



Preventing Miscarriages of Justice: the Reliability of Forensic Evidence and the Role of the Trial Judge as Gatekeeper

Justice Chris Maxwell

Paper presented to the July 2017 Conference of the International Society for the Reform of Criminal Law, San Francisco

Wednesday 12 July 2017

“The obvious risk in a criminal trial when expert evidence is led from a forensic scientist is that a jury will give the evidence more weight than it deserves. To prevent unfair prejudice of that kind, it is essential that the reliability of the expert evidence be established to the Court’s satisfaction before it is led.”¹

In 2015 in *Tuite*, the Victorian Court of Appeal applied the decision of the US Supreme Court in *Daubert v Merrell Dow Pharmaceuticals Inc.*² which defined the “gatekeeper” role of the trial judge as being to ensure that “all scientific testimony ... is not only relevant but reliable”.³ This paper examines the onerous nature of that responsibility, especially given the serious lack of validation in so many areas of forensic science, as revealed by the 2009 Report of the National Academy of Sciences (“NAS”) and the 2016 Report of the President’s Council of Advisors on Science and Technology (“PCAST”).

The problem of unreliable science and wrongful convictions is a matter of profound concern to the criminal justice system. It is simply chilling to read of the hundreds of exonerations (through subsequent DNA testing) of individuals wrongfully convicted in US courts, and imprisoned, on the basis of “junk” science.

Australian judges and practitioners are only just beginning to engage with these issues. We have had nothing like the high level, high quality investigations and reports of NAS and PCAST. But they

¹ *Tuite v The Queen* (2015) 49 VR 196, 200 [11] (“*Tuite*”).

² 509 US 579 (1993) (*Daubert*).



are universal issues, and we have much to learn from North America, both from the experience of judges and from the rich academic literature. I hope that, through this Society, there may be prospects for productive collaboration.

My own awakening to the importance of these issues came about by pure accident, as a result of two unrelated events in the course of my work at the Victorian Court of Appeal. As these incidents illustrate, appellate work provides a unique opportunity for judges to see where the criminal justice system is not working properly. Moreover, I have come to realise, judges have the responsibility, and the authority, to drive the necessary change.

The first event was an appeal against conviction for “baby-shaking” manslaughter. When I reviewed the trial transcript, I was shocked to discover that no one in the trial — neither the judge nor the prosecutor nor defence counsel — seemed to have understood the true import of the forensic pathologist’s report, on which the whole Crown case depended.

As I sought to explain in my judgment, on a proper analysis the expert evidence could not have established that any act of the accused man had caused the baby’s death.⁴ By the time we reached that conclusion, however, and quashed the conviction, the accused man had spent a year in gaol.

Plainly enough, something needed to be done. The problem here seemed to be what is referred to in the literature as the “scientific illiteracy” of the participants in the trial.

I subsequently joined forces with Professor Stephen Cordner, then the Director of the Victorian Institute of Forensic Medicine. By coincidence, it was he who had been the Crown’s expert in that trial. In 2012, we established what has become known as the Forensic Evidence Working Group. The Group’s membership comprises representatives of all of the participants in the criminal justice system, together with experts from the relevant scientific disciplines.

The Group developed a new Practice Note entitled “Expert Evidence in Criminal Trials”, which came into force on 1 July 2014. That Practice Note—

³ Ibid 589.



- a) contains detailed specifications of what an expert report must contain;
- b) establishes procedures to enable defence counsel to confer with a prosecution expert before trial;
- c) enables the trial judge to direct experts to confer and prepare a joint report; and
- d) makes provision for “concurrent evidence” in a case where both prosecution and defence rely on expert evidence.⁵

The latter procedure, known as the “hot tub”, has been used for many years — and with considerable success — in civil litigation in Australia. The basic procedure is for the experts to be seated side-by-side in the courtroom and for them to be questioned by counsel on both sides, by the judge and, where appropriate, by each other. Experience shows that this procedure greatly reduces the time taken in the presentation of expert evidence and, most importantly, makes it easier for the judge to assess the competing views and to be clear about the areas of disagreement.

For criminal trials, however, this is uncharted territory. As far as I am aware, the new procedure has not yet been invoked in a criminal trial.

The second key event for me was a judicial education seminar on evidence-based decision-making, presented by the Judicial College of Victoria. Professor Gary Edmond of the University of New South Wales spoke about the findings of the NAS Report. I was, once again, shocked by what I heard. As you may know, the conclusion of that report was as follows:

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

As I subsequently discovered, Professor Edmond has written widely, both in Australia and abroad, about the lack of proven reliability of much forensic science. I have learnt a great deal from his

⁴ *R v Klamo* (2008) 18 VR 644, 657 [53]–[55].

⁵ The availability of this procedure is now underpinned by s 232A of the *Criminal Procedure Act 2009*.



writing, and was able to draw on it when the Court of Appeal in 2015 had to consider a challenge to the reliability of a new methodology for calculating “likelihood ratios” based on DNA samples.

The decision in Tuite

The accused is facing charges of aggravated burglary, rape, indecent assault and intentionally causing injury. Expert opinion evidence is to be called at the trial about the analysis of DNA samples from the crime scene. The DNA evidence is to be presented in the usual form of a “likelihood ratio”.

In this case, the likelihood ratios have been calculated using a recently-developed software package, known as STRmix, which was introduced into Victoria in March 2013. At a pre-trial hearing, the accused challenged the admissibility of the DNA evidence on the ground that the new methodology was not — or had not been shown to be — sufficiently reliable for use in criminal trials. The methodology was largely untested, it was said, and had not been generally accepted by the forensic science community.

The pre-trial hearing extended over some 22 days, in the course of which the judge heard evidence from the three prosecution experts and one defence expert. In a reserved decision, her Honour rejected the application to exclude the evidence. The Court of Appeal upheld that decision.

We addressed at some length how a judge should go about assessing evidentiary reliability. We concluded that the touchstone of reliability for scientific evidence must be trustworthiness, and trustworthiness depends on validation. For this purpose, we adopted what the US Supreme Court had said in *Daubert*, as follows:

“We note that scientists typically distinguish between ‘validity’ (does the principle support what it purports to show?) and ‘reliability’ (does application of the principle produce consistent results?). Although ‘the difference between accuracy, validity, and reliability may be such that each is distinct from the other by no more than a hen’s kick,’ ... our reference here is to evidentiary reliability — that is,



*trustworthiness. In a case involving scientific evidence, evidentiary reliability will be based upon scientific validity.*⁶

It followed, in our view, that the focus of attention for the purposes of assessing the reliability of scientific evidence should be on proof of validation. Ideally, there should be proof of both in-house validation and independent external validation. But, as the Court noted, the commercialisation of forensic science makes this latter requirement increasingly difficult to satisfy.

We pointed out that the focus on proven validation had a number of advantages. First, and most importantly, it meant that the scrutiny of scientific evidence in the judicial process would apply the rigour which the discipline of science itself demands. As it was put in *Daubert*, evidentiary reliability would be based on scientific validity. Secondly, the trial judge considering scientific evidence would ordinarily be able to assess the sufficiency of validation — based on the published results of validation tests — without needing to acquire particular expertise in the relevant field of science.

Thirdly, validation studies would provide a framework which would assist the judge —and, ultimately, the jury — to evaluate the evidence. Fourthly, this approach would avoid what we considered to be the unworkable imprecision of a “general acceptance” test, and would ensure that new developments and novel techniques were not excluded, provided always that their scientific validity was established to the satisfaction of the court.

The Supreme Court in *Daubert* defined the judge’s “gatekeeper role” as entailing:

*“a preliminary assessment of whether the reasoning or methodology underlining the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue”.*⁷

But how, in practice, is the testing of reliability expected to occur? PCAST made one

⁶ 509 US 579, 590 n9 (1993).



recommendation to the US judiciary regarding the use of scientific validity as a foundation for expert testimony. Recommendation 8 states:

- “(A) *When deciding the admissibility of expert testimony, Federal judges should take into account the appropriate scientific criteria for assessing scientific validity including: ... (1) foundational validity ... and (ii) validity as applied ...*
- (B) *Federal judges, when permitting an expert to testify about a foundationally valid feature-comparison method, should ensure that testimony about the accuracy of the method and the probative value of proposed identifications is scientifically valid in that it is limited to what the empirical evidence supports.”*

Lawyers and judges

As American experience shows, having a clear set of criteria for testing evidentiary reliability is no guarantee that junk science will be kept out of criminal trials. Much of the responsibility for ensuring reliability lies with judges and practitioners, in particular prosecutors. And it turns out that we are a major part of the problem.

In 2016, Chris Fabricant and Tucker Carrington of the Mississippi Innocence Project delivered a powerful critique of what they describe as the courts’ “abdication” of the gatekeeper role.⁸ Their article addresses the two most notorious areas of so-called expertise —bite mark analysis and microscopic hair comparison — which have been productive of miscarriages of justice. They criticise what they describe as the—

“failure of courts to distinguish between magic and science in the first instance and the judicial system’s continuing reflexive reliance on deeply flawed, scientifically invalid precedents to support the admissibility of false and misleading evidence.”⁹

⁷ Ibid 592–3.

⁸ M.C. Fabricant and T. Carrington, “The Shifted Paradigm: Forensic Science’s Overdue Evolution from Magic to Law” (2016) 4 *Virginia Journal of Criminal Law* 1.

⁹ Ibid 7.



Clearly, judges need to be both more active, and better informed, if we are to perform this crucial gatekeeper role. As Judge Nancy Gertner said in 2011 in the *UCLA Law Review*:

*“Until courts address the deficiencies in the forensic sciences — until courts do what Daubert requires that they do — there will be no meaningful change here.”*¹⁰

As has been repeatedly pointed out, if judges are to discharge this function, we need to become more scientifically literate and, to that end, look for sources of expert assistance. The NAS Report said:

“The judicial system is encumbered by judges and lawyers who generally lack the scientific expertise necessary to comprehend and evaluate forensic science in an informed manner.”

Writing about the PCAST Report, Professor Edmond and his colleague, Kirsty Martire, commented:¹¹

“Australia has not undertaken any systematic independent review of its forensic sciences. Moreover, tradition and adjectival rules prevent judges from unilateral engagement with materials, however authoritative, that are not adduced by the parties. Consequently, there are relatively few ways for Australian judges to find out about mainstream scientific perspectives. Australian judges tend to be insulated from, and practically incapable of engaging with, scientific knowledge and the best scientific advice. Apart from judicial education and independent reading, operating as some kind of ‘background’ or ‘framework’, the only means of bringing mainstream scientific perspectives to the attention of judges is through the testimony of forensic practitioners. Notwithstanding formal obligations to serve the court impartially, or to act as a ‘minister of justice’, very few forensic practitioners or prosecutors, have

¹⁰ N. Gertner, “Commentary on the Need for a Research Culture in the Forensic Sciences” (2011) 58 *UCLA L. Rev.* 789, 790.

¹¹ G. Edmond and K. Martire, “Forensic science in criminal courts: the latest scientific insights” (2016)



brought the issues of validation and error to the attention of Australian judges and juries.”

Performing the gatekeeper role

How does the trial judge go about assessing reliability? A number of possibilities suggest themselves. The first is for the judge to undertake her own *Daubert* hearing. That is what the trial judge did in *Tuite*, a process which took some 22 days.¹² In the ephedra case, Judge Rakoff undertook a two week evidentiary hearing, in which he heard from the competing experts and made up his own mind as to which parts of the evidence could be relied on for which purpose.¹³

A second possibility is for the judge to seek the assistance of a court-appointed expert. As Judge Stephanie Domitrovich of Pennsylvania pointed out last year, Federal Rule of Evidence 76 provides for court-appointed experts to assist judges in their gatekeeping role.¹⁴ It has thus been recognised that the current training of the judiciary may not sufficiently prepare judges to perform the role of scientific evaluator.¹⁵

As Judge Domitrovich argues:

*“Somehow trial judges need to acquire the requisite tools and knowledge to assess the reliability of the methods used by forensic scientists and other experts. To do so, they need assistance.”*¹⁶

And again:

“[Judges] must become sophisticated consumers of science who are capable of

42 *Australian Bar Review* 1, 15.

¹² See also *Thompson v Johnson and Johnson Pty Ltd* (1989) Aust. Torts Reports 80–278.

¹³ J Rakoff, “Science and the Law: Uncomfortable bedfellows” (2008) 38 *Seton Hall Law Review* 1379, 1391.

¹⁴ S. Domitrovich, “Fulfilling Daubert’s Gatekeeping Mandate Through Court-Appointed Experts” (2016) 106 *Journal of Criminal Law and Criminology* 35.

¹⁵ See, in the Australian context, *Genetic Institute Inc v Kirin–Amgen Inc (No 2)* (1977) 78 FCR 368.



*understanding the core issues relating to disputed evidence.*¹⁷

Thirdly, and more generally, the forensic sciences themselves need to be encouraged to adopt a “research culture”. This entails:

*“a commitment to conducting, participating in, and relying upon high quality empirical research.”*¹⁸

This idea was first propounded in a 2011 article in the UCLA Law Review by Jennifer Mnookin and others.¹⁹ More recently, Jonathan Koehler and John Meixner from North Western Law School have recommended:

*“a series of scientific studies that may provide guidance to legal decision-makers about the reliability and validity of forensic science conclusions.”*²⁰

Clearly, there is much to be done.

¹⁶ Ibid 41.

¹⁷ Ibid 48.

¹⁸ J. Koehler and J. Meixner, “An Empirical Research Agenda for The Forensic Sciences”, (2016) Journal of Criminal Law and Criminology 1, 31.

¹⁹ J. Mnookin et al, “The Need for a Research Culture in the Forensic Sciences” (2011) 58 UCLA L. Rev. 725

²⁰ Koehler and Meixner (supra n18) 5.