

Pathological Science?

Demonstrable Reliability and Expert Forensic Pathology Evidence

Dr. Gary Edmond[†]

1. Introduction: The Importance of Being *Reliable*

In recent years, a number of public inquiries have highlighted the importance of safeguarding the criminal justice system—and protecting the accused who are tried under it—from the possibility of wrongful conviction.... “[t]he names of Marshall, Milgaard, Morin, Sophonow and Parsons signal prudence and caution in a murder case” ... In the case at bar, we consider once again the need to carefully scrutinise evidence presented against an accused for reliability and prejudicial effect, and to ensure the basic fairness of the criminal process.¹

This paper is focused on jurisprudence and admissibility standards pertaining to expert evidence. Informed by recent theoretical and empirical approaches from the history, sociology, and anthropology of science and medicine, it suggests that judges should impose an explicit reliability standard on expert evidence adduced by the state.

The basic contention is that courts should not admit expert evidence adduced by the prosecution unless there are good grounds for believing that the evidence is reliable. Expressed more precisely, judges should not admit expert evidence adduced by the prosecution unless that evidence is *demonstrably reliable*. This would require the state to demonstrate, on the balance of probabilities, that the techniques and

[†] B.A. (Hons.) University of Wollongong, LL.B. (Hons.) University of Sydney, Ph.D. University of Cambridge. Associate Professor, Faculty of Law, The University of New South Wales, Sydney 2052, Australia. Opinions expressed are those of the author and do not necessarily represent those of the Commission or the Commissioner.

theories used by its experts and the opinions they present in court are reliable. In practice, the state would be expected to undertake some kind of empirical testing to ascertain whether the techniques and theories relied upon by forensic scientists, pathologists, and technicians are valid and accurate. In the absence of testing, judges might consider a range of supplementary, and usually weaker, indicia of reliability to determine whether the evidence is sufficiently reliable for use in a criminal trial.

Such an approach is broadly consistent with recent Canadian evidence jurisprudence, especially the pronouncements in *R. v. J.-L.J.* (2000) and most recently *R. v. Trochym* (2007) and *Re Truscott* (2007).² While the Supreme Court has not made reliability an explicit requirement for admissibility, in the aftermath of *R. v. Mohan* (1994), versions of reliability, along with the need to subject novel evidence to special scrutiny, have assumed increasing prominence in its expert evidence jurisprudence. The Supreme Court and the Ontario Court of Appeal have both demonstrated a willingness to embrace U.S. evidence jurisprudence to help them with the meaning of reliability, but the precise way this should occur and the content of any reliability standard await detailed elaboration. Currently, reliability sits awkwardly with formal concerns about relevance, necessity, and the prejudicial effect of expert evidence. This paper aims to clarify admissibility standards by championing an explicit and distinctive role for *demonstrable reliability*.

Several social, institutional, and procedural advantages arise from a more direct interest in the reliability of expert evidence. The imposition of a substantial admissibility standard will require judges to exclude unreliable expert evidence or expert evidence of unknown reliability. This helps the judiciary to regulate legal processes, but it simultaneously disciplines the agencies responsible for the investigation and prosecution of crime. The exclusion of unreliable expert evidence is likely to contribute to more legitimate outcomes. In terms of practice, the exclusion of unreliable expert evidence may increase the length of some preliminary proceedings but overall is likely to reduce the length of trials, avert the need for trial judges to give complex instructions about questionable evidence, and prevent juries from having to make, quite literally, uneducated guesses. Just as important, the exclusion of unreliable expert evidence obviates the need for defence lawyers to undertake long and technical cross-examinations along with the need to identify and secure the services of rebuttal experts. Instead,

¹ *R. v. Trochym* [2007] 1 SCR 239 at [1]. See also *United States v. Burns* [2001] 1 SCR 283 at [94]-[117].

² *R. v. J.-L.J.* [2000] 2 SCR 600; *R. v. Trochym* [2007] 1 SCR 239; *Re Truscott* [2007] ONCA 575.

exhaustive cross-examination and rebuttal expertise will only be necessary where the prosecution adduces demonstrably reliable expertise. Moreover, the emphasis on reliability means that the defence can question or challenge incriminating expert evidence on its own terms rather than being compelled to impugn the reputation or abilities of experienced experts called by the state. A reliability standard requires the state to demonstrate that its techniques and theories are reliable rather than expect individual defendants to challenge their value each and every time they are adduced. An explicit interest in the reliability of expert evidence also extricates concern about the probative value of evidence from the balancing exercise mandated by the judicial discretions. It should also separate assessment of the reliability of techniques, theories, and opinions from other incriminating evidence and prevent unreliable expert evidence going before the trier of fact in emotive cases, such as those involving the death of babies. Interest in reliability, particularly empirically derived error rates, can also help to ensure that the language used by expert witnesses appropriately reflects the validity and accuracy of the underlying scientific techniques. Finally, demonstrable reliability should reinforce the expectation that forensic pathologists and other experts will be accurate and independent.

An admissibility standard that incorporates demonstrable reliability is really another way of requiring evidence-based forensic science and medicine. To the extent that forensic science and medicine have been historically insulated from more mainstream scientific and biomedical research, imposing expectations that require evidence of testing or other indicators of reliability would seem to be important responses that will assist the courts as well as the investigative institutions and laboratories to improve the standard of expert evidence relied upon in criminal prosecutions, convictions, and appeals. Moreover, the state is in a position to take remedial steps to ensure that expert forensic evidence is subject to a variety of testing and validation procedures.

Overall, imposing a reliability standard on the prosecution (and the state) will help to make criminal trials fairer and outcomes reflect the known value of expert evidence. To the extent that there are losses from supplementing existing admissibility criteria with demonstrable reliability, they emerge from an inability to prosecute individuals using expert evidence that is either unreliable or of unknown reliability. In a rational system of criminal justice, the exclusion of such evidence cannot realistically be considered a loss.

In closing, this paper will briefly review several proposals for refining the preparation and presentation of expert evidence. All of these issues will be examined through the lens of recent theoretical, historical, and empirical studies of science, medicine, and expertise, and in ways that are sensitive to institutional and practical dynamics.

2. Problematizing Popular Approaches to the Sciences and Expertise

To begin, it is useful to challenge some pervasive beliefs about science, medicine, and expertise. This overview is intended to provide a necessarily brief yet relatively sophisticated approach to the messy realities of expertise in order to discourage recourse to idealized and romanticized models of science and medicine.³ Restricted to scientific method, scientific norms, publication and peer review, it provides a useful framework for approaching expert evidence jurisprudence.⁴

The vast majority of literature on law, science, and medicine appears oblivious to decades of research by historians, philosophers, and sociologists specializing in the study of science, medicine, and technology. This is unfortunate because this later body of work challenges many of the conventionally held views about expertise routinely, and somewhat glibly, employed in legal discourse, practice, and proposals for reform (see also Section 11).⁵

By way of example, virtually all contemporary historians, philosophers, and sociologists of science and medicine would reject, as implausible, the existence of an historically stable, prescriptive, and efficacious scientific method doctrine. Consider three representative responses to the most prominent of the philosophical accounts of scientific method—Karl Popper’s doctrine of falsification—which was cited by the

³ David Caudill and Lewis LaRue, *No Magic Wand: The Idealization of Science in the Law* (Lanham: Rowman & Littlefield, 2006); Gary Edmond, “Judicial Representations of Scientific Evidence” (2000) 63 *Modern Law Review* 216–251; Sheila Jasanoff, *Science at the Bar* (Cambridge, MA: Twentieth Century Fund, 1995).

⁴ For a general introduction consider Bruno Latour, *Science in Action* (Milton Keynes: Open University Press, 1987); Steven Yearley, *Making Sense of Science* (London: Sage, 2005); David Hess, *Science Studies: An Advanced Introduction* (New York: New York University Press, 1997); Alan Irwin and Brian Wynne (eds.), *Misunderstanding Science* (Cambridge: Cambridge University Press, 1996); Sheila Jasanoff et al. (eds.), *Handbook of Science and Technology Studies* (Thousand Oaks: Sage, 1995); Harry Collins and Trevor Pinch, *The Golem: What Everyone Should Know about Science* (Cambridge: Cambridge University Press, 1993).

⁵ Gary Edmond and David Mercer, “Experts and Expertise in Legal and Regulatory Settings” in Gary Edmond (ed.), *Expertise in Regulation and Law* (Aldershot: Ashgate, 2004) 1–31.

U.S. Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993) and implicitly endorsed in several appeals to the Supreme Court of Canada.⁶

But what the *Daubert* Court has to offer by way of advice about how to make such determinations is—well, a little embarrassing.... [N]either Popper’s nor Hempel’s philosophy of science will do the job they want it to do. Popper’s account of science is in truth a disguised form of skepticism; if it were right, what Popper likes to call “objective scientific knowledge” would be nothing more than conjectures which have not been falsified. And, though Hempel’s account at least allows that scientific claims can be confirmed as well as disconfirmed, it contains nothing that would help a judge decide whether evidence proffered is really scientific, or how reliable it is. And the most fundamental problem is that the *Daubert* Court (doubtless encouraged by the dual descriptive and honorific uses of “scientific”) is preoccupied with specifying what the method of inquiry is that distinguishes the scientific and reliable from the non-scientific and unreliable. There is no such method.⁷

“There is no logic of discovery”, there is no logic of testing, either; all of the formal algorithms proposed for *testing*, by Carnap, by Popper, by Chomsky, etc., are, to speak impolitely, *ridiculous*: if you don’t believe this, program a computer to employ one of these algorithms and see how well it does at testing theories!⁸

In a nutshell the problem is that all characterizations offered of scientific method at the level of generalization and abstraction favoured by philosophers of science fail to be an account of anything specifically scientific. Hence such stories cannot account for what is special about science.⁹

Historical and sociological research suggests that formal education and socialization into a research tradition or research institution are more important to scientific practice than knowledge of philosophical formulations or formal accounts of method.

⁶ *Daubert v. Merrell Dow Pharmaceuticals, Inc.* 509 US 579 (1993).

⁷ Susan Haack, “An Epistemologist in the Bramble-Bush: At the Supreme Court with Mr Joiner” (2001) 26 *Journal of Health Politics, Policy & Law* 217–48, 231–232; David Stove, *Popper and After: Four Modern Irrationalists* (London: Pergamon, 1982).

⁸ Hilary Putnam, “The ‘Corroboration’ of Theories” in Paul Schilpp (ed.), *The Philosophy of Karl Popper*, vol. 1 (La Salle, IL: Open Court Press, 1974) 221; Imre Lakatos and Alan Musgrave (eds.), *Criticism and the Growth of Knowledge* (1970); David Oldroyd, *The Arch of Knowledge* (1986); Anthony O’Hear, *Karl Popper* (1980); T. Burke, *The Philosophy of Popper* (Manchester: University of Manchester Press, 1983); Alan Chalmers, *What Is This Thing Called Science?* (St Lucia: University of Queensland Press, 1982).

⁹ William Newton-Smith, “Popper and Reliabilism” in Anthony O’Hear (ed.), *Karl Popper: Philosophy and Problems* (Cambridge: Cambridge University Press, 1995).

It is no coincidence that courses on scientific methods are more common in economics, psychology, and sociology than biology, chemistry, and physics.¹⁰

Moreover, philosophers, sociologists, and scientists have been unable to produce criteria that can be consistently operationalized to distinguish between the scientific, the non-scientific, and the pseudo-scientific.¹¹ Larry Laudan, a philosopher of science, summarized the situation in the following way:

The quest for a specifically scientific form of knowledge, or for a demarcation criterion between science and non-science, has been an unqualified failure. There is apparently no epistemic feature or set of such features which all and only the “sciences” exhibit.¹²

Some of the same issues arise in relation to *reliability*. The absence of a universal scientific method or a demarcation criterion (or criteria) capable of distinguishing between science and non-science means that attempts to determine the reliability of knowledge and techniques are not reducible to simple algorithms. Assessments of reliability require sensitivity to their purpose(s) as well as the available evidence. Reliability determinations should always be responsive to the question: Reliable enough for what?¹³

Appeals to *objectivity*—along with impartiality and neutrality—rarely assist in the resolution of technical controversy (see Section 11). The historian of medicine Randall Albury explained that

matters of disagreement between scientific experts are not typically conflicts between objectivity on one side and bias on the other, but conflicts involving two rival conceptions of objectivity—that is, two different ways of assigning relevance to the available data and of interpreting their meaning.... The question of objectivity, then, as it relates to the problem of conflicting advice from scientific experts on matters of social importance, is not properly a

¹⁰ Michael Polanyi, *Personal Knowledge* (1958); Michal Mulkay and Nigel Gilbert, “Putting Philosophy to Work: Karl Popper’s Influence on Scientific Practice” (1981) 11 *Philosophy of the Social Sciences* 389–407; John Schuster and Richard Yeo (eds.), *The Politics and Rhetoric of Scientific Method* (Dordrecht: D. Reidel, 1986).

¹¹ Rachel Laudan (ed.), *The Demarcation between Science and Pseudo-Science* (Blacksburg, VA: Virginia Polytechnic and State University, 1983); Phillip Quinn, “The Philosopher of Science as Expert Witness” in John Cushing et al. (eds.), *Science and Reality* (Notre Dame: University of Notre Dame Press, 1984) 367–86.

¹² Larry Laudan, *Beyond Positivism and Relativism* (Boulder, CO: Westview Press, 1996) 86.

¹³ Thomas Gieryn, *Cultural Boundaries of Science* (Chicago: University of Chicago Press, 1998); Karin Knorr-Cetina, *Epistemic Cultures* (Cambridge, MA: Harvard University Press, 1999); Peter Galison and David Stump (eds.), *The Disunity of Science* (Stanford: Stanford University Press, 1996).

question of deciding in the abstract which expert is more objective. It is a concrete question of which expert's version of objectivity is to be preferred.¹⁴

For some, like the historian Thomas Kuhn, it was the death of scientists rather than the identification of bias, allegations of methodological impropriety, or persuasive demonstrations that changed the commitments held by individual scientists and research communities.¹⁵ We can see this in a prominent historical example from Kuhn's own work on the history of astronomy. After the publication of Nicolaus Copernicus' *de Revolutionibus* (1543), and notwithstanding intervening discoveries by Christopher Columbus and Galileo Galilee, it took generations before the geocentric system was replaced by a heliocentric cosmology. Even the application of the telescope, improved astronomical tools, and more precise measurements did not persuade several technically proficient astronomers, such as Tycho Brahe, to abandon their commitment to an Aristotelian earth-centred universe.¹⁶

In addition, historical and empirical studies have not been able to identify a set of institutional commitments or a professional ethos consistently adhered to by scientists across a field or even within a subdiscipline, let alone norms embraced by all scientists and experts. The sociologist Robert Merton provided an early and influential elaboration of scientific norms and their functions. Merton's account explained the importance of values like *communalism*, *openness*, *disinterestedness*, and *skepticism* to the pursuit of legitimate scientific research.¹⁷ Today, many commentators, lacking Merton's historical erudition and theoretical sophistication, have promoted Mertonian-style norms as some kind of essential prerequisites or description of authentic scientific activity. Not only does this overstate Merton's actual position, but more recent sociological research suggests

¹⁴ Randall Albury, *The Politics of Objectivity* (Maryborough: Deakin University Press, 1983) 42; Robert Proctor, *Value-Free Science?* (Cambridge, MA: Harvard University Press, 1991); Steven Shapin, *A Social History of Truth* (Chicago: University of Chicago Press, 1995).

¹⁵ Thomas Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1970) 144–159.

¹⁶ Thomas Kuhn, *The Copernican Revolution: Planetary Astronomy in the History of Western Thought* (Cambridge, MA: Harvard University Press, 1957). More recent work by the sociologists Harry Collins and Trevor Pinch on the study of gravitational waves and solar neutrinos suggests that the commitments and beliefs held by scientists are not always conditioned by widely accepted experiments and findings. See Harry Collins, *Gravity's Shadow: The Search for Gravitational Waves* (Chicago: University of Chicago Press, 2004) and Trevor Pinch, *Confronting Nature: The Sociology of Solar-Neutrino Detection* (Dordrecht: D. Reidel, 1986). Similarly, the slow demise of forensic expertise based around hair comparison, bullet lead analysis, and some forms of explosive identification also suggest that it can be difficult to eliminate bodies of knowledge and practice even where they have not been empirically validated and/or are subject to serious challenge. See, generally: David Faigman et al., *Modern Scientific Evidence: Forensics* (Eagan: Thomson/West, 2006); Kelly Pyrek, *Forensic Science under Siege: The Challenges of Forensic Laboratories and the Medico-Legal Death Investigation System* (London: Academic, 2007).

¹⁷ Robert Merton, *The Sociology of Science: Theoretical and Empirical Investigations* (Chicago: University of Chicago Press, 1973).

that norms, such as disinterestedness and skepticism, are better understood as part of a complex professional language, susceptible to manipulation and strategic deployment. The following examples help to illustrate some of the practical limitations with scientific norms.

On the basis of a study of NASA personnel, Ian Mitroff explained how experienced scientists routinely derogate from Mertonian-style norms. From his observations and interviews, Mitroff concluded that highly regarded scientists were often passionately committed to pet theories, sometimes in the face of very damaging *evidence*.¹⁸ In practice, NASA scientists tended to be far more skeptical of their rivals' theories and data than their own. And, these scientists were able to provide reasons for preferring their own positions. Some of these reasons included concerns about: relative levels of competence; underlying assumptions; (in)consistency with theory, other results, and interpretations; the reliability of equipment; and so forth. *Inconsistent evidence* was almost never interpreted as some kind of definitive refutation or falsification. In addition, Mitroff's scientists were not as co-operative or forthcoming with results and techniques as those committed to Mertonian norms—like communalism—might have anticipated. Once again the scientists provided (seemingly credible) explanations for their behaviour. Reasons for withholding data and materials included: the need to establish priority claims; previous (or anticipated) failure to reciprocate; protecting the work of graduate students; and waiting for confirmatory studies or actual publication. Mitroff found that scientific practices were explained using a variety of discursive resources. Where activities seemed to contradict popular normative expectations, scientists appealed to a range of exceptions and qualifications that helped to excuse or legitimate what might otherwise have been understood as deviant behaviour. Mitroff characterized these principled or reasoned derogations as *counter-norms*.¹⁹

Interestingly, and perhaps against expectations, derogation from scientific norms and the invocation of counter-norms did not simply correlate with a scientist's standing or professional credibility. Similarly, knowledge derived through secret, non-co-operative, or interested activities was not necessarily seen as pathological or unreliable. Instead, Mitroff's work, like subsequent sociological and anthropological studies of expertise,

¹⁸ Ian Mitroff, *The Subjective Side of Science* (Amsterdam: Elsevier, 1974).

¹⁹ Ian Mitroff, "Norms and Counter-Norms in a Select Group of the Apollo Moon Scientists: A Case Study in the Ambivalence of Scientists" (1974) 39 *American Sociological Review* 579.

suggests a much richer realm of scientific and medical practice.²⁰ Members of specialist communities are often familiar with the personality and temperament of fellow scientists, as well as their previous work, their abilities, their commitment to ideas and theories, and earlier *derogations* from popular normative commitments. Indeed, these are often combined with other evidence and commitments in a complex and morally inflected evaluation of an expert's competence, performance, and findings.

These, however, are not the only limitations with recourse to idealized norms. Others, such as the sociologist Michael Mulkay, have explained how norms themselves can create interpretative complexity.²¹ Vague norms—like openness and skepticism—are unlikely to guide scientific practice or the assessment of knowledge claims, especially in contexts where the behaviour, motivations, and alignments of scientists *as well as* technical issues all form part of the dispute. If we momentarily reflect on the idea of *skepticism*, the significance of under-determination should become clear. Confronted with unexpected experimental results, should a scientist tinker with their equipment and assumptions or simply accept the results even if they are potentially disruptive to widely accepted commitments, background theories, and previous findings?²² This simple example raises question about how a skeptical attitude might be embodied in any given situation. It probably will not surprise lawyers when they are told that norms such as skepticism are capable of being interpreted and applied inconsistently.

Additional limitations with idealized approaches to science and medicine emerge from a closer examination of the research literatures. Superficially, peer review and publication might appear to provide a presumptive indication of reliability. However,

... emphasis on peer review reinforces a myth that says all scientific journals use rigorous expert review in selecting all content and that the peer review process operates according to certain universal, objective, and infallible procedures, standards and goals. Quite the opposite is true, however.... [P]eer review is neither uniform nor totally reliable nor intended as a fraud detection mechanism. It's principal goal—and perhaps what should be its only goal—is to evaluate

²⁰ See also the pioneering laboratory studies by: Bruno Latour and Steve Woolgar, *Laboratory Life: The Social Construction of Scientific Facts* (Beverly Hills: Sage, 1979); Karin Knorr-Cetina, *The Manufacture of Knowledge* (Oxford: Pergamon, 1981); Michael Lynch, *Art and Artifact in Laboratory Science* (1985).

²¹ Michael Mulkay, "Norms and Ideology in Science" (1980) 4–5 *Social Sciences Information* 637–56; Michael Mulkay, "Interpretation and the Use of Rules: The Case of the Norms of Science" in Thomas Gieryn (ed.), *Science and Social Structure (A Festschrift for R.K. Merton)* (New York: New York Academy of Sciences, 1980) 111–25. See also Steve Woolgar, *Science: The Very Idea* (London: Tavistock, 1988) and, for an informative example of the social complexity of scientific practice, consider Steve Epstein, *Impure Science: AIDS Activism and the Politics of Knowledge* (Berkeley: University of California Press, 1996).

²² These difficulties have been described as the *experimenter's regress*. See Harry Collins, *Changing Order* (Chicago: University of Chicago Press, 1985).

manuscripts according to whether they should be accepted or rejected, not to determine their authenticity. The peer review procedures so often touted in political settings as ensuring scientific authenticity, accountability, or authority are simply arbitrary creations.... No set rules govern how, when, or by whom all journal peer review is conducted.... Considering the admitted failing, cumbersomeness, and cost of the peer review system, a skeptic might wonder why it survives.²³

Scientific journals, especially the most prestigious, have many, sometimes competing, obligations. They have interests in rapid dissemination, maintaining broad appeal, providing the most accurate information available, and remaining solvent—through subscriptions and sometimes advertising revenues.²⁴ Increasingly, they have obligations not only to scientists, engineers, and physicians but to security investors and venture capitalists. Nevertheless, submissions are not always thoroughly reviewed, not always written by named authors and only replicated in exceptional circumstances. Perhaps it will not be surprising to find that reviewers tend to be more sympathetic to views that are consonant with their own.²⁵ Moreover, in recent decades commercial competition, private funding, and concerns about liability have provided new incentives for not undertaking and not publishing research, and imposed new contractual restrictions on the dissemination of technologies, data, and results.²⁶

Studies of the meaning of publication (and implicitly peer review) among professional communities reveal a complex state of affairs. In a longitudinal study of physicists the sociologist Harry Collins found that the published literature held a range of inconsistent meanings for different audiences.²⁷ Among small groups of specialists, who were extremely conversant with ongoing controversies in a particular discipline or sub-field, articles that had been published in highly respected peer reviewed physics journals were sometimes dismissed or discounted. By way of comparison, those on the peripheries of the subgroup and further afield—usually

²³ Marcel LaFollette, *Stealing into Print: Fraud, Plagiarism, and Misconduct in Scientific Publishing* (Berkeley: University of California Press, 1992) 119–121; Daryl Chubin and Edward Hackett, *Peerless Science: Peer Review and U.S. Science Policy* (Albany: SUNY Press, 1990) 122; Stephen Lock, *A Difficult Balance: Editorial Peer Review in Medicine* (Philadelphia: ISI Press, 1986).

²⁴ Sheldon Krimsky, “Publication Bias, Data Ownership, and the Funding Effect in Science: Threats to the Integrity of Biomedical Research” in Wendy Wagner and Rena Steinzor (eds.), *Rescuing Science from Politics: Regulation and the Distortion of Scientific Research* (Cambridge: Cambridge University Press, 2006) 60, 72.

²⁵ Sheila Jasanoff, *The Fifth Branch: Science Advisers as Policy-Makers* (Cambridge, MA: Harvard University Press, 1999) 61–83.

²⁶ Sheldon Krimsky, *Science in the Private Interest* (Lanham: Rowman & Littlefield, 2003); Wendy Wagner and Rena Steinzor (eds.), *Rescuing Science from Politics: Regulation and the Distortion of Scientific Research* (Cambridge: Cambridge University Press, 2006).

²⁷ Collins, *Gravity’s Shadow* (2004); Harry Collins, “Tantalus and the Aliens: Publications, Audiences and the Search for Gravitational Waves” (1999) 29 *Social Studies of Science* 163.

unfamiliar with individual scientists (or teams) and the contours of the field—were far less discriminating in their approach to published work.

Revealingly, Collins found that the members of specialist physics communities were most likely to respond to competing theoretical approaches and results, especially those already in print, when external funding was at stake. This was particularly conspicuous when those managing the funding were generalists or had expertise that was considered, by the specialists physicists, to be insufficiently discriminating. Factors such as funding decisions could mobilize interested, but otherwise passive or indifferent, scientists into action and publication. On the basis of this and other sociological research, the meaning of peer review and publication appears to depend on an assortment of technical, institutional, professional, and social factors. The study by Collins suggests that it may be dangerous to approach the simple fact of publication without some sensitivity to a range of tacit and often subterranean dynamics.

Another important dimension to peer review and publication is that their meaning and significance change across fields and over time. Sometimes peer review refers to the refereeing of papers prior to publication. When papers are reviewed before publication, though, the referees are not very consistent in their performances and editors tend to have a range of views on how to respond to comments and criticisms. Sometimes peer review refers to the attention given to papers that have already been published. Peer review is also used, though perhaps in its weakest sense, to describe the informal review of work by a colleague.

Publication also has many valencies. Not all scientific publications, and this even extends to articles featured in refereed journals, are refereed prior to publication. Indeed, it is common for conference proceedings and solicited papers to escape formal refereeing. Many journals feature non-refereed supplements, and few journals devote resources to quantitative reviews of the articles they publish.²⁸ Also, the quality and depth of review varies considerably among publications and referees. Studies suggest that the average reviewer spends just a couple of hours refereeing a journal submission.²⁹

²⁸ M. Cho and L. Bero, “The Quality of Drug Studies in Symposium Proceedings” (1996) 124 *Annals of Internal Medicine* 485.

²⁹ Alfred Yankauer, “Who Are the Peer Reviewers and How Much Do They Review?” (1990) 263 *Journal of the American Medical Association* 1338; Stephen Lock and Richard Smith, “What Do Peer Reviewers Do?” (1990) 263 *Journal of the American Medical Association* 1341.

These insights suggest that the ingredients of popular representations of science, medicine, and other forms of expertise are unlikely to provide the kinds of discriminating criteria that might guide admissibility determinations and meaningful assessments of expert evidence.³⁰ Materials deemed suitable for publication may not, for example, provide a sufficiently reliable basis to convict someone for a serious criminal offence. Rather than build admissibility standards on idealized versions of science and medicine, common-law judges are in a much better position to develop tools and resources specifically suited to legal needs and values. In approaching expert evidence produced by the prosecution, this paper contends that it is desirable to require a demonstration of reliability, preferably supported by some evidence of empirical testing. Placing emphasis on evidence of reliability incorporates a flexible standard that can be tailored to the particular kinds of evidence, the type of litigation, the capabilities of the respective parties, and the exigencies of the case without reifying scientific method, the normative ethos, or the efficacy of peer review and publication.

This, however, takes us too far ahead. Before moving to consider admissibility standards, it is useful to examine the close relationship between legal institutions and the forensic sciences and to review contemporary Canadian expert evidence jurisprudence. As we shall see, judicial confidence in the expert evidence produced by the state has meant that courts have not always excluded unreliable inculpatory evidence or achieved their full potential in holding the state's investigative agencies and their experts to account.

3. Forensic Medicine and the Forensic Sciences as *Law–Science Hybrids*

Most of the fields we are discussing did not grow out of basic science.... There is no systematic, rigorous, empirical research on which the forensic identification sciences' knowledge is built. If called upon to prove their claims, they have little or no data to marshal in their support. Instead [they] point to a guild of mutually self-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.³¹

³⁰ Compare the norms implicit in *The "Ikarian Reefer"* [1993] 20 FSR 563, an English case, and the Guidelines from the Federal Court of Australia reproduced in the Appendix to this paper.

³¹ Michael Saks, "Banishing the *Ipse Dixit*: The Impact of *Kumho Tire* on Forensic Identification Science" (2000) 57 *Washington & Lee Law Review* 879; Michael Risinger, Mark Denbeaux, Michael Saks, "Exorcism of Ignorance as a Proxy for Rational Knowledge: The Lesson of Handwriting 'Expertise'" (1989) 137 *University of*

Since their inception, as the names suggest, forensic science and forensic medicine have been directly and continuously involved in the legal and administrative functions of the state.³² These forms of expertise have evolved in a symbiotic relationship with the criminal justice system. From the judicial perspective that relationship has been characterized by trust rather than scrutiny or accountability.

Discussing forensic pathology, Smith and Wynne explain:

It is not only the court room interaction that shapes knowledges: the institutional integration of a particular expert profession into the legal process already achieves this. Indeed, for forensic science and pathology, the legal process itself has created their particular type of professional interaction and expert knowledge. The social integration of forensic expertise with the law is such that forensic experts have learnt to reconcile themselves to [and we might add insulate themselves from] the regular adversarial scepticism of legal processes, whilst maintaining the normal consensual discourses of scientific expertise whereas other disciplines may manage this by defining the courtroom interaction as ‘unscientific’, this is not so easily available to forensic experts, because the courtroom is their ultimate professional arena.³³

Forensic scientific practice, manuals, protocols, training, and even the language used in reports and court, are all shaped and refined by legal, and especially judicial, requirements and expectations.

Forensic science and medicine are, as the previous extracts suggest, tightly coupled with law and legal practice. In the context of an inquiry into pediatric forensic pathology this is significant, because systematic or institutional failures with forensic science and medicine are simultaneously and inextricably legal problems. Scientific and medical failures reflect not only the limitations of public institutions, organizational arrangements within laboratories, and the competence of individual forensic practitioners, but the unwillingness or inability of the legal system—and here we need to include trial and appellate judges—to successfully manage expert evidence produced by the state.

Pennsylvania Law Review 731–792; Michael Risinger, “Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock” (2000) 64 *Albany Law Review* 99–152; David Faigman, *Legal Alchemy: The Use and Misuse of Science in the Law* (New York: W.H. Freeman and Co., 1999); David Faigman, *Laboratory of Justice* (New York: Henry Holt, 2004); William Thompson, “Analyzing the Relevance and Admissibility of Bullet-Lead Evidence: Did the NRC Report Miss the Target?” (2005) 46 *Jurimetrics* 65–89.

³² Carol Jones, *Expert Witnesses: Science, Medicine and the Practice of Law* (Oxford: Clarendon Press, 1994) 194–223; Tal Golan, *Laws of Nature, Laws of Men* (Cambridge, MA: Harvard University Press, 2004).

One of the reasons for recurrent difficulties with expert evidence in criminal justice is that courts have not subjected institutionalized forensic science and medicine to particularly stringent tests of accountability. The lack of scrutiny means that many of the forensic scientific and medical techniques in use today were originally granted access to the courts during more liberal admissibility regimes when judges were less concerned about admissibility and more deferential to scientific and medical evidence produced by the state. Many forensic techniques, some longstanding, are yet to be subjected to rigorous investigation and validation.³⁴ Common-law judges have often preferred to rely on earlier decisions and legal commentary than undertake a review of the validity or accuracy of widely used and presumptively admissible techniques and theories. With the continuing support of the state and legal institutions, forensic scientific and medical practice have been relatively sheltered from serious scrutiny and the need to test their techniques.

As the extracts from Smith, Wynne, and Saks suggest, forensic medicine and the forensic sciences seem to have operated outside or at the margins of mainstream biomedical and scientific research. To some extent their operations are a function of the expectations placed upon them by police and investigative agencies, the reluctance of courts to impose more appropriate standards, as well as the types of cases and issues forensic experts are required to investigate. The professional marginalization of forensic science and medicine is also a result of the historical unwillingness of governments to adequately resource and regulate them. The close relations between forensic scientists, investigators, police, and prosecutors seem to have fostered a range of pro-prosecution orientations and sympathies.³⁵ In conjunction with unexplicated judicial confidence, these commitments have contributed to a state of affairs that may be undesirable in a system concerned with truth and justice.

These conditions are compounded by other factors. On average, forensic scientists and technicians tend to have quite limited formal qualifications. They are less likely to have completed research degrees, post doctoral fellowships or held university positions than scientists involved in more mainstream biomedical and scientific research. They are also less likely to undertake systematic research and, at least

³³ Roger Smith and Brian Wynne, "Introduction" in Roger Smith and Brian Wynne (eds.), *Expert Evidence: Interpreting Science in the Law* (London: Routledge, 1989) 1–22, 15.

³⁴ Simon Cole, *Suspect Identities: A History of Fingerprinting and Criminal Identification* (Cambridge, MA: Harvard University Press, 2001).

³⁵ Doreen McBarnett, *Conviction: Law, the State and the Construction of Justice* (London: MacMillan, 1983); Pat Carlen, *Magistrates' Justice* (London: Martin Robertson and Co., 1976).

historically, were far less likely to publish any findings. The small volume of research that was pursued tended to be case-based—often reflecting *success* at trial—and published in parochial journals.³⁶

Forensic science and medicine tend to be *applied*. In practice, it is not clear that such an appellation is particularly meaningful. It certainly does not provide an excuse to circumvent the need for rigour, especially the need for empirical validation to demonstrate reliability and individual competence. The *sui generis* nature of many crimes, along with ethical and practical constraints on some lines of inquiry, may occasionally prevent thorough investigation. However, these kinds of constraints and limitations are precisely why judges need to carefully scrutinize the manner in which forensic pathologists, scientists, and technicians develop and support their expert evidence and opinions.³⁷

Historically, forensic science and medicine have relied upon “art” and “experience” in addition to experimental techniques. Where forensic pathologists, or other forensic scientists and technicians, rely upon their experience at trial, they create pronounced difficulties. They produce opinions that may be practically difficult to assess. Unless the expert has been formally censured, is known to have made errors in the past, or his or her opinion is wildly speculative, implausible, or obviously outside their previous experience, it can be incredibly difficult for the defence to meaningfully challenge the expert’s evidence. Furthermore, poor laboratory practices can reduce or eliminate materials such as photos, slides, results, and notes that might be used by defence experts to evaluate findings and conclusions. This raises the question, addressed in more detail below, of whether it should be left to individual defendants to try and demonstrate weaknesses and limitations in the state’s expert evidence.

Expert evidence based on intuition, speculation, and experience—the so-called *ipse dixit*—should be approached with considerable apprehension by judges and prosecutors. Over time, forensic science and medicine should become more conspicuously *evidence-based* (see Section 12). Forensic scientific, medical, and technical evidence should be more *science* than *art*.

³⁶ Stephan Timmermans, *Postmortem: How Medical Examiners Explain Suspicious Deaths* (Chicago: University of Chicago Press, 2006) 19–20.

³⁷ Gary Edmond, “Science in Court: Negotiating the Meaning of a ‘Scientific’ Experiment during a Murder Trial and Some Limits to Legal Deconstruction for the Public Understanding of Law and Science” (1998) 20 *Sydney Law Review* 361–401.

There are many things that might be done to improve the quality of institutionalized forensic science and medicine. These include: better resources and facilities; higher entry-level qualifications and ongoing training; encouraging and rewarding publication; encouraging and rewarding the identification and disclosure of exculpatory evidence; establishing multidisciplinary supervisory groups; encouraging experts to take research sabbaticals in prestigious research institutions; requiring continuous testing of performance; blinding experts to other aspects of the investigation, where possible; making the expert's primary duty to the court even more explicit; and so on. These possibilities, however, are largely beyond the scope of this paper.

Instead, this paper is primarily concerned with legal responses from the courts. Regardless of any institutional or organizational reforms, it is imperative that the courts independently reform their admissibility jurisprudence to insulate themselves against unreliable expert evidence and its deleterious system effects. Judges have the ability to make forensic science and medicine publicly accountable and more reliable. They also have the ability to regulate their own rules and procedures so that concern with truth and justice is not undermined by blind trust or sacrificed to political expediency—especially during high-profile prosecutions and appeals.

Historically, appellate judges and commissioners have been eager to cast responsibility for wrongful convictions on police, investigators, forensic scientists, and (to a certain extent) prosecutors and defence counsel, and to absolve the performances of trial judges, and earlier and differently constituted appellate courts.³⁸ Judges cannot, however, shift the entire responsibility for problems with the forensic sciences because these sciences and technologies have grown up around the courts and have been condoned or sanctioned by them. To blame forensic science and medicine for wrongful convictions trivializes the constitutive role of trial and appellate courts in the recognition and admission of questionable and unreliable forms of incriminating expert evidence (and sometimes the exclusion of defence expert evidence) and the affirmation of convictions.³⁹ Approaching institutionalized forensic

³⁸ For example, Justice May, *Final Report: Return to an Address of the Honourable the House of Commons Dated 30 June 1994 for a Report into the Circumstances Surrounding the Conviction Arising out of the Bomb Attacks in Guildford and Woolwich in 1974* (1994); Viscount Runciman, *Royal Commission on Criminal Justice*, (HMSO, 1993); Justice Morling, *Report of the Commissioner: Royal Commission of Inquiry into the Chamberlain Convictions* (1987).

³⁹ Gary Edmond, "Misunderstanding the Uses of Scientific Evidence in High Profile Criminal Appeals: The Social Construction of Miscarriages of Justice" (2002) 22 *Oxford Journal of Legal Studies* 53–89 and "The Law-Set: The Legal-Scientific Production of Medical Propriety" (2001) 26 *Science, Technology & Human Values* 191–226.

medicine and the forensic sciences as law–science hybrids implicates the judiciary in the production, admission, use, and review of expert forensic evidence.

The practical upshot of all this is that it makes little sense to focus all the critical attention beyond the courts. For, unavoidably, part of the responsibility for legal failures and miscarriages of justice involving unreliable expert evidence is directly attributable to judges. The existing rules of admissibility, judicial discretions, jury directions, limitations placed on the use of evidence, and the availability of review have not been used in ways that might have prevented serious and continuing problems.

Overall, there would seem to be a need to make institutionalized forensic science and medicine more accountable and more reliable. At the same time it would seem important for courts to adopt a more judicious response to the state’s expert evidence by refining their admissibility standards and maintaining a critical distance.

At this point we turn to consider some of the Canadian admissibility jurisprudence delivered in the wake of *Mohan*. Even though this jurisprudence has, in recent years, adopted a more explicit concern with reliability, it has not prevented the admission of expert evidence that has directly contributed to a number of notorious miscarriages of justice.

4. Contemporary Canadian Admissibility Jurisprudence

Supreme Court expert evidence jurisprudence (and the jurisprudence from the Ontario Court of Appeal) is already concerned, if occasionally obliquely, with the reliability of expert evidence. As this succinct overview explains, Canadian courts have embraced indicia of reliability, including some from the U.S., apparently without fully appreciating the value of empirical testing or adequately explaining how to operationalize reliability in the context of admissibility decision making.

Undoubtedly, the most important of the modern decisions on the admissibility of expert opinion evidence is *R. v. Mohan* (1994).⁴⁰ *Mohan* was an appeal to the Supreme Court from the Ontario Court of Appeal. Writing for that Court, Justice Sopinka outlined the four criteria (hereafter the *Mohan* criteria) governing the admissibility of *all* expert evidence. They are:

⁴⁰ [1994] 2 SCR 9. See also *R. v. Béland* [1987] 2 SCR 398.

- (a) relevance;
- (b) necessity in assisting the trier of fact;
- (c) the absence of any exclusionary rule; and
- (d) a properly qualified expert.⁴¹

Basically, *relevance* (a) is governed by logical relevance but also shaped by other factors—such as whether the expert evidence will consume an amount of time and resources disproportional to its probative value, or whether the probative value of the evidence might be “overborne by its prejudicial effect.” Both considerations, to be revisited below, require focusing some attention on the reliability of the expert evidence.

Necessity (b) is linked to assisting the trier of fact. This criterion imposes a standard higher than mere “helpfulness” but one that should not be interpreted too strenuously. Absolute necessity is not required.⁴² Drawing from *R. v. Abbey* (1982), the *Mohan* Court explained that expert evidence may be necessary if it provides relevant information “which is likely to be outside the experience and knowledge of a judge or jury.”⁴³ In his discussion of necessity, Justice Sopinka suggested that some of the distraction and confusion—aspects of prejudicial effect—created by expert evidence “can often be offset by proper instructions.” But he also warned that too liberal an approach to necessity might result in trials becoming “nothing more than a contest of experts” with the trier of fact obliged to act “as a referee.”⁴⁴

The third factor (c) requires the trial judge to consider expert evidence in the context of the rules of evidence, particularly the *exclusionary rules*, governing the trial. Such considerations are almost always *sui generis*—depending on the circumstances of the case and the way it is presented and defended. Placing emphasis on the *expert’s qualifications* (d) requires the evidence to “be given by a witness who is shown to have acquired special or peculiar knowledge through study or experience in respect of the matters on which he or she undertakes to testify.”⁴⁵

Summarizing the four criteria, Sopinka J. explained:

⁴¹ [1994] 2 SCR 9 at 16.

⁴² [2000] 2 SCR 275 at [21].

⁴³ [1994] 2 SCR 9 at 19; *R. v. Abbey* [1982] 2 SCR 109.

⁴⁴ [1994] 2 SCR 9 at 20.

⁴⁵ [1994] 2 SCR 9 at 21.

It appears from the foregoing that expert evidence which advances *a novel scientific theory or technique is 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.⁴⁶

Even though it was not one of the four criteria enumerated by the Supreme Court in *Mohan*, *reliability* features prominently in its jurisprudence and reasoning. In *Mohan*, the issue of reliability arose most conspicuously in relation to necessity and relevance. Applying the admissibility criteria to the case at hand, Sopinka J. explained that in the absence of “*indicia* of reliability, it cannot be said that the [expert] evidence would be necessary.”⁴⁷ On the issue of relevance, the Court accentuated the need to exclude expert evidence if the probative value is outweighed by the prejudicial effect.

Evidence that is otherwise logically relevant may be excluded ... if its probative value is overborne by its prejudicial effect, if it involves an inordinate amount of time which is not commensurate with its value or if it is misleading in the sense that *its effect on the trier of fact, particularly a jury, is out of proportion to its reliability*.⁴⁸

Regardless of whether reliability is treated as an aspect of relevance (or necessity) or treated as a separate exclusionary rule, the Court recognized that the “reliability versus effect factor has special significance in assessing the admissibility of expert evidence.”⁴⁹ Where it is not reliable there is “a danger that the expert evidence will be misused and will distort the fact-finding process.” The Court was anxious, lest

[d]ressed 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.⁵⁰

⁴⁶ [1994] 2 SCR 9 at 21. (italics added)

⁴⁷ [1994] 2 SCR 9 at 34.

⁴⁸ [1994] 2 SCR 9 at 17. (italics added) See also *R. v. Johnston* (1992) 69 CCC (3d) 395. The probative value/prejudicial effect discretion was formalized in the decades following the House of Lords decision of *R. v. Christie* [1914] AC 545.

⁴⁹ [1994] 2 SCR 9 at 16-17; *R. v. Terceira* (1998) 123 CCC (3d) 1 at [29]; *R. v. Dimitrov* (2003) 181 CCC (3d) 555 at [48]; *Report of the Kaufman Commission on proceedings involving Guy Paul Morin* (1998) 315.

⁵⁰ [1994] 2 SCR 9 at 17.

Later in the judgment Sopinka J. 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.”⁵¹

Concern with reliability was consolidated in subsequent appeals to the Supreme Court. In *R. v. J.-L.J.* (2000)—an appeal from the Court of Appeal for Quebec—the Supreme Court suggested that concerns about the reliability of expert evidence may transcend the *Mohan* criteria; at least with respect to novel scientific techniques.⁵² In *R. v. D.D.* (2000)—an appeal from the Ontario Court of Appeal—the Supreme Court explained that even when the criteria are satisfied “the [expert] evidence may be rejected if its prejudicial effect on the conduct of the trial outweighs its probative value.”⁵³ For the Court, “probative value is determined by considering the reliability, materiality and cogency of the expert testimony.”⁵⁴

In *R. v. J.-L.J.* the Supreme Court reiterated the need to subject any novel scientific technique “to *special scrutiny* to determine whether it meets a basic threshold of reliability.”⁵⁵ The Court intimated that *Mohan* had explicitly rejected the *general acceptance* test derived from the U.S. decision in *US v. Frye* (1923) and—“moving in parallel with its [U.S.] replacement”—drew upon the four reliability criteria outlined by the U.S. Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993).⁵⁶ *Daubert*, the leading U.S. federal admissibility determination for expert evidence, provided an authoritative interpretation of the *Federal Rules of Evidence* (1975). It placed emphasis on the reliability of scientific evidence and provided a set of criteria (hereafter the *Daubert* criteria) effectively superseding *Frye*.⁵⁷

The *Daubert* criteria offered the Canadian judges assistance in assessing the reliability (or “soundness”) of expert evidence. Writing for the Court, Justice Binnie explained:

While *Daubert* must be read in light of the specific text of the *Federal Rules of Evidence*, which differs from our own procedures, the U.S. Supreme Court did list a number of factors that could be helpful in evaluating the soundness of novel science:

⁵¹ [1994] 2 SCR 9 at 33.

⁵² [2000] 2 SCR 600.

⁵³ [2000] 2 SCR 275 at [11].

⁵⁴ [2000] 2 SCR 275 at [36].

⁵⁵ [2000] 2 SCR 600 at [35]-[36], citing *R. v. Mohan* (1994).

⁵⁶ Curiously, given this interpretation, *Mohan* does not refer to *Daubert* and only refers to “general acceptance” in passing at the end of the judgment.

⁵⁷ Paul Giannelli, “The Admissibility of Novel Scientific Evidence: *Frye v. United States*, a Half Century Later” (1980) 80 *Columbia Law Review* 1197–250.

- (1) whether the theory or technique can be and has been tested:

Scientific methodology today is based on generating hypotheses and testing them to see if they can be falsified; indeed, this methodology is what distinguishes science from other fields of human inquiry.

- (2) whether the theory or technique has been subjected to peer review and publication:

[S]ubmission to the scrutiny of the scientific community is a component of “good science,” in part because it increases the likelihood that substantive flaws in methodology will be detected.

- (3) the known or potential rate of error or the existence of standards; and,

- (4) whether the theory or technique used has been generally accepted:

A “reliability assessment does not require, although it does permit, explicit identification of a relevant scientific community and an express determination of a particular degree of acceptance within that community.”...

Widespread acceptance can be an important factor in ruling particular evidence admissible, and “a known technique which has been able to attract only minimal support within the community,” ... may properly be viewed with scepticism.

Thus, in the United States, as here, “general acceptance” is only one of several factors to be considered.⁵⁸

Both *Mohan* and *R. v. J.-L.J.* expressed the need to “preserve and protect the role of the triers of fact” against unreliable expert evidence.⁵⁹

In *R. v. J.-L.J.*, once again drawing upon jurisprudence from the U.S., the Supreme Court of Canada emphasized the importance of judicial *gatekeeping*. The “admissibility of expert evidence,” Binnie J. 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.”⁶⁰ The Court was anxious that the “search for truth” in the courtroom should not include “expert evidence which may ‘distort the fact-finding process.’”⁶¹ The Court also linked its concerns with reliability to the rationale underlying the criminal trial. Applying this to the evidence at hand it explained that while penile plethysmograph was “quite useful in therapy because it 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

⁵⁸ [2000] 2 SCR 600 at [33]. (references omitted) *R. v. D.D.*[2000] 2 SCR 275; *R. v. Dimitrov* (2003) 181 CCC (3d) 554 at [38]. The *Daubert* criteria were earlier endorsed by Justice Hill in *R. v. JET* [1994] OJ NO. 3067 (General Division) at [75].

⁵⁹ [2000] 2 SCR 600 at [25].

⁶⁰ [2000] 2 SCR 600 at [28].

offence.”⁶² Both *Mohan* and *R. v. J.-L.J.* 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.”⁶³ Drawing on the influential Scottish case of *Davie v. Magistrates of Edinburgh* (1953), Binnie J. drew attention to the need for experts to furnish “the necessary scientific criteria for testing the accuracy of their conclusions.”⁶⁴

In *R. v. Trochym* (2007) the Supreme Court’s approach in *R. v. L.-J.L.* was affirmed by a divided court.⁶⁵ The majority decision, written by Justice Deschamp, 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.”⁶⁶

The majority judgment in the *Trochym* appeal is the most recent and most emphatic of the Supreme Court’s admissibility decisions. In Deschamp’s J. reasons, questions about the reliability of expert evidence moved to centre stage.

*Reliability 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.*⁶⁷

The majority repeated earlier interest in *Daubert*. Though, rather than 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.”⁶⁸

In addition to reliability, the Court reiterated the trial judge’s gatekeeping responsibility, particularly the need to determine

⁶¹ [2000] 2 SCR 600 at [29].

⁶² [2000] 2 SCR 600 at [35].

⁶³ [2000] 2 SCR 600 at [50].

⁶⁴ *Davie v. Magistrates of Edinburgh* [1953] SC 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 *Makita Pty Ltd v. Sprowles* [2001] NSWCA 305; *Sydneywide Distributors Pty Ltd v. Red Bull Australia Pty Ltd* [2002] FCAFC 157; *HG v. The Queen* (1999) 197 CLR 414 at [39].

⁶⁵ [2007] 1 SCR 239.

⁶⁶ [2007] 1 SCR 239 at [1]. See the extract from *Trochym* at the beginning of Section 1.

⁶⁷ [2007] 1 SCR 239 at [27]. (italics added)

⁶⁸ [2007] 1 SCR 239 at [34].

whether the value or utility of the evidence outweighs its potential costs in terms of consumption of time, potential prejudice to the accused, and confusion caused to the trier of fact. For this reason, a judge should, in exercising his or her role as “gatekeeper”, carefully scrutinize the admissibility of novel scientific evidence. While parties must be able to put forward the most complete evidentiary record possible, admissibility will necessarily be circumscribed where the evidence may “distort the fact-finding process”.⁶⁹

The majority explained that even if hypnosis was useful in therapy, this did not mean it was sufficiently reliable as a source of evidence in a criminal trial.⁷⁰ It also placed emphasis on testing when it stated:

[I]f evidence whose reliability cannot really be tested is admitted and relied upon simply because it is consistent with other admissible evidence, the danger is that a web of consistent but unreliable evidence will lead to a (potentially wrongful) conviction. As a result, given our current understanding of hypnosis, the admission of post-hypnosis memories may render the right of cross-examination illusory, thereby undermining a key aspect of the adversarial process.⁷¹

Unremarkably, the Supreme Court’s jurisprudence, with its evolving interest in the reliability of expert evidence, has been embraced by the Ontario Court of Appeal. In *R. v. Terceira* (1998), for example, the Court referred to “general acceptance” and “reliability.”

The trial judge’s function is limited to an overview of the evidence proffered in order to be satisfied that it reflects a scientific theory or technique that has either *gained acceptance in the scientific community*, or if not accepted, is considered *otherwise reliable in accordance with the methodology validating it...* [T]he *threshold test of reliability* must remain capable of adaptation to changing circumstances and realities. *Reliability* is best determined under the scrutiny of the trial judge as guided by the demands and peculiarities of the case. Simply stated, *the threshold test of reliability* is met when the trial judge, having reviewed certain evidence presented by counsel, feels that the novel scientific technique or theory is *sufficiently reliable* to be put to the jury for its review.⁷²

⁶⁹ [2007] 1 SCR 239 at [54]. (references omitted).

⁷⁰ [2007] 1 SCR 239 at [37].

⁷¹ [2007] 1 SCR 239 at [60].

⁷² (1998) 123 CCC (3d) 1 at [64] (italics added), see also [24]-[26]; *R. v. Johnston* (1992) 69 CCC (3d) 395 and *R. v. B.(K.G.)* (1993) 79 CCC (3d) 257.

On “the threshold issue of reliability,” Finlayson J.A. explained that this involved asking “[I]s the science itself valid?”

The need for reliability, albeit intertwined with probative value, relevance, and necessity, also featured prominently in *R. v. A.K. & N.K.* (1999).

[Reliability] *concerns the validity of the theory which forms the basis of the opinion advanced by the expert.* The evidence must meet a certain *threshold of reliability* in order to have sufficient probative value to meet the criterion of relevance. *The reliability* of the evidence must also be considered with respect to the second criterion of necessity. After all, it could hardly be said that the admission of *unreliable* evidence is necessary for a proper adjudication to be made by the trier of fact.⁷³

The question of whether the prosecution’s expert evidence had been adequately verified emerged in *R. v. Ranger* (2003).⁷⁴ Referring to the *Mohan* criteria and *R. v. D.D.*, the Court of Appeal reiterated the link between relevance, necessity, and prejudicial effect, on the one hand, and reliability on the other:

The first two criteria [relevance and necessity] and the assessment of whether the probative value outweighs the prejudicial effect also *include an inquiry into the reliability of the proposed evidence.* In the case of novel scientific evidence, this latter inquiry is often critical.⁷⁵

The *Ranger* appeal canvassed additional issues. On the subject of expert evidence the Court cautioned: that juries might approach their assessment of expert evidence lacking the appropriate level of skepticism; that the cross-examination of experts might not be particularly effective; and, that the “significant costs in terms of time and money to the parties and strains upon judicial resources” created by expert evidence might outweigh any benefits. They concluded:

Those dangers must be considered in the *balancing process* that forms part of the test for admissibility. Further, the trial judge’s *gatekeeper function* does not end with the ruling on admissibility. The expert evidence must be carefully constrained in its presentation with a view

⁷³ (1999) 137 CCC (3d) 225 at [84]. (italics added)

⁷⁴ (2003) 178 CCC (3d) 375 at [32].

⁷⁵ (2003) 178 CCC (3d) 375 at [48]. (italics added)

to minimizing the associated dangers so that, in the end result, the judge is still satisfied that the probative value of the evidence exceeds its prejudicial effect and is properly admissible.⁷⁶

The Court explained how admissibility determinations pertaining to expert evidence may raise serious issues for the conduct and fairness of a trial.

In *R. v. Dimitrov* (2003), in the aftermath of *R. v. J.-L.J.*, the Court of Appeal embraced the *Daubert* criteria along with the need for the trial judge to “take seriously the role of gatekeeper.” Justices Weiler and Gillese continued:

Novel scientific theories or techniques are subject to “*special scrutiny*”; so, too, is the novel application of established or recognized scientific techniques. *The threshold question* that arises in relation to the admissibility of either is well-established: the court must be satisfied that the evidence proffered is capable of being the subject matter, that the proposed evidence is, indeed, “science”. The burden is on the party putting forth the expert, in this case the Crown, *to establish its reliability on a balance of probabilities.*⁷⁷

In *R. v. Klymchuk* (2005) the Court of Appeal excluded prosecution expert evidence on the basis that:

The Crown did not offer any evidence that Agent Bromley’s opinions ... had been or could be tested according to the generally accepted scientific methodology identified in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* 509 U.S. 579 (1993), and quoted with approval in *R. v. J.(J-L)* ...⁷⁸

Following *R. v. J.-L.J.*, the *Klymchuk* Court explained that the danger of juror overvaluation of expert evidence, from obviously well qualified experts, “animates the gatekeeper function.”⁷⁹ Similar concerns about the potential for expert evidence to mislead “in the sense that its effect on the trier is disproportionate to its reliability” were earlier canvassed in *Dimitrov*.⁸⁰

While most of this jurisprudence is expressly directed toward *novel* scientific theories and techniques or *novel* applications, there is authority from both the Supreme Court and the Ontario Court of Appeal suggesting that the reliability of all

⁷⁶ (2003) 178 CCC (3d) 375 at [62]-[63]. (italics added)

⁷⁷ (2003) 181 CCC (3d) 555 at [37]-[38]. (italics added)

⁷⁸ (2005) 203 CCC (3d) 341 at [36]. See also the argument in *R. v. Ho* (1999) 141 CCC (3d) 270 at [68].

⁷⁹ (2005) 203 CCC (3d) 341 at [54].

⁸⁰ (2003) 181 CCC (3d) 555 at [48]

expert evidence might, at least in theory, be of continuing concern. This issue was raised by the majority in *Trochym*.

Not all scientific evidence, or evidence that results from the use of a scientific technique, must be screened before being introduced into evidence. In some cases, the science in question is *so well established* that judges can rely on the fact that the admissibility of evidence based on it has been clearly recognised by the courts in the past. Other cases may not be so clear. Like the legal community, the scientific community continues to challenge and improve upon its existing base of knowledge. As a result, the admissibility of scientific evidence is not frozen in time.

While some forms of scientific evidence become more reliable over time, others may become less so as further studies reveal concerns. Thus, a technique that was once admissible may subsequently be found to be inadmissible.... Therefore, *even if it has received judicial recognition in the past, a technique or science whose underlying assumptions are challenged should not be admitted in evidence without first confirming the validity of those assumptions.*⁸¹

Subsequently, this paper will explore some of the practical implications raised in this passage in conjunction with the need, also expressed by the majority, to “establish that the underlying science” is “*sufficiently reliable* to be admitted in a court of law.”⁸²

The previous passage, as Justice Bastarache explained on behalf of the dissentients, raised doubts about the status of all expert evidence. For, notwithstanding the opening sentences, prior acceptance does not seem to provide any guarantee of admissibility. The dissenting judgment expressed some, perhaps justifiable, confusion at the majority’s exact position. They were concerned that the majority may have opened “well-established scientific methods” to unnecessary reassessment (see also Section 7).

Moreover Bastarache J. was apprehensive that the majority had “set down a rigid formula where the results must be proved beyond reasonable doubt before scientific evidence can be admitted.”⁸³

While my colleague [Deschamps J.] suggests that not all previously accepted scientific techniques will have to be reassessed under *J.-L.J.*, her 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

⁸¹ [2007] 1 SCR 239 at [31]-[32]. (italics added)

⁸² [2007] 1 SCR 239 at [33]. (italics added)

⁸³ [2007] 1 SCR 239 at [139].

previously accepted techniques being subject to challenge under *J.-L.J.* without establishing a serious basis for the inquiry.⁸⁴

While the standing of previously admitted (or non-novel) techniques and theories seems to be a point of some contention among the justices of the Supreme Court, the Ontario Court of Appeal had previously endorsed the need for continuing assessment. In *Terceira*, Finlayson J.A. indicated that

[i]t is therefore important to keep in mind that *the admissibility of expert opinion evidence is not a question of precedent*. Both general and case-specific appellate pronouncements respecting the admissibility of expert opinion evidence in similar cases *must always be considered in context*.⁸⁵

This approach received support, from a differently constituted bench, in *R. v. A.K. & N.K.* That Court explained “the state of scientific knowledge is fluid.... The fact that a particular theory may have been accepted in the past does not necessarily end the inquiry.”⁸⁶ There is, it would seem, a need for the trial judge—as gatekeeper—to maintain vigilance. The fact that a certain type of expert evidence has been admitted in the past might support a presumption in favour of admissibility but it cannot guarantee prospective admission.

Finally, in this review of expert evidence jurisprudence, a recent decision by the Ontario Court of Appeal is especially illuminating. *Re Truscott* (2007) was a ministerial reference in relation to the notorious case of Steven Truscott.⁸⁷ As a 14-year-old, Truscott was convicted for the 1959 murder of Lynne Harper. Initially he was sentenced to death but that sentence was subsequently commuted to life.⁸⁸ An appeal of the verdict was unanimously dismissed by the Ontario Court of Appeal in 1960 and the conviction was upheld by the Supreme Court of Canada in 1966. Truscott’s case became something of a *cause célèbre* and, after an inquiry by the Honourable Fred Kaufman, Q.C., which reported in 2004, was referred back to the Court of Appeal. In 2007, a panel of five judges heard the appeal, quashed the conviction, and formally acquitted Truscott.

⁸⁴ [2007] 1 SCR 239 at [139]. (references omitted)

⁸⁵ (1999) 137 CCC (3d) 225 at [75]. (italics added) Compare Laurens Walker and John Monahan, “Scientific Authority: The Breast Implant Litigation and Beyond” (2000) 86 *Virginia Law Review* 801–834.

⁸⁶ (1999) 137 CCC (3d) 225 at [86].

⁸⁷ *Re Truscott* [2007] ONCA 575.

⁸⁸ Though Truscott was released on parole in 1974.

The 2007 appeal might be considered revealing because the most persuasive and critical insights emerged out of a review of the evidence produced by the state forensic pathologist. New gastroenterological and etymological perspectives were used to challenge the time of death evidence, which seems to have been a central pillar in the original conviction. In reviewing the forensic evidence about digestion, the degree of rigor mortis, the extent of decomposition, and the rate of insect depredation, an unanimous Court of Appeal accepted the fresh expert evidence produced by the defence. In so doing, they dismissed the pathological evidence originally relied upon by the Crown as well as the expert evidence it assembled for the latest appeal.

The Court invested little confidence in the supplementary forensic pathological evidence 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.” Moreover,

[h]e refused to acknowledge obvious shortcomings in his opinion when these were pointed out in cross-examination. He refused to concede that his opinion rested on faulty assumptions and misperceptions of the available primary evidence in this case.⁸⁹

In a similar way the Court dispensed with the testimony of the Crown’s etymological expert, Dr. Haskell.

Several critical elements of his opinion were based on nothing more than his purported experience, which could not be verified and was not supported by any empirical work. He was unable to demonstrate that his experience had been replicated by other scientists.⁹⁰

And,

[h]e provided no scientific evidence to support this theory.⁹¹

Rather than “authoritative experience and anecdotal case reports,” this Court implicitly endorsed an “evidence-based approach” to expert evidence.⁹² Significantly,

⁸⁹ [2007] ONCA 575 at [166].

⁹⁰ [2007] ONCA 575 at [313].

⁹¹ [2007] ONCA 575 at [349].

the Court was interested in reliability and seemed to expect the Crown to support the evidence produced by its experts.⁹³ As the previous extracts indicate, anecdote, experience, and opinion were all characterized as insufficiently reliable. The Court was particularly attuned to evidence of experiment, testing, and 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 pathological evidence about the time of death "scientifically untenable."⁹⁴

Unlike the vast majority of criminal convictions, the *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 analysis and sustained public interest over so many decades."⁹⁵ Nevertheless, the criminal justice system seems to have encountered considerable difficulty correcting one of Canada's most prominent and longstanding 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 relevant and published empirical research. The challenge of overturning convictions, even those based on unreliable expert evidence, should not be underestimated.

In summary, concern with the reliability of expert evidence plays a conspicuous, if inchoate, role in the expert evidence jurisprudence of Canada. Appellate courts repeatedly refer to "reliability" or "sufficient reliability" without explaining what the terms mean or how they should be applied. Similarly, claims that novel techniques are subjected to "special scrutiny" seem exaggerated. Relying explicitly on *Daubert*, Canadian jurisprudence does not discriminate between the *Daubert* criteria or among other indicia of reliability. It invests confidence in acceptance as well as testing. With the emphasis on gatekeeping, it also makes the trial judge, as opposed to appellate courts, the main bulwark against prejudicial, unnecessary, irrelevant, and unreliable expert evidence.

Mohan represents what might be characterized as a fairly superficial approach to admissibility decision making. The Supreme Court seemed to be more concerned with

⁹² [2007] ONCA 575 at [169], [183], [371]-[372].

⁹³ [2007] ONCA 575 at [233], [307].

⁹⁴ [2007] ONCA 575 at [365].

relevance, necessity, and an expert's qualifications than the intricacies of reliability. To the extent that reliability emerged, it was part of a balancing exercise involving probative value and prejudicial effect. In subsequent decisions, reliability was only occasionally liberated from that role. Moreover, the Supreme Court appears divided on the question of whether the admissibility jurisprudence since *Mohan* should be restricted to novel techniques and applications or whether reliability might play a continuing role. Canadian courts have also encountered difficulty, or have simply been unwilling to consider, how the commitment to fairness along with the relative position of the state and the accused might structure the way they approach admissibility determinations. They have not adequately distinguished between the standards required of expert evidence adduced by the prosecution and the standard required by the defence. Encouragingly, though, such an approach may be implicit in the Ontario Court of Appeal's recent decision in *Truscott*.

While reliability has become more central to admissibility jurisprudence in recent years, especially after *R. v. J.-L.J., Trochym*, and *Truscott*, Canadian courts have not explained how indicia of reliability, like those from *Daubert* and elsewhere, should be weighted or applied. In addition, they have not explained if s.7 of the *Charter of Rights and Freedoms* provides protection from criminal prosecution based on expert evidence that is unreliable or of unknown reliability in circumstances where the validity and accuracy of the evidence might be readily ascertained.

As things stand a vague concept of reliability erratically impacts upon Canadian expert evidence jurisprudence and legal practice.

5. *Mohan Plus*: Making Reliability Explicit and Substantial

Since *Mohan*, references to reliability have become increasingly conspicuous in Canadian expert evidence jurisprudence. This paper aims to make this interest in reliability more conspicuous, more consistent, and more substantial. It is designed to reinforce the importance of reliability and to help clarify its meaning and application.

To secure a minimum guarantee on the quality of forensic medicine, science, and other forms of inculpatory expert evidence, it seems desirable to supplement Canadian admissibility jurisprudence with the formal expectation that the prosecution

⁹⁵ [2007] ONCA 575 at [71].

can only adduce expert evidence if it is shown to be reliable. In other words, expert evidence adduced by the prosecution should be *demonstrably reliable*. The practical upshot is that there should be evidence supporting the reliability of the state's expert evidence. Here attention is not restricted to novel scientific and medical evidence and novel applications but potentially applicable to all expert evidence adduced by the state. If wrongful convictions across the common-law world have demonstrated anything, it is that liberal admissibility standards and judicial complacency have enabled prosecutors to use (and continue to rely upon) expert evidence that is not reliable.⁹⁶ Failure to take the threshold of reliability and the exclusionary discretions seriously opens the possibility that investigations, prosecutions, and convictions will be predicated upon unreliable evidence or evidence of unknown reliability.

Consequently, to the *Mohan* criteria there should be added a formal expectation that expert evidence adduced by the state should be *demonstrably reliable*. The prosecution should be required to show, on the balance of probabilities, that the expert evidence it seeks to adduce is reliable. This is what is meant by *Mohan plus*—the *Mohan* criteria *plus* demonstrable reliability. Such a standard would make the role of reliability in Canadian expert evidence jurisprudence explicit and unambiguous. It would require judges to take the reliability of expert evidence seriously.

Ordinarily empirical testing will provide the most useful indication of reliability. Where a technique or theory is shown to be valid and the levels of accuracy are known, it will be relatively easy for a judge to determine its admissibility. In the absence of testing, judges will have to consider a range of alternative and generally less-conclusive factors. Testing should not be mandatory, but judges should inquire about the failure to test a technique or theory (see Sections 6 and 7). In the absence of testing judges ought to carefully scrutinize the prosecution's expert evidence and the reasoning purportedly supporting it (see Sections 7 and 8).

Distinguishing between the Prosecution and the Defence

It is useful to make differences between the admissibility standards applicable to the prosecution and defence explicit. The requirement that expert evidence should be

⁹⁶ Clive Walker and Keir Starmer (eds.), *Miscarriages of Justice: A Review of Justice in Error* (London: Blackstone Press, 1999); Richard Nobles and David Schiff, *Understanding Miscarriages of Justice: Law, the Media, and the Inevitability of Crisis* (Oxford: Oxford University Press, 2000).

demonstrably reliable should only apply in criminal proceedings and to evidence adduced by the prosecution.

Contemporary Canadian expert evidence jurisprudence applies, with minor qualifications, to the prosecution and defence. The state and the citizenry both have an interest in fair trials and accurate outcomes. However, the state is not in the same position as its citizens. As an exemplary litigant the state should lead by example. This paper contends that the highest admissibility standards for expert evidence should be applied to evidence adduced by the prosecution in criminal proceedings. Frequently, the defence will not be in a position to conduct testing, determine error rates, or publish the results of studies. The defence does not maintain its own experts or investigative institutions and laboratories. It does not routinely sponsor social, scientific, or biomedical research. In consequence, the defence should not be burdened with the same admissibility standards imposed upon the prosecution. Rather, the defence should only have to meet the kinds of standards currently required by the *Mohan* criteria. While trial judges should be reluctant to prevent an accused person from mounting a vigorous defence, they should not be completely indifferent to the reliability of the expert evidence adduced by the defence. After all, *Mohan* itself, in a manner that is difficult to fault, upheld the exclusion of exculpatory expert evidence.

Such a differentiated approach to admissibility might be productive of unfairness—to the state—if there were no safeguards. Perhaps the most practical response to a bifurcated admissibility standard would be to allow the prosecution to adduce rebuttal expert evidence, restricted to the topic or issue raised by the defence—which does not have to satisfy *Mohan plus*—where the accused adduces expert evidence of a kind that merely satisfies the four *Mohan* criteria. Relieving the prosecution of the need to demonstrate reliability *in order to respond to* expert evidence introduced by the defence would prevent the accused from obtaining a strategic or evidentiary advantage by adducing less reliable forms of expert evidence.

Requiring the prosecution to produce demonstrably reliable expert evidence is not intended to reduce the scope for cross-examination or discourage the defence from calling on rebuttal and other types of expert witness. Instead, imposing a demonstrable reliability standard is designed to provide a basis for substantial cross-examination and to illuminate expert disagreement. Merely passing the demonstrable reliability threshold is

not intended to privilege incriminating expert evidence, insulate the state's expert evidence from criticism, or as a guarantee of reliability.

Distinguishing between the Criminal and Civil Justice Systems

This juncture also provides an opportunity to distinguish between criminal and civil justice when thinking about admissibility standards. Onerous admissibility standards thwart the regulatory potential of the civil justice system. The imposition of a practically demanding admissibility standard in areas like tort or product liability may eviscerate substantial civil doctrines and prevent plaintiffs from recovering damages or accessing legal remedies.⁹⁷ Relevantly, plaintiffs, like criminal defendants, are not always in a position to produce demonstrably reliable expert evidence. Plaintiffs in toxic tort litigation, for example, rarely have the foresight to sponsor prospective epidemiological studies before they become ill.⁹⁸

Moreover, in theory, the state does not have the same interest in private disputes as it maintains in the criminal sphere. Here, it is useful to note, the U.S. Supreme Court decisions in *Daubert*, *General Electric Co. v. Joiner* (1997), and *Kumho Tire v. Carmichael* (1999) were all civil cases where the plaintiffs' expert evidence was excluded—in *Daubert* the evidence was excluded by the Ninth Circuit Court of Appeal on remand.⁹⁹ In a curious and undesirable inversion of public policy, U.S. federal judges have occasionally experienced difficulty applying the admissibility standards (i.e., the four *Daubert* criteria) they routinely invoke in civil litigation to expert evidence adduced by the state in criminal prosecutions.¹⁰⁰ Such reticence might be considered disconcerting, especially when the architect of the *Daubert* decision made a principled civil–criminal distinction in the context of an earlier capital appeal:

One may accept this in a routine lawsuit for money damages, but when a person's life is at stake—no matter how heinous his offense—a requirement of greater reliability should prevail.

⁹⁷ Lloyd Dixon and Brian Gill, *Changes in the Standard for Admitting Expert Evidence in Federal Civil Cases Since the Daubert Decision* (Santa Monica: Rand Institute for Civil Justice, 2001); SKAPP, *Daubert: The Most Influential Case You've Never Heard Of* (Tellus Institute, 2003); Carl Bogus, *Why Lawsuits Are Good for America* (New York: New York University Press, 2003); William Haltom and Michael McCann, *Distorting the Law: Politics, Media, and the Litigation Crisis* (Chicago: University of Chicago, 2004).

⁹⁸ Carl Cranor, *Toxic Torts: Science, the Law and the Possibility of Justice* (New York: Cambridge University Press, 2006).

⁹⁹ *General Electric Co. v. Joiner* 522 US 136 (1997); *Kumho Tire v. Carmichael* 526 US 127 (1999).

¹⁰⁰ *Daubert v. Merrell Dow Pharmaceuticals, Inc.* 43 F.3d 1311 at 1317–1318, especially n5 (9th Cir. 1995); *United States v. Llera Plaza I* 179 F. Supp. 2d 492 (E.D. Pa. 2002); *United States v. Llera Plaza II* 188 F. Supp 2d 549 (E.D. Pa. 2002). For some explanation of this approach consider: Gary Edmond, "Legal Engineering:

In a capital case, the specious testimony of a psychiatrist, colored in the eyes of an impressionable jury by the inevitable untouchability of a medical specialist's words, equates with death itself.¹⁰¹

Public–Private Realms: Expert Evidence and Reliability in Child Protection

The standards for expert evidence employed in other types of public–private (or quasi-public) legal processes, such as child protection proceedings, require principled reflection. The interests of children, parents, the state, and the public need to be balanced against the reliability of expert evidence, precautionary orientations and, in many instances, an express intent to relax formal rules of evidence and procedure.

Child protection proceedings represent an area with which I am not particularly familiar. My impression is that where particular techniques, theories, and opinions are routinely employed by an agency, then the agency proffering (or relying upon) the evidence as well as the institution receiving or reviewing it should make serious attempts to ascertain the reliability of the techniques, theories, and opinions. To the extent that the expert opinions are not based on demonstrably reliable techniques, then the relevant institutions should certainly treat the evidence with caution. As in the criminal sphere, the seriousness of any allegations (or suspicions) and the centrality of the expert evidence should not rise above concerns about reliability.

To the extent that an agency or tribunal has a precautionary (i.e., risk averse) mandate, then that may lower the minimum standard for expert evidence to be admitted or relied upon. The child protection context, by changing the liability rules, may widen the scope of inquiry in a manner that impacts on the expert evidence. Questions of causal responsibility, for instance, may be less important in child protection than criminal proceedings. Conversely, findings of injuries that did not necessarily cause death may be probative of risk of abuse. Nevertheless, we should not forget that unreliable expert evidence cannot assist with decision making even in circumstances where children appear to be at serious risk of harm or abuse. More difficult questions emerge in relation to expert evidence of unknown reliability. Here, though, once again, the question of whether such evidence should found or support administrative decisions and interventions deserves serious and explicit consideration.

Contested Representation of Law, Science (and Non-Science) and Society" (2002) 32 *Social Studies of Science* 371–412.

Given the emotive nature of many of these proceedings and the serious implications for children *and* parents, it may be appropriate for forensic pathologists, social workers, psychologists, psychiatrists, pediatricians, and tribunals to adopt a reliability approach. Suspicions and experience may be useful guides to further inquiry, but intervention should, to the extent that it incorporates expert evidence, be based on some minimal standard or reliability. Whether this should be demonstrable reliability (on the balance of probabilities) or some alternative standard depends upon the social and policy values motivating the legislation, the agency, and the style of proceedings (and review). Lower standards of reliability will have a more precautionary effect but they may also remove children or encourage intervention in situations where there is no really compelling evidence of abuse or mistreatment and certainly no expert evidence that might be capable of supporting a criminal prosecution.¹⁰²

A low reliability threshold will generate false positives and a higher standard will produce false negatives. Where the standards should be placed are ultimately policy choices and these decisions along with their rationales should be made explicit.

6. *Daubert* and Its Reliability Criteria

As the Canadian Supreme Court indicated in *R. v. J.-L.J.* and more recently in *Trochym*, and various Ontario Court of Appeal decisions have confirmed, the *Daubert* criteria offer “help” in assessing the reliability of expert evidence. *Daubert* provides a range of resources that might be *selectively* invoked when determining whether expert evidence is sufficiently reliable for legal purposes.

The four *Daubert* criteria represent an eclectic assortment of epistemological, philosophical, and sociological approaches to science and expertise.¹⁰³ According to the majority:

Ordinarily, a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact will be whether it can be (and has been) tested. “Scientific methodology today is based on generating hypotheses and testing them to see if they can be falsified; indeed, this methodology is what distinguishes science from other fields of human inquiry.”... See also C. Hempel, *Philosophy of Natural Science* 49 (1966) (“[T]he statements

¹⁰¹ *Barefoot v. Estelle*, 463 US 880 at 916 (1983), Justice Blackmun dissenting.

¹⁰² Ian Hacking, *The Social Construction of What?* (Cambridge, MA: Harvard University Press, 1999) 125–162.

constituting a scientific explanation must be capable of empirical test”); K. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge* 37 (5th ed. 1989) (“[T]he criterion of the scientific status of a theory is its falsifiability, or refutability, or testability”) (emphasis deleted).

Another pertinent consideration is whether the theory or technique has been subjected to peer review and publication. Publication (which is but one element of peer review) is not a *sine qua non* of admissibility; it does not necessarily correlate with reliability.... The fact of publication (or lack thereof) in a peer reviewed journal thus will be a relevant, though not dispositive, consideration in assessing the scientific validity of a particular technique or methodology on which an opinion is premised.

Additionally, in the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error....

Finally, “general acceptance” can yet have a bearing on the inquiry. A “reliability assessment does not require, although it does permit, explicit identification of a relevant scientific community and an express determination of a particular degree of acceptance within that community.”¹⁰⁴

The *Daubert* criteria provide neither an accurate (or even coherent) characterization of science, medicine, and expertise, nor an especially neat solution to issues of reliability.

At a practical level, however, the four criteria, and especially testing, provide serviceable resources for approaching the question of *legal reliability*. Unlike the mandatory *Mohan* criteria—concerned with the *admissibility* of novel expert evidence—the *Daubert* criteria are specifically focused on the *reliability* of expert evidence. Originally they were intended, or so it would seem, as a set of resources to be applied flexibly by federal judges. In practice, many federal (and state) courts have approached and used the *Daubert* criteria as an inflexible checklist.

Significantly, the criteria are not equally important or discriminating in relation to reliability. Whether expert evidence is sufficiently reliable for legal purposes will depend on the amount of supporting evidence associated with a technique, theory, or opinion. Testing is the most important of the *Daubert* criteria. Whether a technique or theory has survived some kind of testing is particularly important in the context of forensic scientific and medical evidence. Empirical studies of forensic techniques provide very useful resources for making reliability assessments, ascertaining levels of accuracy, and determining practitioner competence. To the extent that techniques and theories have not been thoroughly tested, their validity and accuracy are simply

¹⁰³ Gary Edmond and David Mercer, “Conjectures and Exhumations: Citations of History, Philosophy and Sociology of Science in US Federal Courts” (2002) 14 *Law & Literature* 309–366.

not known. The gravity of this point warrants repetition. In the absence of rigorous empirical testing, the reliability of techniques and theories is uncertain.

Error rates, where they have actually been ascertained, are also important. They indicate that there has been some investigation or testing of a technique, theory, or set of practitioners. Similarly, the absence of testing and information on error rates may also be significant. In the absence of testing, claims about rates of error and expressions of confidence may be worse than useless. Typically, as many public inquiries into wrongful convictions attest, forensic scientists and technicians tend to overestimate—that is, provide inexplicably high expressions of—the accuracy of their techniques, theories, and conclusions. Such confident expressions are frequently presented in circumstances where there has been no testing but in relation to techniques that could, quite readily, be tested. Error rates reinforce the primacy of testing. Where error rates are unknown, then the reliability of the technique and any claims about accuracy are mere conjecture.

The remaining *Daubert* criteria are usually of less import 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 authoritative textbooks, (inadmissible) hearsay, or just speculation. Sometimes it is a combination of these sources. In practice, research is rarely conducted into the distribution of expert opinion. If information on the extent of acceptance was readily available, it would reveal more about the field and orientations within the field than about reliability per se. General acceptance is, perhaps, especially weak in areas such as forensic science and forensic medicine where fields and subspecialization tend to be small and relatively insulated from more mainstream biomedical and scientific research, and the practitioners maintain close and ongoing contacts with police, investigative agencies, and prosecutors.

Similarly, peer review and publication are not particularly good surrogates for reliability. Peer review and publication have a range of meanings and uses, and consequently their value can vary quite dramatically. 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 possess very limited probative value. Alternatively, where peer review involves the testing of new

¹⁰⁴ *Daubert v. Merrell Dow Pharmaceuticals, Inc.* 509 US 579 at 593-94 (1993).

(or older) techniques by members of exogenous specialist communities, and is published in mainstream scientific journals, it should carry considerably more weight. The fact that a technique or theory is mentioned in the literature, however, will generally be less significant than the fact that a technique or theory has survived rigorous testing.

In combination the *Daubert* criteria impose a practically demanding admissibility standard. For this reason they are far better suited to assessing the reliability of expert evidence adduced by the state than evidence assembled by individual and often impecunious defendants (or plaintiffs). Elsewhere, I have characterized the inflexible application of all four of the *Daubert* criteria as *hard Daubert*.¹⁰⁵ Though onerous, the combination of these criteria makes it more likely than any of the criteria individually that a technique or theory is reliable. Such an admissibility standard would seem to be a reasonable requisite for forensic scientific and medical evidence and techniques in regular use—such as identifications from latent fingerprints or DNA analysis.¹⁰⁶ Techniques and theories used routinely should only be admissible if there *has been* extensive testing and the results of that testing *has been* published in peer-reviewed biomedical or scientific journals. As things stand, the institutionalized forensic sciences do not always test their techniques, sometimes rely upon impoverished versions of peer review, and appeal to highly parochial versions of acceptance.

Trial and appellate judges should be vigilantly seeking indicia of reliability. In the absence of testing, demonstrations of reliability will not always be convincing. Where quantified error rates are not available, judges might consider indicia such as publication and professional acceptance (where it is known or can be persuasively demonstrated). In the absence of credible support expert techniques, theories and opinions are mere speculation and should not be relied upon for criminal prosecutions.

7. Testing

¹⁰⁵ Gary Edmond, “*Supersizing Daubert: Science for Litigation and Its Implications for Legal Practice and Scientific Research*” (2007) 52 *Villanova Law Review* (forthcoming).

¹⁰⁶ Mike Lynch, “Science Above All Else: The Inversion of Credibility between Forensic DNA Profiling and Fingerprint Evidence” in Gary Edmond (ed.), *Expertise in Regulation and Law* (Aldershot: Ashgate, 2004) 121–135.

Where many ... forensic science techniques are concerned ... scientists outside the forensic science community are neither interested nor knowledgeable enough to scrutinize the techniques. The claims of handwriting experts, forensic odontologists, and experts on hair and voice identification simply do not interest most scientists, and have been subjected to little empirical validation. Yet within their own domains, these techniques are generally accepted. This is one of the principal failings of *Frye*, and is one of the reasons why it has not prevented dubious evidence from being admitted in United States courts. A testedness requirement would, in theory, do a far better job of screening out unreliable evidence.¹⁰⁷

So far, this paper has presented testing as a particularly useful resource and perhaps the only reliable means of ascertaining the reliability of expert evidence. Notwithstanding this approach, it would be misleading not to acknowledge that testing has limitations and that competent experts regularly disagree over the adequacy of tests or what particular tests demonstrate.¹⁰⁸ It should not be thought that testing will provide some kind of evidentiary or admissibility panacea.¹⁰⁹ It is always possible to challenge a testing regime or attempted replication.¹¹⁰ Nevertheless, on average, it is far better for investigators to have seriously tested, or attempted to test, their techniques, assumptions, and abilities. If nothing else, these efforts embody desirable normative traits such as skepticism, independence, and commitment to the accuracy of verdicts. They may also enable the defence to identify the limits of scientific opinion. It will, again on average, be easier for a judge to accept that expert evidence is relevant, necessary, probative, and reliable where a technique or theory has survived rigorous testing.

It is also important to distinguish between empirical testing and falsification (discussed in relation to *Daubert*, above). As a guide to scientific practice, falsification has encountered insuperable difficulties. Rather than engage in philosophy of science or abstract debates about falsification, it is preferable for judges to approach reliability in more pragmatic ways.¹¹¹ Rather than attend to philosophical

¹⁰⁷ Mike Redmayne, *Expert Evidence and Criminal Justice* (Oxford: Oxford University Press, 2001) 116.

¹⁰⁸ Adwina Schwartz, "A 'Dogma of Empiricism' Revisited: *Daubert v. Merrell Dow Pharmaceuticals, Inc.* and the Need to Resurrect the Philosophical Insight of *Frye v. United States*" (1997) 10 *Harvard Journal of Law and Technology* 149–237; Gary Edmond and David Mercer, "Recognising *Daubert*: What Judges Should Know about Falsificationism" (1996) 5 *Expert Evidence* 29–42.

¹⁰⁹ This is why demonstrable reliability is important as an admissibility threshold and why cross-examination and expert disagreement should be accommodated.

¹¹⁰ Collins, *Changing Order* (1986).

¹¹¹ Brian Leiter, "The Epistemology of Admissibility: Why Even Good Philosophy of Science Would Not Make for Good Philosophy of Evidence" (1997) *Brigham Young University Law Review* 803–19. See also Trevor Pinch, "'Testing—One, Two, Three ... Testing!': Toward a Sociology of Testing" (1993) *Science, Technology & Human*

subtleties, judges should consider whether the technique and theories have been subjected to some kind of empirical test and, if so, what the results were. They should be interested in limitations to testing and what might be considered best practice for similar sorts of tests. They should also incorporate their own common sense to identify obvious weaknesses in the way tests have been framed, conducted, and represented. The invocation of scientific method doctrines and casting of empirical investigations as formal attempts at disproof should not become prerequisites to determinations of legal reliability. Instead, questions of admissibility and reliability should be focused on the more fundamental and legally significant question of whether the expert evidence is demonstrably reliable.

Notwithstanding potential difficulties and controversy around the adequacy and meaning of testing, there are several things we can say. First, most of the techniques and theories relied upon by forensic science and medicine can be tested. In consequence, the absence of testing—and statistical information about error rates, levels of confidence, and individual competence—will often be revealing and, in some instances, damning. The state's failure to test techniques and theories, especially when they could be readily tested, should lead to adverse inferences, exclusion, and, if appropriate, judicial censure. Courts should not condone disinterest in testing by readily admitting testable but untested techniques and theories.¹¹²

Where forensic techniques and theories have undergone some form of testing, then judges should consider that process as part of their admissibility determination. Judges should be guarded where testing is perfunctory, restricted, undertaken entirely in-house, or the results withheld. The more rigorous the testing, the more likely any results will be trustworthy. Where testing is extensive, realistic, independent, public, multidisciplinary, and includes genuine possibilities to show that a technique may not work or to identify its limitations, then judges should be more favourably disposed to positive results. Where techniques and theories are in regular use, lack of rigorous testing and uncertainty about error rates may reveal much about experts and their institutions. Judges should be inclined, and encouraged to ask: Where are the studies supporting these techniques?; and What is the error rate associated with this technique and this particular expert?

Values 25–41; Harry Collins and Trevor Pinch, *The Golem at Large* (Cambridge: Cambridge University Press, 1998).

¹¹² *Re Truscott* [2007] ONCA 575.

Obviously, there are practical and ethical constraints on some kinds of empirical testing. Few of us would want to be stabbed to aid the advancement of forensic medicine. In the absence of testing—including circumstances where testing would be difficult or impossible—judges are confronted with serious dilemmas. Should they allow the prosecution to adduce the untestable opinions of ostensibly independent scientists and pathologists or should they exclude the evidence? This issue may have particular salience in relation to forensic pathology. The inability to conduct clinical trials or double-blind forms of testing does not mean that testing should have no place in forensic science and medicine. My contention is that judges should carefully scrutinize untested expert evidence with a willingness to exclude it. Where it is extremely difficult (or impossible) to undertake meaningful testing *and* the technique, theory, or opinion is obviously based in some published—preferably authoritative and widely accepted—research, then a judge might admit untested expert evidence.

Techniques and theories developed spontaneously in response to the exigencies of particular cases need to be demonstrably reliable. In such circumstances reliability might be supported by grounding the claims in authoritative literatures and research as well as through testing. In circumstances where the issues are idiosyncratic—that is, highly case-specific—there may still be scope for testing. There should always be some evidence that clearly and unequivocally supports the techniques, theories, and expert opinions adduced by the prosecution. In approaching expert evidence developed in highly unusual circumstances, judges should attend to the reliability of the evidence, particularly the basis of the opinion, the way it will be expressed, as well as the ability of the accused to credibly challenge it. In these circumstances it might be especially important to ensure that the defence has adequate resources to evaluate the incriminating expert evidence.

The difficulties facing the defendant in terms of credibly challenging untested expert evidence based on experience or purported widespread acceptance should not be underestimated. One English evidence commentator, Mike Redmayne, puts this forcefully when he suggests that

[u]nless the information gained through testing is available, a jury will often be left in the position of having to defer blindly to an expert's claims. Most forensic science techniques can be tested without undue difficulty, and the state is in an ideal position to perform tests.... To put

the point more bluntly: if the state does not test the scientific evidence with which it seeks to convict defendants, it should forfeit the right to use it.¹¹³

To the extent that the state has not endeavoured to subject the techniques and opinions developed by its experts to some form of robust scrutiny, judges are entitled to respond skeptically.

In *Trochym*, the dissentients expressed concern that the majority had “set down a rigid formula where the results must be proved beyond reasonable doubt before scientific evidence can be admitted.”¹¹⁴ They thought that Justice Deschamp’s reliability standard “really” required a “total consensus by members of the scientific community.” Such a reaction overstates the position advanced in this paper. Requiring that the Crown demonstrate the reliability of its expert evidence in criminal cases should not be confused with requiring it to prove the reliability of such evidence beyond reasonable doubt. It is essential to distinguish between the need to prove that scientific and medical evidence should be established beyond reasonable doubt and the more modest requirement that the prosecution demonstrate some evidentiary foundation, preferably based on empirical validation. The prosecution need only satisfy a court about the reliability of their expert evidence on the balance of probabilities. They will have to produce tightly focused evidence, preferably evidence of testing, that supports the reliability of their techniques, theories, and opinions. That, however, is very different from requiring certainty or proof beyond reasonable doubt.¹¹⁵

In practice, judges will be confronted with diverse evidentiary arrays designed to support the reliability of forensic scientific and medical evidence. They will be confronted with complex decisions about what testing establishes, whether evidence is demonstrably reliable, and what to do with techniques (or technicians) with accuracy rates that are not high. Evidence of testing or the lack of testing will be combined with expert opinions, references to qualifications, authoritative texts, unpublished research, determinations from other courts, and the endorsement of experts and judges from foreign jurisdictions.

¹¹³ Redmayne, *Expert Evidence and Criminal Justice*, 139.

¹¹⁴ [2007] 1 SCR 239 at [139].

¹¹⁵ Not every piece of evidence presented by the prosecution needs to be established beyond reasonable doubt, even though at the end of the day guilt must be proven beyond reasonable doubt. See *R. v. Morin* [1988] 2 SCR 345.

8. Weaker (or Supplementary) Indicia of Reliability

Typically, the most important evidence of reliability will be empirical validation. This refers not to the use of forensic science and medicine in successful prosecutions but to formal tests of techniques and theories in circumstances where the correct answer is known, so that their validity and accuracy can be meaningfully gauged.¹¹⁶

Publication, peer review, and claims about acceptance provide indirect—and usually weaker—evidence of reliability. In the absence of testing, peer review, publication, and acceptance are decidedly more fragile.

In addition to testing there are many factors that might bear upon a trial judge or appellate court's assessment of reliability. The following list provides a sample of these indicia:

What is the error rate—for the technique, as well as the equipment and practitioner?

Has the technique or theory been applied in circumstances that reflect its intended purpose or known accuracy? Departures from established applications require justification.

Does the technique or opinion use ideas, theories, and equipment from other fields? Would the appropriations be acceptable to those in the primary field?

Has the technique or theory been described and endorsed in the literature? This should include some consideration of where and by whom and with what qualifications.

Is the reference in the literature substantial or incidental? Is it merely the author's opinion or something more?

Has the publication, technique, or opinion undergone peer review? Logically, peer acceptance of techniques and theories should take priority over peer review of individual results or applications. Where the reliability of a technique is unknown, positive peer review may be (epistemologically but not sociologically) meaningless.

Is there a substantial body of academic writing approving the technique or approach?

¹¹⁶ Compare *United States v. Haavard* 117 F. Supp 2d 848 (S.D. Ind 2000). Similarly, quality assurance programs and the peer review of results should not count as testing.

To what extent is the technique or theory accepted? Is the technique or theory only discussed in forensic scientific and forensic medical circles? In assessing the extent of *acceptance*, the judge should consider what evidence supports acceptance—opinions based on personal impression or hearsay and incidental references in the relevant literature may not be enough to support claims about wide acceptance. The fact that support comes from earlier judgments rather than scientists or scientific, technical, and biomedical publications will usually be significant.

Is the expert merely expressing a personal opinion (*ipse dixit*)? To what extent is the expert evidence extrapolation or speculation? Is the expert evidence more than an educated guess? Is this clearly explained?

Does the expert evidence actually form part of a field or specialization? Judges should not be too eager to accept the existence of narrow specializations or new fields based on limited research and publication.

Does the evidence go beyond the expert's recognized area of expertise?

In determining the existence of a field or specialization, it may be useful to ascertain whether there are practitioners and experts outside the state's investigative agencies. If so, what do they think?

Is the technique or theory novel? Does it rely on established principles? Is it controversial?

Is the evidence processed or interpreted by humans or machines? How often are they tested or calibrated?

Does the evidence have a verification process? Was it applied? Were protocols followed?

Is there a system of quality assurance or formal peer review? Was it followed?

To what extent is the expert evidence founded on proven facts (and admissible evidence)?

Has the expert explained the basis for the technique, theory, or opinion? Is it comprehensible and logical?

Has the expert evidence been tainted or influenced by inculpatory or adverse information and opinions? Did the expert have close contact with the investigators or were they formally and substantially independent?

Has the expert made serious mistakes in other investigations or prosecutions? Has the expert been subjected to adverse judicial comment?

Does the expert invariably work for the prosecution (or defence)?

Are the techniques or conclusions based on individual case studies or more broadly based and statistical approaches such as epidemiology and meta-analysis?

How confident is the expert? Does the expert express high levels of confidence or quantify certitude in the absence of validation and accuracy studies? Is this a feature of his or her regular practice?

Is the expert willing to make concessions?

How extensive is the expert's education, training, and experience? Are they directly relevant? Judges should look at overall training and experience and, in an age of increasing specialization, not be too eager to allow individuals who are not the most appropriate experts to testify.

Does the expert have a financial interest in the evidence or technique? This extends beyond employment to issues of intellectual property, proprietary interests, managerial roles, and shareholding. *Conflicts of interest* should be disclosed so they can be factored in to assessments of admissibility and weight.

This long list of supplementary indicia of reliability is designed to provide judges, and reformers, with a sense of the many dimensions to expert evidence and reliability. It is far from comprehensive and is certainly not intended as a checklist. As we saw at the beginning of this paper (in Section 2), expertise is too variegated and complex to be subjected to simple categorization or algorithms. The degree of detail should indicate that sometimes superficial scrutiny, such as considering whether a technique or theory has been mentioned in the literature, might not be enough to demonstrate reliability and sustain admissibility. These indicia are advanced because they may provide practical assistance in the determination of *legal reliability*.

The more indicia we include, the harder it will be to satisfy them all. That said, on average, the more indicia satisfied, the more reliable the technique or theory will be. But again, it is important to stress that not all criteria are equally important or discriminating. Generally, whether a technique or theory has been tested will be the most fundamental and most important factor in any assessment of reliability. Validity,

in conjunction with high levels of accuracy, based on competent testing, will normally be sufficient to establish legal reliability.

Judges should be reticent in using these (and other) supplementary indicia to overcome a lack of testing. They should inquire about the failure to test and not simply excuse such failures because the inculpatory expert evidence is important, or vital, to the prosecution's case. Where rigorous empirical studies have been undertaken, the results of these studies will tend—though not invariably—to outweigh the other indicia of reliability. Ordinarily, the results of rigorous empirical testing should be preferred to other evidence no matter how prevalent the view, no matter how authoritative the expert, or how counterintuitive the result. Without more, the fact that a technique or theory has been used by a forensic community for decades and previously admitted into trials will rarely provide a persuasive basis to resist adverse results from validation and accuracy studies.

The meaning and significance of the supplementary indicia can be quite complex. They are, perhaps, more useful as exclusionary resources than positive indicia of reliability. If, for example, an expert relies on a technique that has not been studied, is not discussed in authoritative writings, and is not widely accepted by a relevant community of experts, then it would seem difficult for the evidence to pass the legal threshold for reliability. Things start to get more complicated when evidence of testing is combined with the supplementary indicia. Where, for example, an expert has been actively involved in the development and commercialization of an investigative technique, the fact that they hold a financial interest might be used to impugn the results of any testing—especially if the testing was conducted personally and the details of the study are not disclosed.¹¹⁷ Alternatively, evidence derived from techniques that had purportedly survived testing might still be challenged on the basis that protocols were not followed or the technique was applied in a way that went beyond what the available testing could legitimately support. The fact that techniques, theories, and opinions have been tested will not always conclude the inquiry into legal reliability.

In the absence of evidence of reliability, senior judges should be willing to exclude expert evidence adduced by the state, whether scientific, technical, or medical.

¹¹⁷ See, for example, the discussion of Dr. Sutisno's facial and body identification techniques in: *R. v. Tang* [2006] NSWCCA 167; *R. v. Jung* [2006] NSWSC 658; *R. v. Kaliyanda* (17 October 2006) NSWSC and *R. v. M.* (14 September 2005) NSWDC (unreported).

Confidence in the province's forensic science laboratories and police service are not substitutes for evidence of reliability. It is important to remember that judges should be concerned with evidence of the reliability of particular techniques and theories, not evidence of the eminence of scientists, their performance and credibility, their impressive credentials, their past successes, or the reputation of their institutions. Credentials, training, authority, and experience are all unreliable predictors of reliability. Trial and appellate judges should ask: Where is the evidence that suggests, on the balance of probabilities, that this technique or theory is valid and this particular application accurate?

In practice, making admissibility determinations using *Mohan plus*—that is, based upon the *Mohan* criteria plus demonstrable reliability—will be a complex and demanding activity. Unavoidably, these complexities are part of the difficult work and responsibilities associated with being a judge.

9. Language: Expressions of Confidence and (Un)certainty

One of the more prominent features of the Canadian experience with expert evidence in wrongful convictions concerns the language used by forensic scientists, pathologists, technicians, and police. Several public inquiries into wrongful convictions have expressed concerns about levels of confidence and phrases such as “to a medical certainty,” that a person “cannot be excluded” from a relevant group, that a specimen “could have” originated from a particular source, and that a sample represents a “match.”

In the context of this paper, with its emphasis on demonstrable reliability, the language employed by expert witnesses is particularly important and, arguably, revealing. Here, once again, the significance of testing is pronounced. Forensic scientists, technicians, and pathologists cannot provide credible probability estimates—in any terms—if they have not conducted tests or if there are no empirical foundations for their claims. To the extent that there is no testing or a narrowly identified basis (for an opinion or expression) in authoritative scientific and medical literatures, ideas about accuracy and certitude would seem to be mere speculation. The language employed by forensic scientists should be linked to testing, underlying frequencies, and publicly accessible data sets. Where available, the results of testing assist with the selection of appropriate terminologies and expressions. It is the failure

to have tested or empirically grounded the expert evidence that usually creates the problems.

Moreover, in the absence of testing, even purportedly neutral expressions are potentially misleading. The phrase “may or may not,” for example, implies that a technique or theory has some validity as well as providing an expression of confidence. Even though such expressions are preferable to many extant formulations, in the absence of evidentiary support, they may suggest that an opinion or technique is more reliable or impressive than it is known to be.¹¹⁸ In its more neutral guise, it may raise guilt as a real possibility even though the evidence does not, or cannot, adequately support such a conclusion. Debates over expressions of confidence are less important than the validity of the underlying technique or theory.

The dangers inherent in exaggeration and misrepresentation should only encourage judges to require evidence of testing and empirically derived rates of error. Commenting on hair microscopy in the Driskell Inquiry, Commissioner Lesage, Q.C, indicated that purportedly “scientific’ evidence should not be presented in criminal trial as probative on the issue of identity unless this conclusion has a strong empirical and/or theoretical foundation.”¹¹⁹ For the judge, the first thing to do when confronted with probabilistic expressions is to ask: What evidence supports such a claim or level of confidence? If judges are attentive to reliability, they will sometimes be dissatisfied with the technique or theory regardless of the way the results are expressed.

Without more, expressions of confidence and especially high levels of certitude are merely *ipse dixit*. Restricting assertions of confidence, where there are no rigorous empirical studies, might prevent forensic scientists with pro-prosecution sympathies from using prejudicial terminologies in ways that illegitimately assist the prosecution case.

Significantly, levels of certitude are often surreptitiously indexed to informal information—which may not be admissible or reliable. Such information might include beliefs among investigators, remote hearsay, knowledge of prior convictions or criminal histories, and other investigative biases. Concerns about the misleading

¹¹⁸ Brian Campbell, “Uncertainty as Symbolic Action in Disputes among Experts” (1985) 15 *Social Studies of Science* 429–53; Susan Leigh Star, “Scientific Work and Uncertainty” (1985) 15 *Social Studies of Science* 391–427.

¹¹⁹ *Report of the Commission of Inquiry into Certain Aspects of the Trial and Conviction of James Driskell* (2007) 149, 172; *Report of the Kaufman Commission on Proceedings Involving Guy Paul Morin* (1998) 340–341; *R. v. Bennett* (2003) 179 CCC (3d) 244. (Ont.C.A.).

impression created by internally referencing interpretations of evidence were expressed by the majority in *Trochym*:

[I]f evidence whose reliability cannot really be tested is admitted and relied upon simply because it is consistent with other admissible evidence, the danger is that a web of consistent but unreliable evidence will lead to a (potentially wrongful) conviction.¹²⁰

Consequently, expressions of confidence that are not restricted to efficacious techniques may place contamination upon contamination. Forensic scientists are often intimately, and probably unavoidably, involved in the investigation and prosecution of crimes. While inside information might, quite properly, be used to assist in the investigation of crime, the very same information may simultaneously, and sometimes indirectly or unconsciously, contaminate the expert evidence. Judges (and jurors) should be careful not to mistake such contamination for independent corroboration. Assessing the admissibility of expert evidence—especially the reliability of techniques, theories, and opinions—independently of the other inculpatory evidence will help to reduce cross-contamination.

It is always desirable for expert witnesses to express their opinions in the most neutral or evidence-based manner possible. In practice it can be difficult to control the way expert witnesses actually testify in court or to control the overall impression they convey, even when they use relatively neutral expressions. To minimize dangers, the language used by expert witnesses should always be subservient to questions of reliability and admissibility.

10. Taking Reliability Seriously: The Many Advantages of *Demonstrable Reliability*

This section reviews some of the social, institutional, and logistical benefits of making demonstrable reliability a central feature of expert evidence jurisprudence. The first and most obvious benefit of requiring the prosecution to demonstrate that its expert evidence is reliable is that, to the extent judges take their gatekeeping responsibility seriously, the kinds of techniques, theories, and opinions that have contributed to wrongful convictions in Canada and elsewhere are far less likely to

enter courtrooms and contaminate criminal trials. The exclusion of expert evidence that is not demonstrably reliable makes convictions based on unreliable expert evidence less likely. More legitimate verdicts and enhanced public confidence in the courts are two very important benefits.

One of the institutional advantages flowing from an explicit reliability standard is that it insulates the courts. It prevents communities or cliques of experts from subverting legal processes and, to the extent exercised, places responsibility for expert evidence, especially the lack of admissible expert evidence, upon the state. Historically, the failure to scrutinize inculpatory expert evidence has made courts complicit in wrongful convictions. Imposing a reliability standard will help to extricate judges from responsibility for wrongful convictions, enable the courts to regulate their own processes, and prevent police, investigators, and experts from presenting unfounded claims, educated guesses, speculation, and unadulterated prejudice as credible scientific or medical knowledge.

Imposing an explicit reliability standard on expert evidence adduced by the prosecution reinforces the important, if neglected, role of courts in shaping and holding forensic science and medicine to account. There is little doubt that admissibility standards help to shape forensic scientific and medical practice.

Courts, it is clear, hold the lever that controls whether or not ... research will occur. As long as courts continue to admit [forensic] evidence, government needs are satisfied, and research dollars are unlikely to flow. Historically, courts created the remarkable situation in which we now find ourselves—not knowing what level of confidence to place in forensic techniques which are used daily in legal proceedings.¹²¹

Judges and courts play an important and constitutive role in the standards used in forensic scientific and medical practice. Judges should not take the reliability of evidence generated by the institutionalized forensic sciences on trust. Instead, they should be guided by lawyers and experts in their attempt to ascertain whether the expert evidence adduced by the state is relevant, necessary, and reliable. Imposing a genuine reliability threshold on the state will have appreciable institutional and

¹²⁰ [2007] 1 SCR 239 at [60].

¹²¹ Simon Cole, "Grandfathering Evidence: Fingerprint Admissibility Rulings from *Jennings* to *Llera Plaza* and Back Again" (2004) 41 *American Criminal Law Review* 1189–1276, 1216.

professional ramifications on institutionalized forensic science and medicine and the evidence produced during the investigation of crime.

Another of the advantages with *reliability* is that it provides a flexible legal standard that circumvents the need for legally trained and generalist judges to engage with abstract models of science or philosophical debates. In their everyday practice, there is no need for judges to become embroiled in arcane debates around Popperian falsification or romanticised images of science and expertise.¹²² Rather, continuing a long pragmatic tradition expressly concerned with fairness as well as the veracity of legal decisions, common-law judges should be interested in whether there is some reason for believing that a particular technique, theory, or opinion is, on the balance of probabilities, reliable. Judges need to be satisfied that expert evidence adduced by the prosecution is dependable or trustworthy. On the *voir dire*, they can listen to evidence and arguments. If the prosecution can satisfy the judge that their expert evidence is demonstrably reliable, then it should be admitted; if not, then it should be excluded. These kinds of assessments are familiar to lay judges steeped in the common-law tradition.

In addition, a range of operational and logistical benefits should follow the imposition of an explicit reliability standard. The exclusion of unreliable expert evidence (and some evidence of unknown reliability) will save time and money. By excluding unreliable evidence, some trials will be shorter and some prosecutions will not be initiated. To the extent that prosecutions are not based on questionable scientific, medical, and technical evidence, verdicts are more likely to reflect the known value of the expert evidence. To the extent that courts are willing to exclude expert evidence, the defence will not be obliged to contest apparently—but not necessarily—disinterested forensic scientific and medical evidence adduced by the prosecution and admitted to the trial with the imprimatur of the state (and court). The accused will not be obliged to devote time and resources to challenging unreliable expert evidence through lengthy, and often technical, cross-examinations.¹²³ Moreover, the accused will be less dependent on the quality of the defence lawyer(s) or the resources available to them.

¹²² Consider the dissent of Chief Justice Rehnquist (with Justice Stevens agreeing) in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* 509 US 579 (1993).

¹²³ *R. v. Ranger* (2003) 178 CCC (3d) 375 at [62]: “Other significant dangers include the usual resistance of the expert opinion to effective cross-examination and the reliance by the expert on out-of-court material that would be otherwise inadmissible.”

Where the state's incriminating expert evidence is deemed inadmissible, the defence will not need to call its own rebuttal experts in an attempt to counter unreliable expert evidence. Defendants will not have to challenge—individually and repeatedly—the admissibility of untested or poorly grounded techniques and theories. More importantly, lay juries (and judges) will be spared from having to evaluate expert evidence—among complex assemblages of evidence—simply because the state was unwilling to test its techniques or the competence of its experts. The exclusion of unreliable expert evidence, along with expert evidence where there is a real danger that the evidence is unreliable, will mean that juries will not have to make impossible choices.¹²⁴ Excluding unreliable expert evidence adduced by the state may require explanation, but it will save the trial judge from having to issue guidance or instructions to the jury.¹²⁵ It will also reduce the need for appellate courts to engage in the constitutionally awkward reversal of jury verdicts.

Furthermore, a reliability standard relieves some of the pressure borne by the discretionary exclusions. One of the difficulties for any judge attempting to balance probative value against prejudicial effect is that, in the absence of information about the validity and accuracy of the technique, theory, or opinion, this becomes a very difficult exercise. Where the reliability of the expert evidence is low or uncertain, the potential prejudice will be considerable—especially where the limitations are not clearly identified or explained. Indeed, the potential for prejudice arises from the fact that the jury (or judge) may assign an inappropriate value to the evidence. Overvaluing and misusing unreliable or potentially unreliable evidence are two of the classic dangers associated with prejudicial effect. Focusing directly on the reliability of the expert evidence, however, removes questions about the probative value of the evidence from a calculus where issues of proof are mixed with concerns about fairness to the accused.¹²⁶ A formal reliability standard would assume much of the work currently left to the probative value/prejudicial effect discretion. It would also make the work presently left to the discretion more straightforward and more transparent. A demonstrable reliability standard would not, however, replace the discretion and it would not prevent the trial judge from excluding reliable expert

¹²⁴ *R. v. D.D.* [2000] 2 SCR 275 at [54]: “In cases where there is no competing expert evidence, this will have the effect of depriving the jury of an effective framework within which to evaluate the merit of the evidence.”

¹²⁵ See, for example: Joel Lieberman and Jamie Arndt, “Understanding the Limits of Limiting Instructions” (2000) 6 *Psychology, Public Policy & Law* 677–711.

¹²⁶ *Pfennig v. R.* (1995) 182 CLR 461 at [39]. Justice McHugh's dissent in *Pfennig* has now become the dominant approach in Australia, see for example: *R. v. Ellis* [2003] NSWCCA 319.

evidence in circumstances where the admission of that evidence would be unfairly prejudicial to the accused.

Perhaps the most important advantages flowing from a reliability standard are fairness to the accused and the maintenance of a rational legal process.¹²⁷ The *Charter of Rights and Freedoms* might be interpreted in a way that prevents the state from relying upon unreliable expert evidence in criminal prosecutions. At this stage the Supreme Court has been unwilling to read the need for reliability into s.7 of the *Charter*.¹²⁸ That approach *may* be comprehensible in response to lay evidence, but it becomes far more tenuous when it comes to the use of incriminating expert evidence.¹²⁹ For, in contrast to lay evidence where there are few dependable means with which to assess credibility and reliability, the sciences and biomedicine have established ways of determining the reliability—that is, validity and accuracy—of techniques, theories, and opinions. Can a rational system of justice remain indifferent to criminal prosecutions based upon expert evidence that is either unreliable or not shown to be reliable?

Just as evidence derived by torture, or obtained through duress, has been gradually expunged from our system of evidence and proof, so too expert evidence without a credible empirical foundation should not be relied upon by the state in criminal proceedings.¹³⁰ The expectation that only reliable expert evidence should be used in criminal prosecutions, whether grounded in constitutional or adjectival law, entails the possibility of disciplining the state and its agencies, preventing irrationality, and reducing the number of wrongful convictions. These would seem to be precisely the kinds of values guiding the development of a rational system of proof and presumably the kinds of values that bills of rights were intended to guarantee.

One additional benefit arising from the emphasis on reliability is that, at least in principle, it facilitates the *de novo* review of admissibility determinations by appellate

¹²⁷ William Twining, *Theories of Evidence: Bentham and Wigmore* (Stanford: Stanford University Press, 1985); William Twining, *Rethinking Evidence: Exploratory Essays* (Oxford: Basil Blackwell, 1990).

¹²⁸ *R. v. Buric* [1997] 1 SCR 535; *R. v. Buric* (1996) 106 CCC (3d) 97. Although the case was not primarily concerned with expert opinion evidence, in the Court of Appeal Weiler J.A., who endorsed the opinion of Labrosse J.A., equated proper qualifications and acceptance within the scientific community with “a finding of reliability.”

¹²⁹ Kent Roach, “Unreliable Evidence and Wrongful Convictions: The Case for Excluding Tainted Identification Evidence and Jailhouse and Coerced Confessions” (2007) 52 *Criminal Law Quarterly* 210; David Paciocco, *Charter Principles and Proof in Criminal Cases* (Toronto: Carswell, 1987).

¹³⁰ John Langbein, *Torture and the Law of Proof: Europe and England in the Ancien Régime* (Chicago: University of Chicago Press, 1976, reprinted 2006); Friedrich Spee von Langenfeld, *Cautio Criminalis, or a Book on Witch Trials* (Charlottesville: University of Virginia Press, 1632 reprinted 2003).

courts.¹³¹ Historically, appellate courts have been quick to defer to the many advantages available to the trial judge. On the admissibility and particularly the reliability of expert evidence, there would seem to be few reasons for deference. Where there is a *voir dire* on the admissibility of expert evidence, an expert's demeanour and performance should be granted little, if any, significance.¹³² Trial judges should be looking for evidence of testing or collective experience expressed in authoritative literatures. The burden of demonstrating the reliability of the expert evidence lies with the prosecution. If dissatisfied, in terms of reliability or fairness, trial judges and appellate courts should be willing to exclude the state's expert evidence. Such assessments are not dependent on any special advantages available at the trial.

While it may not be the most economical approach, the benefit of allowing the *de novo* review of expert evidence admissibility determinations is that it may shift some of the responsibility for excluding forensic scientific and medical evidence—including longstanding techniques—onto senior members of the judiciary. *De novo* review enables appellate courts to take the lead on admissibility jurisprudence and actual exclusion.

In closing, we might wonder about evidentiary problems created by the introduction of an explicit reliability threshold. Some might contend that the need to demonstrate the reliability of the state's expert evidence will make crimes more difficult to prosecute. Though worthy of empirical investigation, this would seem to be the wrong way to conceptualize the concern. Instead, we should be asking: Is the state willing to base criminal prosecutions on expert evidence that is unreliable or of unknown reliability? Do we wish to have more professional and more accountable forensic scientific and forensic medical communities? Do we wish to strengthen the separation of the courts from those communities?

To the extent that some incriminating evidence may be excluded from criminal trials or some prosecutions abandoned, the state and its citizens will have sacrificed very little. The state has no interest in prosecuting individuals, however serious the alleged offences, with unreliable forms of expert evidence.

¹³¹ Consider *Re Truscott* [2007] ONCA 575 at [95]: “the rules of evidence governing the admission of evidence in criminal proceedings are shaped primarily to facilitate the search for the truth. That search is not less important and no different when considering the admissibility of evidence offered on appeal.”

¹³² Compare *R. v. D.D.* [2000] 2 SCR 275 at [12]; *R. v. F. (D.S.)* (1999) 43 OR (3d) 609 at 625; *R. v. Ranger* (2003) 178 CCC (3d) 375 at [49]; *R. v. Morin* [1988] 2 SCR 345 at [64].

11. Other Procedures and Reforms Such As Court-Appointed Experts, Pretrial Meetings, and Codes of Conduct

So far, this paper has focused predominantly on admissibility standards and reliability. It might be useful, nevertheless, to briefly consider several procedural alternatives to explain how appeals to impartiality and idealized images of science, medicine, and expertise might not provide particularly effective solutions to perceived problems with expert evidence or help to maintain a beneficial separation between the courts and expert evidence produced by the state.

Court-Appointed Experts

The use of court-appointed experts has long been celebrated as a solution to the problems with expert disagreement, partisanship, cost, and delay.¹³³ It is true that recourse to court-appointed experts, especially if the parties have restrictions imposed on their ability to call (additional) experts, *may* simplify and expedite proceedings.¹³⁴ However, the use of court-appointed experts, particularly their introduction into adversarial systems, may be more problematical than is often assumed. There are a range of apparently mundane procedural issues, all with the potential to impugn judicial independence and even determine the outcome of the litigation.

How, for example, will these experts be selected?¹³⁵ Who picks them, and how? How many experts should be chosen? Should selection be undertaken informally through a judge's social and professional network? Are busy judges in a good position to determine which experts are appropriate in particular cases? Should the relevant professional body provide a list (assuming the relevant type(s) of expertise is non-controversial, the professional body not riven by controversy, and the existence of the field not in issue.)?¹³⁶ If so, is a judge confined to the list provided by a particular profession? What are the implications of disregarding the "official" list? Should

¹³³ John Spencer, "The Neutral Expert: An Implausible Bogey" (1991) *Criminal Law Review* 106; Petra van Kampen, *Expert Evidence Compared* (Antwerp: Intersentia Publishers, 1998).

¹³⁴ Joint or parties' single experts are currently in use in civil proceedings in England and Australia. For a variety of reasons, including the unlikelihood that the parties could agree on a shared expert in criminal litigation—especially where state-employed forensic experts are involved—these reforms would seem to be ill-suited to the adversarial criminal trial.

¹³⁵ Tony Ward, "Experts, juries and witch-hunts: From Fitzjames Stephen to Angela Cannings" (2004) 31 *Journal of Law & Society* 369, 382.

¹³⁶ Andrew Abbott, *The System of Professions: An Essay on the Division of Expert Labor* (1988).

judges select an expert or experts at the cutting edge of an issue—that is, those doing original research, embroiled in professional debates, and in possession of an intimate knowledge of current controversies? Or, should they prefer generalists with no specific knowledge or predetermined opinions on a subject?

Institutional pressures and risks posed by the use of court-appointed experts may lead judges to make conservative selections. Will judicial selection make experts more expensive as risk-averse judges select eminent experts from prestigious institutions? Will the selection of safe and eminent experts raise standards of admissibility and proof and alter the operation of legal doctrines? Will it be even more difficult for criminal defendants to challenge the opinions of eminent and established experts? While on this point it is worth reflecting on the fact that distinguished experts—such as Professor Sir Roy Meadow, once the doyen of British paediatrics—are precisely the kind of expert that judges are likely to select *and trust*.¹³⁷

Will court-appointed experts make the outcome of litigation less predictable as additional experts introduce new opinions after much of the preparatory work, pleadings, and pleas have been finalized? The use of court-appointed experts may actually complicate settlement negotiations and plea bargains and even stimulate more pretrial activity and litigation.

What about costs? Who should pay for the expert(s)? Will the use of court-appointed experts lead to the use of more experts—an additional expert for every special issue? What happens when the different parties want different types of experts? Judicial preferences may be outcome-dispositive.¹³⁸ If several experts are selected, how are different types of potentially incommensurable evidence to be reconciled? Where a number of court-appointed experts are selected, what happens when they disagree? Also, what happens when the court-appointed expert(s) disagrees with the judge's ultimate decision or reasoning?

Further, trial judges may be required to spend time and energy managing the credibility of a court-appointed expert, especially the appearance of *independence* and *impartiality*. Efforts to protect or guarantee the credibility of a court-appointed expert may (appear to) compromise judicial independence. Examples from large-scale civil

¹³⁷ See the treatment of Meadow in *R. v. Clark* [2003] EWCA Crim 1020 and *R. v. Cannings* [2004] 1 All ER 725.

¹³⁸ Gary Edmond and David Mercer, "Litigation Life: Law-Science Knowledge Construction in (Bendectin) Mass Toxic Tort Litigation" (2000) 30 *Social Studies of Science* 265–316.

litigation in the U.S. suggest that court-appointed experts require considerable attention, management, and protection (to sustain the impression of independence). In mass silicone-gel breast implant litigation, a panel of experts selected by a judge was challenged: because of the way they were selected as well as the suitability of their expertise; because private discussions and drafts of their final report were not discoverable; because of prior relations between members of the panel and major medical corporations; and because of their reasoning and conclusions. During the trial, because the panel was continuously challenged, it requested independent representation to defend the increasingly anxious experts' reputations and interests. The unsatisfactory alternative was continuing judicial intervention. The expert panel was eventually provided with its own lawyer who was ultimately paid more than US\$1,000,000.¹³⁹ Interestingly, a judge hearing similar breast implant litigation in another U.S. jurisdiction also appointed a panel of experts. This other panel was constituted by a slightly different assortment of specializations. While the different panels came to roughly similar conclusions in relation to causation, what would happen if court-appointed experts, individually or as a group, disagreed, or two different panels (or experts) reached inconsistent conclusions in similar circumstances? Which experts should a judge select in any subsequent litigation? Should a panel's findings pre-empt subsequent litigation? How should appellate courts respond?

Finally, what happens to public confidence in the courts and in individual judges when the credibility of an expert—especially an expert repeatedly appointed by the same judge or court—is compromised?

All of these questions and scenarios are, at least potentially, manageable, but they suggest that apparently simple solutions might actually be more complex and disruptive to adversarial legal institutions, the practice of judging, and the independence of the judiciary than is routinely suggested. Many of these decisions introduce new and acute risks into traditional adversarial systems. In criminal trials they threaten to erode widespread perceptions of fairness and public confidence; even where decisions are endorsed by appellate courts.

¹³⁹ Laura Hooper, Joe Cecil and Thomas Willging, "Assessing Causation in Breast Implant Litigation: The Role of Science Panels" (2001) 64 *Law & Contemporary Problems* 139; Joe Cecil and Thomas Willging, *Court-Appointed Experts: Defining the Role of Experts Appointed under the Federal Rule of Evidence 706* (1993).

Pretrial Expert Meetings

Another procedural reform, derived from recent changes to civil justice systems in England and Australia, is the expectation that experts will meet before any trial.¹⁴⁰

The goal is to try to reach agreement, narrow the extent of disagreement, and (in some jurisdictions) produce a joint report. The value of pretrial conferences in civil litigation seems to be equivocal. The suitability of their extension to criminal justice would seem to be even less certain.¹⁴¹ For, as we have seen, the state and accused are unevenly matched and the defence is rarely in a position to undertake a pre-emptive critique of forensic expert evidence.

In criminal proceedings, the various experts could be asked to meet before an anticipated trial in order to discuss the reliability of the state's incriminating expert evidence, particularly the value of any empirical testing, the significance of its absence, and whether techniques, theories, and opinions have support in specialist communities and literatures. However, the value of pretrial meetings between experts, especially where lawyers are excluded, would seem to be predicated on the empirically tendentious proposition that, displaced from legal institutions and lawyers, experts are likely to reach consensus guided by shared commitment to universal methods and normative conventions.¹⁴² Hopefully, Sections 2 and 3 of this paper will have shaken some of this naive optimism.

In effect, pretrial meetings would place legally experienced *forensic* experts—what were described in Section 3 as *law–science hybrids*—in a setting with (often) legally inexperienced defence experts.¹⁴³ Not only will this legal shadowland tend to advantage the more legally experienced (and possibly more numerous and personally familiar) forensic experts, it may also disclose defence concerns in ways that allow forensic scientists to retrospectively repair or enhance (the presentation of) their evidence in any trial.¹⁴⁴ Drawing attention to weaknesses in the state's expert evidence (and case) may encourage the Crown to abandon cases, but it is just as likely

¹⁴⁰ *Civil Procedure Rules 1998* (England) and *Uniform Civil Procedure Rules 2005* (NSW).

¹⁴¹ See Edmond, "Judicial Representations of Scientific Evidence," 242–249. For a discussion of pretrial expert meetings associated with the use of *concurrent evidence* procedures drawn from an ongoing empirical study sponsored by the Australian Research Council, see Gary Edmond, "Merton and the Hot-Tub: Expert Witnesses, Concurrent Evidence and Judge-Led Law Reform in Australia" (2008) 27 *Civil Justice Quarterly* (forthcoming).

¹⁴² Even if we were to accept that adversarial legal settings accentuate expert disagreement there can be no doubt that protracted and bitter controversy is a regular feature of specialization and expertise occurring well beyond legal, regulatory, and policy realms.

¹⁴³ Marc Galanter, "Why the 'Haves' Come Out Ahead: Speculations on the Limits of Legal Change" (1974) 9 *Law & Society Review* 95–160. Galanter discusses some of the many advantages that accrue to "repeat players."

to provide forensic scientists with opportunities and incentives to produce additional evidence and to make their evidence *appear* more robust. These kinds of responses may make it harder to mount a successful defence or successfully challenge expert evidence at trial through cross-examination or rebuttal experts. Requiring its experts to meet with the state's forensic scientists may drain the limited resources available to the defence. In consequence, even if meetings were imposed, it may be difficult to persuade the defence to fully divulge their concerns about incriminating expert evidence prior to the *voir dire* or trial.

Pretrial meetings *may* improve the quality of the state's forensic scientific and medical evidence, but they may also contribute to the production of cynical responses and more resilient evidence that may make it harder for the defence to identify limitations or persuade the trier of fact about weaknesses, uncertainties, and improprieties, even with unreliable evidence or evidence of unknown reliability. Diachronic refinements may be designed to superficially address criticisms raised by defence experts (and lawyers) and rhetorically reinforce the incriminating expert evidence, rather than gauge or improve the reliability of the evidence.

If there are endemic institutional and cultural problems with forensic science and medicine, it might not be appropriate to make the state's experts responsible for negotiating and reporting on the reliability of the expert evidence. We should recognize that this is expert evidence that they have produced and in which they maintain serious and ongoing personal, professional, and ideological interests. It may be difficult for forensic scientists to make concessions in pretrial meetings. And, close and continuing professional relations among forensic scientists, pathologists, and technicians may make it difficult for them to question incriminating results or to identify methodological limitations in the work of their peers. Given that many of the problems with forensic scientific and medical evidence in recent decades seem to have been attributed to cultural and institutional problems—including pro-prosecution commitments—as well as incompetence and hubris, we might wonder about the propriety of allowing forensic experts to broker consensus with defence experts removed from the scrutiny of lawyers—especially defence lawyers—and the supervision of courts. To the extent that lawyers should be included in pretrial meetings in order to protect legal interests, prevent inadvertent concessions or legal

¹⁴⁴ Stephen Hilgartner, *Science on Stage: Expert Advice as Public Drama* (Stanford: Stanford University Press, 2000).

mistakes by defence experts, and to supervise the state's forensic experts, any assessment of admissibility (and reliability) may as well be undertaken in public with the oversight of a judge. Regardless of any consensus brokered by the experts, a judge will still be required to determine the significance and extent of agreement, the adequacy of testing, the reliability, and ultimately the admissibility of the evidence. Admissibility determinations and assessments of reliability are, after all, legal rather than technical decisions. Moreover, there are benefits, in terms of transparency and accountability, to having reliability determinations on the public record, especially where the same kinds of techniques, theories, and opinions are used over and over.¹⁴⁵

Codes of Conduct (and Guidelines for Experts)

Over the last decade many jurisdictions in Australia have introduced some kind of code of conduct intended to regulate the behaviour of experts (see the Appendix).¹⁴⁶ These codes may help to reiterate—to the extent that it requires any clarification or reinforcement—the paramount duty owed by expert witnesses to the court. They might also be used by experts, especially experts called in civil litigation or by criminal defendants, to resist importunity from lawyers and clients. However, codes of conduct and guidelines typically provide greater symbolic than practical value. They are unlikely to discipline experts, change the cultures associated with expert witnessing, provide means of identifying impropriety, or provide particularly serviceable means of sanctioning experts. If anything, the Australian codes (such as the influential example in the Appendix) seem to have contributed more to changes in the form of expert reports and testimony than to the substance or reliability of expert evidence.

Though superficially appealing, the practical value of codes and guidelines seems to be limited. These limitations are even more pronounced where experts, such as forensic scientists, are relatively conversant with legal institutions, procedures, rules, and even substantive law. Consider, by way of example, how difficult it might be to discipline experts who are *perceived* to have breached a code (or failed to meet expectations at a pretrial meeting).

¹⁴⁵ For example, transcripts will be available to later defendants to assist with cross-examination.

¹⁴⁶ Ian Freckelton et al., *Australian Judicial Perspectives* (Melbourne: AIJA, 1999). Contrast Gary Edmond, “Judging Surveys: Experts, Empirical Evidence and Law Reform” (2005) 33 *Federal Law Review* 95–139.

... [O]n what grounds are judges to apply sanctions against experts who *breach* their obligation to the court or who are unable to achieve consensus around their opinions? How should judges determine whether reluctance to agree or narrow the grounds of disagreement at an expert conference constitutes legitimate professional differences or obduracy driven by a party's desire for a trial? What in the process divulges this? When is adherence to a particular 'school of thought' partisan and under what circumstances might it be reasonable or objective? What can judges do when experts hold firm opinions about areas characterised as *uncertain* or disagree about the extent or significance of certitude in a field? Could experts be punished—by contempt of court proceedings, for attempting to pervert the course of justice, or even perjury—for steadfastly holding an opinion, or even changing their mind in relation to particular opinions or fresh evidence? When should fresh evidence or assumptions excuse or require such shifts?¹⁴⁷

Flagrant misconduct will be fairly obvious and remediable with or without a code of conduct. More subtle forms of exaggeration, misrepresentation, and omission will prove far more difficult to combat and will not be readily identified through the imposition or adoption of guidelines.

Experts—and especially forensic pathologists, scientists, and technicians—are not under illusions about their primary duty to the court. To the extent that problems are created or accentuated by forensic scientific and medical cultures, a normative vision embodied in a code of conduct is unlikely to transform those cultures—even if it mandates some changes to practices. There may, of course, be little harm with such innovations, unless law reformers and judges actually believe that codes of conduct are capable of overcoming long periods of socialization, professional marginalization, resource deprivation, the ideological proclivities of institutionalized forensic experts, and existing reward structures.

Overview

Most of the benefits attributed to the use of court-appointed experts and these other procedural reforms depend on the kinds of idealized images criticized earlier in this paper (see Section 2). Recourse to purportedly impartial experts, like limiting the number of experts or the scope for disagreement, does not necessarily improve reliability, eliminate controversy, reduce costs, or enhance social legitimacy.¹⁴⁸ Once we abandon quaint and unrealistic commitments to strong forms of impartiality and

¹⁴⁷ Gary Edmond, "After Objectivity: Expert Evidence and Procedural Reform" (2003) 25 *Sydney Law Review* 131–64, 156–157.

¹⁴⁸ Consider Tom Tyler and Yuen Huo, *Trust in the Law* (New York: Russell Sage Foundation, 2002).

recognise that the invocation of scientific norms, peer review, publication, and a universal scientific method might not be as useful or discriminating as lay decision makers routinely assume, recourse to court-appointed experts, attempts to suppress expert disagreement, expecting experts to narrow agreement or adhere to abstract codes may actually be conceived as contrived and even politically disingenuous.¹⁴⁹

Throughout this paper, the emphasis on demonstrable reliability has not been developed in order to limit the scope of cross-examination, reduce the number of rebuttal experts available at trial, or restrict expert disagreement. *Demonstrable reliability* has been advanced as an admissibility standard because it has the potential to make the expert evidence relied upon by the state more reliable. Rather than reduce the number of experts appearing in court or encouraging experts to resolve their differences in (private) pretrial conclaves, there would seem to be social and institutional benefits from making the state's forensic experts publicly accountable and from allowing the parties to explore the limitations of even demonstrably reliable expertise at trial.¹⁵⁰

12. Conclusion: The Emergence of *Evidence-Based Forensics* (EBF)

This paper has endeavoured to explain the value of *demonstrable reliability*. If courts and reformers are genuinely interested in reducing wrongful convictions, improving accuracy, and enhancing fairness, regardless of organizational and structural changes to forensic science and medicine, then refining and enforcing admissibility standards will have a major systemic effect. Requiring demonstrably reliable expert evidence would compel institutionalized forensic science and medicine to reform their approaches to investigation, evidence, and proof. The need for demonstrable reliability is, in reality, just another way of requiring forensic scientific, medical, and technical evidence to be based on solid foundations. Mirroring developments in the mainstream biomedical sciences, it seems desirable for techniques, theories, and opinions relied upon by the state to be *evidence-based*.

In recent decades there has been a very conspicuous turn in biomedical research and medical practice toward evidence-based practice—commonly referred to as

¹⁴⁹ Consider Yaron Ezrahi, *The Descent of Icarus: Science and the Transformation of Contemporary Democracy* (Cambridge, MA: Harvard University Press, 1990); Richard Sclove, *Democracy and Technology* (New York: The Guildford Press, 1995).

evidence-based medicine (EBM). While there have been criticisms of evidence-based approaches to medical research, publishing, and practice, as a general framework for evidence jurisprudence, expecting the state's expert evidence to be evidence-based does not seem unreasonable. In biomedical research and publication, the shift toward evidence-based approaches represented a response to the influence of large pharmaceutical and therapeutic product manufacturers (and the recalcitrance of individual physicians).¹⁵¹ Commercial sponsorship, along with changes to intellectual property regimes and closer ties between investigators and manufacturers, rapidly transformed the culture and practice of biomedical research. In particular, the rise of private research organizations, the manipulation of study designs and results, the use of “ghost authors,” the repeated publication of the same commercially favourable research (known as “redundant publication”), along with the use of “gag clauses” and the suppression of adverse findings (a type of “publication bias”), were all seen to be corrupting biomedical research cultures and the published literature.¹⁵²

In response, biomedical and public health researchers and editors implemented a series of reforms designed to limit the influence of large multinational corporations and modify the culture associated with biomedical research and publication. These reforms are, perhaps, best seen through the activities of the International Committee of Medical Journal Editors (ICMJE). In response to commercial pressures and widespread deregulation, the leading generalist biomedical journals—including the *British Medical Journal*, the *Lancet*, the *New England Journal of Medicine*, the *Journal of the American Medical Association*, the *Canadian Medical Association Journal*, *Annals of Internal Medicine*, and the *Medical Journal of Australia*—changed several of their editorial policies.¹⁵³ They now require all clinical trials and studies to be prospectively registered and the results made publicly available (through collectives such as the Cochrane Collaboration and other accessible databases). They also require submissions for publication to list all those involved in the research and

¹⁵⁰ Jasanoff, *Science at the Bar*, Edmond, “Science in Court.”

¹⁵¹ Stephan Timmermans and Marc Berg, *The Gold Standard: The Challenge of Evidence-Based Medicine and Standardisation in Health Care* (Philadelphia: Temple University Press, 2003).

¹⁵² Philip Mirowski and Robert van Horn, “The Contract Research Organization and Commercialization of Scientific Research” (2005) 25 *Social Studies of Science* 503–548; AMA Council of Scientific Affairs, *Influence of Funding Source on Outcome, Validity, and Reliability of Pharmaceutical Research* (2004); Marcia Angell, *The Truth about the Drug Companies: How They Deceive Us and What to Do About It* (Scribe: Melbourne, 2005); Jerome Kassirer, *On the Take: How Medicine's Complicity with Big Business Can Endanger Your Health* (New York: Oxford University Press, 2004); Merrill Goozner, *The \$800 Million Pill: The Truth Behind the Cost of New Drugs* (Berkeley: University of California Press, 2004).

¹⁵³ International Committee of Medical Journal Editors, *Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication* (2005).

preparation of a paper; disclose all sources of funding and support; declare any contractual constraints; and to disclose any conflicts of interest that any of the researchers or authors may have (or have had). In combination, more disclosure and prospective registration make it more difficult for manufacturers to rely upon unregistered studies, prevent the same (always favourable) studies being counted numerous times in meta-analyses, and enables those drawing on studies to factor conflicts of interest into their analysis of the value of research. For, it is well documented that sponsorship, close associations, and continuing relationships strongly influence research results.¹⁵⁴

The move to evidence-based medicine and the changes in biomedical publication are informative because they were, at least in part, a response to endemic cultural problems. These developments, particularly the endeavours to improve research practices and monitor the relations between researchers and sponsors (or clients), might have particular salience to the reform of institutionalized forensic science and medicine.

Perhaps the most interesting dimension of the reforms to biomedical research and publication—simultaneously reinforcing the earlier discussion of the sciences and the limits of simplistic solutions, such as court-appointed experts—is that well-resourced and highly skilled scientists, medical researchers, and editors have not resorted to neutral advisers, additional peer review, or calls for scientific and biomedical research to conform to abstract philosophical models of scientific method. Rather, the changes to biomedical publication demonstrate that even well-resourced specialists, and here we should include the editorial teams and referees available to members of the ICMJE, such as the *New England Journal of Medicine*, the *Journal of the American Medical Association*, and the *Canadian Medical Association Journal*, encountered difficulty identifying instances of gross error and fraud, let alone more subtle and endemic problems—such as systematic bias, omission, and exaggeration—through the use of peer review.¹⁵⁵ Consequently, technically competent staff, at pre-eminent biomedical journals, have focused their efforts to understand, evaluate, and improve

¹⁵⁴ Justin Bekelman et al., “Scope and Impact of Financial Conflict of Interest in Biomedical Research” (2003) 289 *Journal of the American Medical Association* 454–465; Joel Lexchin et al., “Pharmaceutical Industry Sponsorship and Research Outcome and Quality: Systematic Review” (2003) 326 *British Medical Journal* 1167–1170.

¹⁵⁵ David Michaels and Celeste Monforton, “Scientific Evidence in the Regulatory System: Manufacturing Uncertainty and the Demise of the Formal Regulatory System” (2005) 13 *Journal of Law & Policy* 17; Krimsky, *Science in the Private Interest*; Wagner and Steinzor, *Rescuing Science from Politics*.

the biomedical literature (and associated research cultures) upon empirical studies, socio-economic relations, disclosure, and public accessibility.

These developments have a profound significance for legal practice. Not only do they reinforce the limits to idealized models of scientific practice and superficial approaches to the biomedical literature, they also suggest that lay judges and reformers might develop techniques for improving the quality of the expert evidence used in investigations and prosecutions. Developments in biomedical publication should be reassuring to common-law judges because they suggest that ordinary features of expert practice might be used in the assessment of expert evidence and the reformation of the state's forensic scientific and pathological evidence.

Developments in evidence-based medicine do not provide a complete solution, but to the extent that the state needs to produce demonstrably reliable expert evidence, there is a need for that evidence to be evidence-based. Moreover, judges should be attentive to the cultures and practices associated with the state's investigative institutions. They should be willing to cultivate, like members of the ICMJE, more practical approaches to reliability as well as to identify and counter some of the deleterious social and institutional aspects of knowledge construction in the state's forensic institutions. The institutions responsible for producing forensic scientific and medical evidence should become more attentive to their practices and the potentially detrimental influence of close relations with police and investigators along with generally pro-prosecution sympathies. Just as biomedical editors have sought to identify and disclose potentially damaging conflicts of interest, so judges and the state's scientists, technicians, and pathologists should endeavour to address and better manage the cultural dimensions of forensic scientific and medical practice. An evidence-based turn in forensics will also help to sensitize judges to some of the problems with modern expertise, the difficulties encountered by defendants and defence lawyers, as well as practical responses in other domains—such as public health—also fundamentally concerned with reliability, efficacy, and cost.

Ultimately, in order to uphold fundamental institutional values such as accuracy and fairness (i.e., truth and justice), scientists, technicians, doctors, lawyers, trial and appellate judges will all need to be vitally interested in the reliability of expert evidence. The trial judge should not be the sole arbiter of reliability and admissibility. Members of all these groups must be professionally preoccupied with accuracy and

fairness and should pay close attention to the evidentiary foundations of any techniques, theories, or opinions relied upon or presented as evidence.

A criminal justice system concerned with fairness and the accuracy of decisions cannot afford to admit evidence presented by ostensibly disinterested forensic scientists, pathologists, and technicians unless that evidence is demonstrably reliable. Expressed the other way around, a criminal justice system cannot afford to base determinations of guilt on unreliable expert evidence or expert evidence of unknown reliability. For, once unreliable expert evidence is admitted into the trial, it is extremely difficult, and sometimes impossible, to manage its impact, especially in complex and emotive cases such as those involving the death of a child.

There is little doubt that many of those engaged in forensic science and medicine, and those wedded to crime control, will strenuously object to these proposals. But such objections are, in effect, rejecting the need for demonstrably reliable expert evidence. To the extent that objections are based around difficulties with testing, its limitations or applicability, these are precisely the kinds of issues that can be resolved by a lay judge. Perhaps more troublesome is the danger that, even if demonstrable reliability was confirmed as a prerequisite for the admission of expert evidence, trial and appellate judges—including those formally committed to reliability, fairness, and accuracy—would not take their gatekeeping responsibilities sufficiently seriously. The long and symbiotic relations between courts, police, forensic science, and medicine have created a level of judicial confidence that may prove resilient despite the fact that it has worked to the detriment of all. Any revision to the admissibility standards for expert evidence will require systematic and continuing vigilance from the appellate courts of the type surfacing in *Trochym* and *Truscott*.

If the state's forensic scientists, pathologists, and technicians cannot persuade courts and the public that their techniques and results are reliable, then the courts should not hesitate to exclude their evidence. Regardless of any reforms undertaken to improve the culture of the state's investigative institutions or the competence of their forensic experts, the exclusion of unreliable expert evidence is the most fundamental and practical defence against the ever-present danger of legal mistakes and wrongful conviction. Similarly, an admissibility standard based on a genuine commitment to reliability provides the best hope of disciplining experts and enhancing the legitimacy of the criminal justice system.

Moving into the future, the forensic sciences and forensic medicine should become increasingly evidence-based. Resources and effort should be devoted to demonstrating the validity and accuracy of techniques and theories. Rather than requiring judges to undertake extensive pretrial inquiries into the reliability of forensic science and medicine, most of the resources and efforts should be devoted to enabling investigative agencies and laboratories to establish the value of their techniques and theories before trial, and where possible, prior to investigation.

Evidence-Based Law Reform?

In concluding this paper, and particularly this section on evidence-based forensics, it is important to express the need for empirical research into changes to institutionalized forensic science and medicine and the effects of changing admissibility standards on legal practice and the quality of expert evidence. One of the main difficulties with expert evidence and its reform is the very limited volume of systematic empirical information. Most reform proposals are based on the partial perspectives of judges and, to a lesser extent, scientists, doctors, and lawyers. There is a manifest need for further study and ongoing monitoring that incorporates the empirically based perspectives of social scientists.

Appendix “Guidelines” from the Federal Court of Australia

Codes of conduct may reinforce desirable normative commitments but they are unlikely to exert a pronounced influence on expert cultures and practice. They are probably more useful for those experts—not typically forensic scientists, pathologists, and technicians—who do not routinely appear in and around legal settings. There can be little doubt that forensic scientists, pathologists, and technicians are already aware that they owe a duty to the court and should, consistent with their oath or affirmation, endeavour to be truthful.

As this example suggests, the Federal Court “Guidelines” are largely concerned with the form of expert reports and testimony.

Guidelines for Expert Witnesses in Proceedings in the Federal Court of Australia

Explanatory Memorandum

The guidelines are not intended to address all aspects of an expert witness’s duties, but are intended to facilitate the admission of opinion evidence,¹ and to assist experts to understand in general terms what the Court expects of them. Additionally, it is hoped that the guidelines will assist individual expert witnesses to avoid the criticism that is sometimes made (whether rightly or wrongly) that expert witnesses lack objectivity, or have coloured their evidence in favour of the party calling them.

Ways by which an expert witness giving opinion evidence may avoid criticism of partiality include ensuring that the report, or other statement of evidence:

- (a) is clearly expressed and not argumentative in tone;
- (b) is centrally concerned to express an opinion, upon a clearly defined question or questions, based on the expert’s specialised knowledge;
- (c) identifies with precision the factual premises upon which the opinion is based;
- (d) explains the process of reasoning by which the expert reached the opinion expressed in the report;
- (e) is confined to the area or areas of the expert’s specialised knowledge; and
- (f) identifies any pre-existing relationship (such as that of treating medical practitioner or a firm’s accountant) between the author of the report, or his or her firm, company etc, and a party to the litigation.

An expert is not disqualified from giving evidence by reason only of a pre-existing relationship with the party that proffers the expert as a witness, but the nature of the pre-existing relationship should be disclosed. Where an expert has such a relationship the expert may need to pay particular attention to the identification of the factual premises upon which the expert's opinion is based. The expert should make it clear whether, and to what extent, the opinion is based on the personal knowledge of the expert (the factual basis for which might be required to be established by admissible evidence of the expert or another witness) derived from the ongoing relationship rather than on factual premises or assumptions provided to the expert by way of instructions.

All experts need to be aware that if they participate to a significant degree in the process of formulating and preparing the case of a party, they may find it difficult to maintain objectivity.

An expert witness does not compromise objectivity by defending, forcefully if necessary, an opinion based on the expert's specialised knowledge which is genuinely held but may do so if the expert is, for example, unwilling to give consideration to alternative factual premises or is unwilling, where appropriate, to acknowledge recognised differences of opinion or approach between experts in the relevant discipline.

Some expert evidence is necessarily evaluative in character and, to an extent, argumentative. Some evidence by economists about the definition of the relevant market in competition law cases and evidence by anthropologists about the identification of a traditional society for the purposes of native title applications may be of such a character. The Court has a discretion to treat essentially argumentative evidence as submission, see Order 10 paragraph 1(2)(j).

The guidelines are, as their title indicates, no more than guidelines. Attempts to apply them literally in every case may prove unhelpful. In some areas of specialised knowledge and in some circumstances (eg some aspects of economic "evidence" in competition law cases) their literal interpretation may prove unworkable. The Court expects legal practitioners and experts to work together to ensure that the guidelines are implemented in a practically sensible way which ensures that they achieve their intended purpose.

Guidelines

1. General Duty to the Court²

- 1.1 An expert witness has an overriding duty to assist the Court on matters relevant to the expert's area of expertise.
- 1.2 An expert witness is not an advocate for a party even when giving testimony that is necessarily evaluative rather than inferential.³

- 1.3 An expert witness's paramount duty is to the Court and not to the person retaining the expert.

2. The Form of the Expert Evidence⁴

- 2.1 An expert's written report must give details of the expert's qualifications and of the literature or other material used in making the report.
- 2.2 All assumptions of fact made by the expert should be clearly and fully stated.
- 2.3 The report should identify and state the qualifications of each person who carried out any tests or experiments upon which the expert relied in compiling the report.
- 2.4 Where several opinions are provided in the report, the expert should summarise them.
- 2.5 The expert should give the reasons for each opinion.
- 2.6 At the end of the report the expert should declare that "[the expert] has *made all the inquiries that [the expert] believes are desirable and appropriate and that no matters of significance that [the expert] regards as relevant have, to [the expert's] knowledge, been withheld from the Court.*"
- 2.7 There should be included in or attached to the report; (i) a statement of the questions or issues that the expert was asked to address; (ii) the factual premises upon which the report proceeds; and (iii) the documents and other materials that the expert has been instructed to consider.
- 2.8 If, after exchange of reports or at any other stage, an expert witness changes a material opinion, having read another expert's report or for any other reason, the change should be communicated in a timely manner (through legal representatives) to each party to whom the expert witness's report has been provided and, when appropriate, to the Court.⁵
- 2.9 If an expert's opinion is not fully researched because the expert considers that insufficient data are available, or for any other reason, this must be stated with an indication that the opinion is no more than a provisional one. Where an expert witness who has prepared a report believes that it may be incomplete or inaccurate without some qualification, that qualification must be stated in the report.⁵
- 2.10 The expert should make it clear when a particular question or issue falls outside the relevant field of expertise.
- 2.11 Where an expert's report refers to photographs, plans, calculations, analyses, measurements, survey reports or other extrinsic matter, these must be provided to the opposite party at the same time as the exchange of reports.⁶

3. Experts' Conference

- 3.1 If experts retained by the parties meet at the direction of the Court, it would be improper for an expert to be given, or to accept, instructions not to reach agreement. If, at a

meeting directed by the Court, the experts cannot reach agreement about matters of expert opinion, they should specify their reasons for being unable to do so.

Notes

1. As to the distinction between expert opinion evidence and expert assistance see *Evans Deakin Pty Ltd v Sebel Furniture Ltd* [2003] FCA 171 per Allsop J at [676].
2. See rule 35.3 Civil Procedure Rules (UK); see also Lord Woolf “Medics, Lawyers and the Courts” [1997] 16 CJK 302 at 313.
3. See *Sampi v State of Western Australia* [2005] FCA 777 at [792]-[793], and *ACCC v Liquorland and Woolworths* [2006] FCA 826 at [836]-[842].
4. See rule 35.10 Civil Procedure Rules (UK) and Practice Direction 35—Experts and Assessors (UK); *HG v the Queen* (1999) 197 CLR 414 per Gleeson CJ at [39]-[43]; *Ocean Marine Mutual Insurance Association (Europe) OV v Jetopay Pty Ltd* [2000] FCA 1463 (FC) at [17]-[23].
5. The “*Ikarian Reefer*” [1993] 20 FSR 563 at 565.
6. The “*Ikarian Reefer*” [1993] 20 FSR 563 at 565-566. See also Ormrod “Scientific Evidence in Court” [1968] Crim LR 240.