Judges as Gatekeepers: The Admissibility of Scientific Evidence Based on Novel Theories

The Honourable Mr Justice Sidney N. LEDERMAN*

* Superior Court of Justice, Ontario. The assistance of Emily Marrocco, a law clerk in the Superior Court of Justice, in the research for and preparation of this paper is gratefully acknowledged.

I. FORENSIC SCIENCE AND THE COURTS: AN UNEASY ALLIANCE

The testimony of experts is commonplace in Canadian courtrooms. Although courts have long recognized that an expert's opinions are necessary to provide the court with scientific technical information which is likely to be outside the experience and knowledge of the judge or jury, a certain amount of judicial skepticism has always surrounded such evidence. Perhaps the most strident criticism was expressed by Lord Justice Bramwell who after listening for a considerable time to expert witnesses whose opinions were diametrically opposed, tersely remarked that all witnesses can be classified as follows: "Liars, damned liars and experts."[1]

Although the traditional concern was that the expert testimony might well be influenced by the fact that the expert was being paid by one side and, thus, might give evidence in less than an independent fashion, the new battle ground relates to the danger of experts introducing what has been called pseudo-science or junk science into the courtroom.

This latest concern may have originated in the context of contingency fee toxic tort litigation in the United States. A "toxic tort" is a cause of action that arises when a plaintiff has developed a disease following long-term exposure to a physical agent—either a chemical or a form of energy such as electro-magnetic fields. Typically, it is alleged that the defendant's economic activity resulted in the plaintiff's exposure to the agent, thereby causing the plaintiff harm. Courts essentially have had to determine whether the plaintiff's exposure and subsequent disease are causally related, or whether the exposure and disease are associated merely by chance. For example, did the asbestos inhaled by the plaintiff cause his or her lung cancer? Or did the plaintiffs' exposure to radiation from high-tension power lines or to cellular phones cause their cancer? Did the anti-nausea drug taken by the plaintiff cause the birth defects that occurred thereafter?

---

[1] P. Anisman & R.F. Reid, eds., *Administration Law Issues and Practice* (Toronto: Carswell, 1995) at 196, footnote 14.

Expert testimony is tendered with respect to causal knowledge and furtherance of justice. The expert may employ the causal concepts of science when expressing purely scientific knowledge. Indeed, it is the layman's lack of such specialized knowledge that has been the fundamental justification for the receipt of such expert testimony. But many of these causation theories were new and their reliability has been questioned.

The concern about reliability of expert evidence has extended far beyond tort law to all cases in which scientific evidence is sought to be introduced. Historically, courts have treated science as an objective and dispassionate source of knowledge, but not a source of error. Based on this deferential view of science, courts found it unnecessary to ask the scientific expert, "How do you know?" In reality, however, science is no more objective and free of bias than are other areas of human endeavour. The courts' absolute faith in scientists, therefore, was misplaced.

In Canada, this concern found itself not so much in the context of toxic tort litigation, but rather in criminal law and the proliferation of expert testimony in the field of human behavioural sciences. Liberal admissibility resulted in the introduction of some questionable theories which may have led in some situations to wrongful convictions. Evidence of recovered memory of sexual abuse, years after the event in question, found a theoretical basis in expert testimony; but many experts in the same field have condemned it as false memory syndrome.

Having been embarrassed by permitting evidence of questionable reliability under the guise of scientific infallibility, courts began to look anew at how best to deal with the admissibility of such evidence.

It is not only courts that can be beguiled by pseudo or junk science. Renowned scientists themselves can deceive each other. We only have to think back to 1989 when the hype of "cold fusion" captured everyone's attention. The scientific world appeared to become excited by the discovery of low temperature fusion whereby power generation could be decentralized. The suggested benefits were far-reaching: each home could itself produce its own electricity, probably using a form of water as fuel. Even automobiles might be cold fusion powered. The discovery of cold fusion by two well-known scientists was accepted by many as a major breakthrough. Many scientists said that the phenomenon was real, and if developed, it will have profound effects on society. Only after some time, did other scientists come forward to debunk it and point out that the results are impossible. It was then described as "pathological science" or a hoax.

If scientists themselves have trouble distinguishing between good and bad science, what is expected of judges?

II. THE CANADIAN JURISPRUDENCE

The idea that a trial judge should function as a gatekeeper who assesses the legitimacy of an expert theory to determine whether it may be used in a particular trial is a relatively recent one in Canadian law. Prior to Sopinka J.'s 1994 decision in *R.* v. *Mohan*,[2] Canadian judges applied a fairly elastic standard and admitted evidence that met a low threshold of "helpfulness." All determinations concerning the reliability of the evidence were left to the jury. Under these older cases, such as *R.* v. *Fisher*,[3] a judge merely had to be satisfied that "the average person may not have sufficient knowledge of or experience within human behaviour to draw an appropriate inference from the facts." Judges were not required to rigorously scrutinize the evidence. Instead, the focus was on the capabilities of the jury.

Under the older cases, judges also focused on the issue about which the expert planned to testify. For example, until the rule against receiving testimony on the ultimate issue was laid to rest in *R.* v. *Graat*,[4] judges were concerned with excluding evidence that addressed the ultimate issue for fear that such evidence could usurp the function of the triers of fact. This was a relatively straightforward task, which focused only on the subject matter of the expert testimony, and not on the legitimacy of the field upon which the opinion was based. Once the ultimate issue rule was discarded, judges only had to assure themselves that the expert possessed proper qualifications, and that the evidence would be helpful to a jury. The adversarial tools of vigorous cross-examination and presentation of contrary evidence were the traditional and appropriate methods of attacking frail but admissible expert evidence.

With the decision in *Mohan*, the judicial function changed significantly. The Supreme Court articulated four factors upon which the admissibility of all expert evidence depends: relevance, necessity in assisting the trier of fact, the absence of any other exclusionary rule, and a properly qualified expert. Sopinka J. explicitly rejected the standard of helpfulness, finding it to be too low a threshold for admission. The imposition of these specific criteria had the effect of raising the threshold of admissibility for all

---

2 [1994] 2 S.C.R. 9.

3 [1961] O.W.N. 94, aff'd [1961] S.C.R. 535.

4 (1982) 31 C.R. (3d) 289 at 305 (S.C.C.).

kinds of expert evidence. For evidence based on novel theories, the criteria for admissibility tightened up even more; a judge actually had to assess the reliability of the evidence in order to determine whether it could be admitted. Although the validity or reliability of the evidence is not a specifically enumerated consideration in *Mohan*, Sopinka J. included these ideas under the relevance factor. In assessing relevance, *Mohan* requires a judge to conduct a cost-benefit analysis, weighing the probative value of the evidence against any prejudicial effect. A trial judge must consider whether the evidence is misleading, in the sense that its effect on a trier of fact, particularly a jury, is out of proportion to its reliability. Implicit in this calculation for novel scientific evidence is a determination of whether the expert's opinion is anchored in a type of science that is legitimate and reliable enough to go to the jury.

Since *Mohan* was more concerned with the admissibility of disposition evidence than novel scientific evidence generally, Sopinka J. did not provide many particulars about how to conduct the reliability analysis. He merely commented that:

> "[I]t appears […] that 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."

The Supreme Court chose to rely on neither the scientific community nor juries and the adversary system to determine what expert evidence is valid and reliable. Instead, the Court requires generalist trial judges to conduct a gatekeeping inquiry and reach conclusions concerning the basic threshold of reliability.

The preliminary issue for a trial judge faced with expert scientific evidence based on a novel theory is thus a reliability question.

Sometimes a judge must make this decision only on the basis of a short *voir dire*, held in the middle of a trial, with the added pressure of a jury waiting for the trial to continue. The law requires a judge to function like a gatekeeper, weeding out novel theories that lack a sufficient scientific foundation, either because they are too new, or because they are inherently unscientific and unreliable. To the parties to the dispute, and often to the administration of justice in general, the decision is critical. How, then, should judges who may have no scientific background approach the task of assessing expert scientific evidence based on novel theories?

Two categories of so-called "scientific" evidence have emerged in the cases—evidence grounded in "hard" physical and natural sciences, and evidence grounded in "soft" behavioural sciences and human nature. Evidence based on hard sciences might include evidence on DNA profiling, glass fracture analysis, fingerprinting, or handwriting analysis, while evidence based on "soft" sciences might include psychiatric and psychological assessments of battered wife syndrome, child accommodation syndrome, rape trauma syndrome, or delayed disclosure. Soft science evidence tends to be used to construct a social framework or background for a case, or to dispel longstanding myths about human behaviour.

The leading Canadian and American authorities from the 1990s attempted to lay out tests for the admissibility of expert evidence that could apply to all types of science—novel or established, hard or soft. They attempted to clarify the obligations and functions of the judge as gatekeeper. In reviewing and discussing those authorities, and the way that judges have been applying them, this paper will highlight the problems associated with using these tests to assess truly novel science, and the added problems associated with novel soft science. Sensitive to the reality that science does not nicely lend itself to assessment in an adversarial process, this paper will nonetheless attempt to draw some general conclusions about how expert evidence should be assessed.

III. CANADIAN CASES THAT HAVE CONSIDERED HOW TO ASSESS THE RELIABILITY

A fairly small number of Canadian cases have considered the way that reliability may be assessed in the cost-benefit analysis from *Mohan*. In *R.* v. *Terceira*,[5] the Supreme Court endorsed Finlayson J.A.'s analysis of the admissibility of expert opinion on the then "semi-novel" technique of DNA profiling. The technology was sufficiently mainstream at the time of the trial that the defence conceded that the technology was valid, which obviously significantly narrowed the issue. Finlayson J.A. interpreted *Mohan* as establishing that the "requirement of a basic threshold of reliability is met where the trial judge is satisfied as to the reliability of DNA profiling as a novel scientific technique."[6]

---

5 [1999] 3 S.C.R. 866, aff'g (1998), 123 C.C.C. (3d) 1 (Ont. C.A.).

6 *Ibid.* at para. 29.

Finlayson J.A. declined to enumerate a specific structure that must be adhered to in every case, but noted that the trial judge had to consider the nature of the proposed evidence, its foundation in science, and whether the expert had the necessary expertise to enable him or her to express an opinion in the field. He left it up to the trial judge to determine how far he or she has to go in meeting this threshold of reliability in a particular case. He also invited judges to seek guidance from the approaches taken in the pre-*Mohan* decisions in *R*. v. *Melaragni*[7] and *R.* v. *Johnston.*[8]

Melaragni offers a little more in terms of an approach to determining whether a novel scientific theory is valid. Moldaver J. (as he then was) conducted the reliability assessment by making a number of logical inquiries. First, he articulated the precise hypothesis of the science. I emphasize that this involved a consideration of what the science purports to do generally, not the conclusion that the expert claims the science supports for the particular case. The expert's conclusion has nothing to do with the judge's threshold issues of reliability and admissibility. Rather, the expert's insights into the particular case constitute circumstantial evidence of guilt or innocence, the weight of which must be assessed by the trier of fact.

After pinpointing the hypothesis of the science, Moldaver J. then considered the qualifications of the experts attesting to the validity of the proposition. In this way, he incorporated what is now the fourth *Mohan* requirement for a properly qualified expert into his reliability assessment. He went on to consider whether there were any specific studies or experiments which tested the validity of the proposition, and also analyzed whether the process for applying this novel science can be deemed "scientific," in the sense that it involves measurable and repeatable steps. In assessing expert opinion which aimed to undermine the novel science, Moldaver J. focused his assessment on whether the contrary opinion actually challenged the validity of the science or merely addressed the sufficiency or reliability of the other expert's method and conclusions.

In applying a *Melaragni* type of analysis, a judge must be able to conclude that the science is valid, and not guesswork. In other words, as Moldaver J. put it, a judge must be satisfied that the proposition upon which the expert bases his or her evidence amounts to something more than an invalidated or speculative hypothesis. Any evidence which suggests that the particular expert may not have been able to apply the valid science reliably

---

7 (1992), 73 C.C.C. (3d) 348 (Ont. Ct. Gen. Div.).

8 (1992), 69 C.C.C. (3d) 395 (Ont. Ct. Gen. Div.).

in forming his or her conclusions for the purposes of the particular case goes only to the weight of the evidence, and is a matter for the jury.

Terceira also pointed to *Johnston* as a useful reference point. In that case, the trial judge had to determine the admissibility of expert evidence on DNA profiling, which, even at that time in 1992, was generally accepted in the scientific community. Langdon J. exhibited a sensitivity to the way that scientific disciplines can overlap, and began his analysis by noting the two distinct branches of science involved in DNA profiling—molecular biology and population genetics. He then examined scientific authorities and attempted to draw conclusions about the validity of each branch of science. He made the useful observation that "as a matter of common prudence, no matter how reliable may be the expertise of a particular scientific discipline, one must exercise extreme caution in areas where different disciplines interface." He also considered whether any other courts in Canada or the United States had accepted DNA profiling.

Ultimately, Langdon J. put forth a list of fourteen factors to balance in reliability analysis. Despite the length of the list, he clearly did not intend that the test be unduly complex or restrictive. On his list of indicia of reliability, he included the potential rate of error, the existence and maintenance of standards, the care with which the scientific technique has been employed and whether it is susceptible to abuse, whether there are analogous relationships with other types of scientific techniques that are routinely admitted into evidence, the presence of failsafe characteristics, the expert's qualifications and stature, the existence of specialized literature, the novelty of the technique in its relationship to more established areas of scientific analysis, whether the technique has been generally accepted by experts in the field, the nature and breadth of the inference adduced, the clarity with which the technique may be explained, the extent to which basic data may be verified by the court and jury, the availability of other experts to evaluate the technique and the probative significance of the evidence.

The *Johnston* factors merit some discussion. While they may be helpful to the extent that they give a judge some more specific factors to consider, they do not solve all of the problems associated with the assessment of novel theories. For one thing, it must be remembered that the factors were created in the context of DNA profiling evidence, which already enjoyed acceptance in the community. As a result, the enumerated factors which are associated with peer review could actually be applied. For a truly novel theory, these factors would be inapplicable, and would point towards

exclusion. As has been pointed out by Brad Limpert, in his article, "Beyond the Rule in *Mohan*: A New Model for Assessing the Reliability of Scientific Evidence",[9] some of the *Johnston* factors actually show a bias against truly novel techniques, including the peer review factors, the factor requiring a consideration of the novelty of the technique in relation to more established areas of scientific analysis, and the inquiry into whether there are analogous relationships with routinely admitted scientific techniques.

I further echo Limpert's observation that the *Johnston* test does not provide any indication of how the factors should be balanced or which might be more significant. Further, some of the factors might be unworkable or difficult to interpret, including the requirement for "failsafe characteristics" and the extent to which basic data may be verified by the jury.

In my view, there is an additional potential problem. The third factor, which examines the care with which the technique has been employed and whether it is susceptible to abuse, seems like it may wrongly encourage a judge to venture into questions of weight when determining admissibility. Issues concerning the application of the technique in the formation of the expert's conclusion go mainly to weight. The fact that a technique may have been abused does not automatically go to basic threshold reliability or admissibility, unless it is intertwined with the question of whether the expert is properly qualified to give the evidence in the first place, or unless the opinion is not rationally connected to the data underlying it.

In *R.* v. *J.E.T.*,[10] Hill J. considered the application of the *Mohan* test to what he classified as the novel area of soft science involving the symptomatology of child sexual assaults and the impact of child abuse on child behaviour. He held that one criterion of reliability is whether there exists an acceptable body of evidence of acceptance of the theory to validate the opinion objectively. He articulated a number of additional factors that can be considered when deciding if a technique is sufficiently reliable:

> "Needless to say there is a continuum of reliability in matters of science from near certainty in physical sciences to the far end of the spectrum inhabited by junk science and opinion akin to sorcery or magic. Whether the technique can be demonstrably tested, the existence of peer review for the theory or technique, the existence of publication, the testing or validation employing control and error measurement, and some recognition or acceptance in the relevant

---

9 (1996) 54 U.T. Fac. L. Rev. 65 at 82.

10 [1994] O.J. N° 3067 (QL).

> scientific field all contribute to an assessment of the reliability of the opinion and hence its capacity to outweigh the prejudicial impact of imposing on the jury highly suspect opinion evidence masquerading as science."[11]

Hill J. went on to stress that there will inevitably be conflicting opinions in truly novel areas, where study and experimentation are especially recent. He held that this did not raise the spectre of unreliability, and that consensus among scientists is not required to admit opinion evidence. In my view, this is an important observation, which highlights Canada's longstanding rejection of the old *Frye*[12] test in American jurisprudence. Although a number of Canadian tests for reliability have included peer review type factors, such review cannot be deemed a necessary pre-condition to admissibility. In "hard cases",where judges must assess truly novel scientific areas, there will not be a body of scientific literature to which a judge can turn to validate the field of study. There must be other factors which can assist a judge in making this determination on his or her own. Peer acceptance can be an extremely helpful characteristic in the cases where it is available, but it should not be applied so as to create a bias against the admission of very novel scientific evidence.

Hill J. also found that a significant percentage of error in test results may not, by itself, render evidence inadmissible. On the topic of error rates, I emphasize that there are two potential issues associated with error—one which goes to the basic threshold issue of reliability, and one which goes only to weight. In my view, only the error rate associated with the ability of the technique to realize the hypothesis upon which the field is based goes to threshold reliability, or admissibility. In contrast, the error associated with the particular expert's application of the science goes to weight, and may be left to the jury. As long as evidence meets the other *Mohan* criteria and the opinion is rationally connected to the data, a judge may still admit such evidence.

The proper application of this distinction ensures that the adjudication of the merits of the proposed evidence is properly left to the jury. The judge must not usurp the role of the jury by assessing the weight of the specific expert's opinion on the *voir dire*. If the reliability of the expert's specific examination is used to determine admissibility, then the expert will essentially be providing an opinion on the ultimate issue, the content of

---

[11] *Ibid.* at para. 75.

[12] See *infra* note 14.

which has the stamp of reliability and validity from the judge. In such a scenario, there is nothing left for the jury to decide.

IV. AMERICAN JURISPRUDENCE THAT HAS CONSIDERED WAYS TO ASSESS RELIABILITY

Prior to the leading decision of the United States Supreme Court in *Daubert* v. *Merrell Dow Pharmaceuticals Inc.*,[13] the American approach was anchored in the general acceptance test from *Frye* v. *United States*.[14] According to that test, scientific expert evidence was not admissible unless it had been generally accepted in the particular scientific community to which it belonged. By essentially deferring the assessment of legal reliability to the expert's peers, the test took considerable "heat" off the trial judge. At the same time, the test met with the same criticisms that I have discussed already in relation to the peer review factors in Canadian jurisprudence—namely, it pointed towards the exclusion of potentially useful novel science, since it had not yet been considered by the relevant scientific community. The *Frye* test was also criticized for vagueness, in the sense that it was often difficult to determine the relevant scientific community and/or the level of acceptance within it.

In *Daubert*, the United States Supreme Court revisited the *Frye* test in light of Rule 702 of the Federal Rules of Evidence. The Rule provides that:

> "If scientific, technical or other specialized knowledge will assist the trier of fact to understand the evidence or determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training or education, may testify thereto in the form of an opinion or otherwise."

The rule is broadly worded, both in terms of the qualification of the expert, and in terms of the type of information that can form the subject matter of expert opinion. The *Daubert* Court reached the unanimous conclusion that nothing in the text of Rule 702 established that the general acceptance standard from *Frye* was an absolute pre-requisite to admissibility. The Court held that insistence upon the rigid *Frye* standard would be at odds with the liberal thrust of the Rules and their general approach towards relaxing the traditional barriers to opinion testimony.

---

[13] 113 S. Ct. 2786 (1993).

[14] 293 F. 1013 (D.C. Cir. 1923).

The *Daubert* Court held that Rule 702 imposes two distinct requirements on the assessment of scientific expert evidence—reliability and relevance (in the sense that it is helpful to the jury to understand the evidence or a fact in issue). Again, *Daubert* is clear that reliability refers to the reliability of the methodology upon which the expert's conclusion is based, and not the reliability of the conclusion itself. The trial judge must ensure that an expert's conclusion rests on a reliable foundation or has a reliable basis in the knowledge and experience of the expert's discipline. To this end, *Daubert* offers four factors to consider. A judge must consider whether the theory or technique can be and has been tested; whether the theory or technique has been subjected to peer review and publication; the known or potential rate of error or the existence of standards, and whether the theory or technique has been generally accepted.

There is some clear overlap between the factors. For example, the requirements of testing, peer review, and general acceptance seem very similar to each other and import the old ideas from *Frye*. In this way, while the *Daubert* test is designed to appear more flexible and to be used in a more flexible way, it is not a radical departure from *Frye*. In addition, the factors pertaining to testing and error rates also seem interrelated. Part of testing the validity of a technique surely includes an assessment of error rates.

Essentially, the *Daubert* test seems to call for two types of evaluation by the trial judge. First, a judge can independently evaluate the inherent reliability of the methodology, through an examination of standards, error rates, and tests. Second, the judge can look to the scientific community and consider any review or specialized literature. It does not suggest which of the two approaches should be given more weight, although peer review seems to be more featured in the four factors. In terms of specifics, although the test provides more guidance than the Canadian cases, it does not establish a systematic approach that could be applied across the board in a consistent way.

In the 1997 decision of the United States Supreme Court in *General Electric Co.* v. *Joiner*,[15] the Court began to blur the methodology—conclusions distinction from *Daubert*. In *Joiner*, the plaintiff sued the defendant, manufacturers of PCBs, on the grounds that his exposure to those products promoted his development of lung cancer. The plaintiffs sought to rely on experts who had conducted studies of laboratory animals. The Supreme Court reversed the Court of Appeals, and restored the District

---

[15] 118 S. Ct. 512 (1997).

Court's exclusion of the expert opinion. The Court emphasized the important role of the judge as gatekeeper, and the majority held that the plaintiffs never explained how and why the experts could have extrapolated their opinions from the seemingly far-removed animal studies. The Supreme Court endorsed the view that evidence may be excluded when there is too great an analytical gap between the data and the opinion proffered.

The interesting part about the majority decision is the fact that it does not mention any of the *Daubert* factors or attempt to apply them. It does not even attempt to classify the area of expertise of the proposed experts, in order to assess the methodology. It simply considers whether an expert can formulate an opinion on cancer causation based solely on animal studies.

In my view, *Joiner* should have been approached differently. First, the trial judge should have classified the science in which the proposed experts claim to be experts—something like—experts in cancer causation in humans based on exposure to toxins. Then, the judge should have attempted to assess the reliability of this science as a science. To answer this question, the judge could have looked to the *Daubert* and *Johnston* type factors. The inquiry should have focused on whether the methodologies used in this science generally can yield sufficiently reliable and accurate results. Then, if the judge felt that the science was reliable, the judge could have considered the relevance of the testimony, and whether the expert was personally properly qualified to be an expert in the field. If these hurdles were overcome, the issue could have proceeded to trial.

This point was essentially made in the dissenting opinion of Justice Stevens, who commented that:

> "The reliability ruling [in the District Court] was more complex and arguably is not faithful to the statement in *Daubert* that '[t]he focus, of course, must be solely on principles and methodology, not on the conclusions that they generate.' […] *Daubert* quite clearly forbids trial judges from assessing the validity or *strength* of an expert's scientific conclusions, which is a matter for the jury. Because I am persuaded that the difference between methodology and conclusions is just as categorical as the distinction between means and ends, I do not think the statement that 'the conclusions and methodology are not entirely distinct from one another, […]' is either accurate or

> helps us answer the difficult admissibility question presented by this record."[16]

While I agree with the point made by Justice Stevens in dissent, courts should be able to exclude evidence that is actually not derived from the sound methodology that the expert puts forth as the basis for his or her opinion. Professor Michael Saks noted in "The Aftermath of *Daubert*: An Evolving Jurisprudence of Expert Evidence"[17] that although *Joiner* muddied the distinction between methodology and weight, it espoused a sensible proposition that conclusions may be rejected on the basis that they are an illogical step away from the data. This should be an independent inquiry. Courts should exclude such evidence not because they scrutinized the merits of the expert's particular conclusion in detail, but because the conclusion cannot actually be based on the data. This would be consistent with both *Mohan* and *Daubert*, since they seek to exclude evidence that is not tethered to reliable data. *Joiner* should not be taken to stand for the more general proposition that a particular conclusion must be reliable; rather, a conclusion must be anchored in and derived from a reliable scientific methodology.

The United States Supreme Court recently considered the question of the admissibility of expert evidence again, in the 1999 decision of *Kumho Tire Co.* v. *Carmichael*.[18] In *Kumho,* the plaintiff sued a tire manufacturer when his tire blew out and his passengers were injured. He sought to call a tire failure analyst to give evidence that the tire was defective. The Supreme Court upheld the trial decision to exclude the evidence, but confirmed that the gatekeeping role from *Daubert* was intended to apply to all testimony based on specialized knowledge, whether it is scientific or technical. More importantly for our purposes, they also endorsed a flexible application of the *Daubert* factors. The Court recognized that all of the factors may not apply in every case, and that the gatekeeping inquiry must be tied to the particular facts of each case. The list of factors is meant to be helpful, not definitive. This is an important holding for two reasons. First, it confirms that peer review is not required, which eliminates a potential bias towards truly novel science. The Court specifically noted that "it might not be surprising in a particular case […] that a claim made by a specific witness has never been the subject of peer review, for the particular application at issue may never have previously interested any scientist." Second, the holding confirms that

---

16 *Ibid.* at para. 41.

17 (2000) 40 *Jurimetrics Journal of Law, Science and Technology* at 229.

18 119 S. Ct. 1167 (1999).

judges should conduct independent assessments of the methodology of a novel science, and may look beyond the *Daubert* factors to any other factors that seem suited to the particular facts of the case. Obviously, the extension of *Daubert* will raise new challenges, as courts struggle to assess areas of specialized knowledge that look less and less like typical sciences.

Interestingly, an amendment to Rule 702 has recently been proposed. A revision to Rule 702 proposed by the National Conference of Commissioners on Uniform State Laws would establish a presumption that a methodology is reliable if it has met with substantial acceptance by the relevant scientific community but would be presumed to be unreliable where it is not generally accepted. This would discourage junk science by making it difficult, but not impossible, to introduce theories that have not yet gained significant support.

V. RELIABILITY ASSESSMENT: HARD SCIENCES *VERSUS* SOFT SCIENCES

Although the American and Canadian cases are clearly attempting to provide a framework for assessing the reliability of all novel science evidence, in reality, our courts continue to approach hard and soft sciences differently. For soft science questions, judges tend to focus on a different aspect of the *Mohan* test when making admissibility determinations. Judges have concentrated on whether behavioural science evidence is necessary to assist the trier of fact, rather than on threshold reliability.

Only a small number of cases have attempted to tackle the reliability of soft science. In *R.* v. *Olscamp*,[19] Charron J. (as she then was) reviewed the state of knowledge in the psychology field and held that the soundness and reliability of expert opinion concerning child accommodation syndrome could not be demonstrated. She concluded that if there was any consensus among experts in the field, it was that there was no valid profile from which one could identify abused children. Similarly, in *J.E.T.*, discussed above, Hill J. spoke generally about reliability assessment in a case involving the behavioural indicators of child sexual assaults. In the 1997 Court of Appeal decision in *R.* v. *McIntosh & McCarthy*,[20] Finlayson J.A. made some general comments about assessing the reliability of expert evidence concerning the frailties and inherent weaknesses of eyewitness testimony. He held that courts cannot be overly eager to abdicate their fact-finding responsibility to

---

[19] (1994), 35 C.R. (4th) 37 (Ont. Ct. Gen. Div.).

[20] (1997), 35 O.R. (3d) 97 (C.A.).

experts in the field of behavioural sciences, and that before a witness can be permitted to testify as an expert, the court must be satisfied that the subject-matter of his or her expertise is a branch of study in psychology concerned with a connected body of demonstrated truths or with observed facts systematically classified and more or less connected together by a common hypothesis operating under general laws. The branch should include trustworthy methods for the discovery of new truths within its own domain. He later added that the scientific method requires "the formation of a hypothesis, the testing of the hypothesis using reliable methodology, the examination of the results (usually with statistical analysis) and the formation of a conclusion." Like the hard science cases, *McIntosh* emphasizes the assessment of the reliability of the underlying science and its methodologies, and not an assessment of the application of such methodologies by the expert in question. At the same time, *McIntosh* is not a typical behavioural science case, since the expert was dealing with the limitations of eyewitnesses, most of which are related to the memory, not human psychology.

Most recently, in *R.* v. *A.K.*,[21] the Court of Appeal considered the admissibility of expert testimony concerning the behavioural indicators of child abuse. In discussing the relevance requirement from *Mohan*, Charron J.A. raised a number of questions for judges to consider, including the extent to which the opinion is founded on proven facts, and the extent to which the expert opinion evidence supports the inference sought to be made from it. Charron J.A. also addressed the assessment of reliability for novel soft science theories. She emphasized that reliability concerns the validity of the basis of the expert's opinion, as opposed to the opinion itself. She also discussed the importance of scrutinizing novel areas of recognized fields:

> "Although psychology or sociology are certainly recognized fields of expertise, some theories advanced in courtrooms in recent years within those fields are entirely novel. Further […] the state of scientific knowledge is fluid. The fact that a particular theory may have been accepted in the past does not end the inquiry."[22]

---

21 (1999), 45 O.R. (3d) 641.

22 *Ibid.* at para. 86.

Charron J.A. highlighted the relationship between the factors of reliability and necessity, noting that "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."[23]

Despite these clear attempts to apply standard reliability analysis to soft science cases, it is not difficult to see why judges have hesitated to delve into the issue. Indeed, it is easy to understand why a judge might feel ill-equipped to decide that one social science theory, such as battered woman's syndrome, is more reliable than another theory, such as battered child syndrome. While the Supreme Court of Canada has accepted the scientific validity of battered wife syndrome in *R.* v. *Lavallée*[24] and more recently in *Malott* v. *The Queen*,[25] some American Courts have held that battered child syndrome (which purports to explain why abused children commit criminal acts against their parents) is inadmissible for lack of a scientific foundation. How is a judge to tell the difference? This is especially true for a judge faced with a truly novel soft science theory that has not been the subject of extensive peer review. In addition, after the problems associated with once popular expert evidence on child sexual abuse accommodation syndrome and recovered memory, judges are alert to the potential to be "fooled" by trendy soft science theories.

Although Finlayson J.A. in *McIntosh* talked about examining social sciences to determine whether they employ the scientific method, the bottom line is that many social science theories do not derive from statistical or empirical data. As Steven Skurka and Elsa Renzella noted in "Misplaced Trust: The Courts' Reliance on the Behavioural Sciences",[26] "the behavioural sciences typically do not adopt the formal scientific methodology used in the fields of natural sciences. Instead the analysis will usually involve case comparisons and take on […] 'an anecdotal' quality."

How then, is a judge supposed to conduct reliability analysis of a soft science grounded in social science research methods such as interviews and field work? This is even more difficult and arbitrary than assessing the reliability of novel hard science evidence. How can a judge recognize the problems common to such research, including problems with selection bias, inconsistent interviewing, and lack of long-term follow-up? In her paper,

---

23 *Ibid.* at para. 84.

24 [1990] 1 S.C.R. 852.

25 (1998), 121 C.C.C. (3d) 456 (S.C.C.).

26 (1998) 3 Can. Crim. L. Rev. 269 at 281.

"Lies, Damned Lies, and Statistics: Psychological Syndrome Evidence in the Courtroom After *Daubert*",[27] Krista Duncan considered how *Daubert* should be applied to assess expert soft science evidence. She focused on the peer review factors in the *Daubert* analysis, and concluded that while combat induced post-traumatic stress disorder, rape trauma syndrome, and battered wife syndrome were sufficiently "good science", battered child syndrome had not been adequately tested in the scientific community to be admitted at trials. In my view, this approach essentially equates *Daubert* and *Frye*. At the same time, Duncan's need to resort to *Frye*-type analysis highlights the difficulties associated with coming up with independent reliability indicia for soft science. Other commentators have objected to the admissibility of rape trauma syndrome on the grounds that it is a pseudo-science, whose name pre-judges the very question of whether a rape has occurred.[28]

R. Nichwolodoff[29] noted that most psychologists use an eclectic approach in their practices in which they incorporate various bits of multiple theories, including the suspect ones. He concluded that many psychological techniques are largely unreliable in terms of diagnosis, treatment and behaviour prediction. He echoed the view that many judges are unable to assess the scientific validity of psychological expert opinion evidence, since they do not have experience with concepts such as falsifiability. Moreover, Nichwolodoff highlighted the further problems that can develop when judges attempt to balance and choose between conflicting psychological opinions. In many instances, a psychological expert attempting to stay within the limits of his or her science will seem tentative and less persuasive than other less reliable experts. Nichwolodoff suggested tightening admissibility standards for soft science and limiting psychological professionals to testimony concerning their observations. He further suggested the use of court-appointed *amicus curiae* to assess the credibility of soft science fields tendered as science.

It is obvious that there is no easy answer to the question of how these assessments should be done. Although we have tried to move away from the *Frye* test, when it comes to soft science, it is very difficult to think of ways that a judge could conduct reliability analysis completely on his or her own. The adversarial process does not neatly lend itself to the evaluation of such sciences. As a result, as the Duncan paper illustrates, *Frye* remains. Or,

---

27 (1996), 71 Ind. L.J. 753.

28 R. Nichwolodoff, "Expert Psychological Opinion Evidence in the Courts" (1998) Health L.J. l279.

29 *Ibid.*

judges "skirt" the issue completely and focus largely on the necessity of the evidence. Or, they may rely on other judicial decisions without ascertaining whether those rulings were predicated on a close scrutiny of the science.

VI. ASSESSING RELIABILITY IS NOT THE END OF THE GATEKEEPING FUNCTION

Relevance

Establishing that the evidence meets the basic threshold of reliability is often the greatest hurdle for a judge assessing a novel scientific theory, and the most important component of relevance. Yet, it is important to remember that once the evidence meets this threshold, a judge must move on to consider the other indicia of relevance before deciding whether or not to admit the evidence. A judge must be satisfied that the evidence is relevant to a fact in issue in the proceedings, and that its benefits or probative value outweigh its costs. In other words, the judge must assess both logical and legal relevance. Of course, the typical concern is that evidence dressed up in scientific language might be given extensive weight and misused by a jury.

Necessity

In *Mohan,* Sopinka J. also held that a judge must be satisfied that the evidence is necessary in the sense that it provides information which is likely to be outside the experience and knowledge of a judge or jury. The evidence must be necessary to enable the trier of fact to appreciate matters in issue due to their technical nature, and essential in the sense that a trier of fact will be unable to reach a satisfactory conclusion without it. Although necessity should not be judged by too strict a standard, the application of the principle must be more strict when the evidence approaches an opinion on the ultimate issue. The need for the evidence must always be assessed in light of its potential to distort the fact-finding process.

For novel soft science theories, necessity is often a difficult threshold to overcome. Since the soft sciences tend to provide insight into human behaviour, it can be difficult to establish that expert opinion is needed to help juries make such determinations. Indeed, our system is founded on the idea that juries are effective because of their collective common sense, experience and knowledge of human nature. In *R.* v. *Lavallée*,[30] the Supreme Court first recognized that soft sciences might be used to assist the trier of fact on issues relating to human nature, and to dispel myths and stereotypes.

---

[30] *Supra* note 23.

Following this decision was a flood of cases in which soft science expert evidence was admitted at trials to dispel myths and provide insight into behavioural characteristics, especially in cases involving child behaviour: see *R.* v. *B(G.)*;[31] *R.* v. *Marquard*.[32] In the 1994 decision of *R.* v. *R.H.B.*,[33] the Supreme Court went as far as to allow an expert to testify that a complainant was sexually abused, and to provide reasons why he held such a belief. According to Steven Skurka and Elsa Renzella, this case marked a high point in the Supreme Court's enthusiasm for soft science.

After the revision of admissibility standards in *Mohan*, courts have attempted to scrutinize soft science expert evidence more carefully. After *Olscamp*, discussed above, courts began excluding once popular soft science expert evidence. In *R.* v. *François*,[34] the Supreme Court concluded that the jury was capable of ruling on the validity of a complainant's claim of recovered memory without expert assistance. In *R.* v. *W.S.*,[35] Langdon J. refused to admit evidence about child sexual abuse accommodation syndrome on the grounds that any probative value was outweighed by prejudicial effect, since the evidence functioned as a crutch which handily supported any witness. In *R.* v. *C.(G.)*,[36] expert evidence concerning delayed sexual abuse reporting was ruled reliable but unnecessary. Similarly, in *R.* v. *Deschamps*,[37] Finlayson J.A. found numerous reasons to exclude evidence on late disclosure, finding that it was both unnecessary and irrelevant.

R. v. *A.K.*[38] offers the most recent articulation of the types of questions that judges should ask when considering necessity. Such questions include whether the proposed expert opinion will enable the trier of fact to appreciate the technicalities of a matter in issue, whether the evidence will provide information which is likely to be outside the experience of the trier of fact, and whether the trier of fact is unlikely to form a correct judgement without the assistance of the expert evidence. Relying on *R.* v. *F.(D.S.)*,[39]

---

31 [1990] 2 S.C.R. 30.

32 [1993] 4 S.C.R. 223.

33 [1994] 1 S.C.R. 656.

34 [1994] 2 S.C.R. 827.

35 (1996), 47 C.R. (4th) 354 (Gen. Div.).

36 (1997), 8 C.R. (5th) 21 (Ont. Gen. Div.). Also see: *R.* v. *Villamar* which made a similar finding on the same issue: [1996] O.J. N° 2742 (Gen. Div.) (QL), aff'd [1999] O.J. N° 1923 (C.A.) (QL).

37 Ontario Court of Appeal (October 7, 1998), unreported.

38 *Supra* note 20.

39 (1999), 43 O.R. (3d) 609 (Ont. C.A.).

Charron J.A. recognized that there is no exact way to draw the line between what is within the normal experience of a judge or jury and what is not. She also recognized that expert evidence which touches on credibility presents an even more difficult task for the trial judge in the application of the criterion of necessity. Evidence cannot be ruled necessary if it is solely offered to show that a complainant is telling the truth.

The Absence of Any Exclusionary Rule & Properly Qualified Expert

The final two criteria from *Mohan* cannot be overlooked, but are relatively straightforward. For novel science cases, where the judge does not have the benefit of considering peer review of the science, the qualifications of the expert can take on a special significance.

The current trend seems to be towards tightening admissibility standards, especially in cases involving expert opinion on the subject of human behaviour. The reason for this is best summed up by one noted Canadian legal author as follows:

> "Trials are invariably about human behaviour and "experts" can invariably be found to support a party's characterization of human behaviour. Unless we restrict such evidence to cases where it is truly necessary and demonstrably reliable and unless we exert tight control over where it is allowed, we will not be serving justice. We will be imperilling it."[40]

While the *Mohan* criteria (besides reliability) are relatively easy to apply to novel hard science, novel soft science poses more challenging questions. Before admitting such evidence, trial judges must be prepared to critically examine the proposed evidence through the scientific microscope and consider the dynamics of the trial to determine whether expert assistance is necessary for the jury to evaluate the evidence or draw inferences from it. It is also important to weigh the potential prejudicial effect of the evidence against any probative value and to consider whether any limiting instructions should be incorporated into the charge to complement the admission of the evidence.

---

[40] D.M. Paciocco, "Coping with Expert Evidence About Human Behaviour" (1999) 25 Queen's L.J. 305 at 346.

VII. PROCEDURE

Procedure is obviously an important part of gatekeeping. While Terceira confirmed that the issue of reliability respecting novel scientific theory or technique relates strictly to a question of the admissibility of evidence where proof on a balance of probabilities is an acceptable standard, Canadian law does not lay out a comprehensive set of the rules of procedure and disclosure that should be followed in the assessment of expert testimony. If judges are to perform this difficult task well, they will need a clear and helpful procedure through which they can collect the information, assess the expert opinion, and balance probative value against prejudicial effect. Counsel also need clear direction about the scope and timeline of their disclosure obligations.

In my view, the assessment of expert witness admissibility should be determined prior to trial, in a discovery-like process. Indeed, a similar process is followed in the United States. The *Daubert* Court made reference to a number of Federal Rules of Evidence, including Rule 104(a), which instructs a judge to determine preliminary questions concerning an expert's qualifications and the admissibility of evidence; Rule 703, which permits an expert to base an opinion on facts that would otherwise be inadmissible; Rule 104(b), which instructs the judge insist that each piece of evidence be connected to other evidence to ensure a sufficient foundation; and Rule 706, which explicitly allows a judge to call an expert of his or her own choosing. There have also been disclosure amendments to Rule 26(a) of the Federal Rules of Civil Procedure, which require litigants to disclose the opinions to be expressed by the expert and the basis and reasons therefor.

In Canada, the Rules are less comprehensive and, generally, expert witness assessment occurs once the trial is already in progress. Obviously, the application of the *Mohan* test requires a *voir dire*, which could become quite lengthy. Ideally, this should be done before the trial is underway, to give a judge more time to rule on the admissibility of the theory, and to determine if other evidence is necessary for this determination to be made. This would also assist counsel, who would know in advance whether their expert would be permitted to testify. Pre-trial discoveries or hearings would require advance disclosure by counsel, including disclosure of the methodology upon which the expert anchored his or her conclusions. The absence of a clear pre-trial procedure and clear disclosure obligations compounds the difficulties associated with gatekeeping for Canadian judges.

CONCLUSION

The courtroom is not the ideal place for assessing new science, and the judge is not the ideal person. It is especially difficult for judges to make determinations concerning the reliability of science, and to maintain the blurred distinction between threshold and ultimate reliability, or admissibility and weight. In light of this, some commentators continue to encourage reverting back to a form of the *Frye* test. American Professor Michael Graham[41] continues to submit that judges cannot be amateur scientists, and that they must go back to the "do-able" role of determining whether an explanative theory is sufficiently trustworthy to present to the trier of fact, by examining the theory in the context of the field in which it belongs. In situations where there has been insufficient peer review to make this determination, Professor Graham encourages judges to consider particularized earmarks of trustworthiness by ascertaining whether the theory was derived in a manner consistent with the processes customarily used by experts in the field. In either case, the trial court's function is to look to experts in the particular field for assistance in determining whether the theory is sufficiently trustworthy. Professor Graham submits that gatekeeping cannot involve a judicial assessment of whether a theory actually works; the best a judge can do is decide that there are sufficient assurances of correctness to warrant acceptance by the trier of fact.

In my view, peer review and peer assessment cannot be pre-conditions to admissibility. Such a test would contain a strong bias against truly novel scientific evidence, which may be helpful to the trier of fact. While Canadian judges should be sensitive to the differences between hard and soft sciences when conducting reliability assessments, and careful about loosening admissibility for truly novel soft science, we should not adopt the rigidity of *Frye*.

For soft science, it is better to commence with the easier examination of necessity and to only consider reliability once the usefulness of the evidence is confirmed. In evaluating reliability, courts might consider employing an *amicus curiae*, and might give the peer review factor more consideration than for hard science assessments. Judges must be alert to the research methods used, whether they are capable of verification, and whether the opinion is anchored in facts that can be tested. At the same time, while

---

[41] M. Graham, "The Expert Witness Predicament: Determining 'Reliable' Under the Gatekeeping Test of *Daubert*, *Kumho* and Proposed Amended Rule 702 of the Federal Rules of Evidence" 54 U. Miami Inter-Am. L.Rev. 316.

tightening admissibility for soft science may well be a valid agenda in many cases, judges must be careful not to stray too far into the question of weight, and exclude evidence which could be potentially useful.

For more hard sciences and other forms of technical knowledge, judges are better equipped to make some determinations on their own. We should follow the broad language of American Rule 702 and not focus on classifying the field as scientific or technical or specialized. Instead, trial judges should simply attempt to assess the reliability of the field upon which the expert bases his or her opinion. This should be done through the identification of the hypothesis of the science and the ability of the methodologies in the science to yield accurate results. This might include a consideration of whether the methodology involves logical, systematic steps or whether it seems more like guesswork and speculation. Courts can also consider whether a field addresses an empirical matter and what evidence demonstrates how well it can perform that empirical matter.[42] Another question to consider is whether the field is capable of self-testing and whether any kind of consensus on dependability is emerging. A wide array of factors can be canvassed, none of which are determinative and none of which have mandatory application. Judges should be careful not to stray into the question of weight and must not become preoccupied with the expert's own conclusions, with weaknesses in the expert's personal application of the science, or with our own opinion of whether the expert's ultimate opinion seems valid. The distinction between methodology and conclusion must be maintained, in spite of *Joiner*. At the same time, courts must be satisfied that the opinion is rationally supported by the data. Finally, in considering contrary expert opinion which purports to challenge the validity of an expert's field of science, trial judges should attempt to ensure that it is truly contradicting the validity of the scientific foundation underlying the opinion before it is used to exclude the evidence.

Gatekeeping is not an easy function to perform. Onto the already difficult job of scrutinizing the needs, capabilities and reactions of the jury, gatekeeping further mandates judges to scrutinize science. For all scientific evidence, the difference between threshold and ultimate reliability is fuzzy at best. For novel science, and especially novel soft science, it is even fuzzier. In fleshing out reliability criteria and applying the remaining *Mohan* factors to perform the gatekeeping function, judges must be wary of one serious potential consequence. As we attempt to ensure that experts do not usurp the

---

42 *Supra* note 16 at 239.

function of the jury, we must guard against using the guise of the gatekeeper to usurp that same function ourselves.

With this new gatekeeping function, when dealing with evidence about which there is limited agreement in the scientific community, trial judges will have to reconcile genuine disputes among scientists. When different courts review scientific methodology with different degrees of scrutiny, inconsistent findings over admissibility are bound to occur. Some commentators have suggested that the problems of inconsistency among trial judges require that we establish specialized science courts or scientific appellate/advisory panels.[43] For many reasons, these suggestions may not be practical or would create further problems of their own.

Nevertheless, at the present time, in the absence of guidelines to achieve uniform scientific adjudication, trial judges, in exercising this screening responsibility, will simply have to try to learn all they can about the particular science or technology in issue.

---

[43] J.W. Osborne, "Judicial/Technical Assessment of Novel Scientific Evidence" (1990) U. of Ill. L. Rev. 497.