert evidence (eg DNA profiling), greater (or repeated) exposure to specific kinds of scientific or technical evidence, in conjunction perhaps with the freedom to ask questions and request further information from counsel and witnesses.¹⁵¹ If judges are better informed than laypersons and lawyers then somewhat curiously this has not manifested in requests for: validation studies; indicative error rates; proficiency tests; adherence to standards; empirically-driven expressions; or attention to contextual bias. Revealingly, judicial experience and the experience of the law have not translated into admissibility standards that require the kinds of information that would enable decision-makers — whether judges or a jury — to rationally evaluate incriminating opinions.

Australian courts have been remarkably insensitive to a chorus of authoritative scientific advice in recent years.¹⁵² Confronted with advice and recommendations for change, senior judges have tended to rehearse their confidence in lay juries and the effectiveness of existing rules and safeguards based on

¹⁴⁸ Though, when asked, Australian judges were inclined to rate their abilities as superior to those of jurors: *ibid* 67.

¹⁴⁹ Kovera and McAuliff, above n 125; Chris Guthrie, Jeffrey J Rachlinski and Andrew J Wistrich, 'Inside the Judicial Mind' (2001) 86 *Cornell Law Review* 777, 782–4; Stephan Landsman and Richard F Rakos, 'A Preliminary Inquiry into the Effect of Potentially Biasing Information on Judges and Jurors in Civil Litigation' (1994) 12 *Behavioral Sciences & the Law* 113. See also Sophia I Gatowski et al, 'Asking the Gatekeepers: A National Survey of Judges on Judging Expert Evidence in a Post-*Daubert* World' (2001) 25 *Law and Human Behavior* 433.

¹⁵⁰ These are not common features in formal legal education.

¹⁵¹ See, eg, *R v Bornyk* [2013] BCJ No 2313 (British Columbia Supreme Court). But see *R v Bornyk* [2015] BCJ No 94 (British Columbia Court of Appeal).

¹⁵² *Tuite v The Queen* [2015] VSCA 148 (12 June 2015) [52] (Maxwell ACJ, Redlich and Weinberg JJA) is the only Australian judicial reference to the *NAS Report*.

long tradition and their own experience. Bizarrely, the very rules, safeguards and procedures that have not generated useful information or exposed the magnitude of problems are conceived and presented as appropriate responses to weak, speculative and unreliable opinions proffered by the state. In the end, judicial faith in the abilities and good sense of jurors, and even their constitutional primacy as fact-finders, matter little if the jury is not placed in conditions conducive to rationally fulfilling this most important of civic responsibilities.

Neither judicial experience nor available studies by lawyers and psychologists provide grounds for believing that jurors are capable of identifying problems and evaluating expert evidence without clear explanations of the value and limitations of techniques and the proficiency of the analyst. There is no reason to believe that in the absence of formal evaluation, even where they are informed of its absence, juries (or judges) are capable of compensating for, or overcoming, their essential ignorance.

V IRRATIONALITY, AGNOSTICISM AND POST-ENLIGHTENMENT JUDGING

If the jury system is to be maintained, and it should be so, the machinery of assisting the jury to reach a correct verdict on the evidence needs considerable change and improvement.¹⁵³

The jury was entitled to the best evidence; in its absence arguments as to their competence to assess forensic evidence simply do not get to first base.¹⁵⁴

Even if the thesis underpinning this article is expressed in a form that appears too strong, the basic contention should raise serious anxieties for legislators, prosecutors and judges as well as those asked to serve as jurors. For, there is little doubt that forensic science and medicine evidence is routinely presented in criminal proceedings (and appeals) in ways that do not enable rational evaluation. In many cases forensic analysts rely upon techniques that have not been tested, and in most cases the real chance of error or mistake is not

¹⁵³ Edward Charles Splatt Report, above n 45, 57, quoted in Law Reform Committee, Parliament of Victoria, *Jury Service in Victoria: Final Report (1997)* [2.187].

¹⁵⁴ Richard W Harding, 'Jury Performance in Complex Cases' in Mark Findlay and Peter Duff (eds), *The Jury under Attack* (Butterworths, 1988) 74, 88. See generally Edmond and Roberts, above n 13.

presented in a way that accurately reflects known or potential risks. In these circumstances, presenting opinions as ‘expert’ is substantially unfair to the defendant, and invariably raises the danger that the evidence will be misunderstood or that in the absence of pertinent information fact-finders will invoke alternative criteria of more limited utility, defer to the analyst or rely on the apparent consistency with other strands of evidence. This is all highly undesirable.¹⁵⁵

Our legal system asks a great deal of the citizens selected to be jurors. It requires them to sit in judgement and, with increasing frequency, expects them to consider a range of scientific and technical evidence as part of their evaluation of the case. We have not, however, made credible efforts to present forensic science and medicine evidence in ways that take account of the dependent and vulnerable position of fact-finders, and have consistently failed to provide the kinds of information and insights that would facilitate rational evaluation. We have assumed that juries are capable of understanding and evaluating *expert* opinion based on what transpires during conventional adversarial proceedings. This assumption is misplaced.

Trial judges have not done enough to facilitate explanation and comprehension and appellate judges, perhaps unwittingly, have tended to overlook the main issues by reiterating their confidence in adversarialism, antiquated models of expertise, trial safeguards and jury verdicts. This reactionary and predominantly passive response is both inappropriate and inadequate. Appellate courts ought to provide trial judges with the resources to become responsible gatekeepers. Trial judges should require prosecutors and forensic analysts to present forensic science and medicine evidence in ways that embody both the value of the opinion and the limitations with the evidence based on what is *known*. There is a need to recognise that in most cases fundamental epistemic problems and oversights cannot be remedied through cross-examination, rebuttal witnesses or judicial directions and warnings. There is, in consequence, a need to exclude opinion evidence

¹⁵⁵ The contention that we have relatively few wrongful convictions because so few are exposed is not a persuasive response to the concerns raised in this article. This article has endeavoured to explain that our trial and appellate courts have been inattentive to the kinds of information required to assess expert opinions. In consequence, the claim that few wrongful convictions or unfair trials have been observed involves circular reasoning — basically reiterating the value of existing approaches rather than thinking about the implications of this article, not only for our current practices but our ability to actually gauge factual errors and substantial unfairness.

that is not actually based on identifiable ‘specialised knowledge’. ‘Training, study or experience’, however impressive or extensive, cannot overcome the absence of *knowledge*.

In concluding, it is important to emphasise that the basic implication is not that lay fact-finders cannot be relied upon to cope with complex forensic science and medicine evidence, methodological criticisms, or statistical and probabilistic evidence. Rather, the concern is that we rarely place juries (and judges) in positions that are conducive to the rational assessment of expert opinion evidence and the combination of such evidence with other strands of evidence — both incriminating and exculpatory. We must provide jurors (and judges) with the kinds of information that will allow them to succeed as responsible, rational fact-finders (and gatekeepers).¹⁵⁶ This would, at the very least, make trials and appeals more consistent with espoused criminal justice principles — particularly truth and justice. Until we have had an opportunity to assess jury performance in these improved conditions, we ought to remain agnostic on the question of jury competence.

Of course, in the end it might prove too difficult for randomly assembled groups of laypersons to cope with the increase in the volume and sophistication of scientific and technical evidence, especially if they are left to make their assessment in the context of an adversarial proceeding without basic (let alone balanced) instruction. If this proves difficult or unsatisfactory, we may have to reconsider the continued use of laypersons as fact-finders or revisit our principled preference for understanding and evaluation over deference. To the extent that we genuinely value citizen participation, however, the first step must be to provide our jurors and judges with the information necessary to responsibly perform the most important of their civic responsibilities.

¹⁵⁶ See *Tuite v The Queen* [2015] VSCA 148 (12 June 2015) [11] (Maxwell ACJ, Redlich and Weinberg JJA).