This article examines the admissibility of forensic science and medicine in criminal proceedings. In Part II, we explain how reliability-based admissibility standards in the United States have been unevenly applied to expert evidence in civil and criminal cases and have not prevented wrongful convictions. In Part III, we review a recent Consultation Paper (and report) issued by the Law Commission of England and Wales. Though focused on the need for ‘sufficiently reliable’ expert opinion evidence, we challenge its contemplation of easier admissibility for experience-based forensic sciences and techniques traditionally admitted. In Part IV we examine the evolving law on the admissibility of expert evidence in Canada. In response, we argue that while front-end reforms to the organization and practice of forensic science and medicine, advocated by the Goudge Inquiry and the American National Academy of Sciences, appear more promising than reliance on the adversary system, the gatekeeping role of trial judges should be strengthened. In the concluding section, we contend that threshold reliability standards should be grounded in criminal-justice system values, emerging empirical insights about the weakness of the adversarial trial and be sensitive to the particular evidence and its use, rather than applied mechanically using simplistic models of science and expertise.

Keywords: wrongful conviction/expert evidence/threshold reliability/National Academy of Sciences/law and science/Goudge

1 Introduction

Throughout the common-law world, there is growing appreciation that a great deal of forensic-science and medical evidence is not reliable. DNA exonerations reveal that flawed or exaggerated expert evidence is a

* Professor, School of Law, Director, Expertise, Evidence & Law Program, University of New South Wales, and Consultant to the Goudge Inquiry

** Professor, Faculty of Law, University of Toronto, and former Director of Research for the Goudge Inquiry

† This research was supported by the Social Sciences and Humanities Research Council of Canada. This research was also supported by the Australian Research Council (DP0771770 and FT0992041) and an earlier version was presented to the School of Law, UNSW. Thanks to Emma Cunliffe, David Hamer, Mehera San Roque, two anonymous referees and the editor for comments.
frequent factor in wrongful convictions.\(^1\) A 2009 report from the American National Academy of Science (NAS)\(^2\) concludes that, with the exception of nuclear DNA evidence, most forensic evidence lacks a scientific basis and adequate regulation and quality control. It proposes an ambitious federal program of research together with independent scientific governance and certification but expresses scepticism about the ability of judges to generate the much-needed improvements to the forensic sciences through legal ‘gate-keeping.’

The NAS’s scepticism about judicial gate-keeping has support in the emerging empirical research on the effects of the admissibility jurisprudence associated with *Daubert v Merrell Dow Pharmaceuticals, Inc* (1993)\(^3\) – a decision purporting to place a premium on ‘reliability’ and the need for gate-keeping based on criteria such as testing, peer review, error rates, and general acceptance. As will be explained in this article, *Daubert* emerged out of a perceived crisis in tort and product-liability litigation, where large, well-resourced corporate defendants (i.e., ‘repeat players’) sought to have the expert evidence adduced by ‘unworthy’ plaintiffs excluded.\(^4\) Remarkably, the new and potentially onerous admissibility standards associated with the invocation of *Daubert* in civil litigation have not been applied with any rigour to forensic science and medicine adduced by the state in criminal proceedings, even while DNA and other exonerations have demonstrated how forensic science has contributed to wrongful convictions. Just how the goals and principles underlying criminal justice align with these developments is one of the important issues explored in this article. We question whether the modern adversarial criminal trial currently embodies fundamental criminal-justice values including those that prioritize protection of the innocent from conviction on the basis of unreliable evidence. We also question whether our judges (and prosecutors, forensic scientists, and defence lawyers) are sufficiently cognizant and concerned about unsettling empirical and experimental findings suggesting pervasive problems with incriminating forensic-science and medical evidence.

---

1 We use the term ‘wrongful conviction’ interchangeably with ‘miscarriage of justice,’ even though the overlap is not complete. See generally Kent Roach & Gary Trotter, ‘Miscarriages of Justice in the War against Terror’ (2005) 109 Penn State L Rev 967 at 1032ff.
3 *Daubert v Merrell Dow Pharmaceuticals, Inc*, 509 US 579 (1993) [*Daubert*].
In Canada, the Goudge *Inquiry into Pediatric Forensic Pathology* (2007–2008) was the third public inquiry in a decade to examine the role of forensic science and medicine in wrongful convictions. Like the NAS report, the Goudge Inquiry recommended better training, research and governance for forensic science and medicine, but it also recommended that judges should assume a more robust gate-keeping role – determining the ‘threshold reliability’ of all forensic science. Likewise, the Law Commission of England and Wales (Law Commission) issued a consultation paper with the preliminary proposal that judges should only admit expert-opinion evidence after determining that it is ‘sufficiently reliable.’ The Law Commission also looked to *Daubert* as an inspiration for stricter, reliability-oriented admissibility standards. Recent Canadian decisions in the wake of *Daubert* and high-profile public inquiries into wrongful convictions have begun to accept the idea that judges should play a more pronounced gate-keeping role in determining ‘threshold reliability’ and admissibility. Outside of the United States, there is considerable optimism about the ability of stricter admissibility standards to improve the forensic sciences and the safety of criminal convictions.

We are supportive of more demanding standards for the admissibility of incriminating expert evidence. Indeed, we go beyond current legal practice and proposals for reform to argue for *demonstrable reliability* whenever the state adduces expert evidence to support a criminal conviction (or induce a plea). Our proposed approach is grounded in criminal-justice values, such as the presumption of innocence, the need for the state to prove guilt beyond a reasonable doubt, and the goal of preventing the conviction of the innocent. It incorporates emerging evidence about the practical limitations of the modern adversarial trial (and its various safeguards) and, in doing so, is substantively concerned with fairness. We also rely on the fact that many forensic-science and medical techniques used to obtain criminal convictions

---


could have their validity and reliability assessed and, in this regard, recognize the superior resources and personnel available to the state. At the same time, we would recommend that expert evidence adduced by the defence need only satisfy a basic reliability threshold but would require that judges apply admissibility standards in a robust contextual fashion even should our asymmetrical proposal, which places higher standards on the state, not find favour.

Although we support stricter reliability standards, we accept that the American experience with *Daubert* offers important cautionary insights that Commonwealth reformers ignore at their peril. The American experience suggests that more attention needs to be paid to the ecology or political economy of the administration of admissibility standards and the limitations of modern adversarial trials and appeals. In particular, there is a need to ensure that lawyers representing those accused of crime (as opposed to lawyers representing large corporate defendants in tort and product-liability suits) have the necessary resources and skills to effectively challenge expert evidence offered by the state and that judges have the necessary doctrinal and intellectual support to make difficult decisions to exclude or place effective limits on incriminating expert-opinion evidence.

A second cautionary insight from the American experience is that judicial recourse to *Daubert* can produce ‘romanticized’ or idealized models of science which are displaced from the way many forms of forensic science and medicine regularly offered in criminal proceedings are currently practised. American criminal courts have consistently characterized such practices as ‘non-science’ and applied *Daubert* less stringently when determining their admissibility. Similarly, the English Law Commission appears prepared to make exceptions for experience-based expertise and Canadian courts have also recently recognized that demands for validation and precise error rates may be inappropriate for much expert evidence. Accepting the need to apply admissibility standards flexibly, there is a danger that experience-based (and other) exceptions to *Daubert*-style criteria will be used to legitimate the status quo and circumvent the need for actually assessing reliability.

Although our proposed approach to admissibility borrows from the emphasis on reliability, testing, peer review, publication, the determination of error rates, and the need for more rigorous gate-keeping promoted in the *Daubert* decision (though most conspicuous in civil proceedings), it differs from the model of science presented in *Daubert* and championed by most of the prominent critics of the forensic sciences by being more contextual. Rather than invoke philosophically driven,

---

10 Consultation paper, supra note 6 at 62; *R v Abbey* 2009 ONCA 624 [*Abbey*].
idealized models of science, with which to contrast extant forensic-science and medical practice, our approach is instead grounded in the specific values and purposes of criminal justice, empirical studies of experience with *Daubert* and wrongful convictions, an authoritative review of the forensic sciences in the United States, and empirically sensitive approaches to science and expertise drawn from post-Kuhnian science studies.¹¹

This more contextual approach to admissibility jurisprudence has the virtue of directly addressing the issues that are prompting present concerns; namely, the epistemic content of particular forensic sciences and their contribution to wrongful convictions. It also relieves some of the traditional criminal-trial ‘safeguards’ (such as prosecutorial restraint, cross-examination, opposing experts, judicial instructions, the participation of lay jurors, and appellate review) of responsibilities that they have not been able to discharge consistently. It is also consistent with recent Canadian jurisprudence recognizing that concerns about ‘threshold reliability’ and wrongful convictions apply to many forms of evidence and are no longer limited to novel scientific techniques.¹²

This article is divided into five parts. In Part II, we consider the American experience with the application of *Daubert* and the contributions of flawed forensic science and medicine to wrongful convictions. The United States is the logical starting point for a study of law and science if only because of its longer experience with purportedly stricter admissibility standards and with notorious DNA exonerations (from various innocence projects). As we explain through an overview of empirical legal research and a review of the NAS report, the American experience is far from a success story. Canadian, British, and Australian reformers ignore the hard lessons from America at their peril.

Part III reviews the recent Consultation Paper issued by the Law Commission. The Law Commission’s paper illustrates how a superficial understanding of *Daubert* tends to generate countervailing pressures to make exceptions for expert evidence based on experience, emerging technologies, and older techniques never shown to be reliable. The Law Commission’s proposal exemplifies the apparent reluctance

---


¹² *Trochym*, supra note 7.
(or inability) of many law reformers (and practicing judges and lawyers) to engage with critical non-legal literatures or contemplate change to traditional forms of practice in order to preserve fundamental legal principles underpinning the accusatorial criminal trial.

In Part IV, we examine the Canadian experience through a review of the evolving law on the admissibility of expert evidence and the recommendations of the Goudge Inquiry. It is our contention that Justice Goudge correctly advocated more demanding admissibility standards – with the judge playing a more active gate-keeping role – and that this is largely consistent with emerging Canadian jurisprudence. Nevertheless, American experience suggests that defence lawyers and judges must, as Justice Goudge recommends, be given more resources if they are to make effective use of new admissibility standards. ‘Reliability’ is unlikely to work as an admissibility threshold without legal mobilization and extra-legal reform. In many ways, the ‘front-end’ reforms to the organization and practice of forensic science and medicine advocated by both Goudge and the NAS appear more promising than reliance on the adversary system. Nevertheless, Canadian lawyers and judges cannot afford to ignore domestic experience with wrongful convictions or abandon gate-keeping to others. Review of recent cases illustrates the potential for judicial gate-keeping to screen all forms of expert evidence. This includes not only excluding unreliable expert evidence and expert evidence of unknown reliability but placing constraints on the language used by the expert and the need to disclose ongoing controversies where evidence is sufficiently reliable for admission.

In the fourth and concluding section, we argue that more rigorous standards for admissibility should be grounded in criminal-justice-system values rather than in simplistic models of science or expertise. Recourse to such values may help to prevent gaping exceptions for long-standing techniques (so-called ‘grandfathering’) and expertise based on experience including, we fear, most of the forensic sciences.13 We would require the state to show that its expert evidence is demonstrably reliable. This approach assumes a degree of asymmetry that places a greater burden on the state to demonstrate the reliability of expert evidence than on the accused. Although such asymmetry is novel and controversial, it is consistent with fundamental criminal-justice values – specifically, the presumption of innocence and the state’s unique obligation to prove guilt beyond reasonable doubt.

doubt and informed by empirical insights. Our proposed demonstrable-reliability standard would give the main producers of forensic science – state institutions or private consultants hired by the state – incentives to devote more resources to attempts to refine scientific techniques as well as to quality assurance.

Our proposed approach to admissibility is informed not only by the values and context of criminal justice but also by the content of the particular form of expertise that is offered as testimony. This means that judges and lawyers must not rely solely on the idealized and abstract vision of proper science implicit in *Daubert* but should also attend to the limitations of techniques and the controversies in the particular field from which the expert evidence derives. Such an approach imposes formidable burdens on lawyers and judges, again reaffirming our position that they must be given the necessary tools to conduct such analysis and that it would be dangerous to rely solely or primarily on judicial reform of the forensic sciences. That said, the complicity of the legal system in the dismal state of much forensic science and medicine can no longer be ignored.

II **Hard lessons from America**

Commonwealth jurisdictions have much to learn from the American experience with forensic-science evidence, admissibility decisions, and wrongful convictions. Unfortunately, studying the United States can be daunting, given the complexity of its law and the number of jurisdictions. While there is a danger that a superficial understanding of American law and practice will mislead, it is, nevertheless, important to provide an overview of the major developments in relation to expert evidence over the last few decades in order to gain the benefit of its experience and the results of empirical research.

A THE *DAUBERT* TRILOGY AND THE CIVIL JUSTICE REVOLUTION

Significantly, all of the modern expert-evidence decisions from the Supreme Court of the United States emerged out of disputes about the quality and representativeness of expert evidence adduced in civil litigation during the 1980s and 1990s. Starting in the 1980s – in response

---


15 Though see also *United States v Scheffer*, 523 US 303 (1998) and the recent decision of *Melendez-Diaz v Massachusetts*, 129 S Ct 2527 at 2529, 2531–2 (2009), where, in an appeal over the confrontation clause in relation to expert reports, the US Supreme
to the increasing prominence of mass torts and massive damage awards, to insurance crises, to corporate mobilization, to intense media scrutiny of legal institutions, and to the politicization of those institutions – trial and appellate courts became less receptive toward plaintiffs’ expert evidence. In the 1990s, the Supreme Court repeatedly intervened through authoritative interpretations of the Federal Rules of Evidence (FRE).

The first and most important of these interventions was Daubert v Merrell Dow Pharmaceuticals, Inc (1993). In Daubert, the Supreme Court was confronted with the question of whether the general acceptance of a principle or technique, an older admissibility standard derived from the case of Frye v US (1923), had survived enactment of the FRE in 1975. In deciding Daubert, the Supreme Court concluded that ‘general acceptance’ no longer governed the admission of expert evidence. The majority explained that for admission via Rule 702, ‘scientific’ evidence must be both ‘relevant and reliable.’ ‘Relevance’ has a logical basis, but the majority provided four criteria (henceforth, the Daubert criteria) to help trial judges assess the ‘reliability’ of scientific evidence.

Court seems to recognize the existence of pervasive problems with the forensic sciences.

16 The litigation over Agent Orange and Bendectin seems to have been broadly influential; see Peter Schuck, Agent Orange on Trial: Mass Toxic Disasters in the Courts (Cambridge, MA: Harvard University Press, 1986); Joseph Sanders, Bendectin on Trial (Ann Arbor, MI: University of Michigan Press, 1998); Michael Green, Bendectin and Birth Defects (Philadelphia, PA: University of Pennsylvania Press, 1998) [Sanders]. Interestingly, changes to admissibility standards may have been a response to difficulties with tort reform; see Lucinda Finley, ‘Guarding the Gate to the Courthouse: How Trial Judges Are Using Their Evidentiary Screening Role to Remake Tort Causation Rules’ (1999) 49 DePaul L Rev 335.


18 Supra note 3.

19 293 F 1013 (DC Cir 1923).


21 Daubert, supra note 3 at 589–95. Rule 702 of the FRE was designed to govern the admissibility of expert opinion evidence in US Federal Courts (as an exception to the general prohibition on opinion evidence provided by the exclusionary Rule 701). The original version of Rule 702 read, ‘If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise’; Fed R Ev702 (1975) [FRE 1975].

22 Some commentators include the existence and maintenance of standards controlling the use of a technique or method as a fifth criterion. See e.g. Lloyd Dixon & Brian Gill,
The criteria are whether the theory or technique (1) has been tested (referring to Karl Popper’s concept of falsification); (2) has been published and/or peer reviewed; (3) has a known or potential rate of error; and (4) is ‘generally accepted’ in the relevant specialist community (from *Frye*). These criteria, which according to the majority would ‘ordinarily’ be helpful, effectively replaced (or subsumed) the *Frye* standard and a range of other approaches to admissibility emerging across the various circuits. Notwithstanding references to the need for flexibility, the majority placed emphasis on the trial judge’s ‘gate-keeping responsibility.’

In dissent, Chief Justice Rehnquist (with Justice Stevens agreeing) was critical of the criteria selected:

I defer to no one in my confidence in federal judges; but I am at a loss to know what is meant when it is said that the scientific status of a theory depends on its ‘falsifiability,’ and I suspect some of them will be, too. I do not doubt that Rule 702 confides to the judge some gatekeeping responsibility in deciding questions of the admissibility of proffered expert testimony. But I do not think it imposes on them either the obligation or the authority to become amateur scientists in order to perform that role.

---

23 *Daubert*, supra note 3. A strong distinction between ‘methodology’ and ‘conclusions’ suggests that the majority may have initially accepted that (Popperian) falsification embodied ‘the scientific method.’ This distinction was subsequently read down in *General Electric Co v Joiner*, 522 US 136 (1997) [*Joiner*].

24 For example, *DeLuca v Merrell Dow Pharmaceuticals, Inc*, 911 F (2d) 941 (1990); *United States v Downing*, 755 F (2d) 1224 (1985); *Brock v Merrell Dow Pharmaceuticals, Inc*, 874 F (2d) 307 (1989), modified on reh’g, 884 F (2d) 166 (1989); *Christopherson v Allied Signal Corp*, 939 F (2d) 1106 (1991).

After the Supreme Court decision, the case was remanded to the Ninth Circuit Court of Appeal where the plaintiffs’ biomedical evidence on causation – suggesting that the anti-nausea drug Bendectin (also marketed as Debendox) may have caused birth defects – was deemed inadmissible when assessed against all the *Daubert* criteria and (in addition) on the basis that it was considered to be ‘science for litigation’ or ‘junky.’

Six years after *Daubert*, the Supreme Court heard another appeal on the admissibility of expert evidence. In *Kumho Tire Co v Carmichael* (1999), the Court addressed the admissibility standard for non-scientific forms of expert evidence. This was the other portion of Rule 702 concerned with ‘technical, or other specialized knowledge.’ Firmly endorsing the *Daubert* decision – no dissent this time – the Court explained that the four criteria *might* be flexibly applied to all admissibility determinations, including those involving non-scientific expert opinions. In *Kumho*, the majority applied the *Daubert* criteria to the ‘technical, or other specialized knowledge’ before them. In a powerful demonstration of gate-keeping and the continuing need for judicial vigilance, the Court upheld the decision to exclude the plaintiffs’ expert engineering evidence because it did not satisfy the four *Daubert* criteria.

In *Joiner*, yet another civil appeal concerned with biomedical evidence of causation, the Supreme Court underscored the importance of gatekeeping by effectively insulating the trial judge’s admissibility determinations from appellate review. After *Joiner*, appellate courts were required to identify an abuse of the trial judge’s discretion to activate their power of review. The Supreme Court explained that there was to be no ‘particularly stringent’ review of the exclusion of plaintiffs’ expert evidence. Review is dependent upon the trial judge’s decision’s

---


27 FRE 1975, supra note 21.

28 *Kumho*, supra note 9. Justice Stevens dissented over the majority’s application of the *Daubert* criteria.


30 *Joiner*, supra note 23. In contrast, we contend that expert evidence, and especially admissibility determinations, should be reviewed *de novo*. There seem to be few reasons to consider the demeanour of the expert (and few advantages available to the trier of fact) where the inquiry is into evidence of reliability. See the recommendations of the Law Commission, *Expert Evidence in Criminal Proceedings in England and Wales* (London: Her Majesty’s Stationery Office, 2011) at 5.94 [Law Commission, *Expert Evidence*].
being palpably mistaken.  

Rather than begin a revolution, *Daubert*, *Joiner*, and *Kumho* consolidated a more widespread tightening of admissibility standards in response to perceived problems with civil litigation (and stalled tort reform).  

Thus, trial judges were encouraged to act as vigilant gatekeepers and appellate courts were prevented from interfering unless trial judges clearly abused their wide discretions. In consequence, there was a chilling effect as plaintiffs struggled to have their expert evidence admitted. It has become normal to have a pre-trial admissibility hearing for expert evidence (now called *Daubert* hearings) upon which the fate of a civil suit frequently depends. Plaintiffs unable to assemble *Daubert*-proof evidence are often left without a viable cause of action.

B  *DAUBERT* ‘IN-ACTION’: THE CRITERIA AND RELIABILITY IN CIVIL AND CRIMINAL CASES

When it was handed down, *Daubert* was widely – though not universally – interpreted as imposing a more demanding, implicitly scientific approach to the admissibility of expert evidence. Relatively quickly, trial (i.e.,

31 *Joiner*, supra note 23 at 141.


34 In 2000, largely in response to Supreme Court jurisprudence and pervasive concerns about the quality of expert evidence in civil litigation, the FRE were revised. Rule 702 now states: ‘If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case’; Fed R Ev 702 (2000).

plaintiff) lawyers realized that America’s highest court had provided support, emphatically reinforced in *Joiner* and *Kumho*, to the emerging exclusionary trend, particularly of expert evidence about causation, adduced by plaintiffs in (mass) tort and product-liability suits. The handful of commentators concerned about the quality of the forensic sciences initially welcomed the decision (like the corporate defendants in the civil sphere who were championing more onerous admissibility standards to keep plaintiffs out of the courts) on the basis that it provided judges with appropriate means of evaluating expert evidence. Many forensic scientists (and prosecutors), by contrast, were initially alarmed. This is not surprising, given the general dearth of quality controls, lack of formal qualifications among practitioners, and limited research base. In practice, however, *Daubert*, *Joiner*, and *Kumho* primarily contributed to the exclusion of expert evidence in civil proceedings. Their impact on criminal justice, particularly on the practice of forensic science and medicine, has been far less significant than almost anyone anticipated.

In trying to make sense of *Daubert*, *Joiner*, and *Kumho* and changes to admissibility standards in the United States, it is important to reiterate that these were all appeals from civil cases in the federal courts. It is no coincidence that the *potentially onerous* admissibility standard(s) associated with *Daubert* emerged in the civil sphere in the context of widespread concerns about the explosion in litigation, various insurance crises, excessive damage awards, plaintiffs’ recourse to ‘junk science,’ highly politicized judicial elections, the influence of a Republican dominated judiciary (in the federal courts), and aggressive corporate campaigns aiming to reform tort law. Armed with this background, we can begin to consider ambiguous; see e.g. *Expert Admissibility Symposium – Reliability Standards: Too High, Too Low, or Just Right?* (2003) 33 Seton Hall L Rev.


37 See e.g. ‘Scientific Evidence,’ supra note 35. A prominent example of a pro-defendant, post-*Daubert* contribution to the controversy in the civil sphere is Norman Foster & Peter Huber, *Judging Science: Scientific Knowledge and the Federal Courts* (Cambridge, MA: MIT Press, 1997).

the illuminating empirical research on *Daubert*. Although much of this research focuses on civil litigation, the results provide useful contrasts to the few empirical studies of criminal proceedings, to subsequent discussion of DNA exonerations, and to pessimism about *Daubert* drawn from the NAS review of the forensic sciences.

Probably the best-known and most influential empirical study, by Lloyd Dixon and Brian Gill, examined *Daubert*’s effects on civil proceedings.39 This study systematically reviewed 399 written decisions (by 263 federal judges across 71 federal court districts) issued between 1980 and 1999 involving challenges to the admissibility of expert evidence. Dixon and Gill found that, since *Daubert*, the ‘standards for reliability have tightened’ and judges ‘have become more watchful gatekeepers.’40 Successful admissibility challenges increasingly led to successful requests for summary judgment, with nearly 90 per cent of these judgments going against plaintiffs.41

These findings are generally consistent with work by Carol Krafka et al.42 They surveyed all active federal district court judges in 1991 (responses = 335) and 1998 (responses = 303) about their most recent experience with expert evidence in civil litigation. These surveys inquired about types of cases and experts, the nature and frequency of issues addressed by expert witnesses, the handling of admissibility issues and the exclusion of expert evidence, and (after the 1991 survey) changes in the response to expert evidence. Krafka’s team concluded that judges were more likely to have excluded expert evidence in their last civil trial in 1998 (i.e., after *Daubert* and *Joiner*) than in 1991. And the major change in procedure was the increased use of pre-trial (*Daubert*) hearings. Just over half of the respondents used pre-trial hearings in 1991, whereas more than three quarters reported using them in 1998.43

Krafka et al found that judges in 1998 reported that they were more likely to scrutinize expert testimony before trial and were less likely to admit it. Judges said that they limited or excluded some of the testimony

---


40 Dixon & Gill, ‘Changes in the Standards,’ ibid at 274, 291.

41 Ibid at 294.

42 Ibid at 327.
proffered by experts in 41% of the referenced 1998 cases, compared with only 25% of the 1991 cases. A third of judges claimed to admit expert evidence less often in 1998 than before Daubert. At the same time, while more judges referred to ‘unreliability’ as a reason for exclusion in the 1998 survey, Krafka et al found that few judges referred to the Daubert criteria. The ‘nonfalsifiable’ nature of an underlying theory and ‘an unknown or too-large error rate were cited in less than 2% of cases. The bases for limiting or excluding testimony do not appear to have been greatly affected by Daubert. Judges who excluded testimony in the recent survey did so most often because it was not relevant, the witness was not qualified, or the testimony would not have assisted the trier or fact. These reasons are similar to the reasons most frequently cited by judges in 1991, and they do not reflect the factors cited in Daubert.

A more recent study by Edward Cheng and Albert Yoon also focused on civil litigation sought to determine whether the adoption of Daubert made ‘any difference to the way scientific evidence is handled in practice.’ Understanding Daubert’s effects on the state courts is important because the vast majority of tort and product-liability suits are conducted in (and around) them. The federal court’s diversity jurisdiction enables parties from different states to force lawsuits into a federal court. There, the suit is governed by the substantive law of the relevant state, but by federal rules of evidence and procedure. The major legal advantages accruing through removal tend to be adjectival. Because most tort and product-liability suits involve expert evidence, Cheng and Yoon reasoned that defendants’ decisions to remove civil suits would reflect any perceived advantage associated with Daubert. Cleverly, these authors proposed a metric to compare changes in the rates at which defendants removed cases from state courts – using a version of ‘general acceptance’ (i.e., Frye) – to federal courts. They were also able to compare transfer rates in states that switched in the years after 1993 from ‘general acceptance’ to a standard based on Daubert.

Their results suggested that ‘whether or not a jurisdiction follows the Daubert standard has no statistically significant effect on the removal rate.’ In consequence, they reasoned that ‘debates about the practical merits and drawbacks of adopting a Frye versus a Daubert standard are largely superfluous.’ Among the ‘ramifications’ of this work, Cheng

44 Ibid at 322.
45 Ibid at 323.
46 Ibid at 330.
48 Ibid at 503.
49 Ibid at 503.
and Yoon concluded that ‘the power of the Supreme Court’s decision was not so much in its formal doctrinal test, but rather in its ability to create greater awareness of the problems of junk science.’50 This was generally consistent with claims by David Faigman, Elise Porter, and Michael Saks that *Daubert* cast a long shadow over civil litigation and was as important in awareness raising as in the provision of a specific admissibility standard.51

In combination, these studies may be more subversive than the various authors suggest. For it may be that formal standards (such as the *Daubert* criteria) are not the most important factor in admissibility decision making. This might mean that we need more than just bare reliability to enhance the performance of our courts. Such a reading of the empirical studies of civil litigation is strangely consistent with the few studies of forensic science offered by the state in criminal cases. *Daubert* is applied in both criminal and civil proceedings, but has discernibly different effects.

Jennifer Groscup et al undertook a study of federal (n = 372) and state (n = 321) criminal appellate court decisions between 1988 and 1998 involving expert evidence.52 Using a coded analysis of the judicial discussion of expert evidence and the importance of particular factors shaping the assessment of admissibility, these researchers concluded that there was greater scrutiny (i.e., ‘discussion’) of expert evidence but no change in the proportion of evidence admitted (or excluded) after *Daubert*. The ‘lack of discussion devoted to the four *Daubert* criteria’ was considered ‘perplexing.’53

Daubert seemed to suggest that an expert’s methods, not only his or her credentials, should be evaluated. However, the most important rating on sources of expertise were the experience and education of the proffered expert. Sources of knowledge which indicate methodological reliability were rated as less important . . . judges may be unable to determine what factors are important in assessing scientific reliability, particularly when research is the basis for an expert opinion.54

These researchers concluded that the *Daubert* criteria were not being used to evaluate expert testimony in criminal proceedings and that satisfying them was not an accurate predictor of admissibility.55

---

50 Ibid at 503–4: Cheng & Yoon recommend adopting a universal admissibility standard and interventions aimed at improving the ‘judiciary’s understanding of scientific concepts and processes.’
51 Faigman, Porter, & Saks, supra note 20.
53 Ibid at 353.
54 Ibid at 357.
55 Ibid at 363–5.
Additional research adds a curious twist to these findings. Michael Risinger found that criminal defendants very often lose their (Daubert) challenges to expert evidence adduced by the state, but prosecutors often win their challenges to defendants’ proffers of expert evidence.\footnote{Michael Risinger, ‘Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock’ (2000) 64 Alb L. Rev 99 at 135ff [Risinger, ‘Navigating’]; see also Erica Beecher-Monas, ‘Reality Bites: The Illusion of Science in Bite-Mark Evidence’ (2008) 30 Cardozo L. Rev 1369.} Risinger also observed how early challenges to novel forensic-science evidence (as in \textit{People v Marx}\footnote{\textit{People v Marx}, 126 Cal Rptr 350 (Cal Ct App 1975).} in response to bite mark evidence) are not necessarily effective methods for exploring weaknesses (or the speculative ways new methods might be deployed thereafter), yet operate as permissive resources or precedent in subsequent proceedings.

These readings of \textit{Daubert} and its progeny bring trends in the criminal sphere into sharper relief. Notwithstanding the provision of criteria and the willingness to exclude the expert evidence adduced by plaintiffs, trial and appellate judges have been relatively unresponsive to defence attempts to impugn incriminating expert evidence. Expert evidence favouring the compensation of plaintiffs has been held to consistently higher standards than expert evidence implicating defendants in criminal activities. This may suggest judicial distrust of the civil jury as well as some sympathy for corporate defendants.\footnote{Edmond, ‘Legal Engineering,’ supra note 29.} In criminal proceedings, it is defence experts who appear to be distrusted and tend to be held to higher standards than experts employed or engaged by the state.\footnote{Risinger, ‘Navigating,’ supra note 56 at 131–2.}

Overall, this empirical evidence might be considered unsettling. It seems to suggest that admissibility practice is driven by ideological commitments. That is, impressions of the litigation landscape and anxiety about crime, concerns about institutional legitimacy, and pressures on time and resources seem to be at least as important as attention to admissibility criteria and interest in the reliability of expert evidence, factual rectitude, and justice.\footnote{See also Harris, \textit{Black Robes, White Coats} (New Brunswick, NJ: Rutgers University Press, 2008) at 60–7, 130ff.} These trends seem to be reinforced by reviews of DNA exoneration cases.

C. WRONGFUL CONVICTIONS, FORENSIC SCIENCES, AND THE LIMITS OF THE ADVERSARIAL TRIAL

In recent decades, innocence projects have dramatically illustrated problems with both the adversarial criminal trial and the forensic sciences
in the United States and elsewhere. To the extent that these projects have drawn on DNA evidence as an exclusionary mechanism, they offer an unprecedented opportunity to review trials and appeals where factually innocent persons were convicted of the most serious criminal offences. They provide partial yet highly informative insights into the limitations of criminal justice processes, such as investigations, decisions to prosecute, plea bargains (based on expert evidence never seriously scrutinized), the defence afforded to the accused (particularly those who are not wealthy and especially those who are indigent), the value of trial safeguards (such as cross-examination, opposing experts, judicial directions and warnings, onerous burdens of proof, appellate review, and so on), and lay decision making.

The role of exonerating DNA evidence means that the most prominent of the innocence project cases involve convictions for sexual assault or sexual assault–murder. In these cases, there is often some kind of biological material (usually residual seminal fluid or blood) that can be compared with the DNA of the accused. Most of the Innocence Project successes have involved cases that pre-date the development or widespread application of DNA profiling. Some of those fortunate enough to have original evidence retained have been able to use new technologies to sever their connection with the crime. This process, then, allows the original decisions leading to prosecution, the presentation of evidence at trial, and arguments on appeal to be critically reviewed against a putatively correct answer. Without wanting to suggest that DNA evidence is infallible, DNA exclusion will very often mean that those convicted of sexual assault or murder are factually innocent. There has never been such a reliable means of systematically reviewing convictions and criminal justice processes.


Forensic-science evidence appears to have played some active role in a large percentage of wrongful convictions (see Figure 1). In the majority of cases, the incriminating forensic-science evidence was exaggerated or simply wrong (see Table 1) and compounded with other faulty evidence, often mistaken eyewitness identifications of the accused. In many cases, weak or invalid forensic-science evidence was presented as substantial corroboration in circumstantial cases (and quite possibly accepted as such by the trier of fact). Yet, in many cases, the techniques relied upon by scientists, technicians, and police officers (and prosecutors) had never been shown to be reliable. Significantly, in these cases, admissibility standards, judicial discretions to exclude evidence, and appeals did not identify this oversight and/or did not lead to the exclusion of the expert evidence. Jurors (and judges) either did not hear about limitations or did not, in the overall context of the case, consider them sufficiently problematic to produce reasonable doubt. Although the data is American, many of the forensic-science techniques featuring in wrongful-conviction cases are routinely used in criminal justice systems around the world, and many of the problems with the (mis)presentation of results appear to be ubiquitous and continuing.

Detailed analysis, by Brandon Garrett and Peter Neufeld, of the trials of those wrongfully convicted, illustrates dangers with forensic-science evidence and the presentation of this evidence in criminal trials, as well as the inability of criminal justice processes (including those jurisdictions with a Daubert-inspired admissibility standard) to adequately deal with flawed or exaggerated forensic-science evidence. On the basis of a

---

**FIGURE 1**

---

review of 137 trial transcripts from wrongful conviction cases, Garrett and Neufeld concluded that in 82 cases (or 60 per cent of their sample) the prosecution presented invalid forensic-science evidence (see Table 1). That is, in 60 per cent of their sample the prosecution presented testimony that mis-stated empirical data or was wholly unsupported by empirical data.\(^6\) The unreliable (and misleading) testimony was not limited to one or two ‘bad apples’ (although these were undoubtedly included in their sample), but was presented by 72 forensic scientists, from 52 laboratories and hospitals spread across 25 states.\(^6\) The incriminating forensic-science evidence, ultimately discredited by DNA exclusion, included serological evidence along with hair, fibre, bite-mark, and voice-, shoe- and finger-print comparisons.\(^6\)

In terms of the actual trials, Garrett and Neufeld found remarkably few cases where defence counsel ‘cross-examined analysts concerning invalid testimony.’\(^6\) It is perhaps not coincidental that defence counsel rarely obtain their own expert. Moreover, in ‘the few cases in which invalid forensic science was challenged, judges seldom provided relief. Courts do not typically review testimony after finding the underlying

<table>
<thead>
<tr>
<th>Type of forensic analysis</th>
<th>Cases with trial transcripts</th>
<th>Cases involving invalid science testimony</th>
<th>Percentage of cases with trial transcripts involving invalid science testimony</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serology</td>
<td>100</td>
<td>57</td>
<td>57</td>
</tr>
<tr>
<td>Hair comparison</td>
<td>65</td>
<td>25</td>
<td>57</td>
</tr>
<tr>
<td>Soil comparison</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Finger-print comparison</td>
<td>13</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Bite mark comparison</td>
<td>6</td>
<td>4</td>
<td>67</td>
</tr>
<tr>
<td>Shoe-print comparison</td>
<td>3</td>
<td>1</td>
<td>33</td>
</tr>
<tr>
<td>DNA testing</td>
<td>11</td>
<td>3</td>
<td>27</td>
</tr>
<tr>
<td>Voice comparison</td>
<td>1</td>
<td>1</td>
<td>100</td>
</tr>
</tbody>
</table>


\(^6\) Garrett & Neufeld, supra note 64 at 9, 16.

\(^6\) Ibid at 9.

\(^6\) Interestingly, these findings include cases where DNA evidence was mishandled, misinterpreted, or fabricated.

\(^6\) Garrett & Neufeld, supra note 64 at 10–1.
methodology reliable and permitting the forensic analyst to take the stand.'69 The authors concluded that ‘[t]he adversarial system largely failed to police the invalid testimony during these trials,’ with defence counsel rarely cross examining or being granted funding to retain experts so that ‘the defense is typically an unarmed adversary that lacks expert assistance.’70

Case reviews by Garrett and Neufeld find support in other recent scholarship. The work of Keith Findley, in particular, resonates with the emerging problems with forensic-science evidence. Along with colleagues, Findley has endeavoured to identify and explain disparities in the respective abilities of police and forensic scientists, on the one hand, and the accused, on the other, to develop and assess evidence. He characterizes these limitations as the ‘adversary imbalance.’71 Generally, criminal defendants, including those facing many years in prison, are in a worse position to obtain and develop evidence than those in civil courts disputing over money.72 And those who are factually guilty often have an advantage over the innocent on account of their familiarity with the facts in issue. According to Findley,

[T]he criminal justice system skews the risk of error against the innocent defendant by giving the prosecution far superior access to forensic science and expert witnesses . . . [F]ar from skewing the risk of error to guard against convicting the innocent, the investigative, evidence-collection, and analytical stages of the process give almost all advantages to the prosecution. If the presumption of innocence truly puts the risk of error on the government rather than the accused, that presumption will have to be effectuated in ways that compensate for the defendant’s inherent disadvantage in the initial stages of a criminal case.73

Findley’s work is particularly helpful in explaining how the political economy of criminal justice affects the practical administration of admissibility standards. His conclusions about the adversary imbalance makes sense of empirical research that suggests that Daubert has had a far

69 Ibid at 2.
70 Ibid at 10–1; see also 33, 97. Not all of the (mis)conduct identified by Garrett & Neufeld was manifested at trial. In addition to misrepresentations by prosecutors and forensic scientists, they reported the non-disclosure of exculpatory materials, serious mistakes in laboratories, and outright fabrication; ibid at 76.
72 Findley, ibid at 901.
73 Ibid at 906–7.
more conspicuous impact in civil proceedings involving well-resourced corporate defendants than in American criminal justice practice.

Affirming preliminary work by Risinger, Findley found that, notwithstanding the very different obligations vis-à-vis the state and the accused, Daubert seems more likely to exclude expert evidence adduced by the accused than expert evidence offered by the prosecution. Standards of admissibility, purportedly symmetrical between the state and the accused, play out in an asymmetrical fashion that allows better resourced prosecutors to challenge expert evidence offered by the defence – in terms of admissibility and probative value – but that rarely results in the state’s expert evidence being excluded. For Findley, ‘these admissibility patterns are perverse, at least if we take seriously our professed commitment to protecting the innocent as a highest-order value.’

Disparities between the prosecution and defence are accentuated by the reluctance of many American trial judges, notwithstanding permissive decisions such as *Ake v Oklahoma*, to expend scarce local resources on defence experts. Moreover, even where funding is available, Findley and Garrett and Neufeld both describe the difficulty of obtaining the services of a critical expert: ‘Where does one find a toolmark examiner, finger-print analyst, or bullet lead composition scientist? The only place these “experts” exist – because the only place these “sciences” exist – is in the government crime laboratories or spin-off private laboratories whose roots are in the law enforcement community.’

Attempts to question the reliability of finger-print evidence represent a conspicuous version of this problem. Simon Cole (AB Princeton, PHD Cornell), a sociologist of science specializing in the history and sociology of finger-print evidence, has encountered aggressive admissibility challenges when he has been enlisted to testify about latent finger-print evidence. In most jurisdictions, it is difficult to obtain critical expert evidence and even more difficult to get it into court if the expert is not from the same professional group. This applies even where the proposed witness is a highly qualified research scientist or statistically sophisticated social scientist critical of the techniques or even expertise claimed by the

74 Ibid at 928.
76 The United States Supreme Court’s decision in ibid empowers trial judges to appoint defence experts but has been read narrowly only to require their appointment when essential to the defence; see Paul C Giannelli, ‘Ake v Oklahoma: The Right to Expert Assistance in a Post-Daubert, Post DNA World’ (2004) 89 Cornell L Rev 1305.
77 Findley, supra note 71 at 943.
prosecution’s expert(s). Cole, for example, has studied finger-printing for
more than a decade and has written numerous scholarly books and
articles on the subject but is not trained as a finger-print examiner.
Courts have been inconsistent in their responses to his well-documented
and apparently credible concerns about the reliability of latent-finger-
print evidence.79

Early experience with DNA evidence, before the techniques were stabil-
ized and consensus reached on the population genetics and calculation
of probabilities, illustrates how these kinds of problems occur even in
areas where there are pre-existing and independent scientific fields.
Initially, academic and research biologists were reluctant to become
involved in challenges to incriminating DNA evidence, apparently believ-
ing that forensic applications of these established biological techniques
were empirically grounded and reliable. When a few biologists, statist-
cicians, and population geneticists (e.g. Eric Lander and Simon Ford)
did enter the fray on behalf of the defence circa 1990, they and their con-
cerns were derided as academic or theoretical – insensitive to practical
investigative considerations. Eventually, these interventions, along with
persistent efforts by an elite group of defence attorneys (including
Peter Neufeld, Barry Scheck, and William Thompson) persuaded a
handful of judges that there might be problems with DNA evidence.
Judicial anxieties and the exclusion of some DNA evidence led to insti-
tutional mobilization and the investment of hundreds of millions of
dollars of public money to develop a more empirically rigorous plat-
form.80 Significantly, in contrast to DNA evidence, for most of the forensic
sciences there are no independent research communities; that is, there
are few, if any, practitioners who are not linked to law-enforcement or gov-
ernment agencies. This, as we shall see, has not prevented criticism, but it
has made it more difficult to persuade judges that expertise from beyond
forensic-science and investigator communities might provide valuable
and even superior forms of insight into the limitations with the pro-
duction, presentation, and reception of forensic-science evidence.

The results of DNA exonerations disclose very real limitations with incrim-
inating forensic-science evidence and contemporary legal practice. The
American experience suggests that the role of the forensic sciences in
wrongful convictions should not be dismissed as rare or isolated events

79 See e.g. Simon Cole, Suspect Identities: A History of Fingerprinting and Criminal
Identification (Cambridge, MA: Harvard University Press, 2001); ‘More than Zero:
Accounting for Error in Latent Fingerprint Identification’ (2005) 95 J Crim L &
Criminology 985; ‘Is Fingerprint Identification Valid? Rhetorics of Reliability in
80 For a very accessible history, see Jay D Aronson, Genetic Witness: Science, Law, and
Controversy in the Making of DNA Profiling (New Brunswick, NJ: Rutgers University
Press, 2007) [Aronson].
associated with a few ‘bad apples.’ Although biological samples are only available in a minority of crimes, the Innocence Projects have identified a significant number of serious cases where the forensic sciences contributed to wrongful convictions. If we look at the breakdown of the problematic forms of evidence recurring in such wrongful convictions, we find that forensic-science evidence is a regular feature in a high proportion of the cases in this sample (see Figure 1). Rather than acting independently (and efficiently) to identify weaknesses or mistakes, such as mistaken eyewitness identifications, the forensic sciences seem to have substantially contributed to mistaken decisions to prosecute and convict and to the resilience of convictions on appeal. There are few reasons to believe that such contributions are not representative.

Perhaps ironically, given its troubled beginnings, DNA evidence presents an unprecedented, if unanticipated and limited, opportunity to consider the operation of a wide range of disciplines and procedures associated with criminal justice, particularly the fairness of the adversarial criminal trial and its ability to protect the innocent. It provides an opportunity to evaluate forensic-science evidence and adversarial-trial safeguards in action across a large number of sexual assault and sexual assault–murder convictions. The DNA exonerations demonstrate that forensic science evidence frequently contributes to mistaken decisions and that trial safeguards, such as admissibility standards, cross-examination, defence counsel, defence experts, judicial directions and warnings, burdens of proof and appeals, all have limited value in generating doubt even in circumstances where the accused are factually innocent.

While discussing the limitations of the adversarial trial, particularly trial safeguards, it is appropriate to mention another strand of relevant empirical research, less directly focused on actual criminal trials, but equally troublesome for those defending current legal practice as a credible response to incriminating expert evidence. Most of the conclusions drawn from the review of wrongful-conviction cases find independent support in experimental psychological research.

On the (in)effectiveness of cross-examination, for example, ‘studies have found little or no ability of cross-examination to undo the effects of an expert’s testimony on direct examination.’ Research on the

81 This raises interesting questions, not explored here, about the degree to which the forensic sciences in these cases may have been biased by faulty lay evidence or the assumptions of investigators.


limited efficacy of judicial directions and warnings is notorious and has led to numerous efforts, across virtually all common-law jurisdictions, to modify the form of words and timing of judicial instructions although, given the magnitude of the problems with directions, this would seem to be tinkering at the margins. Perhaps more disconcertingly, in most of the published empirical research, judges have performed very much like jurors when surveyed about their technical scientific literacy. The contention that judges are consistently better than juries with scientific evidence has limited empirical support. Overall, the experimental research ‘results should give pause to anyone who believes that the traditional tools of the adversarial process (e.g., cross-examination, opposing experts, instructions) will readily undo the effects of misleading expert testimony.’

Experimental work in the social sciences casts doubt on the ability of the modern adversarial criminal trial (and appeal) to handle unreliable expert evidence and thus supplements the findings of the DNA


86 McQuiston-Surrett & Saks, supra note 83 at 451.
exonerations. It reinforces the need for incriminating expert evidence to be reliable because the various trial safeguards, along with lay jurors, trial, and appellate judges, have not performed well in response to prosecutions and convictions incorporating unreliable expert evidence. It is important to impose a reliability threshold because of the subtle and synergistic ways in which unreliable or exaggerated expert evidence may contaminate criminal trials. The fact that a trial becomes substantially unfair when the state adduces and relies upon insufficiently reliable expert evidence – regardless of the overall strength of the case – tends to be overlooked. The fact that it seems to be incredibly difficult to effectively challenge even the weakest incriminating forensic-science evidence adduced by the state reinforces the importance of admissibility decision making.

D CRITICS AND CRITIQUES OF THE FORENSIC SCIENCES

The DNA exonerations examined above provided part of the impetus for the NAS inquiry into the forensic sciences. In addition, a growing chorus of criticism directed at the scientific status of the forensic sciences since the mid-1980s provided an important foundation for the NAS report. Initially, criticism tended to focus upon particular individuals, institutions, or techniques. Frequently, they arose out of scholarly involvement in isolated cases, often leading to detailed study of specific techniques where a range of scientists, social scientists, and technically literate lawyers were scandalized by the low quality of incriminating expert opinions relied upon at trial and by the unwillingness of appellate courts to intervene. Handwriting and ‘voice-print’ evidence are conspicuous examples.87 The emergence and refinement of DNA profiling accentuated disparities among the range of practices and activities loosely assembled under the rubric of forensic science and medicine. Gradually, critics recognized that many of the issues and problems they were encountering were not isolated or limited to the most marginal of the forensic science disciplines. Rather, they seemed to be endemic to all forensic sciences.

Michael Saks, Jonathan Koehler, William Thompson, Peter Neufeld, Barry Scheck, David Faigman, Michael Risinger, Margaret Berger, Erica Beecher-Monas, Paul Giannelli, David Kaye, and Simon Cole have been

among the most persistent, visible, and consistently critical voices.\textsuperscript{88} In their own way, each of these authors has called for a ‘revolution’ in the forensic sciences to raise them from their ‘subnormal status’ as ‘police crafts’ to a position closer to ‘an idealized conception of “real” science.’\textsuperscript{89} Generally, critics of the forensic sciences have portrayed the \textit{Daubert} criteria as an accurate description of the essential ingredients of authentic (or normal) science and have used such representations to make pejorative comparisons with institutionalized forensic science and medicine.\textsuperscript{90} The significance of particular models of science and expertise will re-emerge in our conclusion. For the moment, we are primarily concerned with influential critiques that placed the institutionalized forensic sciences under considerable pressure.

We can obtain a sense of the force of the interventions by considering a few representative publications. In one review, for example, Saks and Faigman provocatively characterize many areas of contemporary forensic science as ‘the non-science forensic sciences’ because ‘their primary claims for validity rest on anecdotal experience and proclamations of

\begin{itemize}
\end{itemize}

\begin{itemize}
\end{itemize}
success over time’ – sometimes linked to successful prosecutions.91 The ‘non-science forensic sciences’ include comparison of finger-prints, handwriting, voice prints, tool marks, firearms, tire prints, shoe prints, hairs, and textiles and are based on claims of individualization and uniqueness that are questionable in the absence of ‘systematic, rigorous research bases’ (see Table 1).92

All of the critics use the empirical underpinnings and probabilistic approaches associated with contemporary DNA evidence as a benchmark against which to generate unfavourable contrasts with other forensic sciences – especially those involving comparisons.93 It is their contention that, had forensic sciences concerned with individualization emerged from ‘normal’ (and implicitly authentic) science, their ‘techniques probably would resemble DNA typing, with its measurement of attributes, sampling of variation in populations, and statistical bases. Error rates, probability levels, confidence intervals, and so on would be a natural part of what developed.’94

For critics, the failure to satisfy even minimum standards is compounded by the way forensic-science practitioners conduct their work. They often point to the existence of ‘a guild of mutually reassuring examiners who have come to believe in the truth of their claims, often sounding more like a faith-based religion than a data-based science.’95 In many areas of forensic science and medicine, there are few, if any, procedures protecting practitioners from a wide range of undesirable influences and pressures. Many forensic scientists are at risk of being influenced (whether they appreciate it or not) through exposure to information that is not necessary for their specific analysis.

The effects of bias and forms of suggestion on perception, memory, and interpretation represent an important area of inquiry, notorious among research psychologists. Here, influential and unsettling work by Itiel Dror looms large. Dror and his colleagues have produced several influential papers demonstrating that highly trained forensic scientists,

94 Saks & Faigman, supra note 91 at 153.
95 Ibid at 157.
and even those experts using long-standing standards and protocols (such as experienced latent-finger-print examiners) encounter extreme difficulty overcoming contextual influences and biases when it comes to interpreting evidence. Dror's studies reinforce the importance of structural independence (over impartiality or personal attempts to be impartial) and explain why blinding is, where possible, highly desirable.96 Where protocols and standards are not grounded in demonstrably reliable methods and techniques or where personal proficiency is not assessed, dangers would seem to be not merely acute but, as the review of criminal trials associated with DNA exonerations suggests, difficult to contest, expose, and remedy.97 Together, academic critiques of forensic science, combined with the refinement of DNA profiling and the effects of DNA exonerations, laid the principal foundations for the important NAS report on forensic science.98

E THE NAS REVIEW: STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES

In February 2009, a Committee of the National Research Council (of the NAS) released its long awaited report into the condition of forensic science in the United States.99 Written by eminent scientists and engineers along with a few lawyers, the NAS report offers a useful overview of problems confronting forensic science and medicine in the United States. Although the inquiry associated with the report was broadly focused, particularly


98 See Michael Saks & Jonathan Koehler, ‘The Coming Paradigm Shift in Forensic Identification Science’ (2005) 309 Science 892 [Saks & Koehler, ‘Paradigm’]; Harry Edwards, ‘Solving the Problems That Plague the Forensic Science Community’ (2010) 50 Jurimetrics Journal 5. Congress appropriated funds for the NAS study in late 2005, the same year that an influential commentary was published in Science and after flaws in finger-print analysis was underlined by the arrest and then release of Brandon Mayfield in connection with the Madrid bombings; see US, Office of the Inspector General, A Review of the FBI's Handling of the Brandon Mayfield Case (Washington, DC: US Department of Justice, 2006). Note that, before the Mayfield incident, the FBI claimed that its error rate in finger-print identification was ‘essentially zero’; US v Crisp, 324 F (3d) 261 at 269 (4th Cir 2003) [Crisp].

99 NAS report, supra note 2 at S-1.
on organizational reform, our interest is primarily in its treatment of the research base underlying forensic science and medicine.

The committee commenced its inquiry recognizing that ‘significant improvements [were] needed’ but was surprised by ‘the consistency of the message’ it received.\textsuperscript{100}

The forensic science system, encompassing both research and practice, has serious problems that can only be addressed by a national commitment to overhaul the current structure that supports the forensic science community in this country. This can only be done with effective leadership at the highest levels of both federal and state governments, pursuant to national standards, and with a significant infusion of federal funds.\textsuperscript{101}

According to the report, the forensic sciences are characterized by ‘fragmentation and inconsistent practices.’\textsuperscript{102} There are ‘great disparities among existing forensic science operations in federal, state and local law enforcement jurisdictions and agencies.’\textsuperscript{103} The disparities covered funding, equipment and instrumentation, accreditation and oversight, and the availability of skilled personnel. Federal programs tend to be better staffed and funded even though state and local jurisdictions are responsible for most criminal-law enforcement. In consequence, ‘the depth, reliability, and overall quality of substantive information arising from the forensic examination of evidence available to the legal system vary substantially across the country.’\textsuperscript{104} Fragmentation, according to the report, is ‘compounded because operational principles and procedures for many forensic science disciplines are not standardized.’\textsuperscript{105} Standardized training and certification of practitioners are lacking and protocols for the conduct of forensic science work are not routinely enforced.

In addition, and perhaps most notoriously, the committee cast doubt on the ability of many forensic techniques to ‘support conclusions about “individualization.”’\textsuperscript{106} Effectively endorsing the concerns of the critics, the committee explained that

\textit{[w]ith the exception of nuclear DNA analysis, however, no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source … The simple reality is that the interpretation of forensic

\begin{itemize}
\item \textsuperscript{100} Ibid at P-1, S-5.
\item \textsuperscript{101} Ibid at xx.
\item \textsuperscript{102} Ibid at 6.
\item \textsuperscript{103} Ibid at 5.
\item \textsuperscript{104} Ibid at 6.
\item \textsuperscript{105} Ibid.
\item \textsuperscript{106} Ibid at 87. See also Saks & Koehler, ‘Individualization,’ supra note 92; Cole, ‘Forensics,’ supra note 92.
\end{itemize}
evidence is not always based on scientific studies to determine its validity. This is a serious problem.\textsuperscript{107} The committee was concerned about the lack of empirical research behind many of the forensic techniques and approaches. Research on ‘the limits and measures of performance and … the impact of sources of variability and potential bias … is sorely needed.’\textsuperscript{108} The NAS expressed fundamental concerns about whether the forensic sciences have a reliable scientific base and reliance ‘on human interpretation that could be tainted by error, the threat of bias, or the absence of sound operational procedures and robust performance standards.’\textsuperscript{109}

Although the NAS report is clearly inspired by the kind of idealized vision of science associated with the critics’ interpretation of Daubert, the committee concluded – consistent with the empirical studies examined above – that Daubert ‘has done little to improve the use of forensic science evidence in criminal cases.’\textsuperscript{110} More controversially, the report asserts that ‘judicial review, by itself, will not cure the infirmities of the forensic science community.’\textsuperscript{111} Rather than focus on legal institutions and trial procedures – such as rigorous application of reliability standards for expert evidence and vigorous cross-examination – the report made ambitious recommendations about the need for structural independence and a dramatic increase in federal funding for research, accreditation, and the regulation of forensic sciences in the United States.\textsuperscript{112}

F OVERVIEW OF THE AMERICAN EXPERIENCE

The NAS report endorsed the critics’ contention that much of what is presented as forensic science, including the incriminating opinion evidence routinely proffered in court, is not genuine science. The report is striking in its sceptical response to the legal legacy of Daubert. In some ways, one might have expected the NAS to embrace Daubert, given its focus on issues such as the need for testing, peer review, and the determination, if

\textsuperscript{107} NAS report, supra note 2 at 7–8, 87 [emphasis added].
\textsuperscript{108} Ibid at 8.
\textsuperscript{109} Ibid at 9, 87 [emphasis added].
\textsuperscript{110} Ibid at 106.
\textsuperscript{111} Ibid at 12, 110 [emphasis added].
possible, of quantifiable error rates. Instead, the NAS reveals itself to be conversant with the limitations of Daubert (and the judiciary) in the context of criminal justice.

The report concludes that courts cannot assume the primary role in reforming the forensic sciences. Lax admissibility practices, along with consistently poor performance by lawyers and judges, have meant that there are few incentives for institutionalized forensic scientists to undertake research or inquiry into the validity and accuracy of forensic science techniques (or the investigators and prosecutors drawing upon them to qualify their use). The forensic sciences have been allowed to carry on largely without validation studies, without efforts to determine error rates or undertake credible proficiency testing. Indeed, the absence of research and data and, in many cases, credible standards and protocols actually allows forensic scientists to tailor or conjure evidence and opinions in ways that are consistent with the state’s case or adverse to the accused (rather than indicative of factual guilt or innocence).

There are, however, serious questions about whether (or how quickly) the forensic sciences can respond to such a ‘root and branch’ critique. Imposing practices and standards from DNA profiling (or analytical chemistry) upon other types of potentially reliable scientific and non-scientific forms of expertise may not be appropriate, especially with respect to forensic sciences that do not have a base in the universities and are unlikely to attract funding for the type of intensive research that would be necessary to provide validation. There is also a danger that imposing standards based on idealized models of science – idealization common to Daubert, the forensic science critics, and now the NAS report – may encourage the generation of exceptions capable of accommodating many forms of expert-opinion evidence based on experience, intuitive appeal, or perceived need.

Notwithstanding these problems, as emerging commentary from the United States explains, lawyers and judges should not abdicate responsibility for the quality of forensic-science evidence, if only because of concerns about ensuring the integrity and reliability of criminal-justice-system outcomes. Indeed, in the short term, legal personnel and institutions may be required to play a more prominent role.

113 Findley, supra note 71 at 944.
114 Some disciplines, such as forensic medicine, may be advantaged relative to other disciplines, such as the comparison sciences, because the former has established linkages to universities and research traditions.
At this juncture, it is illuminating to consider current proposals for reform in the United Kingdom. In 2009, responding to a damaging series of child-death prosecutions (barely a decade after the fall-out from miscarriages of justice resulting from IRA terrorist attacks) and one of the most liberal admissibility regimes in the common-law world, the Law Commission of England and Wales released Consultation Paper No 190 describing four proposals for the revision of its admissibility standards for expert-opinion evidence in criminal proceedings.116 The commission’s preferred approach is to adopt a reliability standard.117

The Consultation Paper explains,

Our view is that reform is needed. We provisionally propose:

(1) that there should be a new statutory test for determining the admissibility of expert evidence in criminal proceedings, which would apply whether the evidence is tendered by the prosecution or by the accused;


(2) this new test would provide that expert evidence is admissible only if the court is satisfied that the evidence is sufficiently reliable to be admitted; and
(3) in determining whether or not the test is satisfied the court would in all cases have to refer to a statutory list of guidelines for assistance.\(^{118}\)

The new statutory provisions (from ‘(1)’ and ‘(3)’ in the previous extract) should resemble the following:

(1) The opinion evidence of an expert witness is admissible only if the court is satisfied that it is sufficiently reliable to be admitted.

(2) The opinion evidence of an expert witness is sufficiently reliable to be admitted if:
   a) the evidence is predicated on sound principles, techniques and assumptions;
   b) those principles, techniques and assumptions have been properly applied to the facts of the case; and
   c) the evidence is supported by those principles, techniques and assumptions as applied to the facts of the case.

(3) It is for the party wishing to rely on the opinion evidence of an expert witness to show that it is sufficiently reliable to be admitted.\(^{119}\)

According to the Consultation Paper, ‘corroboration through testing … is without doubt the best basis for determining the evidentiary reliability of expert evidence which can be tested.’\(^{120}\) The proposed statutory provision ‘demands that expert witnesses called to testify on forensic scientific techniques must demonstrate that the techniques in question are a reliable basis for their opinion evidence. More to the point, it should ensure that convictions (and acquittals) are not based on unwarranted inferences drawn from unreliable evidence.’\(^{121}\) The influence of Daubert and the revised Rule 702 should be obvious.

Revealingly, the Consultation Paper appears practically indifferent to the empirical research suggesting that US ‘judges have hardly any understanding of the first and third of the Daubert factors’ and that the ‘factors are neither accurately nor consistently being applied in the state courts

---

118 Consultation paper, supra note 6 at 3–4, 50 [emphasis added].
119 Ibid at 50 [emphasis added]. Since the writing of this article, the Law Commission has produced its final report. The report is largely consistent with the consultation paper, especially the emphasis on ‘sufficiently reliable.’ The original distinction between scientific and non-scientific expertise has been read down although the proposals retain considerable judicial discretion in relation to classification and admissibility; see Law Commission, Expert Evidence, supra note 30 at 148, discussed in Gary Edmond, ‘Is Reliability Sufficient? The Law Commission and Expert Evidence in International and Interdisciplinary Perspective’ (2012) 16 International Journal of Evidence & Proof [forthcoming].
120 Consultation paper, supra note 6 at 38 [emphasis in original].
121 Ibid at 51.
which follow the guidelines.’ 122 These are mentioned in passing, without any attempt at engagement or response. The Law Commission is content to rely on somewhat disingenuous assertions that ‘criminal trial judges are in a better position than juries’ to obtain the necessary information and that American judges, more than fifteen years after Daubert was decided, are ‘becoming more sophisticated in their assessment of the reliability of expert evidence as they gain more experience.’ 123

Even though the terms of reference were focused on expert evidence in criminal trials, the Law Commission does not engage with the empirical research – outlined in the first part of this article – suggesting that Daubert has not been applied to evidence offered by plaintiffs in civil cases with nearly the same rigour as it has to expert evidence challenged by the prosecution in criminal proceedings. Nor does it focus on the need for the defence to have the necessary training and funds effectively to challenge expert evidence offered by the prosecution.

With respect to criminal-law values, the Law Commission appeals to ‘truth’ over ‘fairness’ in recommending a symmetrical approach to admissibility. The commission elides the case for imposing stricter standards on the state, notwithstanding its superior resources and the burden and quantum of proof imposed in criminal cases, by raising concerns that the accused may be ‘permitted to distort the truth-seeking function of the trial’ by introducing ‘inherently unreliable but impressive-looking expert evidence, particularly pseudo-science.’ 124 The Law Commission’s approach discounts a more asymmetrical admissibility test that would demand a basic threshold of reliability for expert evidence offered by the accused while imposing a higher standard on expert evidence offered by the state. The commission’s approach would appear to embody a deep suspicion of those accused of crime and seems to overlook the fact that empirical studies and the NAS report both suggest that the state systematically relies upon ‘impressive looking’ but ultimately suspect expert evidence.

A EXPERIENTIAL EXCEPTIONS: THE BACKLASH AND MORE ONEROUS ADMISSIBILITY STANDARDS

In addition, the Law Commission recommends dividing expertise into the categories ‘scientific’ and ‘experienced-based’ and exempting experienced-based expert witnesses from the proposed statutory test predicated

---

122 Ibid at 44.
123 Ibid at 46
124 Ibid at 62.
upon the Daubert criteria.\textsuperscript{125} Whereas the admissibility standard for scientific evidence would focus on testing and margin of error, peer review, and the ‘scientific validity’ of opposing views, the experience-based standard would be based on the expert’s standing in the professional or expert community, the relevance of the expert’s experience, the expert’s impartiality and whether his or her opinion had previously been shown to be in error.

An example of such reliability guidelines is provided in relation to handwriting expertise:

\[\text{T}he\ reliability\ of\ an\ expert\ witness’s\ testimony\ on\ forensic\ document\ examination\ (to\ determine\ whether\ or\ not a\ document\ is\ a\ forgery)\ would\ be\ determined\ on\ the\ basis\ of,\ amongst\ other\ things,\ the\ witness’s\ experience,\ the\ number\ of\ standard\ points\ of\ comparison\ used\ and\ a\ detailed\ description\ of\ the\ process\ by\ which\ the\ expert\ reached\ his\ or\ her\ opinion.\textsuperscript{126}\]

Recall from early interventions by several critics that widespread concerns with forensic science originally emerged out of adverse encounters with handwriting-comparison (and document-examination) evidence. While the need for sensitivity to the kind of expertise and claims being made is obvious, the manner in which the proposal itself explains this distinction introduces some of the very problems which the NAS and the critics sought to remedy.

Excusing handwriting analysis and other forensic comparison practices such as finger-print evidence as experienced-based effectively insulates the practitioners from scrutiny and reduces the need actually to assess or improve their practices. None of the traditional desiderata of handwriting comparison – such as ‘standard points,’ length of experience, and detailed description of process – matter if the technique is unreliable. Revealingly, these considerations may appear persuasive – as they apparently do to the authors of the consultation paper. They reveal very little, however, about whether handwriting experts can actually do as claimed and how often their opinion is accurate. Once again, such abilities could be readily gauged through empirical assessment such as blind trials and proficiency testing, but creating admissibility exceptions for experience – in the absence of a genuine interest in reliability – will spare some of the most problematic forms of incriminating expert opinion from having to demonstrate actual capabilities (and limitations).\textsuperscript{127}

---

\textsuperscript{125} For this and the paragraph that follows, see ibid at 53–6. A recent appeal, illustrating the importance of experience (and some gradual sensitivity to the discourse of reliability associated with DNA evidence, at least) is Weller, supra note 116.

\textsuperscript{126} Consultation paper, supra note 6 at 58.

\textsuperscript{127} Such testing helps to assess value and strength as well as limitations; see e.g. Bryan Found & Doug Rogers, ‘The Probative Character of Forensic Handwriting
Curiously, and inconsistent with its having espoused the need for evidence of reliability and its earlier celebration of the scientific method (as testing), the Law Commission further weakens the significance of any reliability threshold by suggesting that

[T]he question whether or not an expert witness’s evidence is sufficiently reliable to be admissible is not a fact which is susceptible of proof. Rather, it is more akin to the type of question which requires the judge to form a judgment in the light of all the available evidence such as:

1. The question whether the admission of prosecution evidence would have such an adverse effect on the fairness of the proceedings that the court ought not to admit it; or
2. The question whether or not the accused will receive a fair trial if he or she is to be tried many years after the time when the offence charged was allegedly committed.128

Interest in reliability is likely to be tempered by considerations that reinforce the historically liberal approach to admissibility decision making in England and Wales. Rather than require evidence of reliability, the trial judge is to consider ‘sufficient reliability’ in relation to the overall case and traditional concerns about fairness. Without considering the actual reliability of incriminating opinion evidence – something purportedly not ‘susceptible of proof’ – the trial judge is, nevertheless, expected to assess its ‘adverse effect’ and potential for making the trial unfair.129

Interestingly, the exception proposed by the Law Commission is not unlike the way judges in the United States have read down the Daubert decision to temper its rigour in response to incriminating expert evidence. In appeals in state and federal courts, many types of forensic science have been admitted without evidence of testing or accuracy on the basis that they are primarily ‘technical and specialized knowledge’ and therefore beyond the application of the Daubert criteria. Prosecutors routinely contend that the Daubert criteria apply only to scientific evidence and so other types of incriminating expert evidence should not be required to satisfy them.130

---

128 Consultation paper, supra note 6 at 63, 65.
129 This is unlikely to lead to exclusion. For an account of the treatment of ‘unfair prejudice’ in Australia, see Edmond, ‘Specialised Knowledge,’ supra note 71.
130 Kumho, supra note 9, tends to be applied more flexibly in the criminal justice sphere to temper the rigour of admissibility standards for the state’s incriminating expert evidence.
A prominent example, from the United States, is the appeal in *US v Starzecpyzel*. There, in response to a *Daubert* challenge to hand-writing evidence, a federal judge accepted that ‘forensic document examination . . . cannot, after *Daubert*, be regarded as “scientific . . . knowledge.”’ However, the evidence was not on that basis excluded, the judge concluding that identification from hand-writing did not have to satisfy *Daubert*. This illustrates how when a forensic science was found to stand on a weak foundation, the threshold of admission was lowered to accommodate this weakness . . . Ironically, then, fields that initially gained entry to the courts by declaring themselves to be ‘sciences’ now sought to remain in court by denying any connection with scientific methods, data, or principles.

B OVERVIEW OF THE LAW COMMISSION’S APPROACH
The Law Commission’s consultation paper was released several months after the NAS report but there are no references to the report. Rather than engaging with the extensive critical literatures and empirical studies in order to develop more contextually appropriate rules and procedures, the consultation paper blithely dismisses criticism of the model of science promoted in *Daubert* as epistemological radicalism (or worse). Unlike the NAS, the Law Commission says very little about the need to reform the production of forensic-science evidence. This overlooks one of the main points in the NAS report; namely, the need for scientific leadership and supervision.

The Law Commission makes only passing reference to the need for training of judges and lawyers. It does not refer to the importance of legal-aid funding both for experts and for challenges to forensic science offered by the Crown. The commission also remains resolutely within a symmetrical framework, stressing issues such as disclosure of expert reports by the defence and the need for defence experts to be accredited. The framework encourages a standardized response to the reliability of expert evidence offered by the state and by the accused, even though it is only the former that has contributed to the types of miscarriages of justice that have motivated both the NAS and the Law

132 *Starzecpyzel*, ibid at 1036.
133 Saks & Koehler, ‘Paradigm,’ supra note 98 at 894.
134 Consultation paper, supra note 6 at 37ff.
Commission to focus on the role of forensic and other forms of expert evidence in criminal trials. The commission’s focus seems to accord with the current ethos in Britain of realigning the criminal justice system in favour of the state and victims.  

The Law Commission’s consultation paper fails to engage with the actual production of forensic science, the emerging and converging empirical record, or the problems involved in creating a two-tier approach to admissibility that might create broad exceptions for non-scientific forms of opinion evidence (or expert opinions currently admitted), even those that are susceptible to empirical evaluation and improvement. It does little to address the dangers of wrongful convictions caused, or contributed to, by forensic science and medicine.  


136 Consultation paper, supra note 6 at 2.16–24, 3.17. It is useful to contrast current English practice and the Law Commission’s proposal (and final report) with responses from the May Inquiry: Report of the Inquiry into the Circumstances Surrounding the Convictions Arising out of the Bomb Attacks in Guildford and Woolwich in 1974 (final report) (HC 1993–4, 449) (Chair: Sir John May); and the Runciman Royal Commission into the miscarriages of justice associated with the IRA prosecutions (e.g. the Birmingham six, Guildford four, Maguire seven, and Judith Ward): UK, HC, The Royal Commission on Criminal Justice Report, Cm 2263 in Sessional Papers (Chair: Viscount Runciman) (London: Her Majesty’s Stationary Office, 1993); and recent experience with controversial expert-opinion evidence in paediatric death cases. These experiences, in some respects, mirror the role of DNA exonerations in the United States. They cast doubt on the reliability of forensic science and medicine and illustrate how flawed expert opinion frequently plays a part in wrongful convictions; see e.g. Paul Roberts & Chris Willmore, The Role of Forensic Science Evidence in Criminal Proceedings (London: Her Majesty’s Stationary Office, 1993); Clive Walker and Keir Starmer, eds, Miscarriages of Justice: A Review of Justice in Error (London: Blackstone Press, 1999); Richard Nobles & David Schiff, Understanding Miscarriages of Justice: Law, the Media, and the Inevitability of Crisis (Oxford: Oxford University Press, 2000). The commission, to its credit, uses the case of Sally Clark and other baby-death prosecutions, such as Cannings, as prominent examples but simultaneously gives short shrift to the IRA miscarriages of justice and accompanying reform recommendations; see R v Clark, [2003] EWCA Crim 1020; Cannings, supra note 116; though compare R v Kai-Whitewind, [2005] EWCA Crim 1092 and Harris, supra note 116. Similarly, the Law Commission’s report self-consciously builds on recommendations made by the House of Common’s Science and Technology Committee that there be a reliability-based admissibility test but, in effect, waters this down by proposing exceptions and reiterating the value of existing standards, protections, and judicial abilities; see UK, Science and Technology Committee, Forensic Science on Trial (HC 2005, 96-1) (London: Her Majesty’s Stationary Office, 2005).
Canada provides an interesting case study of admissibility jurisprudence having evolved to require greater attention to considerations of reliability based largely on domestic experience with wrongful convictions. Through the work of public inquiries, the Canadian experience also demonstrates the potential of combining reform proposals that attempt to modify both the practice of specific forensic sciences and standards of admissibility for all expert evidence. Canada, however, could benefit from recognizing the limits of *Daubert* ‘in-action’ and the very real dangers posed by the failure to develop emerging reliability standards, especially for forms of expert opinion that may not readily conform to the criteria associated with the *Daubert* decision.

In the mid-1990s, Canadian admissibility jurisprudence focused on whether expert evidence was (1) relevant; (2) necessary to assist the trier of fact; (3) caught by an exclusionary rule; and (4) proffered by a properly qualified expert. This four-part test, proposed in *R v Mohan*,\(^ {137} \) encouraged the use of preliminary hearings that focused perfunctorily on a proposed witness’s *curriculum vitae*. *Mohan* only explicitly addressed the question of reliability to the extent that, where the expert evidence involved ‘a novel scientific theory or technique’ it should be ‘subjected to special scrutiny to determine whether it meets a basic threshold of reliability and whether it is essential in the sense that the trier of fact will be unable to come to a satisfactory conclusion without the assistance of the expert.’\(^ {138} \) The Court expressed anxiety, lest, ‘dressed up in scientific language which the jury does not easily understand and submitted through a witness of impressive antecedents, this evidence is apt to be accepted by the jury as being virtually infallible and as having more weight than it deserves.’\(^ {139} \) Justice Sopinka explained that ‘the threshold test of reliability’ will ‘generally ensure that the trier of fact does not give [expert evidence] more weight than it deserves.’\(^ {140} \) The Court did not advert to the danger of wrongful convictions and made no mention of the *Daubert* case decided in the previous year. *Mohan*’s focus on novel science was consistent with the view that established (forensic) science is inherently reliable. In this way, *Mohan* neglected the iterative nature of scientific practice, the role of courts in processes of social legitimation, and the possibility of emerging controversy – even within established sciences.\(^ {141} \)
In some respects, the immediate fallout from Mohan was consistent with the American experience with Daubert. The test for admissibility was likely to be used to exclude novel science that the accused, rather than the prosecutor, sought to adduce. In Mohan, the Court held that a paediatrician facing four counts of sexual assault was not entitled to call expert evidence from a psychiatrist suggesting that three of the incidents fitted the psychological profile of a paedophile, while the fourth incident did not. The Court considered that the profiles ‘were not seen as sufficiently reliable to be considered helpful.’

A year earlier, in R v Marquard (1993), the Court had deferred to a trial judge who allowed an expert qualified in child abuse to express an opinion about the nature of a child’s burns and an expert on burns to give an opinion about child abuse. Again, this decision was made without advertting to the dangers of wrongful convictions or the use of expert evidence to shore up a weak case.

The next major admissibility case to reach the Supreme Court was R v J-LJ (2000) where the Court reiterated the need to subject any novel scientific technique ‘to special scrutiny to determine whether it meets a basic threshold of reliability.’ The Court intimated that Mohan had explicitly rejected the general acceptance test (derived from US v Frye) and – ‘moving in parallel with its [US] replacement’ – drew upon the criteria enumerated by the US Supreme Court in Daubert. Justice Binnie reproduced the Daubert criteria and discussed testing, peer review, potential error rate, and general acceptance. The ‘admissibility of expert evidence,’ he explained, ‘should be scrutinised at the time it is proffered, and not allowed too easy an entry on the basis that all of the frailties could go at the end of the day to weight rather than admissibility.’

The Court was anxious that the ‘search for truth’ in the courtroom should not include ‘expert evidence which may “distort the fact-finding process.”’ Applying this to the evidence at hand, it explained that while penile plethysmography was ‘quite useful in therapy because it

142 Mohan, supra note 137 at 46.
144 [2000] 2 SCR 600 at paras 35–6 [J-LJ].
145 Ibid at para 35, quoting Mohan, supra note 137 at 25 [emphasis added].
146 J-LJ, supra note 144 at para 33 [references omitted; emphasis added]; R v DD, [2000] 2 SCR 275 [DD]; R v Dimitrov (2003), 181 CCC (3d) 554 at para 38 [Dimitrov]. The Daubert criteria were earlier endorsed by Justice Hill in R v JET, [1994] OJ no 3067 (QL) at para 75 (Ont Gen Civ).
147 J-LJ, supra note 144 at para 28.
148 Ibid at para 29.
yields some information about a course of treatment,’ the technique ‘is not necessarily sufficiently reliable to be used in a court of law to identify or exclude the accused as a potential perpetrator of an offence.’\textsuperscript{149} Both Mohan and \textit{J-LJ} placed emphasis on the need – for the defence in both cases – to satisfy the trial judge that ‘underlying principles and methodology . . . were reliable and, importantly, applicable.’\textsuperscript{150} Justice Binnie drew attention to the need for experts to furnish ‘the necessary scientific criteria for testing the accuracy of their conclusions.’\textsuperscript{151} Both Mohan and \textit{J-LJ} expressed the need to ‘preserve and protect the role of the triers of fact’ against unreliable expert evidence and expert evidence of exaggerated probative value.\textsuperscript{152}

Although the Court had excluded defence evidence in both Mohan and \textit{J-LJ}, its more demanding approach was also applied to prevent the prosecution from offering expert evidence from a psychologist about the reasons why children might delay reporting sexual abuse in \textit{R v DD}. In a 4–3 decision, the majority in that appeal held that the expert’s evidence was not necessary. They explained that mere helpfulness to the jury was not sufficient for admissibility, given the potential for expert-opinion evidence to distort the fact-finding process. The three judges in dissent were of the opinion that the evidence was relevant and necessary and that appellate deference should be paid to the trial judge’s decision.\textsuperscript{153}

While the increasing recognition of the \textit{Daubert} criteria in \textit{J-LJ} was promising, \textit{R v Trochym} (2007)\textsuperscript{154} provided a potential turning point in Canadian admissibility doctrine. The majority decision, written by Justice Deschamps, began with a reference to a number of high-profile wrongful convictions and the resultant ‘need to carefully scrutinize evidence presented against the accused for reliability and prejudicial effect, to ensure the basic fairness of the criminal process.’\textsuperscript{155} One thing

\textsuperscript{149} Ibid at para 35.
\textsuperscript{150} Ibid at para 50.
\textsuperscript{151} Drawing upon \textit{Davie v Magistrates of Edinburgh}, [1953] Sess Cas 34 at 40. This approach has been given rhetorical if not always practical recognition by the New South Wales Court of Appeal, the Federal Court of Australia, and the High Court of Australia; see \textit{Makita Pty Ltd v Sprowles}, [2001] NSWCA 305; \textit{Sydneywide Distributors Pty Ltd v Red Bull Australia Pty Ltd}, [2002] FCAFC 157; \textit{HG v The Queen} (1999), 197 CLR 414 at para 39; \textit{Dasreel Pty Ltd v Haechar} [2011] HCA 21’ after ‘39’ though as part of the same sentence. footnote 151.
\textsuperscript{152} \textit{J-LJ}, supra note 144 at para 25.
\textsuperscript{153} \textit{DD}, supra note 146.
\textsuperscript{154} Supra note 7. The Ontario Court of Appeal in a series of decisions had earlier focused on the issue of whether expert evidence is ‘sufficiently reliable to be put to the jury for its review.’ \textit{R v Tereceira} (1998), 123 CCC (3d) 1 at para 64; \textit{R v AK} (1999), 137 CCC (3d) 225 at para 84 [AK]; \textit{R v Ranger} (2003), 178 CCC (3d) 375 at para 48; \textit{Dimitrov}, supra note 146 at paras 37–8; \textit{R v Klymchuk} (2005), 203 CCC (3d) 431 at para 36.
\textsuperscript{155} \textit{Trochym}, supra note 7 at para 1.
that seems to have conspicuously changed was the Supreme Court’s willingness to explicitly recognize the reality of wrongful convictions. When Mohan and J-LJ were decided, the Supreme Court allowed fugitives to be extradited to face the death penalty. In 2001, however, the Court abandoned this position, holding that experience with wrongful convictions in Britain, Canada, and the United States now rendered it unconstitutional to extradite a fugitive without assurances that the death penalty would not be applied. A logical implication of these concerns is heightened attention to the actual reliability of expert evidence offered by the prosecution. This is consistent with our argument that admissibility decisions should be informed by criminal-justice values.

Justice Deschamps, in her majority reasons, moved concern about the reliability of expert evidence to centre stage. The majority stated that ‘[r]eliability is an essential component of admissibility. Whereas the degree of reliability required by courts may vary depending on the circumstances, evidence that is not sufficiently reliable is likely to undermine the fundamental fairness of the criminal process.’ And the majority reiterated an earlier interest in Daubert. Rather than being merely helpful the Daubert criteria were now characterized as ‘establishing a framework for assessing the reliability of novel science and, consequently, its admissibility in court.’ At the same time, the majority adopted a more historicized and sophisticated view of science that departed from Mohan’s assumption that only ‘novel science’ is problematic. Justice Deschamps warned that admissibility was not ‘frozen in time’ because scientific knowledge includes controversies that continue to evolve. The majority recognized

157 Trochym, supra note 7 at para 27 [emphasis added].
158 Ibid at para 34.
159 Consider, for example, the revised response to ‘voice-prints’ in NAS, Voice Identification, supra note 87, and the contentious history of DNA evidence and ongoing issues with interpretation in Aronson, supra note 80; Thompson, ‘Painting the Target,’ supra note 62.
161 Justice Bastarache expressed concerns that the majority’s ‘guidance that science which is “so well established” need not be reassessed is so vague that it opens the door to most if not all previously accepted techniques being subject to challenge under J-LJ without establishing a serious basis for the inquiry’; Trochym, ibid at para 139 [references omitted]. The dissenting judgment expressed perhaps justifiable confusion about
that all science is novel in the sense that it has iterative elements. It also
applied its test broadly to include not only expert evidence but eyewitness
identifications that had purportedly been enhanced by hypnosis.\textsuperscript{162}
Admissibility decisions in \textit{Trochym} were predicated on criminal-justice
values that appear to apply to all evidence.

Applying this approach to the facts of the case, the majority explained
that, even if hypnosis was useful in therapy and had been admitted in
earlier proceedings, this did not mean that it was sufficiently reliable as
a source of evidence in a modern criminal trial.\textsuperscript{163} The majority also
expressed concern that, if the reliability of the testimony ‘cannot really
be tested,’ then the right to cross-examine might become ‘illusory,
thereby undermining a key aspect of the adversarial process.’\textsuperscript{164} Again
this represents an attempt to use criminal-justice values – in this case
the right of full answer and defence including cross-examination – to
shape admissibility determinations. The accused in \textit{Trochym} subsequently
pled guilty to manslaughter,\textsuperscript{165} but the case also prompted a withdrawal of
charges against Robert Baltovitch in another case where there were con-
cerns that hypnosis-induced identification evidence may have led to a
miscarriage of justice.\textsuperscript{166}

The \textit{Trochym} majority seems to be proposing the idea, previously
accepted by the Ontario Court of Appeal, that ‘the admissibility of
expert opinion evidence is not a question of precedent.’\textsuperscript{167} The majority in
\textit{Trochym} held that compliance with long-established common-law guide-
lines designed to limit and record any influence that a hypnotist might
exert on a witness were insufficient because they were ‘based on an
assumption that the underlying science of hypnosis is itself reliable in

\textsuperscript{162} Ibid at paras 13, 24.
\textsuperscript{163} Ibid at para 37.
\textsuperscript{164} Ibid at para 60 [emphasis added].
\textsuperscript{165} The fact that \textit{Trochym} was initially convicted with the prosecution adducing and
relying upon apparently unreliable evidence does nothing to weaken our argument.
Even the factually guilty deserve a fair trial.
\textsuperscript{166} \textit{R v Baltovich}, 2008 Can Lii 19209 (ON SC), excluding hypnosis induced testimony; see
also \textit{R v Baltovich} (2004), 191 CCC (3d) 289 (ON CA), describing some of the post-
hypnosis evidence but refusing to exclude it on that basis alone.
\textsuperscript{167} \textit{AK}, supra note 154 at para 75 [emphasis added]. Compare Laurens Walker & John
Monahan, ‘Scientific Authority: The Breast Implant Litigation and Beyond’ (2000)
86 Va L Rev 801–34.
The context of judicial proceedings. The Court noted that half the states in the United States had revisited the issue in light of evolving views by experts about the reliability of such evidence. The Court discussed competing theories concerning the reliability of post-hypnosis evidence and cited much legal literature on the topic. Trochym, we contend, exemplifies a more contextual approach to admissibility because of its recognition of the particular danger of wrongful convictions, because of its engagement with competing theories about hypnosis, and because it embodies concern about the reliability of the technique in terms of the frailties of the adversarial criminal trial. In combination, this recognition means that the majority judgment was more substantively (and reflexively) engaged with expert evidence than was Mohan or even J-L.

The three judges who dissented in Trochym stressed both the need for deference to the decision of the trial judge, as in the American case law (associated with Joiner), and the fact that post-hypnosis evidence had been admitted for three decades and thus did not constitute novel science. Justice Bastarache argued that the majority improperly took judicial notice of the dangers associated with post-hypnosis testimony from reported American cases. Leaving aside questions about the scope of judicial notice (and legislative facts), the approach adopted by the majority of relying on reported cases and secondary literatures in the law reviews underscores the difficulties that many judges experience in being adequately informed by evidence called at the voir dire about the reliability of various forms of evidence.

The viability of the Trochym approach to the admissibility of evidence in future cases will depend in no small part on the ability of counsel and experts to assist judges. This will require adequate funding of the defence. If this does not occur, judges will be obliged to inform themselves about the frailties of particular proffers of forensic science and, as

168 Trochym, supra note 7 at para 27.
169 Ibid at paras 29-30.
170 Ibid at para 132.
171 Ibid at para 142.
173 The accused’s lawyer in Trochym was James Lockyer, founding director of the Association in Defence of the Wrongfully Convicted. His factum on appeal featured extensive references to secondary literatures.
the minority suggest, this process may push the limits of judicial notice and independence. Alternatively, it may be more convenient for judges to subvert reliability issues by using heuristics such as experience, qualifications, prior admission, and crude attempts to temper reliability through demarcating scientific from non-scientific opinions. It is also possible that judges could avoid the exclusionary implications of *Trochym* and instead rely on warnings to juries and the ability of the accused to obtain evidence about the unreliability of the state’s expert evidence. *Trochym* opens the possibility of a more demanding contextual approach to the admissibility of expert evidence that should result in the exclusion of expert evidence that is unreliable or of unknown reliability. The question that remains, however, is whether lawyers and judges will be able to live up to the promise of *Trochym*.

Concern about wrongful convictions in *Trochym* also raises the question of whether the Court would be prepared to re-visit the existence of a Charter right not to have unreliable expert evidence and expert evidence of unknown reliability admitted in a criminal proceeding. In *R v Buric*, the Ontario Court of Appeal held that there was no right under section 7 of the Charter to the exclusion of unreliable evidence despite recognizing that the conviction of the innocent would offend the principles of fundamental justice. Much water has passed under the bridge since 1997, and the Court’s recognition of the risks of wrongful conviction has likely strengthened the case for such a right. While the Court has resisted attempts to have other forms of suspect evidence, such as jailhouse-informer testimony, excluded, it is important to emphasize that the reliability of many forms of expert opinion is readily susceptible to determination. A right not to be confronted with unreliable or speculative expert-opinion evidence would seem to embody the state’s obligation to prove guilt beyond reasonable doubt, demonstrate a commitment to fairness, and lend constitutional legitimacy to the idea of asymmetry in the laws of evidence. For only the accused could invoke such a right.

175 *R v Brooks*, [2000] 1 SCR 237; *R v Duguay* 2007 NBCA 65. For arguments that judges should be more willing to exclude this suspect form of evidence and less willing to defer to the ability of the jury to follow warnings and determine the credibility of the evidence, see Roach, ‘Unreliable Evidence and Wrongful Convictions: The Case for Excluding Tainted Identification, and Jailhouse and Coerced Confessions’ (2007) 52 Criminal Law Quarterly 210. For a recent decision that contemplates the continued admission of jail-house informer testimony while also demonstrating the need for warnings and full rebuttal of such dubious evidence, see *R v Hurley* 2010 SCC 18.
The Goudge Inquiry was established in 2007 in response to concerns about the reliability of expert evidence offered by forensic pathologist Dr Charles Smith in a number of paediatric death cases. The bulk of its 169 recommendations relate to various means, including better training and quality assurance, of improving the practice of forensic pathology. The Ontario government responded to these recommendations with new legislation increasing the oversight of death investigations and requiring the province’s Chief Forensic Pathologist to maintain a register of qualified forensic pathologists and create best practices guidelines. The vast majority of the Goudge Inquiry’s recommendations concern the need for ‘front-end’ reforms to improve the quality of forensic pathology. The report calls for more research and greater engagement with the universities. In this respect, recommendations by the NAS are similar, including a common focus on the scarcity of properly trained forensic pathologists and the dangers of allowing medical doctors and pathologists without forensic training to give evidence in criminal trials.

The Goudge report differs from the NAS report, however, in concluding that a stricter approach by trial judges to determining the scope and limits of an expert’s area of expertise and the ‘threshold reliability’ of the expert’s evidence could help to improve the reliability of forensic science and medicine and the value of incriminating expert evidence. Relying on both Trochym and Daubert, the Goudge Inquiry recommended that judges should assess ‘the reliability of the proposed witness, the field of science, and the opinion offered in the particular case’ regardless of whether the science was novel. It also explained that ‘the gate-keeper function need not face the trial judge with only a binary choice – to admit scientific evidence fully or exclude it completely’ and contemplated that the trial judge could restrict or ‘edit’ part of the expert’s testimony because of concerns both about the lack of threshold reliability and about the need to weigh the prejudicial effect of the evidence against its probative value. Restriction and editing would require attention to the language that the expert proposed in order to avoid confusion or misrepresentation. In some cases, experts would be required to situate their opinion within the range of a controversy within a field. Implicit in such recommendations is the notion that judges should be well informed

176 Coroner’s Amendment Act, SO 2009, c15.
177 The two reports are compared in Roach, ‘Forensic Science,’ supra note 112.
178 Goudge Inquiry, supra note 5 at 496.
179 Ibid at 499–500, recommending that ‘judges should consider whether there are parts of the proposed expert evidence that are sufficiently reliable to be admitted and others that are not or which must be modified to be admitted.’
about the particular area of expertise. To this end, the Goudge
Commission recommended increased judicial education, increased joint
training in forensic pathology for prosecutors and defence lawyers, and
additional legal-aid funding for defence lawyers in cases involving forensic
pathology.\textsuperscript{180} Although the Goudge Commission has already helped to
reform the practice of forensic pathology, the verdict on whether it, in
combination with the Supreme Court’s decision in \textit{Trochym}, will improve
judicial gate-keeping is unclear. The record so far, as two recent decisions
of the Ontario Court of Appeal help to illustrate, seems to be mixed.

B \textit{RE TRUSCOTT} (2007)
The first decision, \textit{Re Truscott},\textsuperscript{181} resulted from a ministerial reference in
relation to the notorious case of Steven Truscott. As a fourteen-year-old,
Truscott was convicted for the 1959 murder of Lynne Harper. Forensic
science and medicine were instrumental in placing Truscott in the vicinity
at the time Harper was alleged to have been killed. Initially condemned
to death, Truscott’s sentence was subsequently commuted to life.\textsuperscript{182} In
1960, an appeal was unanimously dismissed by the Ontario Court of
Appeal; and in 1966, the conviction was upheld on a reference by the
Minister to the Supreme Court of Canada, over one dissent. A critical
issue in the case was the time of Harper’s death, it being accepted by
all that Truscott had only a narrow window of opportunity to commit
the crime. After an inquiry by the Honourable Fred Kaufman QC reporting
in 2004, the Minister of Justice referred the case back to the Court of
Appeal. In 2007, a panel of five judges heard the appeal, quashed the
conviction, and formally acquitted Truscott.

During the 2007 appeal, new gastroenterological and entomological
evidence were used to challenge the original time-of-death estimates. In
reviewing the forensic evidence about digestion, the degree of \textit{rigor mortis}, the extent of decomposition, and the rate of insect depredation,
a unanimous Court of Appeal accepted the expert evidence produced
by the defence. In so doing, they dismissed the forensic pathology evi-
dence originally relied upon by the Crown as well as the expert evidence
it had assembled for the latest appeal.

Significantly, the Court placed little confidence in the forensic pathol-
ogy adduced by the Crown. Unlike the defence experts, Dr Spitz – who
had ‘been a forensic pathologist for more than fifty years’ – was ‘unable
to cite any recent scientific literature that would support [his] view.’\textsuperscript{183} In

\begin{itemize}
  \item \textsuperscript{180} Ibid at 502, 468, 460–3.
  \item \textsuperscript{181} Supra note 7.
  \item \textsuperscript{182} Truscott was released on parole in 1974.
  \item \textsuperscript{183} \textit{Truscott}, supra note 7 at para 166.
\end{itemize}
a similar way, the Court dispensed with the testimony of the Crown’s ento-
\[\text{mological expert, Dr Haskell.} \]

Several critical elements of his opinion were based on nothing more than his pur-
\[\text{ported experience, which could not be verified and was not supported by any} \]
\[\text{empirical work. He was unable to demonstrate that his experience had been} \]
\[\text{replicated by other scientists . . . He provided no scientific evidence to support} \]
\[\text{this theory.}^{184} \]

Rather than ‘authoritative experience and anecdotal case reports,’ this Court implicitly endorsed an ‘evidence-based approach’ to expert evi-
\[\text{dence.}^{185} \] The Court was interested in reliability and seemed to expect the Crown to support the opinion evidence produced by its experts.\textsuperscript{186}

As the previous extracts indicate, anecdote, experience, and bare opinion were all characterized as insufficiently reliable. The Court was particularly attuned to evidence of experiments and testing and to whether an expert’s opinions had support in authoritative literatures. Attentive to whether the expert evidence was grounded in studies and publication, the Court found the Crown’s evidence about the time of death ‘scientifically untenable.’\textsuperscript{187}

Unlike the vast majority of criminal convictions, the \textit{Truscott} case has received sustained, and frequently sympathetic, attention across five decades. Truscott’s defence also benefited from developments in forensic science and medicine. According to the Court of Appeal, ‘probably no other case in Canadian history has engaged the same level of judicial analy-
\[\text{sis and sustained public interest over so many decades.}^{188} \] Nevertheless, the criminal justice system seems to have had considerable difficulty correcting one of Canada’s most prominent and enduring miscarriages of justice in circumstances where it was not dealing with entirely novel scientific and medical evidence, and some of the Crown’s experts expressed what might be considered a disconcerting ambivalence toward published empirical research. The challenge of overturning convictions, even those based on unreliable expert evidence, should not be underesti-
\[\text{mated.}^{189} \] At the same time, the Court of Appeal’s approach in \textit{Truscott} does demonstrate the potential of a more contextual and attentive approach to the admissibility and weight of forensic evidence. Assisted

\textsuperscript{184} Ibid at paras 313, 349.
\textsuperscript{186} \textit{Truscott}, supra note 7 at paras 233, 307.
\textsuperscript{187} Ibid at para 365.
\textsuperscript{188} Ibid at para 71.
\textsuperscript{189} Scheck, Neufeld, & Dwyer, supra note 61.
by able counsel, the judges engaged with the details of both forensic pathology and entomology to a degree that made them aware of limitations and controversies in the particular fields. That said, the *Truscott* decision involved an extraordinary reference and did not require the judges to exclude expert evidence on reliability grounds. With some exceptions, Canadian courts, even in the wake of *Trochym*, the Goudge Commission, and wrongful convictions caused in part by faulty forensic expert evidence, are still reluctant to exclude expert evidence offered by the Crown.190

C R V ABBEY (2009)
The Ontario Court of Appeal grappled with Supreme Court authority and the recommendations of the Goudge Inquiry in the recent case of *R v Abbey*.191 The appeal considered whether a sociologist – rather than a forensic scientist – specializing in street gangs and their cultures could testify about the meaning of a tear-drop tattoo obtained by the accused just a few months after he was alleged to have committed a gang-related murder. The trial judge had excluded Dr Totten’s evidence because of concerns about his opinion’s being qualitative, along with his failure to determine an error rate, use random sampling and peer review, and validate his method. The Court of Appeal overturned this attempt to impose *Daubert*-style admissibility criteria on the expert’s opinion, concluding that, because the sociologist ‘did not pretend to employ the scientific method,’ it was a misguided ‘attempt to place the proverbial square peg into the round hole.’192 Instead, the Court of Appeal proposed a more flexible approach to the reception of non-scientific expert evidence.193

The Court of Appeal contemplated that, in all cases, judges should assess the threshold reliability of proposed evidence, including the balance between probative value and prejudicial effect, after determining that expert evidence was *prima facie* admissible according to the four-part *Mohan* test. Drawing from the Goudge Inquiry, the court stressed that ‘admissibility is not an all or nothing proposition’ and requires the trial judge to pay careful attention to the limits of the witness’s expertise

190 But see *Dimitrov*, supra note 146 at para 56, setting aside a murder conviction and excluding a barefoot impression as insufficiently reliable; *R v Wood* (2007), 79 Alta LR (4th) 358 (CA), excluding drug-recognition evidence offered by Crown in an impaired driving case; *R v Klymchuk* 2008 OAC 210, excluding crime reconstruction evidence offered by Crown; *R v Hughes* 2009 ONCA 268, excluding a sniffer dog’s detection of fire accelerator offered by Crown.

191 Supra note 10.

192 Ibid at paras 108–9. We agree that the trial judge did not appear to understand some of the methodological issues and their relevance to social-science research.

193 Ibid at para 109.
and the content and language in which the opinion was expressed. 194 Justice Doherty explained that Dr Totten could not testify that the tear-drop tattoo meant that the accused had murdered someone around the relevant time. He could, however, explain to the jury that tear-drop tattoos had a range of meanings within gang culture including the murder of a rival gang member but also time spent in prison or the loss of someone close. 195 This willingness to constrain the scope of the expert opinion was presented as a case-specific ‘cost-benefit’ response designed to ensure threshold reliability, control for any misleading or prejudicial effects, and prevent the expert departing from his appropriate area of expertise. 196 Such an approach has the potential to limit extravagant and unreliable claims made by experts. 197 At the same time, it can be a ‘second-best’ strategy that attempts to mitigate the frailties of expert evidence that ought to be excluded on grounds of unreliability or unknown probative value.

Although we agree with the general framework set out in Abbey and its focus on reliability – as well as the cost and benefits of admitting expert evidence – the actual decision is controvertible both in terms of the response to threshold reliability and in finding the incriminating expert evidence necessary. While there is little doubt that Dr Totten possessed appropriate formal qualifications and extensive experience studying gang culture, had interviewed hundreds of gang members (primarily from Ottawa), could speak about the general significance of tear-drop tattoos across North America, made many appropriate concessions, and would be restricted to commenting on potential meanings, he was not familiar with Abbey’s gang. In terms of qualitative methods, this oversight introduced real risks, especially those of extrapolating from general knowledge about gangs to the behaviour of particular gangs and individuals.

We accept that the sociologist’s opinion was not a guess and that, to some extent, it was even empirically predicated. Nevertheless, there was a real danger in admitting evidence about the meaning of tattoos among other groups, even if it was left to the jury to draw the appropriate inferences. The main problem was that the meaning of similar tattoos among other gangs was of questionable relevance to the meaning within the particular gang and to the specific individual. This problem

194 Ibid at para 63.
195 Ibid at paras 67, 27, 103.
196 Ibid at para 79.
197 William Thompson has recently called for a ‘finer-grained analysis of admissibility’ which does not focus on the reliability of whole disciplines but rather on whether ‘the specific conclusions offered have an adequate scientific basis’; see William Thompson, ‘The NRC’s Plan to Strengthen Forensic Science: Does the Path Ahead Run through the Courts’ (2009) 50 Jurimetrics Journal 35 at 48.
and its implications are hardly mentioned by the Court of Appeal.\textsuperscript{198} Significantly, the opinion could have been made demonstrably reliable through, for example, interaction with members of the accused’s gang.

Consequently, Dr Totten’s opinion was not demonstrably reliable and relying upon his experience introduced unfairly prejudicial evidence about the behaviour of other gangs and speculation about its relevance to the particular gang and the accused’s behaviour. In reviewing the decision to recommend admitting the opinion about the tattoo and leave its precise meaning to the jury, it is useful to reflect upon how difficult it would be to challenge a mistaken opinion and to engage in effective cross-examination. It will be recalled that a concern about an inability meaningfully to test reliability (e.g. through cross-examination) was critical to the majority’s opinion in Trochym.\textsuperscript{199} In Abbey, there was no real possibility of a defence expert rebutting extrapolation from the general to the particular. The only option was formal methodological criticism from defence counsel and/or another experienced qualitative researcher (in circumstances where methodological limitations did not dissuade the Court of Appeal). There was a serious risk in effectively challenging the expert evidence (even if it was wrong or misleading) where the ‘two’ alternative meanings had been excluded by the Crown (the accused had not been in prison and no close friend or family member had died around the relevant time) and where the opinion might appear to independently corroborate the evidence of other gang members.\textsuperscript{200} The probative value of the evidence was simply unknown (though, had it been reliable, it might have been compelling), but the danger to the accused if the opinion was mistaken or the extrapolation inapposite was extreme. Overall, the possibility of cross-examination, the apparent simplicity of the evidence (e.g. purportedly only three possible meanings and a proximate temporal relationship), and the trial judge’s ability to restrict the way the opinion is expressed (just short of the implicit inference), do not overcome reliability problems, the very real risk of unfair prejudice to the accused, and the difficulty of effectively contesting the evidence.

The reliability problems and the difficulty of conveying subtle methodological dangers (and logical flaws) during an adversarial murder trial were compounded by the question of necessity. Given the witnesses available to the Crown, Dr Totten’s opinion hardly seems necessary. For example, members of the actual gang were available to comment on

\textsuperscript{198} Abbey, supra note 10 at para 139, suggesting that research on American gangs might also be applicable to Canadian gangs.

\textsuperscript{199} Supra note 7 at para 60.

\textsuperscript{200} Compare the approach to anthropological evidence about Aboriginal beliefs in the Australian Native title case of Milirrpum v Nabalco Pty Ltd (1971) 17 FLR 141.
their understanding of a tear-drop tattoo within the gang – and even in relation to Abbey. This meant that the inductive opinion of Dr Totten (and a police officer with general experience of gang culture) was not necessary. If members of the actual gang were able to testify about the meaning or range of meanings they might ascribe to such a tattoo, what value was the opinion of a sociologist who had studied other gangs? The evidence of Dr Totten and the police officer should have been excluded according to the Mohan factors; particularly on the grounds that the expert evidence was not necessary. Dr. Totten testified at the retrial as to the possible meanings of the tattoo and Mr Abbey was convicted of first degree murder.

Abbey also reveals how experience-based and non-science exceptions to Daubert might exempt a great deal of incriminating expert evidence from having to demonstrate actual reliability. The Court of Appeal emphasized that the expert’s opinion came from ‘specialized knowledge gained through extensive research, years of clinical work and his familiarity with the relevant academic literature. It was unhelpful to assess Dr Totten’s evidence against factors that were entirely foreign to his methodology.’ This decision may provide incentives to forensic scientists (and others) to cheerfully accept the critique of their work as ‘non-science’ and to attempt to explain a lack of testing on the basis that their disciplines, like Dr Totten’s, cannot produce experiments and precise error rates. The Court of Appeal’s blunt conclusions that ‘the Daubert factors are not applicable’ because Dr Totten’s opinion ‘was not scientific . . . not novel . . . [and] was not a theory’ creates a danger of encouraging de facto exceptions for experience-based expertise. Such an approach could dissuade judges from questioning why expertise has not been tested in other ways and from exploring the potential for error, exaggeration, or alternative theories and explanations. Even in cases where

---

201 Such comments and opinions are not without their own risks, but there are few prima facie reasons not to admit them. Moreover, their admission is consistent with best-evidence principles. For who would be better to speak about gang culture than members of the gang? See e.g. Dale Nance, ‘The Best Evidence Principle’ (1988) 73 Iowa L Rev 227.

202 ‘Man Convicted 4 Years after Acquittal,’ The Toronto Star (29 March 2011) GT3. The necessity requirement for expert evidence in Mohan is grounded in legal values that might play out differently in cases where the Crown adds an expert to a strong case and in cases where the accused bases a defence on expert testimony. This might provide another legal basis for an asymmetrical approach to the admissibility of evidence in criminal trials as will be argued in Part V of this article.

203 Abbey, supra note 10 at para 108.

204 Ibid at paras 104, 116. This danger is perhaps increased by the fact that the Court of Appeal only made passing reference to Trochym, supra note 7, which, as discussed earlier, casts doubt on the distinction between novel science and other forms of questionable evidence. See also Law Commission, Expert Evidence, supra note 30 at 5.79.
validation and the determination of error rates are not possible, judges should, as the Goudge Commission stressed at length, search for alternative *indicia* of reliability.\footnote{Goudge Inquiry, supra note 5 at 487–96.}

**D OVERVIEW OF THE CANADIAN EXPERIENCE**

The Canadian experience is a testament to the dynamic nature of the doctrine surrounding the admissibility of expert evidence driven, in part, by the impact of experience with wrongful convictions. *Mohan* represents what might be characterized as a fairly traditional common-law approach to admissibility decision making. The Supreme Court seemed to be more concerned with an expert’s qualifications than with the intricacies of reliability. To the extent that reliability emerged, it was part of a balancing exercise involving probative value and prejudicial effect when novel science was adduced. The most significant change in Canada came when a majority of the Supreme Court in *Trochym* indicated that reliability would become more central to admissibility. This case is steeped in criminal-justice values, most notably concerns about wrongful convictions, and it also applies threshold-reliability concerns beyond the realm of novel science and even beyond the realm of scientific evidence and expert evidence. The Ontario Court of Appeal’s decision in *Truscott*, while not entirely representative, demonstrates that judges are capable of evaluating and dismissing expert evidence offered by the state on reliability grounds.\footnote{See also *Henderson*, supra note 116. These cases, particularly in the counsel and experts available to the panel of appellate judges, could hardly be considered representative or particularly indicative of what is possible in ordinary trial courts.} The Ontario Court of Appeal’s subsequent decision in *Abbey* sets the stage for discipline-specific examination of both threshold reliability and the prejudicial and misleading effects of expert evidence in individual cases. That said, *Abbey*-style applications could, in the absence of effective engagement by lawyers and judges, degenerate into a wide, experience-based exception for many types of incriminating expertise where judges are insufficiently conversant with social-science (or other) methods and their limitations. The majority’s decision in *Trochym* appropriately, given recent experiences, recognizes the danger that flawed expert evidence may contribute to wrongful convictions, but Canadian courts have not taken the additional step, also rooted in criminal-justice values and arguably the Charter of requiring all expert evidence offered by the prosecution to be demonstrably reliable.

Canadian judges, more than judges in other common-law jurisdictions, seem to have a sense of the limitations of trial safeguards and the need for continuing vigilance in criminal proceedings. Nevertheless, as things stand, a vague concept of reliability affects Canadian legal practice
somewhat erratically. Like their counterparts in other common-law jurisdictions, Canadian judges remain dependent upon the ability of counsel and the quality of evidence and argument and constrained by the extent of their own methodological and technical sophistication.

V Contextualism and an exclusionary ethos

A CONTEXTUAL (OR REFLEXIVE) JURISPRUDENCE AND PRACTICE

Across the anglophone world, expert evidence and expert disagreement are widely perceived as perennial problems.207 Experience in the United States and beyond demonstrates that judges and law reformers seem to encounter real difficulties with forensic science, medicine, and other forms of incriminating expert evidence. Empirical and experimental studies suggest that judges struggle to apply admissibility standards to evidence developed by state-employed forensic scientists (and consultants) with the same kind of exclusionary rigour that they routinely apply to the complex technical opinions proffered by plaintiffs’ experts in tort and product-liability litigation. Admissibility decisions, it would seem, are informed by the values and commitments of judges, particularly their impressions – however skewed – of the socio-legal order. Institutional and ideological pressures seem to converge to make judges overly sceptical of experts engaged by plaintiffs and criminal defendants and overly credulous when it comes to incriminating opinions espoused by state-employed forensic scientists and other experts.

In response, we advocate a more principled and empirically credible approach to admissibility jurisprudence and practice, which we describe as ‘contextualism.’208 Rather than focus upon features of idealized versions of science209 such as are associated with essentialized readings


of *Daubert*, it is our contention that admissibility decisions should be primarily shaped by legal principle and criminal-justice values. This approach aims to revitalize admissibility practice through recourse to legal fundamentals while taking account of a range of highly relevant contextual considerations. In consequence, a more contextual approach requires that admissibility practice should be reflexively indexed to legal-system goals and principles, to rules and procedures, to the type of proceeding and the parties, to the insights that wrongful convictions and empirical legal studies provide about limitations with trial and appellate processes, to revelations about the current condition of particular disciplines within forensic science and medicine, to a fundamental pre-occupation with evidence of reliability (i.e., demonstrable reliability), to empirical accounts of expertise and an awareness that there is no simple way to define science or to identify reliable knowledge (especially where techniques are not readily susceptible to testing), to a willingness to exclude or qualify incriminating evidence that is insufficiently reliable (however central to the case), and to an awareness that, historically, judges have been too accommodating of incriminating expert evidence.210

It might be argued that our approach, with its focus on wrongful convictions, places too much emphasis on what could be conceived as a temporary social problem.211 After all, the use of *Daubert* as an exclusionary tool in the realm of civil justice was inspired by a widespread belief in the existence of a ‘litigation explosion’ – linked to a rapid increase in tort litigation, to excessive damage awards, to a toleration for ‘junk science,’ and to rapid rises in the cost of public-indemnity insurance. Empirical evidence (eventually) challenged these pervasive impressions but has exerted a less conspicuous influence on subsequent civil-justice practice and reform. This experience should encourage us to reflect on how impressions may have influenced and might continue to influence civil-justice admissibility decision making as well as the possibility of reforming the reception of expert evidence in criminal proceedings.

Problems with forensic science and medicine are unlikely to be a passing fad. As we have endeavoured to explain, there are good


211 DNA exonerations will presumably decline as standardized DNA testing and probability estimates are used routinely in preliminary investigations. At the same time, the cases revealed by the Goudge Inquiry, supra note 5, underline how other miscarriages of justice may be discovered without reliance on DNA.
reasons – drawn from fundamental principles of the adversarial criminal trial, long-standing criticism of the content of many forensic sciences, empirical evidence from trials and wrongful convictions, and the express recommendation of peak scientific organizations – for a more aggressive approach to the admissibility of incriminating expert evidence. The convergence of opinion from a range of sources and confirmed by empirical study warrants a credible response. The problem of wrongful convictions, in particular, engages long-standing and fundamental legal principles. Jurists since John Fortescue (c 1394–c 1480) and William Blackstone (1723–1780) have continuously expressed a strong preference for allowing the guilty to go free rather than risk convicting the innocent.212 There would, in consequence, seem to be a premium placed on factual accuracy and skewing risks away from defendants. In the criminal sphere, the accused should derive any benefit from doubt while the state assumes (or should assume) most of the burdens and risks.

If we compare these long-cherished commitments with actual legal practice, empirical evidence (largely, though not entirely, from the United States) suggests that, contrary to principle, trial and appellate judges tend to screen expert evidence adduced by plaintiffs in tort and product-liability suits more aggressively than they do expert evidence adduced by the state in criminal proceedings. In practice, the most exclu- sionary orientation (or ethos) seems to be reserved for plaintiffs suing large corporations, whereas the state’s incriminating forensic sciences seem to benefit from institutional trust and an attendant inclusionary ethos. Revealingly, the trend is basically the same across the United States in both Daubert and non-Daubert jurisdictions.213 And notwithstanding variation in their rules and jurisprudence, similar patterns are discernible or apparently desirable in England, Australia, and New Zealand.214

Common law judges appear curiously receptive to the state’s forensic-science evidence, particularly where other courts (whether in the United

213 Cheng & Yoon, supra note 47.
214 See, for example, Robb, supra note 116; Atkins, supra note 116; Weller, supra note 116; and R v Luttrel, [2004] EWCA Crim 1344; R v Tang (2006) 65 NSWLR 681; R v Li (2003), 140 A Crim R 288; R v Madigan, [2005] NSWCCA 170; and Evidence Act 2006 (NZ), 2006/69 s 25 (2006). And there is little reason to think things are much better in inquisitorial systems, where the court is involved in the appointment of the expert, or that inquisitorial-style reforms will substantially address the underlying reliability issues. See e.g. Laura Hooper, Joe Cecil, & Thomas Willging, ‘Assessing Causation in Breast Implant Litigation: The Role of Science Panels’ (2001) 64 Law & Contemp Probs 139; Gary Edmond, ‘Merton and the Hot Tub: Scientific Conventions and Expert Evidence in Australian Civil Procedure’ (2009) 72 Law & Contemp Probs 159; Edmond, ‘Judicial Representations,’ supra note 11.
States or elsewhere) have previously admitted incriminating expert opinions derived using similar techniques. These responses seem inappropriate. In principle, admissibility standards should be highest where the state is relying upon expert evidence to prosecute and less onerous where private citizens and corporations engage experts in civil disputes that are normally considered private. Many Anglo-American judges seem to have departed from, indeed inverted, a more principled approach to admissibility and the respective goals of civil and criminal justice.

B ACTUAL RELIABILITY RATHER THAN ‘SCIENCE,’ ‘NON-SCIENCE,’ OR DAUBERT

In many ways, the questions of whether Daubert embodies the essentials of genuine science and whether we can develop useful means of demarcating science from other types of knowledge and experience are distractions. Rather than getting bogged down in sterile and intractable attempts at demarcation or efforts to characterize particular attributes as essential ingredients of the modern sciences, we suggest that the Daubert criteria do not embody the essence of all authentic scientific activity and, importantly, that the more fundamental issue for criminal investigations and trials is the actual reliability of incriminating expert opinions. Regardless of whether some technique, method, or ability is characterized as science-based or experience-based, given the high stakes of the criminal trial (and its obvious deficiencies), the admissibility of incriminating expert opinion should be dependant on evidence of its


216 Some of these issues are raised by Margaret Berger, ‘Upsetting the Balance between Adverse Interests: The Impact of the Supreme Court’s Trilogy on Expert Testimony in Toxic Tort Litigation’ (2001) 64 Law & Contemp Probs 289; Berger, ‘Eliminating,’ supra note 14; Risinger, ‘Preliminary,’ supra note 14.

reliability that is empirical and demonstrative. The Supreme Court of Canada’s decision in *Trochym* has great potential in this regard. It rejects the idea that threshold-reliability concerns should be limited to novel science or left to weight.

In the context of the adversarial trial and in response to incriminating expert evidence, the *Daubert* criteria, especially testing, can provide, in many and perhaps most criminal cases, serviceable resources for approaching the question of whether incriminating expert evidence is sufficiently reliable to admit. Unlike the mandatory *Mohan* factors – concerned with the *admissibility* of novel expert evidence – the *Daubert* criteria are specifically focused on *indicia of reliability*. Significantly, the *Daubert* criteria are not equally important or discriminating in relation to reliability. Testing, broadly conceived, is most important. Whether a technique or theory has survived some kind of empirical assessment – such as a validation study – is particularly important in the context of forensic science and medical evidence. Experimental evaluation of forensic science and medical techniques provides very useful information for reliability assessments, ascertaining levels of accuracy and determining practitioner competence. Testing can also include research on the possible effects of bias or confirmation bias. In the absence of testing, conservative estimates, claims about rates of error, and expressions of confidence may all be worse than useless. Research on wrongful convictions attests to how forensic scientists, forensic pathologists, and prosecutors routinely exaggerate the accuracy of techniques, theories, and conclusions – particularly in the absence of testing.

The remaining *Daubert* criteria are usually less important than the results of testing. In theory, general acceptance provides a useful proxy but is no substitute for rigorous validation and accuracy studies. Evidence about the extent of acceptance is often extrapolation from a few books or articles or impressionistic hearsay. In practice, research is rarely conducted into the distribution of expert opinion. Even if information on the extent of acceptance were readily available, it would probably reveal more about the field and orientations within the field than about reliability *per se*. ‘Acceptance’ is, perhaps, especially weak in areas such as forensic science and medicine, where communities tend to be small and, historically, insulated from more mainstream biomedical and scientific research and the practitioners maintain close contacts with police, investigative agencies, and prosecutors.

---

Similarly, peer review and publication are not especially good surrogates for ascertaining reliability. Peer review and publication have a range of meanings and uses, but most scholars recognize that they rarely provide definitive support for techniques or vouch for the reliability of results. That a theory or technique is mentioned in the literature will generally be of less significance than its having survived rigorous independent testing (with a low level of error). Where peer review describes the appraisal of a particular result by a colleague employed in the same institution and not blinded to the investigation or result, any positive review is likely to have very limited value.

Overall, scientific, medical, and other forms of expertise are too variegated and complex to be subjected to simple categorization or reliability algorithms. Nevertheless, judges should be reluctant to use supplementary indicia to overcome the failure to test or check. They should inquire about the failure to test and not simply excuse such failures because the incriminating expert evidence is perceived as important, or even necessary, to the viability of the prosecution case. Where techniques are not readily susceptible to testing, judges should identify evidence relevant to methods and conclusions that justifies admission. They should, as in Abbey, consider standard methods used by qualitative researchers but should also consider whether additional research or information is required. In some circumstances, placing restrictions on what the expert can say and on his or her level of confidence may suffice. Such restrictions, though, should not be used as an admissibility compromise. Incriminating expert opinions, however circumscribed, should always be demonstrably reliable.


220 Similarly, the fact that the defence might be able – or even intends – to call rebuttal experts should not provide a basis for admission or an easing of the standards for admission for the state’s incriminating expert evidence.

221 If not empirically predicated, expressions of opinion are very likely to mislead and are always potentially complex; see McQuiston-Surrett & Saks, supra note 83; ‘Atkins v The Emperor,’ supra note 116.
Similarly, confidence in the state’s forensic-science laboratories and police service are not substitutes for evidence of reliability. Judges should be concerned with evidence supporting the reliability of particular techniques, theories, and opinions, not evidence of the eminence of scientists, their performance and credibility, their impressive credentials, past successes, or the reputation of their institutions. Credentials, training, authority, and experience are all important pre-requisites but are unreliable predictors of reliability. There are many well-documented cases of eminent experts straying well beyond their area of expertise.222

When determining the admissibility of expert-opinion evidence, it is important to ascertain the reliability of the methods and technique before any resulting opinions are buttressed with supplementary evidence. The fact that there is other admissible and even compelling evidence pointing to the guilt of the accused does not necessarily add to the reliability of incriminating expert evidence. In some cases, such as Abbey, it may be unnecessary (and even misleading) for the court to hear speculative expert evidence. Assessment of the reliability of expert evidence for admissibility purposes should be independent of (or indifferent to) any other incriminating evidence. Otherwise, there is a real risk of compounding errors.223 The overall strength of a case tells us nothing about the validity or reliability of the techniques relied upon by forensic scientists and other experts.224 And, as DNA exonerations attest, successful prosecution does not test or validate a technique and forensic-science errors and exaggerations can compound other mistakes, such as faulty eyewitness identifications.225

While we are deeply concerned about the reliability of much existing forensic-science evidence, we are not sympathetic to the way forensic-science critics, the NAS, and the Law Commission have characterized ‘science.’226 Our contention, which in relation to most forensic-science evidence is functionally equivalent with the position recommended by critics and the NAS, is that the most demanding admissibility standards should be applied to forensic-science evidence adduced by the state

222 Prominent examples include Dr Charles Smith, Sir Roy Meadow, & Sir Bernard Spilsbury; see the Goudge Inquiry, supra note 5; General Medical Council v Professor Sir Roy Meadow [2006] EWCA Civ 1390; William Eckert, ‘Sir Bernard Spilsbury’ (1981) 2 American Journal of Forensic Medicine 179.

223 Indeed, judges should be particularly attentive to whether experts, particularly forensic scientists and police, were ‘blinded’ to extraneous information when collecting evidence and forming their opinions; see Dror, Charlton, & Peron, supra note 96; Risinger et al, supra note 97.


225 Compare United States v Harvard, 117 F Supp (2d) 848 (SD Ind 2000).

226 Instead, our approach draws upon empirical studies by scientists, experts, laboratories, and research institutions; see the works cited at note 209 supra.
and that, in most cases, the *Daubert* criteria represent a useful default condition. This approach is not based on an empirically suspect appeal to what all authentic science and expertise is like (or should be like), or optimistic predictions, like those made by the NAS, about the ability of all of the forensic sciences to imitate the techniques and standards of reliability associated with DNA evidence. Rather, it is based on matching expectations derived from criminal-justice principles with evidence of reliability. It is the importance we ascribe to the presumption of innocence, to avoiding the conviction of the innocent, the premium placed on fairness and accuracy, the difficulty of challenging incriminating expert evidence in the context of the accusatorial criminal trial, along with the very real limitations of supposed trial safeguards (e.g. warnings to juries), that encourage us to recommend the imposition of fairly onerous standards in the determination of whether the state’s forensic scientists (and other experts) should be entitled to express incriminating opinions in criminal proceedings.

Our orientation, and the need to distinguish it from the model of science advocated by critics, the NAS, and the Law Commission may seem trivial or pedantic, but it has profound implications. Emphasis on idealized features of science can generate, as is evident in American practice and the Law Commission’s proposals, experience-based (and other) exceptions that could subsume most of the forensic sciences. Although the NAS is undoubtedly correct about the need for better research, training, and governance in the forensic sciences, its implicit model of reproducing standards of reliability associated with DNA analysis may prove to be unrealistic, especially for many of the comparative forensic sciences with a limited research base and few connections to university-based scholarship. By abandoning philosophically oriented models of science, our approach also facilitates the development of judicial techniques that engage with evidence and proof more broadly. Here, it is important to recall that cases such as *Trochym* did not involve forensic science but rather testimony from eyewitnesses that might have been influenced by hypnosis. Relating admissibility decisions to legal values could support different admissibility standards in civil and criminal proceedings and be responsive to concerns that *Daubert* standards in civil litigation have been, perhaps, an unjustified barrier to holding corporations accountable.

227 Commitment to universal models of authentic science and the contention that the *Daubert* criteria embody those models might be used to justify the imposition of onerous admissibility standards in civil litigation that work against civil-justice goals and the possibility of broader socio-legal regulation. Conversely, as exemplified in the Law Commission’s recommendations and American trial practice in the wake of *Daubert* (e.g. *Starzecpyzel*), onerous admissibility standards may encourage attempts at circumvention and derogation.
accountable for dangerous products, unsafe workplaces, and environmental degradation. Our approach also represents an attempt to encourage judges and lawyers to think about how institutional pressures, ideology, and the distribution of resources among parties may be driving their admissibility practices.

Recognition that there is no proper model or universal ingredient applicable to all the activities that might be described as science (let alone expertise) provides judges with opportunities and simultaneously obligations to develop adjectival law in more socially responsible and epistemologically legitimate ways. Judges can, for example, develop models of expertise that are tailored to and embody legal values and principles and take into account experience with wrongful convictions and a broad-based critique of forensic science and medicine. In this respect, the majority's judgment in *Trochym* is particularly encouraging in its open acknowledgement of the risk of wrongful convictions and its linking concerns about the unknown reliability of expert evidence to the ability of the accused to engage in meaningful cross-examination. These more contextual responses to incriminating expert evidence pose very real challenges for experts, lawyers, and judges but offer the advantage of being consistent with espoused values.

We are not arguing for the strict application of the *Daubert* criteria to all proffers of expert evidence nor derogation from those criteria according to ‘non-science’ and ‘experience-based’ classifications. Rather, we are advocating greater attention to evidence of reliability in every case of incriminating expert opinion. *Daubert*-style criteria will have an important contribution to make but are unlikely to be universal in their scope and application. The main problems, obviously, are how we encourage judges, prosecutors, and forensic scientists to take seriously (or more seriously) actual reliability and legal values that require proof of guilt beyond a reasonable doubt and how we prevent the erosion of reliability standards for (purportedly) non-scientific, experience-based, and established forms of expertise.

C MOTIVATING THE EXCLUSIONARY ETHOS

At the heart of this article lies the question of who should bear the risk of unreliable expert evidence in criminal prosecutions?228 Traditionally, the state has been conceived as the proper bearer of risk – in theory, if not always practice – and this seems to be consistent with contemporary principle. However, the admission of expert evidence that is unreliable or of unknown reliability effectively dilutes the state’s burden. In effect, the

---

accused is required to negate incriminating expert-opinion evidence or suffer its prejudicial evidentiary consequences in relation to the overall case. The accused also bears the significant risk that the trier of fact will not understand or follow judicial warnings about suspect forms of admissible evidence. Experience with wrongful convictions suggests that the accused often has difficulty obtaining the services of an expert and effectively challenging the state’s evidence (even where it is unreliable or fraudulent). The continued use of non-technical fact-finders and reviewers – which may make sense overall – has the result that the criminal justice system is poorly organized to redress deficiencies with expert evidence (especially at trial). 229

Unfortunately, there is no simple way to identify reliable knowledge or distinguish scientific from non-scientific activity. The contextual approach advocated in this article encourages judges to develop and apply standards tailored to particular expertise and particular controversies. To be sure, such an approach is more labour-intensive than an insistence on the need to satisfy the Daubert criteria across the board. It also requires recognition that there are many serious problems and a need for vigilance with respect to almost all expert evidence offered by the state. It does, however, have greater potential to achieve optimal reliability standards than a more simplistic two-tier strategy where demonstrable reliability is only required for a few advanced ‘scientific’ disciplines, and experience (along with the possibility of exploring limitations and concessions at trial) is sufficient for all other types of expertise, however controversial. Contextualism is not a basis for arbitrary flexibility or slackening rigour to admit ostensibly reliable incriminating expert opinion. Evidence of reliability should be fundamental to the admission of all incriminating expert evidence, however characterized. A contextual approach should foster an exclusionary ethos. Where judges are not confident about the reliability of incriminating expert evidence, it should not be admitted (regardless of what assessment a jury might make). In some cases, contextualism may require the judge to engage with the substance of the proposed opinion evidence to the extent of excluding those parts of an expert’s testimony that are not demonstrably reliable. 230

229 We have not embarked on an assessment of jury capabilities, although the subject warrants serious attention. It is our contention that, regardless of their capabilities and aptitudes, it is not incumbent upon juries to decide whether particular proffers of expertise are reliable. There is no constitutional or charter requirement that disputes over the reliability of expert evidence should be resolved by juries. Rather, we suggest that these are issues for admissibility and judicial gate-keeping. Juries, whatever their merits and limitations, should only be allowed to consider incriminating expert evidence that is demonstrably reliable.

We recommend asymmetrical approaches to the admissibility of expert-opinion evidence between civil and criminal trials and between the state and the accused.231 It is our contention that the different values at stake should result in a lower admissibility standard for expert evidence in most civil proceedings. There should, in consequence, be a difference between the standard imposed in civil proceedings (on both the plaintiff and defendant) and what is required of the state in criminal proceedings. This is the first part of the asymmetry. The second dimension is the imposition of asymmetry between the state and the accused.232 In criminal proceedings, where the state is obliged to prove guilt beyond a reasonable doubt, the prosecution should not be allowed to adduce and rely upon incriminating expert-opinion evidence unless that evidence is demonstrably reliable. Conversely, we propose a lower admissibility threshold for expert evidence adduced by the accused.

It is not our intention to promote some kind of admissibility free-for-all. Our proposed asymmetrical standard will not enable the accused to rely upon manifestly unreliable expert evidence. We are vitally concerned about the accuracy of the criminal-trial process along with fairness (and the need to avoid convicting the innocent). The accused should be entitled to adduce expert evidence that impugns the state’s case or engenders doubt about the value of incriminating expert evidence.233 There is no obvious need, to the extent that this evidence is responsive to the state’s evidence, to hold it to the same admissibility standard as that imposed on the accusing proponent.

What would an asymmetrical standard of demonstrable reliability mean in practice? While it would not require the accused to satisfy the

231 This discussion of admissibility standards should not be confused with methodological symmetry, a theoretical postulate employed by sociologists of scientific knowledge; see e.g. David Bloor, *Knowledge and Social Imagery* (Chicago: University of Chicago Press, 1991).

232 Those who are critical of such an approach should recall that empirical research suggests that an asymmetrical approach is already in place, albeit privileging the state and discriminating against criminal defendants. Existing practice would seem very difficult to reconcile with criminal-justice values and evidence-law principles.

233 Courts should be more responsive to defence attempts to elicit the concerns of highly trained individuals, particularly those of statisticians and psychologists, proffering critical opinions about incriminating evidence and even the limitations of investigative and legal processes. Where those individuals are active scholars, like many of the critics, publishing their concerns in mainstream journals, courts should be reluctant to exclude or disregard their insights. Indeed, these were the kinds of perspectives that helped to bring the breadth of the problems with modern forensic sciences to the attention of the NAS. Courts should, in consequence, allow a scholarly critic like Professor Simon Cole to testify against incriminating individualization evidence from latent-fingerprint examiners; see the discussion in Part II.C of this article.
same admissibility standard as the state, it would still require some weaker indication of reliability (beyond relevance). Moreover, where a trial judge admits expert evidence *unilaterally* adduced by the accused, courts could confer upon the state the ability to respond *in kind* and/or with more established expertise that is critical of the particular evidence and/or techniques (*i.e.*, rebuttal evidence). Such responsive provisions might prevent the kind of concerns about allowing the accused to ‘distort the truth-seeking function of the trial’ by adducing ‘unreliable but impressive-looking expert evidence, particularly pseudo-science’ apparently worrying the authors of the Law Commission’s Consultation Paper. If the state were able to respond in kind, without having to demonstrate reliability (in such cases), there would seem to be few reasons to approach expert evidence adduced by the accused in a particularly exclusionary manner. Where the probative value of the accused’s expert evidence is low or uncertain, it could be regulated through relevance, a weaker reliability standard (e.g. soft _Daubert_) and trial efficiency concerns.

For pragmatic reasons, we could contemplate tempering the strength of our asymmetrical commitments. While we believe that criminal

---

234 Such ‘mirror image’ mechanisms are used in a number of jurisdictions; see e.g. the treatment of character evidence in the _Evidence Act 1995_ (Cth) s 110 (Australia) and the discussion in Stephen Odgers, _The Uniform Evidence Acts_, 9th ed (Sydney, Australia: Thomson Reuters, 2010) at paras 1.3.9020.

235 Returning to the Canadian cases (discussed in Part IV of this article), on reflection, we would be reluctant to allow the accused to rely on penile plethysmography, speculative syndrome evidence, or even hypnosis. In some cases, the accused might be able to persuade a judge that the use of evidence derived by such techniques (e.g. hypnosis) is necessary and is required by the right of full answer and defence. The asymmetry here would, consistent with our approach, have to be justified by an appeal to specific legal principles and rights that are themselves ‘asymmetrical’ in the sense that they can only be claimed by the accused. Were this evidence to be admitted, the state would be entitled to respond without having to satisfy demonstrable reliability for expert evidence that challenged the hypnosis evidence. The exception is not intended to be more expansive or encourage weak responses but rather to facilitate the accused’s full answer and defence and effective challenge by the state.

236 After all, this is how incriminating expert evidence is currently ‘managed.’

237 Even Laudan, a prominent critic of exclusionary rules (on the grounds of ‘double counting’ and their tendency to produce too many false acquittals), seems ambivalent about a reliability threshold for expert evidence, in addition to the overarching burden of proof imposed upon the state in criminal prosecutions; see Laudan, _Truth_, supra note 228. Interestingly, Laudan remains agnostic but also uninformed about many of the very real limitations with trial procedures and safeguards. Laudan’s text constitutes a recent contribution to a long tradition championing the inclusion of all relevant evidence, which has its modern origins in the works of Bentham and Thayer: Jeremy Bentham, _Rationale of Judicial Evidence_, ed by John Stuart Mill (London: Hunt and Clark, 1827) 477–645; James Bradley
justice systems should entrench different admissibility standards for expert evidence adduced by the state from those for expert evidence adduced by those accused of crime. The most important single reform would be to raise admissibility standards across the board. This would prevent the state adducing and relying upon and exposing the fact finder to expert evidence that is unreliable or of unknown probative value. Such an approach would improve legal practice, even if it prevented the accused from adducing some weaker forms of expert opinion as part of a defence. In many cases, it would prevent the accused from having to challenge expert evidence and demonstrate unreliability in the context of an adversarial trial. It would also prevent other incriminating evidence from artificially reinforcing the (apparent) probative value of incriminating expert evidence.

In the end, both asymmetrical and symmetrical approaches – to the extent that they impose upon the state the need for demonstrably reliable expert evidence – would help to encourage the institutionalized forensic sciences and medicine (and many individual experts) to develop standards, undertake research, and improve their performance. Along with supervising the expression of expert opinions, admissibility standards would appear to represent the limit of judicial abilities to meaningfully influence, let alone regulate and supervise, forensic science and medicine.

In closing, we are left to wonder how any system of criminal justice can be considered fair or rational when the accuser is able to adduce incriminating expert opinions without having to provide persuasive evidence that the experts can actually do what they claim (or how often they are right); where there is no expectation that evidence be obtained in circumstances that eliminate well-documented risks of bias; and where there is no need for judges to take account of empirical evidence, proliferating criticisms of trial safeguards, and past failures. The inability to credibly regulate incriminating expert evidence (in conjunction with the fact that much forensic-science evidence seems to be questionable) means that

Thayer, A Preliminary Treatise on Evidence at the Common Law (Boston: Little Brown, 1898) 264–6. This article, along with recent and more theoretically oriented work by Ho, supra note 8, and Stein, supra note 228, reinforces the importance of regulating (and even excluding) some types of relevant evidence.

238 This might appear as a reasonable trade-off against the greater dangers raised by states’ continuing reliance upon questionable expert evidence, but there is a need for caution in overly exclusionary approaches to expert evidence that raises doubt or genuinely challenges incriminating evidence – whether expert or not.

admissibility decisions are far more important than most judges and commentators seem to appreciate. Greater contextual awareness derived from empirical studies and converging critiques should encourage judges to develop more principled responses to the current crisis.

Drawing on fundamental principles and the emerging evidence, our lawyers and judges must assume greater responsibility for the supervision and, if necessary, exclusion of unreliable expert evidence.240 Regardless of what happens beyond the courts, prosecutors and judges must prevent unreliable expert evidence and expert evidence of unknown reliability from contaminating criminal trials and unfairly prejudicing those accused of crime.
