NATIONAL JUDICIAL INSTITUTE
INSTITUT NATIONAL DE LA MAGISTRATURE

Science Manual for Canadian Judges

Last updated July 1, 2018

Published by the National Judicial Institute

275 Slater Street, Suite 2000
Ottawa, ON
K1P 5H9
Canada
Phone/Tél. : 613-237-1118 Fax / Téléc.: 613-237-6155
Email: info@nji-inm.ca
Website: www.nji-inm.ca

Note: The scale and clock images used throughout Chapter 2 appear courtesy of Microsoft.

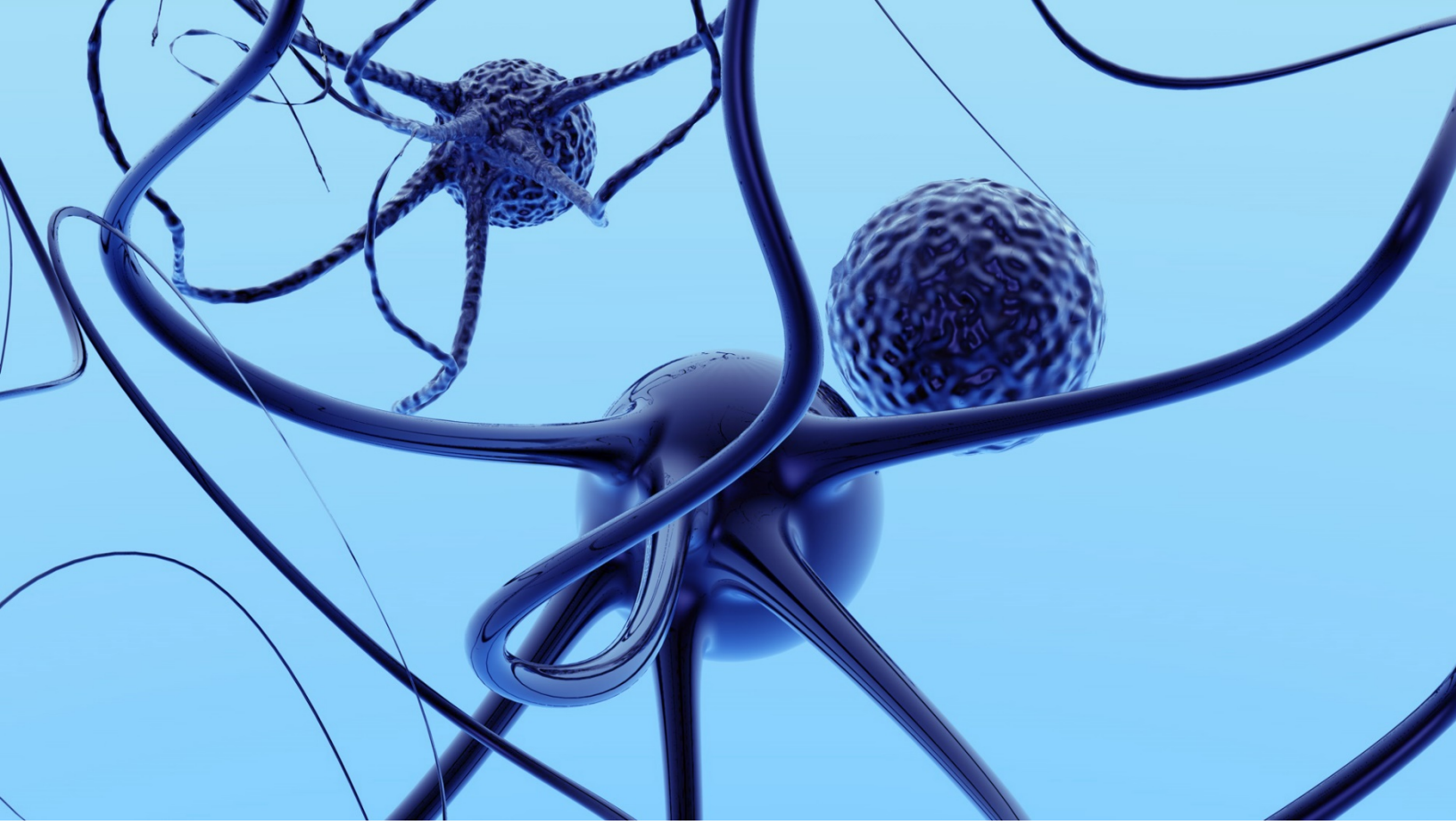
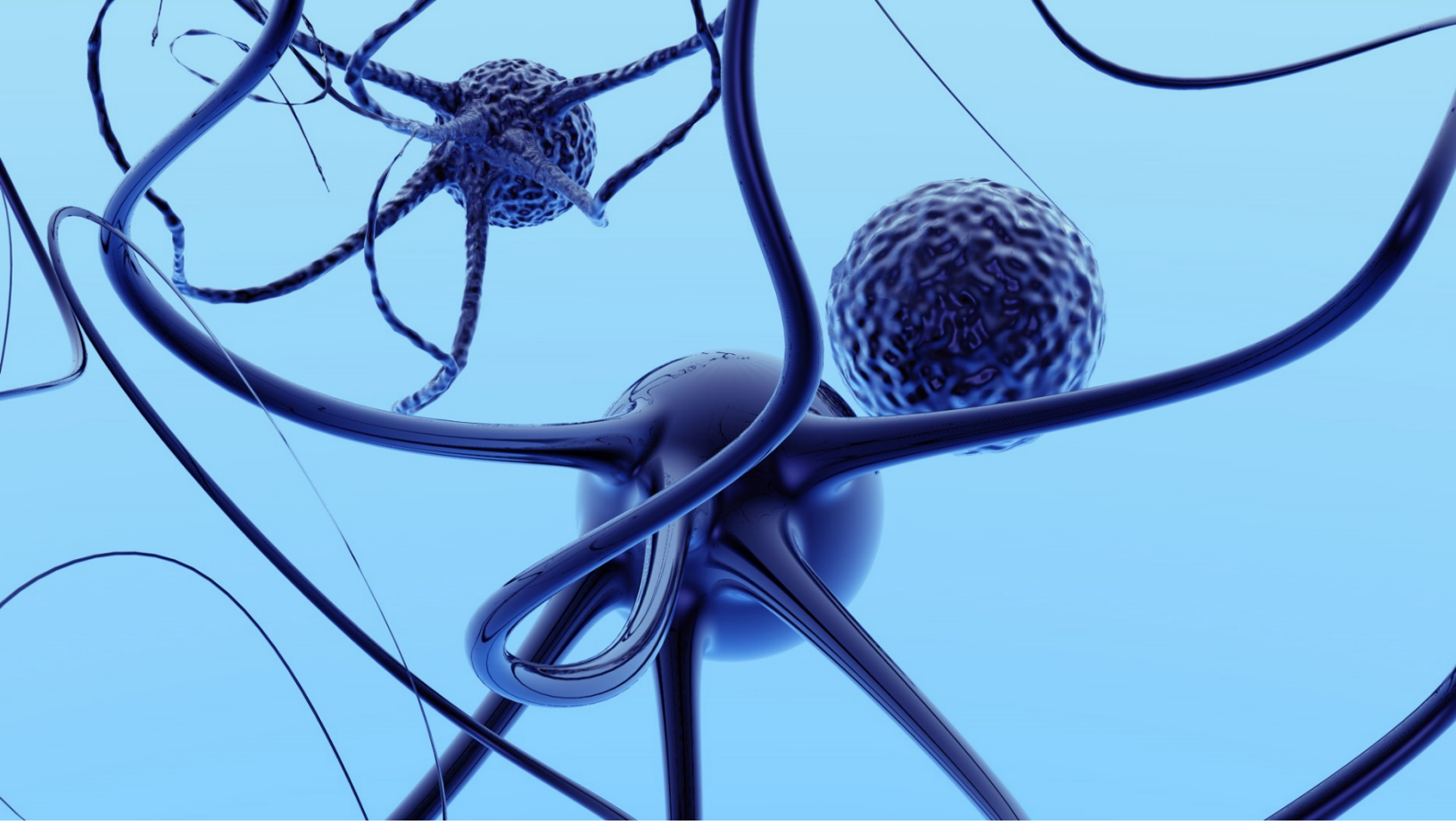


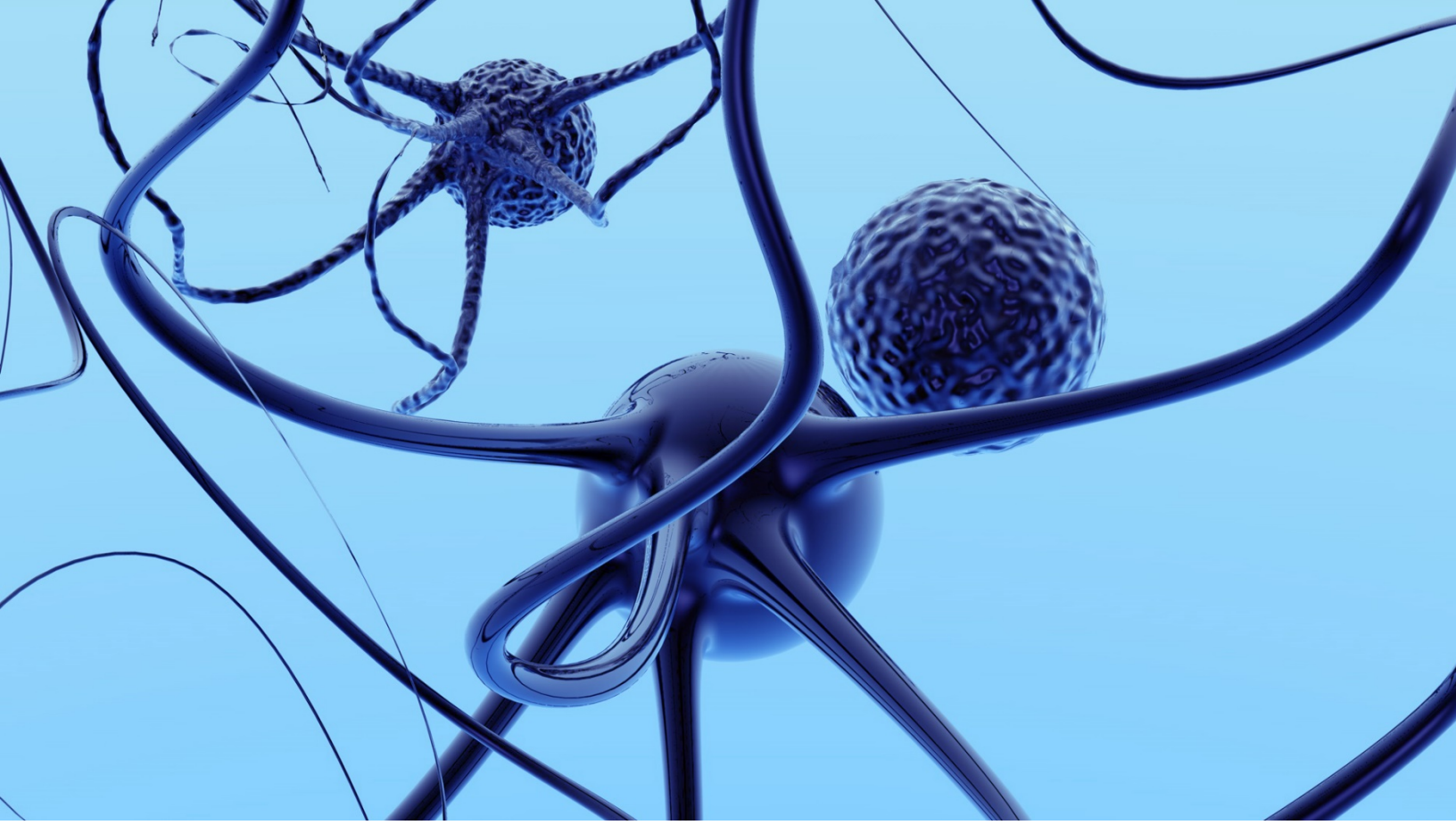
Table of Contents



Acknowledgements

Acknowledgements		
Planning Committee	Justice Robert L. Barnes Justice Leo D. Barry Justice Miriam Bloomenfeld Justice Colin L. Campbell Justice Stephen T. Goudge Justice C. Adèle Kent Justice Judith A. Snider	Justice André Wery Professor Nathalie Chalifour Professor Adam Dodek Professor C. Scott Findlay Professor Heather McLeod-Kilmurray Professor Hamish Stewart Professor Amy Zarzeczny
Introduction	Chief Justice Beverley McLachlin, P.C., Chief Justice of Canada	
Chapter 1	Professor Catherine Piché Professor Hamish Stewart	*Peer review* Justice Christopher E. Hinkson Justice David Michael Paciocco Justice Daniel W. Payette
Chapter 2	Professor Nathalie Chalifour Professor Scott Findlay Professor Heather McLeod-Kilmurray	*Peer review* Professor Brian Baigrie Justice Leo Barry Justice Bruce Cohen Justice Garrett A. Handrigan Justice Ellen Murray
Chapter 3	Mr. W. Ian C. Binnie Justice C. Adèle Kent Justice Stephen T. Goudge Justice André Wery Sarah Berger Richardson Flora Lê	*Peer review* Justice Frederick Ferguson Justice Elizabeth A. Hughes Justice J. Brian Riordan Justice Anthony J. Saunders Justice Earl C. Wilson Justice Leo D. Barry Justice Miriam Bloomenfeld Justice Colin L. Campbell

Chapter 4	Professor Adam Dodek	*Peer review* Christopher E. Hinkson Justice David Michael Paciocco
From the National Judicial Institute	Katherine Ellis Stéphane Émard-Chabot Flora Lê Lynn Leblanc Susan Lightstone	Kevin McGowan Kathryn Poliquin Sarah Berger Richardson Denis Thériault
Others	Canadian Judicial Council Joe Cecil & the Federal Judicial Center American College of Trial Lawyers Environment Canada	Bill Kretzel Sue Liko Westminster Abbey Strategies, Inc.



Introduction to the *Science Manual for Canadian Judges*

The Right Honourable Beverley M. McLachlin, P.C.
Chief Justice of Canada (retired)

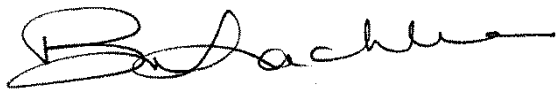
Scientific developments impact justice and the courtroom as never before. From advances in DNA testing to computer credibility assessment techniques, judges are faced with increasingly complex scientific evidence in every type of case that they handle, whether in the context of trial work, case management or settlement conferencing. The pace of scientific developments will only continue to accelerate.

PHILIPPE LANDREVILLE PHOTOGRAPHE INC.

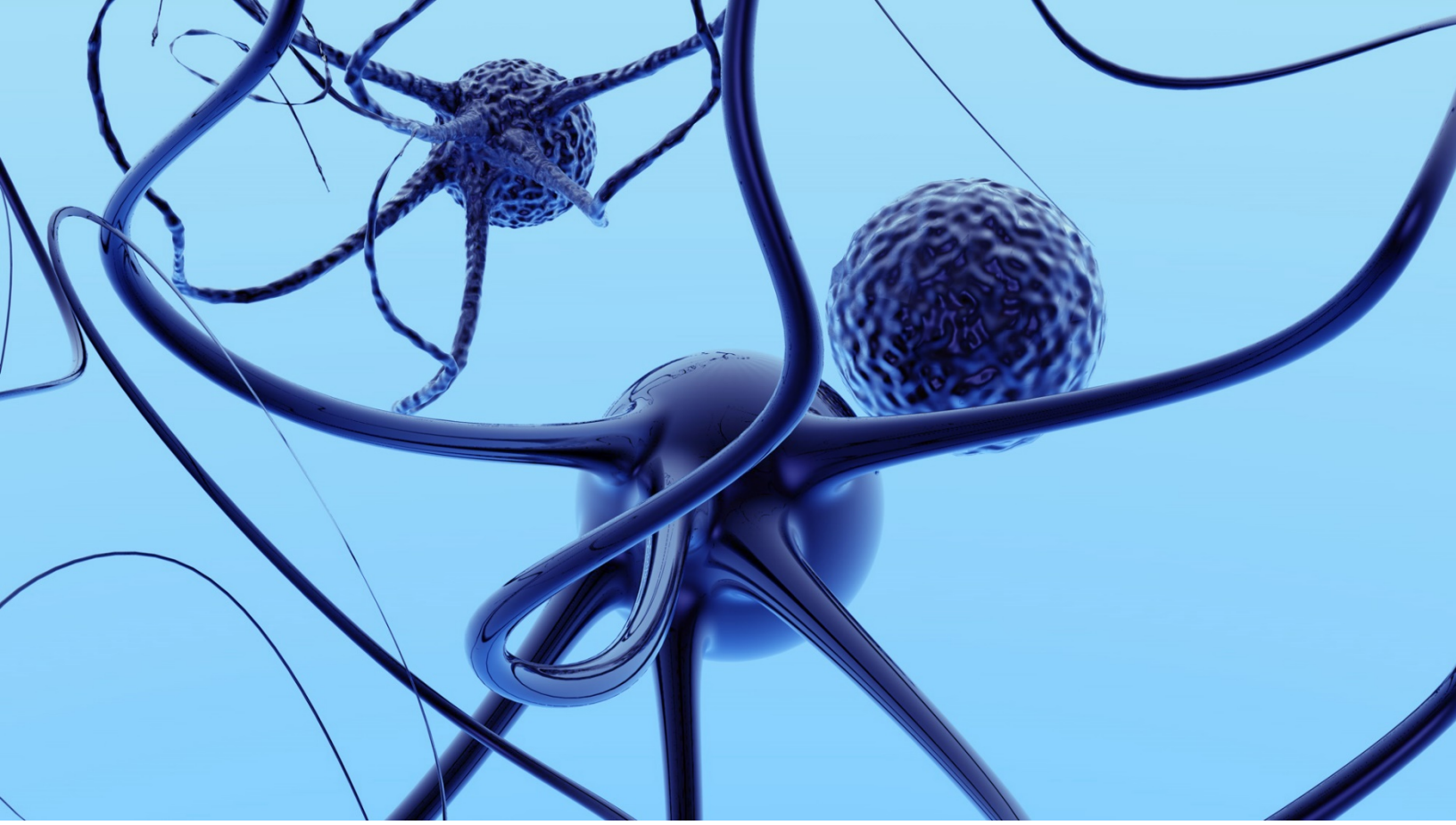
Without the proper tools, the justice system is vulnerable to unreliable expert scientific evidence. Thus, for over a decade, the National Judicial Institute has been delivering science programming to the judiciary, with the goal of equipping judges to deal with this challenging subject. Yet this is not enough. More must be done to assist judges to grapple effectively with complex scientific evidence.

The admission and assessment of unreliable scientific evidence can lead to injustices in both civil and criminal cases. In criminal cases it may produce wrongful convictions. Justice Stephen Goudge's *Report of the Inquiry into Pediatric Forensic Pathology in Ontario* highlighted the systemic failings of the criminal justice system in managing its use of pediatric forensic pathology in investigating and adjudicating pediatric death cases. Lending his voice to the growing concern about judges managing expert opinion evidence, Justice Goudge recommended the creation of a science manual for the judiciary.

This *Science Manual* responds to the calls to find better ways to incorporate the insights of science and scientific developments into our court processes. Within the increasingly science-rich culture of the courtroom, the judiciary needs to discern "good" science from "bad" science, in order to assess expert evidence effectively and establish a proper threshold for admissibility. Judicial education in science, the scientific method, and technology is essential to ensure that judges are capable of dealing with scientific evidence, and to counterbalance the discomfort of jurists confronted with this specific subject matter. This practical manual, written for judges, is an important step forward as we continue to engage with science in the courtroom in a way that strengthens our legal system and its fundamental values.



The Right Honourable Beverley M. McLachlin, P.C.
Chief Justice Of Canada



User's Guide

Inspired by the American Federal Judicial Center's *Reference Manual on Scientific Evidence*, the *Science Manual for Canadian Judges* has been prepared to help judges appreciate and critique expert evidence. The intent is also to stimulate judges to ask incisive questions, to understand accepted theories and matters of controversy in the scientific community, and to evaluate the reliability of expert evidence and experts' qualifications. To this end, a note on scope is in order. The *Science Manual* has been drafted specifically for use by the judiciary, with an eye to meeting the unique needs of judges and providing them with the tools to manage expert evidence in the courtroom. It is not an exhaustive review of all scientific developments or of every scientific discipline, nor has it been written for a scientific audience.

The *Science Manual* is divided into four chapters.

In **Chapter One**, *The Legal Framework for Scientific Evidence*, Professors Hamish Stewart and Catherine Piché provide an outline of the rules concerning the admissibility of opinion evidence, and locate scientific, technical, and other expert evidence within those rules. The ***Mohan***[1] criteria of relevance, necessity, absence of any exclusionary rule, and a properly qualified expert are

[1] *R v Mohan*, [1994] 2 SCR 9 [*Mohan*].

introduced. This is followed by a discussion of the two-step process developed in ***R v Abbey*** (2009)[2] and adopted by the Supreme Court in ***White Burgess Langille Inman v Abbott and Haliburton Co.***[3] under which the trial judge first assesses the preconditions to admissibility, and then acts as a gatekeeper to protect the trier of fact from insufficiently probative evidence.

In **Chapter Two**, *Science and the Scientific Method*, Professors Scott Findlay and Nathalie Chalifour introduce topics ranging from probability to statistics and inferential strength, in order to draw the reader's attention to some of the overarching principles of the scientific method. In doing so, Findlay and Chalifour provide an overview of the questions judges must ask to understand both science as a discipline, and the scientific method. These inquiries will assist judges to make legal decisions that fall within the boundaries of scientifically sound knowledge.

Chapter Three, *Managing and Evaluating Expert Evidence in the Courtroom*, was drafted following consultations with several experienced judges who were asked to share their practical advice on the process of receiving and weighing expert evidence. The first section highlights the importance of assessing the necessity of proposed expert evidence and establishing what should be included in expert reports, as well as addressing issues of the pre-trial and trial management of experts and their evidence, and outlining the innovative process of hot-tubbing. The second section reviews the judge's gatekeeper function and introduces various tools that may assist judges in discharging this challenging task.

In **Chapter Four**, *Ethics of the Expert Witness,* Professor Adam Dodek discusses the ethical implications that arise as a result of an expert witness being retained, instructed and paid for by one of the parties, while simultaneously being expected to assist the court by providing an independent and unbiased opinion about matters within the witness's field of expertise. Key points raised in this chapter include determining what constitutes improper expert witness conduct. The chapter also briefly addresses the issue of whether any appropriate additional measures are required when an expert is found to have violated their duty to the court.

The preparation of this *Science Manual* has benefited from input provided by a wide variety of sources, including members of the judiciary and the legal and scientific academic communities, all of whose contributions were peer-reviewed.

---

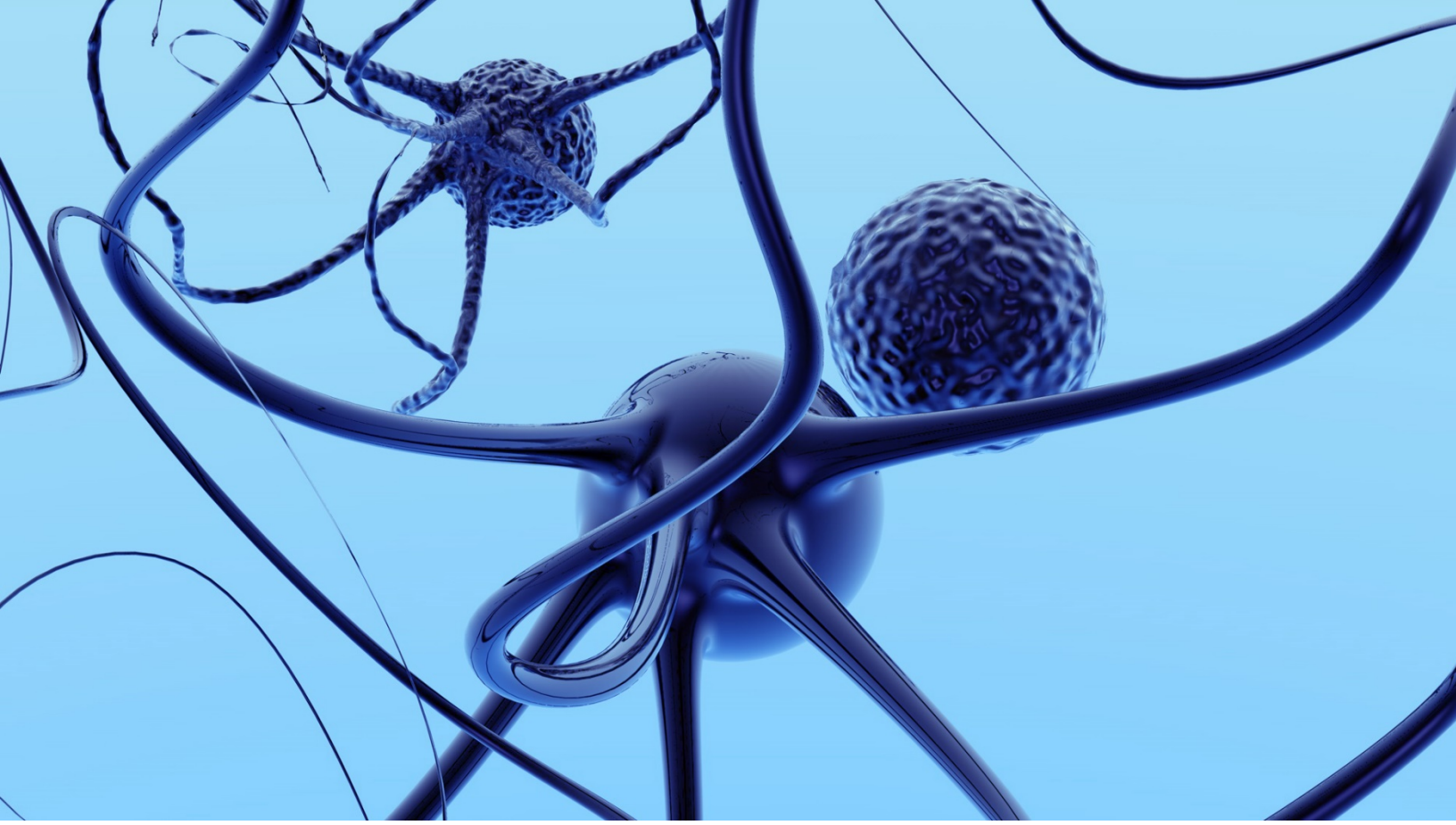
2 *R v Abbey*, 2009 ONCA 624, 97 OR (3d) 330, leave to appeal refused 2010 SCCA No 125.

3 *White Burgess Langille Inman v Abbott and Haliburton Co.*, 2015 SCC 23, [2015] 2 SCR 182.

As the use of expert opinion in the courtroom evolves over time, so too will the *Science Manual*. The National Judicial Institute will update the *Science Manual* on a regular basis, with additional content compiled to address criminal, civil and family law contexts. Subsequent revised editions will serve to ensure the information provided remains current and of practical assistance to all levels of courts, including provincial courts and federal courts.

We welcome your comments and suggestions as this discussion moves forward. Please direct your inquiries or comments to:

Peter Aadoson, Legal Writer, Education Resources
National Judicial Institute
613-237-1118, ext. 227
paadoson@nji-inm.ca



Chapter 1
The Legal Framework for Scientific Evidence

Professor Hamish Stewart
Faculty of Law
University of Toronto
Toronto, Ontario

Professor Catherine Piché
Faculté de droit
Université de Montréal
Montréal, Québec

1. Introduction: Scientific Evidence as Opinion Evidence

This resource has been prepared to provide Canadian trial judges with an understanding of basic principles of science and thus assist them in deciding whether to admit and how to use scientific evidence. This introductory chapter serves to place these questions in the context of the rules of evidence. Witness testimony about scientific matters is a kind of expert opinion evidence. The first steps, therefore, are to outline the rules concerning the admissibility of opinion evidence in general, locate scientific, technical, and other expert evidence within those rules, and address how the rules apply to scientific evidence in particular.

2. Opinion Evidence: The General Rule

The traditional common law trial rule, also applicable in Québec courts, requires witnesses to testify about their personal knowledge of facts in dispute; they are not supposed to offer opinions.[1] Their testimony should pertain to what they know from their own experience rather than be an expression of their views about matters that lie outside their knowledge. The rationale for this exclusionary rule is that the opinion of a witness usually has only minimal probative value but potentially significant prejudicial effects.

Two exceptions apply to the general exclusionary rule for opinion evidence. First, the opinion of lay (non-expert) witnesses is admissible under certain circumstances, typically where the lay witness's opinion is essentially an inference of the kind that lay persons commonly and reliably draw: "he was drunk," "the coat was old and shabby," etc. The common law recognized a number of specific matters on which a lay witness could give an opinion. In ***R v Graat***,[2] the Supreme Court of Canada took a principled approach to these categories and held that a lay witness could give an opinion or impression where "the facts from which a witness received an impression were too evanescent in their nature to be recollected, or too complicated to be separately and distinctly narrated,"[3] that is, where the lay opinion (or inference)

---

1 *R v K.(A.)* (1999), 45 OR (3d) 641 (CA) [*K.(A.)*]; *Civil Code of Québec*, CQLR c CCQ-1991, art 2843 (CCQ); Catherine Piché, *La preuve civile*, 5ième éd. (Montréal: Éditions Yvon Blais, 2016) at para 523 [Piché]. Common law rules and exceptions are applicable in Québec civil law to the extent that they are consistent with provisions of the CCQ, the *Code of Civil Procedure*, CQLR c C-25.01 (*CCP*) and applicable Québec statutes.

2 *R v Graat*, [1982] 2 SCR 819 [*Graat*]. See also *Aménagement Vert-plus de l'île inc. c Scottish & York Insurance Co. Ltd.*, 2006 QCCS 5792, [2006] JQ no 14323 at para 12.

3 *Graat*, [1982] 2 SCR 819 at 837, quoting from Rupert Cross, *Evidence*, 5th ed. (London: Butterworth, 1979) at 448.

amounted to a compendious statement of fact,[4] and where the probative value of the witness's expression of opinion in this form outweighed the potential prejudicial effects, such as unfair surprise, excessive consumption of time, and usurpation of the function of the jury.[5]

The second exception to the general exclusionary rule provides for the admissibility of opinions of expert witnesses, provided the opinion and the qualifications of the expert satisfy certain criteria. The primary focus of this introduction is on the specific criteria for admissibility of expert opinion testimony and their application to scientific evidence.

3. EXPERT OPINION

3.1. General Criteria for Admissibility

3.1.1. The Test for Admissibility of Expert Opinion Evidence

Like all opinion evidence, expert opinion evidence is presumptively inadmissible. The admissibility of expert opinion evidence is determined by application of a two-stage test, as confirmed in 2015 by the Supreme Court of Canada in ***White Burgess Langille Inman v Abbott and Haliburton Co.***[6]

The first step (the threshold stage) requires the party putting the proposed expert forward to establish that the evidence satisfies the "threshold requirements of admissibility."[7] These requirements are the four "*Mohan* factors," described below, and an additional requirement: "in the case of an opinion based on novel or contested science or science used for a novel purpose, the reliability of the underlying science for that purpose."[8] At this first step, the evidence is assessed on a yes/no basis and if it falls short of any of the threshold preconditions, it should not be admitted. The *Mohan* factors are:

---

4 *Graat*, [1982] 2 SCR 819 at 841.

5 On the application of the *Graat* principles in civil litigation in Québec, see Piché, *La preuve civile*, 5ième éd. (Montréal: Éditions Yvon Blais, 2016) at para 523; *Aménagement Vert-plus de l'île inc. c Scottish & York Insurance Co. Ltd.*, 2006 QCCS 5792, [2006] JQ no 14323 at paras 11-13; *Robinson c Films Cinar inc.*, 2008 QCCS 4536, [2008] JQ no 9395.

6 *White Burgess Langille Inman v Abbott and Haliburton Co.*, 2015 SCC 23, [2015] 2 SCR 182 [*White Burgess*].

7 *White Burgess* at para 23.

8 *White Burgess* at para 23.

(a) relevance;
(b) necessity in assisting the trier of fact;
(c) the absence of any exclusionary rule; and
(d) a properly qualified expert."[9]

The second step (the gatekeeping stage) involves the discretionary weighing of the benefits, or probative value, of admitting evidence that meets the preconditions to admissibility, as against the "costs" of its admission, including considerations such as consumption of time, prejudice and the risk of causing confusion.

This test was explained in more detail by the Court of Appeal for Ontario in ***R v Abbey*** (2017):[10]

> Expert evidence is admissible when:
>
> (1) It meets the threshold requirements of admissibility, which are:
> a. The evidence must be logically relevant;
> b. The evidence must be necessary to assist the trier of fact;
> c. The evidence must not be subject to any other exclusionary rule;
> d. The expert must be properly qualified, which includes the requirement that the expert be willing and able to fulfil the expert's duty to the court to provide evidence that is:
> i. Impartial,
> ii. Independent, and
> iii. Unbiased.
> e. For opinions based on novel or contested science or science used for a novel purpose, the underlying science must be reliable for that purpose,
> and
> (2) The trial judge, in a gatekeeper role, determines that the benefits of admitting the evidence outweigh its potential risks, considering such factors as:
> a. Legal relevance,
> b. Necessity,
> c. Reliability, and
> d. Absence of bias.

Should the costs at this stage be found to outweigh the benefits, the evidence may be deemed inadmissible, notwithstanding that it met all the *Mohan* factors.

[9] *R v Mohan*, [1994] 2 SCR 9 at para 17.
[10] *R v Abbey*, 2017 ONCA 640 at para 48.

In civil cases in Québec, rules and procedures for the reception of expert opinion evidence are established by the ***Code of Civil Procedure*** (*CCP*) and the ***Civil Code of Québec***. To be considered admissible in Québec courts, expert opinion must be *necessary* for the trial judge's understanding of the case, and *relevant* to the facts and burden of proof. Art 2843 of the CCQ expressly states that the opinion of an expert is a kind of testimony. Art 293 of the *CCP* provides that the report of an expert stands in lieu of his or her testimony and sets out procedural requirements. [11] Expert testimony is thus exceptional under the New Code, becoming possible where the parties need clarifications on points covered in the report, or the expert's opinion about new evidence, or with the authorization of the court, as per art 294 *CCP*.[12]

Importantly, many provisions of the *CCP* reference the principle of proportionality. Art 18 of the *CCP* codifies a proportionality principle which requires the judge and parties to ensure the means of proof are proportionate, in terms of the cost and time involved, and the nature and complexity of the matter.[13] In establishing a case protocol under art 148, the parties must indicate whether they will be seeking expert opinions, and whether these opinions will be joint. The parties must provide reasons if they decide not to seek joint experts. Art 158 (2) allows a judge to assess the purpose and usefulness of seeking expert opinion in light of the proportionality principle.[14] A judge can impose limits on expert testimony under art 158(2) or order expert evidence on his or her own initiative under art 234. He or she may also order experts who filed contradictory reports to conciliate or join their reports, or to file an additional report within a given delay, the whole as per art. 240(2).

[11] Piché, *La preuve civile*, 5ième éd. (Montréal: Éditions Yvon Blais, 2016) at para 552.

[12] Piché, *La preuve civile*, 5ième éd. (Montréal: Éditions Yvon Blais, 2016) at para 553.

[13] *9217-4887 Québec inc. c Yves Rocher Amérique du Nord inc.*, 2016 QCCS 5123, [2016] JQ no 14298 at para 25.

[14] This principle was previously codified in art 4.2. See *Groupe immobilier de Montréal c Société en commandite immobilière l'Assomption*, [2003] JQ no 12168 (CA). Also see *St-Adolphe d'Howard (Municipalité de) c Chalets St-Adolphe Inc.*, 2007 QCCA 1421, [2007] JQ no 11795 [*St-Adolphe-d'Howard*] where the Court of Appeal rejected expert evidence which tended to prove a fact that was not related to an issue of law or fact of the litigation.

The critical questions of relevance, reliability, and impact on the trial process are very similar in Québec civil law and in common law. Courts in Québec have regularly cited the *Mohan* criteria.[15] More recently, courts in Québec have begun citing the test in *White Burgess* when dealing with expert testimony.[16] Importantly, the trial judge's role as a gatekeeper emphasised in *Abbey* (2009) and *White Burgess* is compatible with the proportionality criteria in arts 18 and 158(2). Further, art 22 of the *CCP* echoes the Supreme Court of Canada's ruling that experts must be objective and impartial, putting their duty to the court above their duty to the parties. Finally, the fact that an expert is employed by one of the parties may affect his or her credibility, but does not render the expertise inadmissible, as per the Supreme Court of Canada case of ***Mouvement laïque québécois v Saguenay (City)***.[17]

The discussion below follows the two-stage analysis confirmed in *White Burgess*, but includes references to both common law and civil law authorities where relevant and required.

3.1.2. General Principles

3.1.2.1. In What Circumstances Do the Expert Opinion Evidence Rules Apply, and At What Stage of the Proceedings?

The expert evidence rules apply where a witness purports to offer an opinion that requires special training or education to arrive at regardless of whether the case is a civil suit,[18] an administrative hearing governed by the law of evidence,[19] or a criminal trial.

---

[15] See *e.g.*, *J.G. c Nadeau*, 2013 QCCS 410, [2013] JQ no 804; *Pelchat c Zone 3 inc.*, 2013 QCCS 78, [2013] JQ no 197; *Widdrington (Estate of) v Wightman*, 2011 QCCS 1786, [2011] JQ no 3882. Rules of evidence developed at common law in criminal cases are to be prudently applied in civil cases particularly when a disposition of the CCQ already deals with the issue a hand. However, when compatible with the CCQ, Québec Courts have applied such rules of evidence: see *France Animation S.A. c Robinson*, 2011 QCCA 1361, [2011] RJQ 1415 at paras 73-74 (applying *Mohan)*; *Hôtel central (Victoriaville) c Compagnie d'assurances Reliance*, [1998] JQ no 1854 at para 9 (CA).

[16] See *e.g.* See *e.g. Berthiaume c McNamara*, 2017 QCCS 1706 at para 109*; Croteau c Sherbrooke (Ville de)*, 2017 QCCS 1395, [2017] JQ no 3844 at para 37*; Perron c Charl-Pol Saguenay inc.*, 2017 QCCS 740 at para 9; and *Ludmer v Canada (AG)*, 2016 QCCS 3119 at para 17. See also *R v Bingley*, 2017 SCC 12 at para 13.

[17] *Mouvement laïque québécois v Saguenay (City)*, 2015 SCC 16, [2015] 2 SCR 3 at para 105; *Croteau c Sherbrooke (Ville de)*, 2017 QCCS 1395, [2017] JQ no 3844 at para 40.

[18] *Drumonde v Moniz* (1997), 105 OAC 295; *Dulong v Merrill Lynch Canada Inc* (2006), 80 OR (3d) 378 (Sup Ct).

[19] *Deeman v College of Veterinarians of Ontario*, 2008 ONCA 600, 298 DLR (4th) 305.

Note that the "gate-keeping" duty in regards to expert evidence is ongoing. In ***Masterpiece Inc. v Alavida Lifestyles Inc***.[20] Justice Rothstein stated that "Courts must fulfil their gatekeeper role to ensure that unnecessary, irrelevant and potentially distracting expert and survey evidence is not allowed to extend and complicate court proceedings."

3.1.2.2. Expert Opinion in Different Adjudicative Contexts

Even though the test for admissibility of expert evidence applies in all kinds of litigation, the intensity with which the gatekeeping function should be applied varies according to circumstances, including who is offering the expert evidence and in what kind of case. In criminal cases the gatekeeping function continues to apply for expert evidence called by the accused, but in the interests of protecting the right to make full answer and defence, a less-rigorous approach than would be applied to Crown evidence is appropriate.[21]

In civil cases the intensity of the gatekeeping function must also be conducted with sensitivity to the stage of an action. In deciding whether to certify class action proceedings, for example, the plaintiffs will not yet have the advantage of disclosure and the foundation for their expert evidence may therefore necessarily be preliminary. This makes a less rigorous application of the expert evidence standards appropriate.[22]

3.1.2.3. Proper Subject Matter for Expert Opinion

Although prior decisions are useful as illustrations, whether expert opinion evidence will be admitted is not a matter of strict precedent. A brand of expert evidence admissible in one case may not be admissible in another. Admissibility is to be determined on a case-by-case basis, because the needs of the case, or even the quality of the evidence presented about the expertise will vary. Appellate courts have recognized that the trial judge is in the best position to determine whether the requirements of the two-stage test have been met and the decision is entitled to deference.[23]

The proper subject matter for expert opinion will be discussed further under the heading of **Necessity in Assisting the Trier of Fact** below, at p 26.

---

[20] *Masterpiece Inc. v Alavida Lifestyles Inc.*, 2011 SCC 27, [2011] 2 SCR 387 at para 76.

[21] *R v Bell*, [1997] NWTR 45 (CA); *R v B.M.*, [1998] OJ 4359 (ONCA); *R v Seaboyer*, [1991] 2 SCR 577 at 611.

[22] *Pro-Sys Consultants Ltd v Infineon Technologies AG*, 2009 BCCA 503, 98 BCLR (4th) 272.

[23] *R v D.S.F.* (1999), 43 OR (3d) 609 (CA); *R v Pearce*, 2014 MBCA 70, 310 Man R (2d) 14 at para 67.

3.1.2.4. Legal Expertise

Expert evidence concerning domestic law,[24] or an expert opinion that amounts to argument on a legal issue,[25] is inadmissible. Legal argument is to be conducted by the parties or their representatives and the legal issues are to be decided by the trial judge. However, expert evidence concerning foreign law is essential where foreign law is a fact in issue.[26]

3.1.2.5. The "Ultimate Issue"

It was once said that expert evidence could not address the "ultimate issue" in the case. This is no longer the rule.[27] But the concern that motivated the rule remains: juries may be overly impressed by an expert whose evidence on a central factual issue may be perceived to essentially dictate the outcome of the case. This danger can be controlled by ensuring strict application of the principle that the opinion offered as evidence should be *necessary* to assist the trier of fact, particularly relative to the degree to which the issue in question relates directly to the outcome the jury must decide.[28] Under Quebec law, this question is analysed in light of art 18 *CCP* and the proportionality principle, as well as the expert's duty to enlighten the court, as per art 22 *CCP*.

---

[24] *R v Century 21 Ramos Realty Inc.* (1987), 58 OR (2d) 737 (CA), leave to appeal refused, [1987] SCCA No 175. See also *Roberge v Bolduc*, [1991] 1 SCR 374.

[25] See, for example, *Mathias v Canada* (1998), 144 FTR 106, [1998] FCJ No. 330 *sub nom. Squamish Indian Band v. Canada* (FC) at paras 7-10; *Côté c Gagnon*, [2005] JQ no 117 (SC); *Fournier c Lamonde*, [2004] RDI 267 (Que CA). Also see Charles Gonthier, "Le témoignage d'experts : à la frontière de la science et du droit" (1993) 53 R du B 187 at 196, where the author explains that expert testimony lies at the border between science and the law, but that the expert cannot usurp the judge's function as a master of the law [*Gonthier*].

[26] *Canadian National Steamships Co. v Watson*, [1939] SCR 11; see also art 2809 CCQ. For an extension of this principle to international arbitration law, see *Holding Tusculum B.V. c S.A. Louis Dreyfus & cie*, [2006] QJ no 4878 (SC) [*Holding Tusculum B.V.*].

[27] *R v Mohan*, [1994] 2 SCR 9 at para 25; *Holding Tusculum B.V.*, [2006] JQ no 4878 at para 27. See also Piché, *La preuve civile*, 5ième éd. (Montréal: Éditions Yvon Blais, 2016) at para 545 and *Hydro-Québec c Steel Components Produits métalliques inc.*, 2014 QCCS 4261 at para 22 and 30; *Giroux c Baillargeon*, 2013 QCCS 3990, at para 26.

[28] *R v Sekhon*, 2014 SCC 15, [2014] 1 SCR 272 at paras 75-76.

3.2. Stage One: The Four Preconditions (*Mohan* factors)

The *Mohan* factors are: logical relevance, necessity in assisting the trier of fact, the absence of any exclusionary rule; and a properly qualified expert. Each of these conditions must be met in order for an expert's evidence to meet the threshold for admission.

3.2.1. Logical Relevance

An expert's opinion, like any other piece of evidence, must be found logically relevant by the trial judge; that is, the opinion must "have a tendency as a matter of human experience and logic to make the existence or non-existence of a fact in issue more or less likely than it would be without that [opinion] evidence."[29] For purposes of comparison with the scientific method discussed in **Chapter 2** it might be helpful to restate this test as follows: evidence is relevant to a fact in issue if the probability of that fact's being true, given the evidence, is different from the probability of the fact's being true, without the evidence. See also *The Logical Relevance of Expert Scientific Opinion* on 54 in **Chapter 2.**

The logical relevance of expert evidence is decided by asking whether:

- it relates to a fact in issue at trial; and
- it tends to prove or disprove the fact in issue.

e.g.	In ***R v Sekhon***[30] the accused was charged with importing 50 – one kilogram bricks of cocaine. The cocaine was hidden in a truck he was driving over the border. The Crown called a police officer with over 33 years of experience involving over 1,000 drug investigations to give evidence on the customs and practices of the drug trade. In his testimony he stated that in the course of his work he had never encountered drug couriers who did not know they were transporting drugs. The Court found this impugned testimony to be both not necessary – the trial judge was in as good a position to make this determination – and was not

[29] *R v Abbey*, 2009 ONCA 624 at para 82. The test for relevance is the same in civil cases in Québec: see Piché, *La preuve civile*, 5ième éd. (Montréal: Éditions Yvon Blais, 2016) at paras 212 and 215; *R v Cloutier*, [1979] 2 SCR 709 at pp. 731 & 733; *Carrier c Québec (PG)*, 2013 QCCS 2211; *St-Adolphe d'Howard*, 2007 QCCA 1421 at para 16.

[30] *R v Sekhon*, 2014 SCC 15.

	relevant. What other couriers may or may not have known had no relevance to what Mr. Sekhon knew.

3.2.2. Necessity in Assisting the Trier of Fact

The necessity requirement exists both to prevent superfluous or redundant evidence from being presented, and to ensure that the problems generally associated with expert evidence are not created needlessly. In ***R v D.(D.)***[31] the Supreme Court of Canada adopted the following passage:

> As the *Mohan* Court explained, the four-part test serves as recognition of the time and expense that is needed to cope with expert evidence. It exists in appreciation of the distracting and time-consuming thing that expert testimony can become. It reflects the realization that simple humility and a desire to do what is right can tempt triers of fact to defer to what the expert says. It even addresses the fact that with expert testimony, lawyers may be hard-pressed to perform effectively their function of probing and testing and challenging evidence because its subject matter will often pull them beyond their competence, let alone their expertise. This leaves the trier of fact without sufficient information to assess its reliability adequately, increasing the risk that the expert opinion will simply be attorned to. When should we place the legal system and the truth at such risk by allowing expert evidence? Only when lay persons are apt to come to a wrong conclusion without expert assistance, or where access to important information will be lost unless we borrow from the learning of experts. As *Mohan* tells us, it is not enough that the expert evidence be helpful before we will be prepared to run these risks. That sets too low a standard. It must be necessary.[32]

A factual matter is a proper matter for expert opinion evidence if it requires knowledge that extends beyond the ordinary knowledge the court may possess as a layperson or group of laypersons (*i.e.*, whether the court consists of a judge alone or of a judge and jury).[33] The expert opinion evidence must be necessary to assist the trier of fact in the sense that should the evidence not be heard, the trier of fact would be unable to appreciate facts of a technical nature or "unlikely to form a correct judgment about [the matter], if unassisted by

[31] *R v D.(D.)*, 2000 SCC 43, [2000] 2 SCR 275.
[32] D. Paciocco, *Expert Evidence: Where Are We Now? Where Are We Going?* (1998), at 16-17; adopted in *R v D.(D.)*, 2000 SCC 43 at para 57.
[33] *R v Mohan*, [1994] 2 SCR 9 at para 23.

persons with special knowledge."[34] There are many examples in Canadian case law. The mechanics of DNA identification,[35] psychiatric diagnosis relevant to a defence of mental disorder,[36] the psychological dynamics of a woman subject to ongoing abuse from her male partner,[37] evidence of a custom,[38] property valuation,[39] and the harmful effects of an allegedly dangerous substance[40] have all been recognized as proper subjects for expert evidence.

On the other hand, assessments relating to credibility or reliability that a jury can make on its own, or with the help of an instruction from a trial judge, are not usually proper subjects for expert evidence. The reluctance of the Supreme Court of Canada to allow expert evidence relating to the credibility of witnesses is illustrated by ***R v Béland***[41] and ***R v Marquard***.[42] In *Béland*, the two accused testified they had not been involved in an alleged conspiracy to commit robbery and offered to repeat this testimony under a polygraph test, the results of which would be introduced as evidence at trial to support their truthfulness. The Supreme Court of Canada upheld the trial judge's decision to exclude this evidence, commenting that it was relevant only to credibility, which is a matter for the trier of fact to decide.[43] In *Marquard*, by way of contrast, the Court indicated that expert evidence concerning credibility might be admissible where it concerned "features of a witness' evidence which go beyond the ability of a lay person to understand," or "the psychological and physical factors which may lead to certain behaviour relevant to credibility," provided the expert remained within the bounds of his or her expertise.[44] The expert may not, however, comment on the credibility of any witness in particular. On the facts of *Marquard*, the expert's opinion "explaining why children may lie to hospital staff about the cause of their injuries"[45] was admissible, but the expert in question had gone farther and had commented impermissibly on the credibility of the child witness herself.

---

34 *Village of Kelliher v Smith*, [1931] SCR 672 at 684, quoting from Thomas Beven, *Negligence in Law*, 4th ed. by William James Byrne and Andrew Dewar Gibb (London: Sweet and Maxwell, 1928) at 141. See also *R v D.(D.)*, 2000 SCC 43 at para 17.

35 *R v Terceira* (1998), 38 OR (3d) 175 (CA), aff'd [1999] 3 SCR 866.

36 *R v Stone*, [1999] 2 SCR 290.

37 *R v Lavallée*, [1990] 1 SCR 852.

38 *R c Joyal* (1990), 55 CCC (3d) 233 (Que CA).

39 *City of Saint John v Irving Oil Co.*, [1966] SCR 581.

40 *Altobelli c Isolation Raymond St-Pierre et fils inc.*, [1985] CA 528 at paras 20-21, [1985] J.Q. no 733 (Que CA).

41 *R v Béland*, [1987] 2 SCR 398 [*Béland*].

42 *R v Marquard*, [1993] 4 SCR 223 [*Marquard*].

43 Contrast the result reached under the C.C.P. in *Hôtel central (Victoriaville) c Compagnie d'assurances Reliance*, [1998] JQ no 1854 at para 9 (CA).

44 See, for example, *R v Bell*, [1997] NWTR 45 (CA), admitting an expert's opinion evidence concerning "the functioning of human memory and the recollection of traumatic events"; compare *R v Semchuk*, 2011 BCSC 1543 at para 24, aff'd 2012 BCCA 389, referring to an expert's evidence concerning repressed memory or "dissociative amnesia."

45 *Marquard*, [1993] 4 SCR 223 at para 53.

Matters that may exceed the level of knowledge of the jury as a group of laypersons, but can be addressed by means of a jury instruction, are also not proper matters for expert evidence. In ***R v McIntosh***,[46] the Court held that the trial judge had properly rejected expert evidence concerning the frailties of eyewitness identification, as this was a matter on which a trial judge could instruct a jury. Similarly, in *D(D)*, the Supreme Court of Canada considered expert evidence tendered by the Crown concerning the possible reasons for delayed disclosure of sexual abuse. In a 5:4 decision, the Court held that the evidence was inadmissible:

> Distilling the probative elements of [the expert's] testimony from its superfluous and prejudicial elements, one bald statement of principle emerges. In diagnosing cases of sexual abuse the timing of the disclosure, standing alone, signifies nothing. Not all victims of sexual abuse will disclose the abuse immediately. It depends upon the circumstances of the particular victim. I find it surprising that a Canadian jury or judge alone would be incapable of understanding this simple fact. I cannot identify any technical quality to this evidence that necessitates expert opinion.[47]

The trial judge should not have allowed the expert to testify, but should have provided a jury instruction to the same effect.[48]

Necessity is to be judged according to whether the particular kind of evidence being offered meets the necessity requirement, not according to whether other experts have already filled the need for expert testimony or whether other evidence could support the same conclusions the expert would offer.

	The fact that a lay witness purports during their testimony to identify the accused, for example, does not make DNA identification evidence unnecessary. In ***R v Klymchuk***[49] it was held that even though police officers could adequately describe indications that the scene was staged to look like a break-in, this did not render the evidence of a crime scene specialist unnecessary.

[46] *R v McIntosh* (1997), 35 OR (3d) 97 (CA).

[47] *R v D.(D.)*, 2000 SCC 43 at paras 64-68. But see *Protection de la jeunesse – 1121*, [2000] RJQ 982 (CQ jeun), a Québec case where the court admitted an expert report about the credibility of extrajudicial declarations of child witnesses that were victims of sexual abuse.

[48] *R v D.(D.)*, 2000 SCC 43 at paras 64-68.

[49] *R v Klymchuk* (2005), 205 OAC 57 at para 61 (CA).

The necessity rule can result in the exclusion of evidence that is dependent on expert interpretation, where no expert witness has been offered. This is because evidence that cannot be interpreted without expert assistance is irrelevant if no expert witness is provided.

	In ***R v Manjra***[50] the trial judge admitted forensic evidence that no DNA was found on the external skin of the complainant several hours after an alleged sexual assault. The accused then urged that the absence of such DNA contradicted the complainant's allegation. The Ontario Court of Appeal held that this forensic evidence should not even have been admitted without expert evidence describing whether DNA would be expected to persist on external skin. The significance of the absence of DNA after so much time was not suitable for judicial notice and depended on expert interpretation. Without that expert evidence to guide relevant inferences, the evidence itself was not relevant and was therefore inadmissible.

3.2.3. A Qualified Expert

The person who gives the opinion must be an expert in the sense that he or she is qualified, by virtue of training or experience, or both, to give the opinion.[51] The party calling the witness must demonstrate his or her qualifications.

The witness's expertise must extend to the matter that he or she is to testify about;[52] a number of proposed expert opinions have been rejected on the ground that, although the witness was an expert, his or her expertise did not extent to the matter proposed to be covered in his or her testimony. Put another way, the trial judge should ensure that the expert's testimony stays within the scope of his or her expertise.[53]

---

50 *R v Manjra*, 2009 ONCA 485, 250 OAC 257.

51 *R v Mohan*, [1994] 2 SCR 9 at para 27; Piché, *La preuve civile*, 5ième éd. (Montréal: Éditions Yvon Blais, 2016) at para 536. For requirements in Québec, see *CCP*, art 535, al. 2 &22.

52 *R v Abbey*, 2009 ONCA 624 at paras 62-66.

53 Ontario, Inquiry into Pediatric Forensic Pathology in Ontario, Report, vol. 3 (Toronto: Queen's Printer, 2008) at pp. 471-5 [Inquiry into Pediatric Forensic Pathology in Ontario, vol. 3].

3.2.3.1. How is Expertise Established?

"Expertise" is established when the expert witness possesses special knowledge and experience going beyond that of the trier of fact, relating to the specific subject matter on which the expert is being offered. The witness must therefore be 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."[54]

"The admissibility of such [expert] evidence does not depend upon the means by which that skill was acquired. As long as the court is satisfied that the witness is sufficiently experienced in the subject-matter at issue, the court will not be concerned with whether his or her skill was derived from specific studies or by practical training, although that may affect the weight to be given to the evidence."[55] Nor is it necessary for the witness to have the best qualifications imaginable.[56] So long as the witness has specialized knowledge, deficiencies in those qualifications go to the weight of the evidence, but not its admissibility.[57]

In ***McPherson v Bernstein (No. 2)***[58] a general surgeon was found qualified to offer an expert opinion on the standards of care expected by obstetrician/gynaecologists because he had researched and published about cancer treatment and breast cancer diagnosis, the relevant issue.

While expertise is a modest standard, it is important that the expert have the kind of special knowledge and experience appropriate to the subject matter.

In ***R v Thomas***[59] a nurse designated as a sexual assault examiner after a brief non-rigorous course and who had several years of experience conducting pelvic exams was not qualified to offer an opinion about whether observed injuries were more consistent with sexual assault than consensual activity. She did not have the requisite special knowledge

[54] *R v Mohan*, [1994] 2 SCR 9 at para 27.
[55] Sopinka, Lederman and Bryant, *The Law of Evidence in Canada* (1992), at 536-37, adopted in *Marquard*, [1993] 4 SCR 223 at para 35.
[56] *R v Pompeo*, 2014 BCCA 317, 359 BCAC 278 at para 62.
[57] *R v Peng*, 2009 ONCA 921, 258 OAC 154.
[58] *McPherson v Bernstein (No. 2)* (2005), 76 OR (3d) 133 (Sup Ct).
[59] *R v Thomas* (2006), 207 CCC (3d) 86 (Ont Sup Ct).

	because she was unfamiliar with the relevant body of literature, had done no relevant research, she had not undertaken a systematic evaluation of the results of her own examinations, and her duties as a sexual assault examiner did not include diagnosing sexual assault from injuries. Similarly, in ***McMillan v Regional District Health Board***[60] an orthopaedic surgeon, a man of obvious expertise, was not shown to have the specific expertise needed to determine the cause of a rotator cuff injury.

The qualifications of the proposed witness are canvassed during the *voir dire.*

	In ***R v Bingley***,[61] the Supreme Court cautioned against wasting judicial resources: "Where it is clear that all the requirements of a common law rule of admissibility are established…the trial judge is not obligated to hold a *voir dire* to determine the admissibility of evidence. To so require would be otiose, if not absurd, not to mention a waste of judicial resources."

The precise area of expertise of the witness should be defined at that time, for "it is trite law that expert witnesses should not give opinion evidence on matters for which they possess no special skill, knowledge or training, nor on matters that are commonplace, for which no special skill, knowledge or training is required."[62] In the absence of objection, however, it is not necessarily an appealable error to permit experts to offer opinions outside their established qualifications where their expertise, although not presented during a *voir dire*, is nonetheless clear.[63]

> Important as the initial qualification of an expert witness may be, it would be overly technical to reject expert evidence simply because the witness ventures an opinion beyond the area of expertise in which he or she has been qualified. As a practical matter, it is for opposing counsel to object if the witness goes beyond the proper limits of his or her expertise. The objection to the witness's expertise may be made at the stage of initial qualification, or during the witness's evidence if it

60 *McMillan v Regional District Health Board*, 2000 SKQB 85.
61 2017 SCC 12 at para 28.
62 *Johnson v Milton (Town)* (2008), 91 OR (3d) 190 at para 50 (CA).
63 *Marquard*, [1993] 4 SCR 223.

becomes apparent the witness is going beyond the area in which he or she was qualified to give expert opinion. In the absence of objection, a technical failure to qualify a witness who clearly has expertise in the area will not mean that the witness's evidence should be struck. However, if the witness is not shown to have possessed expertise to testify in the area, his or her evidence must be disregarded and the jury so instructed.[64]

In ***R v Dominic***,[65] the Alberta Court of Appeal confirmed that expertise gained through experience is sufficient to meet the *Mohan* criteria. This form of expertise may include expertise gained through "anecdotal" reports, such as interactions between a police officer and drug users.

The Court also confirmed that courts should not be overly technical in determining the scope of an expert's area of expertise. The question is not the precise topics the expert was qualified to testify on, but whether the witness has expertise in the area in question. Though the police expert had not been specifically qualified as an expert on *why* cocaine users buy certain amounts at one time, his observations on the point were within his expertise relating to *how much* cocaine a user typically buys at one time.[66]

The opposing side may accept a witness as an expert in a given area and waive the need for a *voir dire*. However, acceptance of expertise does not add weight to the expert's evidence.[67]

3.2.3.2. Requirement of Independence and Impartiality

Experts owe a duty to the court to provide evidence that is fair, objective, and non-partisan. This common-law duty has been codified in the Rules of most of the provinces, which provide explicit guidance related to the duty of expert witnesses, typically requiring mandatory attestations by the experts that they are providing an objective opinion. Cited by the Supreme Court of Canada as perhaps the most succinct and complete statement of the expert's duty to the

[64] *Marquard*, [1993] 4 SCR 223 at para 37. See Piché, *La preuve civile*, 5ième éd. (Montréal: Éditions Yvon Blais, 2016) at para 536.
[65] 2016 ABCA 114.
[66] *R v Dominic*, 2016 ABCA 114 at para 43.
[67] *R v Strickland*, 2013 NLCA 65, 343 Nfld & PEIR 283 at para 19.

court is the Ontario ***Rules of Civil Procedure***, RRO 1990, Reg 194, Rule 4.1, which reads in part:

> *Rule **4.1.01** (1) It is the duty of every expert engaged by or on behalf of a party to provide evidence in relation to a proceeding under these rules,*
>
> *(a) to provide opinion evidence that is fair, objective and non-partisan…*

In the case of *White Burgess*,[68] Justice Cromwell, writing for a unanimous Supreme Court of Canada, clarified that independence and impartiality can be considered at the threshold stage. Prior to *White Burgess*, there had been conflicting lines of authority as to how this issue should be handled, with some decisions suggesting that a lack of independence and/or impartiality would go only to the weight given to an opinion, rather than to admissibility.

In *White Burgess*, the Court confirmed that "a proposed expert's independence and impartiality **goes to admissibility and not simply to weight**" (emphasis added) and that the analysis of a witness's independence and impartiality is properly to be undertaken under the "qualified expert" prong of the *Mohan* test.

Absent a challenge to an expert's independence and impartiality, an attestation or testimony recognizing and accepting the duty will generally be sufficient to establish that this threshold is met. The onus will be on the party opposed to admission of the expert's evidence to show that the expert is unable and/or unwilling to comply with the duty to the court to provide fair, objective and non-partisan evidence. Note that this is not to be assessed based on the *appearance of* bias: Exclusion at the threshold stage will only occur in cases where it is "very clear" that the proposed expert is unable or unwilling to fulfill their duty to the court. Where the inquiry concludes that there are concerns, but they fall below clear unwillingness or inability to do comply with the duty to the court, the evidence should be admitted and the concerns weighed during the second stage of the test:

> In my view, expert witnesses have a duty to the court to give fair, objective and non-partisan opinion evidence. They must be aware of this duty and able and willing to carry it out. If they do not meet this threshold requirement, their evidence should not be admitted. Once this threshold is met, however, concerns about an expert witness's independence or impartiality should be considered as part of the overall weighing of the costs and benefits of admitting the evidence. This common law approach is, of course, subject to statutory and related provisions which may establish different rules of admissibility.[69]

---

[68] *White Burgess*, 2015 SCC 23.

[69] *White Burgess*, 2015 SCC 23 at para 10. See also Piché, *La preuve civile*, 5ième éd. (Montréal: Éditions Yvon Blais, 2016) at para 537.

Although the Court predicts it will be rare that this threshold is not met, it does provide numerous examples of cases where an expert's evidence was ruled inadmissible due to the nature of the relationship between expert and party, while stressing that the analysis is contextual rather than precedent-driven.[70]

In ***Bruff-Murphy v Gunawardena***,[71] the Court of Appeal for Ontario reaffirmed that expert bias is relevant to admissibility inquiries at both the "threshold requirements" and "discretionary gatekeeper" stages of the qualification analysis. Though an expert's bias may not rise to the level requiring disqualification at the threshold stage, it must still be considered by the trial judge in the cost-benefit analysis. Where the risks outweigh the benefits, the trial judge should exclude the testimony.

Under Quebec law, expert bias may limit the opinion's probative force without rendering the report inadmissible.[72]

Note that the statutory and common-law requirements of impartiality and independence do not bar communication between counsel and expert during the report-writing stage. See ***Conseil scolaire francophone de la Colombie-Britannique v British Columbia (Education)***,[73] and ***Moore v Getahun***,[74] 2015 ONCA 55. In *Moore*, Sharpe J.A. for the Court of Appeal of Ontario found that

[70] An expert's interest in the litigation or relationship to the parties has led to exclusion in a number of cases: see, e.g., *Fellowes, McNeil v Kansa General International Insurance Co.* (1998), 40 OR (3d) 456 (Gen Div) (proposed expert was the defendant's lawyer in related matters and had investigated from the outset of his retainer the matter of a potential negligence claim against the plaintiff); *Royal Trust Corp. of Canada v Fisherman* (2000), 49 OR (3d) 187 (Sup Ct) (expert was the party's lawyer in related U.S. proceedings); *R v Docherty*, 2010 ONSC 3628 (expert was the defence counsel's father); *Ocean v Economical Mutual Insurance Co.*, 2010 NSSC 315, 293 NSR (2d) 394 (expert was also a party to the litigation); *Handley v Punnett*, 2003 BCSC 294 (expert was also a party to the litigation); *Bank of Montreal v Citak*, [2001] OJ No 1096 (Sup Ct) (expert was effectively a "co-venturer" in the case due in part to the fact that 40 percent of his remuneration was contingent upon success at trial: para 7); *Dean Construction Co. v M.J. Dixon Construction Ltd.*, 2011 ONSC 4629, 5 CLR (4th) 240 (expert's retainer agreement was inappropriate); *Hutchingame v Johnstone*, 2006 BCSC 271 (expert stood to incur liability depending on the result of the trial). In other cases, the expert's stance or behaviour as an advocate has justified exclusion: see, e.g., *Alfano v Piersanti*, 2012 ONCA 297, 291 OAC 62; *Kirby Lowbed Services Ltd. v Bank of Nova Scotia*, 2003 BCSC 617; *Gould v Western Coal Corp.*, 2012 ONSC 5184, 7 BLR (5th) 19.

[71] 2017 ONCA 502, application for leave to appeal dismissed without reasons, [2017] SCCA No 343.

[72] *Croteau c Sherbrooke (Ville de)*, 2017 QCCS 1395 at para 39; *Roy c Québec (PG)*, 2016 QCCA 2063 at para 9.

[73] *Conseil scolaire francophone de la Colombie-Britannique v British Columbia (Education)*, 2014 BCSC 851, 314 CRR (2d) 139.

[74] *Moore v Getahun*, 2015 ONCA 55, 124 OR (3d) 321.

cross-examination is an effective check against the risk of expert reports which have been subject to undue influence from counsel, writing at para 62 that:

> [I]t would be bad policy to disturb the well-established practice of counsel meeting with expert witnesses to review draft reports. Just as lawyers and judges need the input of experts, so too do expert witnesses need the assistance of lawyers in framing their reports in a way that is comprehensible and responsive to the pertinent legal issues in a case.

3.2.4. The Absence of an Exclusionary Rule

Expert opinion evidence must not be barred by any other exclusionary rule. Probably the two most common examples of other exclusionary rules are those governing character evidence and hearsay.

3.2.4.1. Expert Opinion and Character Evidence

Expert opinion evidence that the accused has a propensity to commit the crime charged would not normally be admissible as part of the Crown's case, even if it otherwise satisfied the criteria for expert evidence, because of the general rule that the Crown cannot lead evidence of the accused's bad character unless the evidence satisfies the "similar fact" rule or unless the accused puts his character in issue. In ***Morin***,[75] the Supreme Court of Canada held that expert evidence as to the accused's propensity to commit the crime charged would be admissible as evidence of identity only in the most exceptional circumstances, where it had a very high degree of probative value similar to that required of similar fact evidence. If the evidence did not have that high degree of probative value, it would infringe the general rule excluding bad character evidence. Perhaps unsurprisingly, there do not appear to be any reported cases subsequent to *Morin* where expert evidence has been admitted on this basis.

In cases where the accused seeks to lead expert evidence to support the inference that he is not the kind of person likely to have committed the offence, *Mohan* requires the accused to demonstrate a very high degree of probative value. The evidence must tend to exclude the accused from the category of persons who could have committed the offence, in one of three ways: (i) the crime is an extraordinary or distinctive one that could only have been committed by a person with certain characteristics that the accused lacks; (ii) the crime is an ordinary one, but the accused has an extraordinary or distinctive characteristic indicating that he is unlikely to have committed it; or (iii) the crime and the accused are extraordinary or distinctive but in different ways. The Court in *Mohan* asked whether psychiatry had established a "standard profile" for offenders, suggesting that expert evidence on this topic

[75] *R v Morin*, [1988] 2 SCR 345.

would not be admissible unless the accused could show that he did not fit the standard profile.

3.2.4.2. The Basis of an Expert Opinion and the Rule Against Hearsay

The expert's opinion is typically based on information gathered from a variety of sources, possibly including information provided by the parties, examination of a person, interviews with a person, and discussions with professional colleagues. If all of the information that the expert relies upon is not proved through admissible evidence at the trial, the opposing party may justly complain that the trier of fact is being asked to accept an opinion without the basis for the opinion having been properly proved, or that the opposing party is trying to circumvent the rule against hearsay by improperly leading evidence through the expert rather than calling witnesses with knowledge of the facts in issue.

The Supreme Court of Canada's approach to this problem has varied over the years. In ***Abbey* (1982)**,[76] the accused offered a defence of insanity to a charge of importing cocaine. The only defence witness was a Dr. Valance, who gave an opinion about the accused's mental state based on an examination 10 weeks after his arrest, an interview with his mother, a review of a report prepared by another psychiatrist, and discussions with colleagues. Neither the accused nor his mother testified. When Dr. Valance examined him, the accused described various incidents of unusual behaviour in the weeks leading up to the commission of the offence. The Court held that, to the extent that these statements showed the basis of the psychiatrist's opinion, they were admissible through the psychiatrist; but they could not be accepted as evidence that the incidents described by the accused actually occurred, as they were hearsay for that purpose. The Court went so far as to hold that "[b]efore any weight can be given to an expert's opinion, the facts upon which the opinion is based must be found to exist."[77]

This aspect of the decision in *Abbey* (1982) was criticized for placing upon the party leading expert evidence the impossible burden of proving *all* the facts underlying the expert's opinion.[78] In *Lavallée*, the Court revisited the issue. The accused was charged with murdering her partner and claimed self-defence. A psychologist testified for the defence that the accused was suffering from

---

[76] *R v Abbey*, [1982] 2 SCR 24.

[77] *R v Abbey*, [1982] 2 SCR 24 at 46.

[78] See, for example, Stanley Schiff, *Evidence in the Litigation Process*, 2d ed. (Toronto: Carswell, 1983) at 480. In Québec, however, *Abbey* (1982) was read, consistent with earlier decisions of the Supreme Court of Canada, as holding merely that the hearsay basis of an expert's opinion would affect its weight: see *Kansa General Insurance Co. c Quincaillerie Roger Lambert ltée*, [1994] RRA 881 (Que CS), (appealed from C.A. Montréal, nº 500-09-001681-949, and later settled on January 16, 2001), 883 and *Paillé c Lorcon inc.*, [1985] CA 528 at paras 16-19, [1985] JQ no 631. See Piché, *La preuve civile*, 5ième éd. (Montréal: Éditions Yvon Blais, 2016) at para 548.

"battered woman's syndrome" as a result of the victim's repeated assaults. The psychologist's opinion was largely based on interviews with the accused, describing the abuse; while there was considerable independent evidence that this abuse had occurred, the accused herself did not testify. So it seemed, based on *Abbey* (1982), that the expert's opinion was entitled to "no weight" because all the facts underlying it had not been proved. The Supreme Court of Canada rejected this conclusion, holding that the fact that the opinion was based in part on hearsay evidence affected its weight but not its admissibility. Sopinka J., in a concurring judgment, drew a useful distinction between a) "evidence that an expert obtains and acts upon within the scope of his or her expertise" and b) "evidence that an expert obtains from a party to litigation touching a matter directly in issue."[79] To accord no weight to opinions based on information of type a) "would be to ignore the strong circumstantial guarantees of trustworthiness that surround it." But an opinion based on information of type b) should have little weight unless the underlying evidence is independently proved.

Thus, the fact that an expert's opinion is based on facts that are not proved except through the expert's testimony about what others have told him does not make the opinion inadmissible; however, the trier of fact is entitled to give less weight to the opinion to the extent that those facts are hearsay; that is, to the extent that they must be true to support the opinion.[80]

In ***R v Pascoe***[81] the evidence was excluded because of the danger it would be used solely to show that the accused was, because of his character, the kind of person to commit the alleged crime. And, in *Mohan* itself, the expert was attempting to suggest that, because of the accused's sexual propensities, he was not the kind of person to commit the offences in question. The rule relating to such proof requires, however, that the kind of crime being prosecuted must be such as would be committed only by persons with distinctive behavioural characteristics, yet the evidence did

[79] *R v Lavallée*, [1990] 1 SCR 852 at 899. This distinction was later adopted by a unanimous court in *R v S.A.B.*, 2003 SCC 60, [2003] 2 SCR 678 at para 62. See also *Eli Lilly Canada Inc. v Apotex*, 2007 FC 455, [2008] 2 FCR 636 at paras 173-188, discussing the effect of an expert's reliance on a study that he had not himself conducted on the admissibility of his opinion. (The expert's opinion was excluded on another ground.)

[80] See also *Portnoy c J.L.*, 2009 QCCS 5390 at para 26-29; *Jo c Directeur de la protection de la jeunesse*, J.E. 2002-309 (Que SC); *Hydro-Québec c Moteurs électriques Dupras inc.*, [1998] JQ no 4320 at para 20 (SC); *Re Lee Estate*, 2001 BCSC 1320, 95 BCLR (3d) 164; Gonthier, "Le témoignage d'experts : à la frontière de la science et du droit" (1993) 53 R du B 187 at 190.

[81] *R v Pascoe*, [1997] 5 CR (5th) 341 (ONCA).

	not establish this to be true for the offences with which Dr. Mohan was charged.

3.3. Stage Two: The Gatekeeper Function

3.3.1. General Considerations

Even if the expert evidence satisfies all four preconditions, the trial judge must still exercise a gatekeeper function in deciding whether the value to the trial process of the testimony outweighs the costs and dangers associated with opinion evidence. The trial judge must consider the reliability of the evidence, its probative value in supporting the inferences for which it is offered, and the potential costs it may exact on the trial process.[82]

On the cost side, the Supreme Court of Canada has identified a number of potential dangers:

> There is a danger that expert evidence will be misused and will distort the fact-finding process. Dressed up in scientific language which the jury does not easily understand and submitted through a witness of impressive antecedents, this evidence is apt to be accepted by the jury as being virtually infallible and as having more weight than it deserves.[83]

Moreover, "expert opinion evidence can also compromise the trial process by unduly protracting and complicating proceedings. Unnecessary and excessive resort to expert evidence can also give a distinct advantage to the party with the resources to hire the most and best experts."[84] On this point, Québec judges refer to the proportionality principle, codified in arts 18 and 158(2) of the *CCP*, when evaluating the reliability and probative value of expert evidence.[85] They are then able to highlight potentially abusive costs involved with expert evidence.

---

[82] See notably *Masterpiece inc. c Alavida lifestyle inc.*, 2011 SCC 27; *Développement FMV inc. c Lévis (Ville de)*, 2008 QCCA 2033, [2008] JQ no 10576; *Iko Industrie c Produits pour toitures Fransyl ltée*, 2007 QCCA 576, [2007] JQ no 3529.

[83] *R v Mohan*, [1994] 2 SCR 9 at para 19; compare *R v Abbey*, 2009 ONCA 624 at para 90.

[84] *R v Abbey*, 2009 ONCA 624 at para 91.

[85] *A. (A.) c K. (M.)* (*sub nom Droit de la famille - 162708*), 2016 QCCA 1816 at para 15; *Cimi inc. c CNH Canada ltée*, 2011 QCCS 1560, [2011] JQ no 3439, applying art 4.2 of the previous version of the *CCP*.

But the most difficult and important aspect of the gatekeeper function is undoubtedly on the benefit side, particularly in the assessment of the reliability of the expert opinion. The Supreme Court has instructed that the trial judge should address any concerns about impartiality or independence that are not compelling enough to require exclusion as part of this assessment.[86] This assessment inevitably requires the trial judge to "intrude into territory customarily the exclusive domain of the jury,"[87] in that the trial judge must go beyond applying the four rule-like preconditions and must consider the value of the evidence in supporting the inference for which it is offered. Moreover, the trial judge, though not himself or herself an expert in any subject other than law, cannot simply accept the expert's proposed opinion at face value but must evaluate it: "Reliability concerns reach not only the subject matter of the evidence, but also the methodology used by the proposed expert in arriving at his or her opinion, the expert's expertise and the extent to which the expert is shown to be impartial and objective."[88] **Chapter 2** has been prepared to assist judges in making such assessments of reliability.

R v Abbey (2017)[89] provides a good example of a judicial assessment of the reliability of an expert's methodology. The accused was charged with first degree murder. The Crown alleged that the murder was gang-related. At trial, the Crown tendered the opinion of a sociologist, who specialized in the study of street gangs, about the meaning of a teardrop tattoo on the face of a gang member. In ***Abbey*** (2009), the Court of Appeal had held that this opinion was admissible; however, on closer examination of the expert's methodology, the court found it insufficiently reliable. The expert stated that in six studies involving interviews with a total of 290 gang members, 71 had been convicted of homicide-related offences, had teardrop tattoos, and provided a consistent explanation for the meaning of the tattoo.[90] However, the court was unable to determine how the expert had arrived at those numbers and found that the studies themselves contained no discussion of teardrop tattoos.[91] Moreover, the raw data underlying the studies was no longer available.[92] In these circumstances, the opinion was so unreliable that its probative value was minimal and its prejudicial effect was great, in particular because its reliability could not be assessed; it therefore should not have been admitted.[93]

[86] *White Burgess*, 2015 SCC 23, at para 45.
[87] *R v Abbey*, 2009 ONCA 624 at para 89.
[88] *R v Abbey*, 2009 ONCA 624 at para 87.
[89] *R v Abbey*, 2017 ONCA 640, 350 CCC (3d) 102.
[90] *Abbey*, 2017 ONCA 640 at para 66.
[91] *Abbey*, 2017 ONCA 640 at paras 90-92.
[92] *Abbey*, 2017 ONCA 640 at paras 94-95.
[93] *Abbey*, 2017 ONCA 640 at para 121.

In civil non-jury trials, the gatekeeper function should, in principle, apply in the same way as in jury trials. However, in practice, there is a tendency for the gatekeeper function to merge with the trial judge's assessment of the weight and credibility of the expert's evidence. This is exactly the case in Québec, where there are no civil jury trials, and where the judge evaluates both admissibility and probative value.[94] When Québec judges analyze the weight to be given to the expert evidence, they examine its probative value, based notably on credibility (and impartiality), qualifications and expertise, the methods used and opinions held in the report.[95]

3.3.2. Types of Expertise

In considering the gatekeeper function, it may be useful to distinguish between three types of expertise:

- scientific expertise;
- technical expertise;
- other expertise based on specialized knowledge and experience.

See also *What Distinguishes Scientific and Technical Evidence?* at 56.

These categories overlap: technical knowledge depends on science (*e.g.*, engineering relies on the laws of physics), and science depends on technology (*e.g.*, data gathering depends on the proper operation of devices such as microscopes, telescopes, and mass spectrometers).[96] And this categorization may not be exhaustive. Nonetheless, as a first step in considering how to exercise the gatekeeper function, it may be useful to ask whether the proffered opinion evidence is plausibly characterized as falling into one of these three categories because that function may have to be exercised differently where the evidence is "scientific" rather than "technical" or "other" expertise.

---

[94] See Piché, *La preuve civile*, 5ième éd. (Montréal: Éditions Yvon Blais, 2016) at paras 547ff and 557, where the author indicates that the probative value of expert evidence is both a prerequisite to admissibility and one aspect of the trial judge's required assessment of the weight and credibility of the evidence presented. See also art 2845 CCQ.

[95] *Mathieu c Beauceville (Corp. de la Ville de)*, [1991] JQ no 2584 (SC); Gonthier, "Le témoignage d'experts : à la frontière de la science et du droit" (1993) 53 R du B 187.

[96] *Kumho Tire v Carmichael*, 526 US 137 at 148 (1999) [*Kumho Tire*].

3.3.2.1. Scientific Expertise

Daubert v Merrell Dow Pharmaceuticals [97] is the leading American case on the admissibility of scientific opinion evidence. The plaintiff brought an action for damages for birth defects allegedly caused by the drug Bendectin. The trial court and the appeal court had refused to admit the plaintiff's proffered expert evidence concerning the causal connection between the drug and the injury because the experts' methods were not "generally accepted" by the relevant community of experts; the result was summary judgment for the defendant, as the plaintiffs had raised no serious issue as to causation. The Supreme Court of the United States refocused the criteria for admissibility from the question of whether the expert's methodology was "generally accepted" in his or her field (the *Frye* test),[98] to the question of whether the expert's methodology was scientifically valid. The Court, relying on the views of Karl Popper and Carl Hempel, understood the scientific method as the process of forming hypotheses and attempting to falsify those hypotheses by testing them against data;[99] in addition to the question of falsifiability, the Court suggested that peer review and publication, "the known or potential rate of error," and general acceptance all had a bearing on whether the expert's method was scientifically valid. But the "overarching subject" of the inquiry "is the scientific validity—and thus the evidentiary relevance and reliability—of the principles that underlie a proposed submission."[100]

Daubert has been influential in Canadian law. In ***R v J.-L.J.***,[101] the accused was charged with sexual offences involving very young boys. The defence sought to tender expert evidence concerning the results of a "penile plethysmograph test," designed to measure the accused's degree of sexual arousal to various sounds and images. The results would support his position that he was not sexually interested in young boys and so was less likely to have committed the offence in issue. The Supreme Court of Canada upheld the trial judge's decision to exclude the test results. Binnie J., for the Court, referred to the *Daubert*

---

97 *Daubert v Merrell Dow Pharmaceuticals* (1993), 509 US 579 [*Daubert*]. The case was, strictly speaking, an interpretation of Rule 702 of the *Federal Rules of Evidence*. See also *R v J.-L.J.*, 2000 SCC 51, [2000] 2 SCR 600 at para. 33 [*J.-L.J.*)].

98 *Frye v United States* (1923), 293 F 1013 (DC Cir).

99 The Court cites Karl Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge*, 5th ed. (London: Routledge, 1989) and Carl Hempel, *Philosophy of Natural Science* (Englewood Cliffs, NJ: Prentice-Hall, 1966). This view of scientific methodology is itself controversial (one might say "not generally accepted") among philosophers of science. For an excellent though not recent overview of the issues, see Ian Hacking, *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science* (Cambridge: Cambridge University Press, 1983). See also *Scientific Hypotheses* on p. 52.

100 *Daubert*, 509 US 579, at pp. 594-5. The Court remanded the case to the 9th Circuit Court of Appeals for further proceedings. That Court once again granted summary judgment to the defendant, holding that the plaintiff's proposed expert evidence was inadmissible under the revised approach: *Daubert v Merrell Dow Pharmaceuticals* (1995) 43 F (3d) 1311 (9th Cir).

101 J.-L.J., 2000 SCC 51.

factors, and held that the test was insufficiently reliable for the purpose offered. The test was a therapeutic rather than a diagnostic tool and had a very high rate of false negatives;[102] it therefore had little probative value to exclude the accused from the group of possible offenders.

In ***R v Trochym***,[103] the Supreme Court of Canada revisited the reliability of a previously accepted scientific technique on the basis that would normally be used to assess a new technique. The accused was charged with murder. The testimony of one of the Crown's witnesses placed the accused at the crime scene at a time when, on the Crown's theory, the killer had returned to "stage" the victim's body. But the testimony of this witness had been enhanced by hypnosis. Although hypnotically enhanced testimony had previously been admitted, the Court reconsidered its admissibility in light of the factors from *Daubert*. It was "difficult to assess" the accuracy of hypnotically enhanced testimony;[104] the scientific evidence suggested several weaknesses of hypnosis as a tool for enhancing memory recall;[105] the literature indicated potential errors linked to three factors: the risk of "confabulation," the reduction in the subject's critical capacity while hypnotized, and concerns about "memory hardening" following hypnosis.[106] The majority concluded that witnesses should not be permitted to testify on any matter on which their testimony had been hypnotically enhanced.[107]

In *Mohan* itself, still the leading case on the admissibility of expert evidence in Canada, an expert's opinion was rejected on similar grounds (though *Daubert* was not explicitly cited). The accused physician was charged with four counts of sexually assaulting teen-aged female patients. The defence sought to call a psychiatrist who would have testified that the accused was not a member of the limited class of "sexual psychopaths" who would have committed these offences. The Supreme Court of Canada upheld the trial judge's ruling excluding this evidence. The Court found that the expert's opinion was not sufficiently reliable, in that there was nothing in the record to indicate "general acceptance" of the view that only a limited class of persons could have committed the offence or that "the profile of a pedophile or psychopath has been standardized to the extent that it could be said that it matched the supposed profile of the offender depicted in the charges."[108] *Mohan* should not be read as holding that a standardized profile is always required for expert psychiatric evidence to be admitted; but since the probative value of the evidence in *Mohan* itself depended on its ability to exclude the accused from

[102] *J.-L.J.*, 2000 SCC 51 at paras 35 and 51-53.
[103] *R v Trochym*, 2007 SCC 6, [2007] 1 SCR 239.
[104] *R v Trochym*, 2007 SCC 6 at para 38.
[105] *R v Trochym*, 2007 SCC 6 at para 40.
[106] *R v Trochym*, 2007 SCC 6 at paras 42-44.
[107] *R v Trochym*, 2007 SCC 6 at paras 62-66.
[108] *R v Mohan*, [1994] 2 SCR 9 at para 46.

the group of possible offenders, the absence of any standard profile of the typical offender reduced its probative value significantly.

The *Daubert* factors should also apply to opinion evidence based in the social sciences, since many social scientists use exactly the same methods that gave rise to the *Daubert* court's image of science: formulating falsifiable hypotheses and testing them against the data. The Ontario Court of Appeal decision in ***K.(A.)***,[109] though it references *Daubert* only briefly, may illustrate the point. Two accused were charged with numerous sexual offences arising out of the alleged sexual abuse of young family members. The credibility of the complainants was challenged on the basis of testimonial factors such as inconsistency, poor memory recall, and so forth. The Crown called a social worker who gave the opinion that these behaviours were consistent with abuse; the basis for the opinion was the so-called "Child Sexual Abuse Accommodation Syndrome" (CSAAS). The accused were convicted, but the Court of Appeal ordered a new trial on the ground that the expert's testimony had exceeded the proper bounds of his expertise; rather than merely stating that behaviours such as delayed disclosure were not unusual,[110] the expert had in effect stated that the behaviours indicated that these children were more likely to have been abused. The CSAAS was not reliable enough to indicate the presence of abuse; it could only act as a possible explanation for facts that might appear inconsistent with abuse. Indeed, the trial record showed that "there is no scientific basis to draw an inference that a child has been sexually abused from the fact that the child exhibits certain behavioural symptoms."[111] The Court's holding might be re-phrased as follows: If the hypothesis was that "the behaviours consistent with a diagnosis of CSAAS indicate child abuse," then the record showed that this hypothesis was not scientifically valid.

3.3.2.2. Technical Expertise

Technical expertise typically does *not* involve the generation of new scientific knowledge or the testing of falsifiable scientific hypotheses against data; instead, it involves specialized knowledge of a particular area of application of scientific principles. For example, most engineers are not research scientists who explore new scientific hypotheses; instead, they apply established scientific and engineering principles to practical, technical problems. The reliability of their methods must still be considered in determining the admissibility of their opinion, but some of the *Daubert* factors may have to be adapted. In ***Kumho Tire***,[112] a leading American case on point, the plaintiffs were injured in a motor vehicle accident and alleged that the accident had been

[109] *K.(A.)*, (1999), 45 OR (3d) 641 (CA).
[110] The expert's opinion on that limited issue would likely now be inadmissible in light of *R v D.(D.)*, 2000 SCC 43.
[111] *K.(A.)*, (1999), 45 OR (3d) 641 (CA) at para 62.
[112] *Kumho Tire*, 526 US 137.

caused by the blow-out of a defective tire. The plaintiffs' proposed expert was an engineer with extensive experience in the design of tires and the analysis of their failure. The trial judge rejected the plaintiffs' argument that the *Daubert* criteria were inapplicable to "technical" rather than "scientific" evidence. Indeed, he applied those criteria rather strictly, concluded that the expert's opinion met none of them, and so excluded his evidence. In the absence of any evidence that the tire was defective, the defendants' motion for summary judgment succeeded. On a subsequent motion for reconsideration, the trial judge held that the *Daubert* inquiry was flexible and adaptable to the circumstances, but confirmed his original decision. The Supreme Court of the United States ultimately upheld the trial judge's decision on the motion for reconsideration. While agreeing with the defendants that *Daubert* was applicable to all proffers of expert opinion evidence,[113] the Court also held that some of the *Daubert* factors might not apply, or might apply differently, to different kinds of expert opinion. For example, the absence of peer review might not be particularly significant in some cases because "the particular application at issue may never previously have interested any scientist"; on the other hand, the idea of an error rate might apply to "an engineering expert's experience-based methodology" by inquiring how often the methodology "has produced erroneous results."[114] The trial judge's decision to exclude the evidence was upheld on the ground that his analysis was properly directed at the reliability of the proposed expert's methodology.

3.3.2.3. Other Expertise

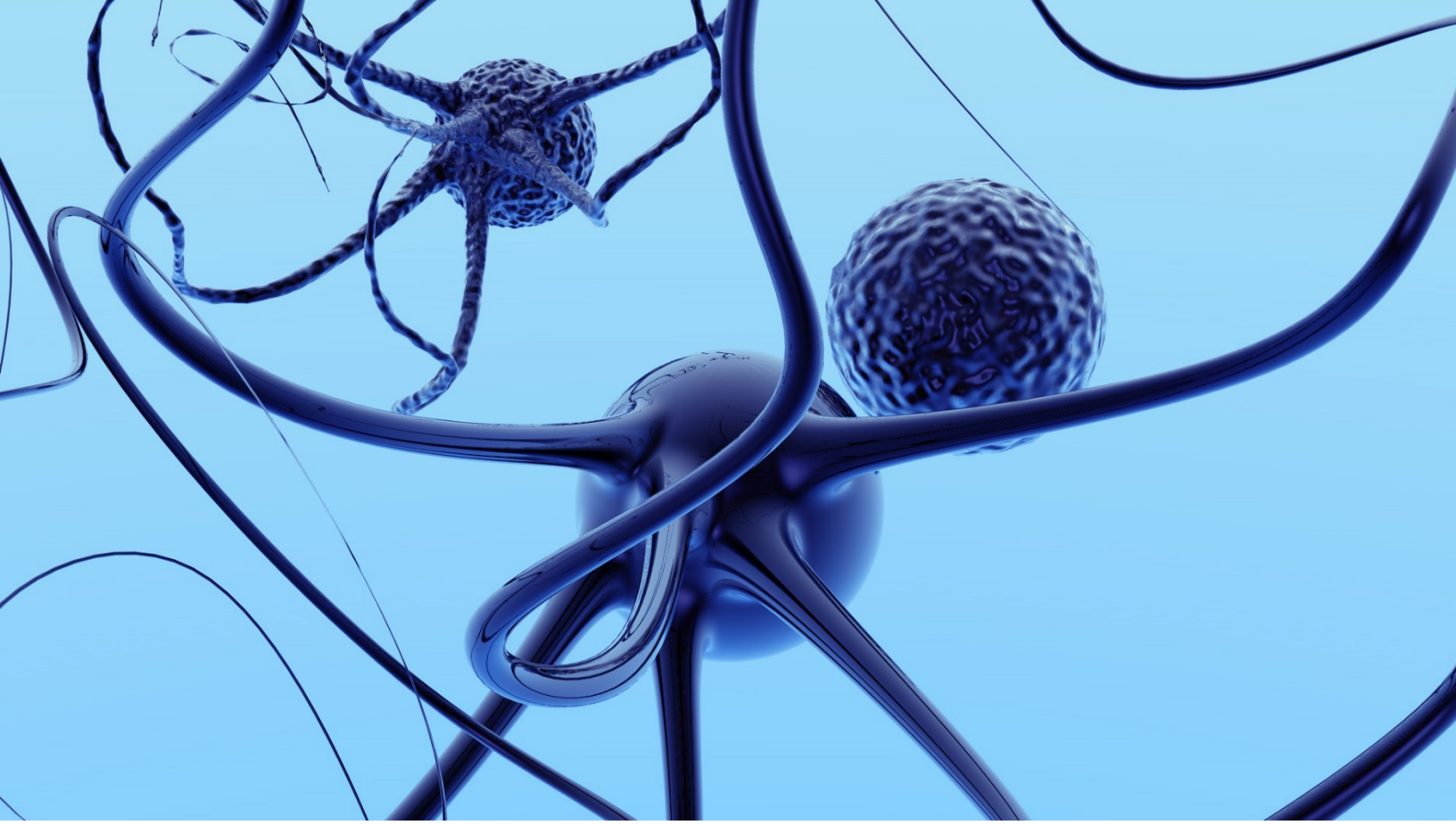
Some witnesses are experts not because they are scientists, physicians, or psychiatrists, or because they have knowledge of technological matters, but simply because they have specialized knowledge that is unlikely to be available to the trier of fact. A witness might, for example, be an expert in foreign law by virtue of his or her training and experience as a lawyer in a foreign jurisdiction;[115] a witness might be an expert on Canadian history by virtue of his or her education and scholarship in that field;[116] a witness might be an expert on the prices of illegal drugs and the practices of the drug trade by virtue of his or her experience as an investigator of, or a participant in, that trade. Although all the usual standards of argument and evidence are applicable in such areas of expertise, they are not "scientific" as that term is understood in *Daubert* because they do not involve the formulation and testing of hypotheses; nor are they particularly technical in the way that, for example, evidence about the proper methods for building a safe structure would be.

[113] *Kumho Tire*, 526 US 137 at pp. 147-149.
[114] *Kumho Tire*, 526 US 137 at 151.
[115] Where foreign law is a fact in issue, expert evidence of foreign law is required.
[116] Compare *Samson Indian Band and Nation v Canada*, [2001] 2 CNLR 353 (FCTD) at para 24.

4. CONCLUSION

The foregoing has briefly outlined the law governing the admissibility of scientific evidence in Canadian trial proceedings. Scientific evidence is one type of expert opinion evidence. Like any other expert opinion evidence, scientific evidence is admissible only if it speaks to a relevant issue, if it is necessary to assist the trier of fact in the sense of being beyond ordinary lay experience, if it can be delivered by a properly qualified expert, if it does not infringe other rules of evidence, and if its value to the trial process outweighs its negative effects on the trial process. What is particular about scientific evidence, according to the leading authorities, is that it is expert opinion evidence that depends on the application of the scientific method—formulating hypotheses and testing them against data—to factual matters that are in dispute in litigation.

For further discussion of evaluating expert opinions, see **Chapter 2**, below, and **Chapter 6** of the National Judicial Institute's *Evidence Electronic Bench Book* (2016).



Chapter 2
Science and the Scientific Method

Professor C. Scott Findlay, Ph.D.
Department of Biology
University of Ottawa
Ottawa, Ontario

Professor Nathalie Chalifour
Faculty of Law
University of Ottawa
Ottawa, Ontario

1. INTRODUCTION

The principles and tools of science are increasingly invoked in legal disputes. In such cases, the trier of fact need not become a scientist nor resolve scientific debates, but he or she must be capable of developing an informed understanding of the science in question. The purpose of this chapter is to provide judges with the tools to understand science in order to assist them in making legal decisions based on appropriate consideration and weighing of any scientific evidence presented in a given case.

The disciplines of law and science share important similarities. These commonalities provide a practical starting point for the judge to understand both the scientific method itself, and its implications to informed legal decision-making.

First, the concept of weight of evidence in science is similar in many respects to its legal counterpart. In both settings, the outcome of a weight-of-evidence assessment by the trier of fact is a binary decision: in a legal setting, either a finding in favour of (or not) of the prosecution/plaintiff; and in the scientific context, the (provisional) conclusion that the hypothesis under consideration is or is not true.

Second, in both science and law, the starting point is often a predisposition that stands unless overturned by contradictory evidence sufficient to exceed some defined standard of proof. In criminal or civil proceedings, the predisposition is that the defendant is innocent or not responsible respectively, and the standard of proof differs between the two: "beyond reasonable doubt" for the former, "balance of probabilities" for the latter. In statistical hypothesis-testing – one of the tools commonly employed by scientists – the predisposition is that there is a particular hypothesis (the null hypothesis) that is assumed to be true unless sufficient evidence is adduced to overturn it. But in statistical hypothesis-testing, the standard of proof has traditionally been set very high such that, in general, scientists will only (provisionally) reject the null hypothesis if they are at least 95% sure it is false. Third, in both scientific and legal proceedings, the setting of the predisposition and the associated standard of proof are purely normative decisions, based ultimately on the perceived consequences of an error in inference.

For instance, in criminal proceedings, the accused is presumed innocent until proven guilty beyond a reasonable doubt simply because we are persuaded that the consequences of convicting an innocent defendant are worse than the consequences of failing to convict a guilty defendant. Thus the law establishes both an exculpatory predisposition and a high standard of proof. In statistical hypothesis-testing, even though the predisposition is inculpatory, we nonetheless insist on a (notionally, at least) high standard of proof, again because as scientists

we believe that for the progress of science, the consequences of rejecting a true null hypothesis are worse than accepting a false null hypothesis.

These similarities between the law and science notwithstanding, the two disciplines are marked by several substantive – and consequential – differences. Perhaps most importantly, in legal proceedings, the trier of fact is concerned with the *comparative* probative value of adduced evidence. That is, the *absolute amount* of evidence does not figure into the decision, as the standard of proof relates to the *relative* difference between the two sides. For example, in a civil proceeding, if the evidence adduced by the plaintiff is weightier than that brought forth by the defendant, a judge is obliged to find in favour of the plaintiff. Similarly, in criminal proceedings, unless the evidence presented by the prosecution is very much weightier than that adduced by the defense, the judge is obliged to find in favour of the defendant.

Science advances by testing and retesting scientific hypotheses. As such, the number of tests of the hypothesis – that is, the absolute amount of evidence one way or the other – matters. Hypotheses that have been subjected to many independent tests, and come through with flying colours, are more likely to be true than those subjected to few tests. Similarly, hypotheses subjected to many independent tests and found to consistently fail are more likely to be false.

Another significant difference between the law and science is temporal in nature. All science is, by definition, provisional. Whereas scientific conclusions are subject to perpetual revision, the law must resolve disputes finally and quickly. Although certain legal findings can be revisited in courts of appeal, a court's assessment of evidence in a given case serves the single purpose of resolving the dispute in question. In other words, while science is advanced by broad and wide-ranging considerations of many different hypotheses, the purpose of evaluating evidence in law is not the exhaustive search for cosmic understanding, but rather the particularized resolution of legal disputes.[1]

This chapter will hopefully help judges determine the admissibility of expert scientific opinion evidence and ultimately, its probative value. Science as a discipline will be explored in detail – warts and all. Hypothetical examples drawn from civil, criminal and family law are used to illustrate how the rules of evidence apply to scientific evidence in legal disputes.

[1] *Daubert v Merrell Dow Pharmaceuticals* (1993), 509 US 579 [*Daubert*].

2. Distinguishing Expert Scientific Opinion from Other Expert Opinion

In the context of expert scientific opinion, application of the ***Mohan***[2] criteria requires that the trier of fact first determine whether the proffered opinion is indeed scientific, or something else. The reason is straightforward: the criteria for determining whether the opinion is necessary or relevant, or whether the author of the opinion is indeed qualified to give it, depend on whether the opinion is scientific or otherwise. As pointed out in **Chapter 1**, several of the *Daubert* [3] criteria for the admissibility of scientific evidence are likely to be inappropriate in the context of technical opinion evidence.

2.1. What Is Science?

Expert scientific opinion pertains, by definition, to science. Science is simply a way of seeking to understand the world. It is often regarded as a coldly analytical, systematic process, and it certainly attempts to be so, but not always successfully. Science is, after all, a human enterprise and, as such, is prone to all the infelicities, peccadillos, errors, biases and serendipities that plague (or bless) any human undertaking.

Science proceeds by posing familiar questions: What? Where? When? Who? How? and Why? The first three questions characterize observations: what occurred (or did not occur), where it occurred, and when it occurred. Although questions of the Who? variety did not traditionally fall within the purview of the physical or natural sciences, they have always been important in the social sciences. And increasingly, questions of the Who? variety may be found in physical or natural science settings – forensic scientific evidence being a notable case in point. But the real soul of scientific knowledge lies in the answers to How? and Why? questions. These answers, tentatively advanced and subsequently refined or discarded, are known as causal hypotheses.

[2] *R v Mohan*, [1994] 2 SCR 9 [*Mohan*].
[3] *Daubert*, 509 US 579.

2.2. What Are Scientific Facts?

According to rules of evidence, witnesses can testify about facts based on their own experience, but cannot offer opinions. As noted in **Chapter 1**, one important exception to this exclusionary rule is expert opinion, provided that the opinion satisfies certain criteria of admissibility.

Similarly, expert scientific evidence has two critical structural elements that both the expert and the trier of fact should distinguish:

1) "facts"; and
2) inferences drawn from these facts concerning the truth or falsity of scientific hypotheses. Just as witnesses may differ in their accounts of the "facts," so too may scientists. Even if scientists agree on the facts, they may well differ in the inferences they draw from them.

In determining the probative value of expert scientific testimony then, the trier of fact must be concerned with the reliability of both the facts themselves, and the inferences drawn therefrom by expert scientific witnesses.

In science, the term "fact" has two meanings. Scientists often use the term "fact" to refer to empirical observations. It is, for example, a fact that, at present, one of the authors of this chapter weighs 78 kilograms, at least on Earth. All observations can be located in both space and time, and have one or more measurable properties
(*e.g.*, colour, size, frequency, etc.). In science, measurements of one sort or another are made on an object called "the unit of observation," namely, that which is measured. The unit of observation can range from very small (*e.g.*, subatomic particles) to very large (*e.g.*, distant galaxies), just as they can be made over very short (*e.g.*, billionth of a second) or very long periods of time (*e.g.*, billions of years).

Scientists also use the term "fact" to distinguish scientific theories or hypotheses for which the evidence is so overwhelming that it is unanimously – or as close to unanimously as one ever gets in science – accepted. For example, Darwin's theory of evolution by natural selection is regarded (by scientists, at least) as both a theory and a fact;[4] a theory insofar as the supporting evidence is overwhelming, and a fact insofar as the evolution of populations (say, of fruit flies or bacteria) under selection can be observed directly in the laboratory.

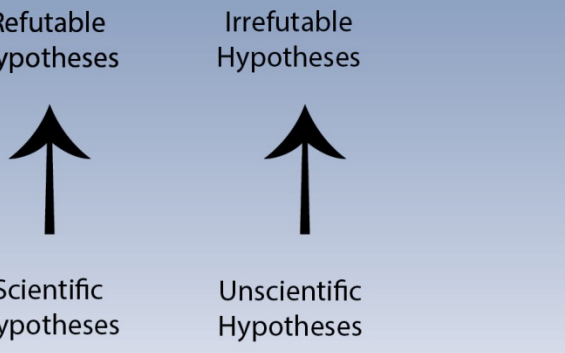
[4] Stephen Jay Gould, "Evolution as Fact and Theory" (1981) 2:5 Discover 34.

But even theories that have attained the Empyrean status of fact still cannot be considered absolutely certain. Rather, as the late paleontologist Stephen Jay Gould remarked, "in science, 'fact' can only mean 'confirmed to such a degree that it would be perverse to withhold provisional assent'."[5] Science was, after all, conspicuously lacking from Benjamin Franklin's very short list of life's certainties.

2.3. Scientific Hypotheses

As noted above, expert scientific opinion includes statements about facts, and statements about the implications of these facts for one or more hypotheses. All hypotheses can, in principle, be classified as: a) true or false; and b) scientific or non-scientific (Figure 1). Science is a method for attempting to establish the truth (or falsity) of scientific hypotheses. Because only scientific hypotheses are amenable to this mode of inquiry, science can say nothing directly about the truth of non-scientific hypotheses.

Figure 1
Karl Popper's Criterion of Refutability



According to Popper, scientific hypotheses are those capable of refutation, whereas non-scientific theories cannot be refuted. Non-scientific hypotheses may be true – but science provides no means by which to establish their truth or falsity.

How do we distinguish scientific from non-scientific hypotheses? This question has been hotly – indeed acrimoniously – debated by both philosophers of science and scientists themselves. Most practising scientists subscribe, at least in principle, to Sir Karl Popper's view: scientific hypotheses are those that could, at least in principle, be falsified.[6] This criterion of falsifiability (or, equivalently,

[5] Gould, "Evolution as Fact and Theory" (1981) 2:5 Discover 34.

[6] Karl R. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge* (London: Routledge, 2002). Unsurprisingly, Popper's concept of falsifiability as the sine qua non of scientific hypotheses has

refutability or testability) means that there are at least some observations that would be considered inconsistent with the hypothesis and hence, would lead one to conclude that the hypothesis is false (see also *Appendix 1: The Logical Structure of Popper's Criterion of Falsifiability* at 126).

Note that the criterion of falsifiability applies to observations in principle, not observations in practice. The history of science is littered with examples of hypotheses which, at the time they were proposed, were rigorously testable in principle but not in practice. For example, rigorous tests of Einstein's general theory of relativity, advanced in 1915, were achieved only in 1959, when technological advances allowed for sufficiently high precision in the measurement of light deflection in weak gravitational fields. In other words, a hypothesis can be considered scientific if it is refutable, even if only in principle.

There may well be causal hypotheses that are true, but not scientific. For example, the diversity of life on Earth could reflect the intervention of a (possibly Divine) intelligent designer. This hypothesis, which could be true, is not scientific because there are no observations that are, even in principle, inconsistent with it. Its validity can, therefore, only be inferred indirectly: if there are competing scientific hypotheses for which there is strong support, this suggests that any unscientific alternate is unlikely to be true. In the case of the Earth's biological diversity, there is a competing scientific hypothesis (Darwin's theory of natural selection and adaptive radiation) for which there is a large body of evidentiary support.

Read more… (Please refer to Appendix 1 at 105)

- *Descriptive Versus Causal Scientific Hypotheses*
- *Descriptive Hypotheses as Description of Patterns Versus Tests of Scientific Hypotheses*
- *How Does One Distinguish Causal Hypothesis-Driven Science from Descriptive Science?*
- *The Logical Structure of Popper's Criterion of Falsifiability*

been the subject of decades of debate by philosophers of science and analytic philosophers. For an accessible and endearingly lucid treatment (see *e.g.*, Martin Gardner, "A Skeptical Look at Karl Popper" (2001) 25:4 Skeptical Inquirer at 13-14, 72). These debates aside, the principle objection to Popper's criterion of falsifiability is that while there is good evidence that most scientists embrace it in principle, the practice of science often falls short of the Popperian ideal. For example, Thomas Kuhn argued that in practice, scientists more often expend considerable energy defending their pet hypotheses or theories from falsification, often by the addition of ad hoc ancillary premises (Thomas s Kuhn, *The Structure of Scientific Revolutions*, 1st ed. (Chicago: University of Chicago Press, 1962)). Needless to say, the debate continues today, with no appreciable decline in enthusiasm.

2.4. The Logical Relevance of Expert Scientific Opinion

For both ***Mohan*** and ***Abbey*** (2009),[7] logical relevance is an important criterion of admissibility. In science, the concept of logical relevance is replaced with scientific relevance with respect to a particular (scientific) hypothesis.

Evidence (that is, observations, study results, etc.) is relevant to a particular scientific hypothesis if the probability of the hypothesis being true, given the observations or results, is different than the probability of it being true in their absence. Three criteria may be employed, in the context of judicial decisions on admissibility, to determine scientific relevance:

1) the hypothesis (or hypotheses) for which the evidence is, at least notionally, relevant, must be clear and, ideally, made explicit;
2) the hypothesis in 1) must be demonstrably scientific, *i.e.*, capable of refutation, at least in principle; and
3) the information or data adduced as evidence in the experiment must result in an appreciable change in the probability that the hypothesis in question is true (or, equivalently, false). That is, if the probability of the hypothesis being true, given the opinion, is the same as it being true in the absence of the opinion, the opinion is scientifically irrelevant.

[7] *R v Abbey*, 2009 ONCA 624, 97 OR (3d) 330, leave to appeal refused 2010 SCCA No 125 [*Abbey*].

IN THE COURTROOM

Several examples may be helpful for judges to illustrate the concept of scientific relevance.

- In ***Abbey*** (2007),[8] the Crown attempted to elicit evidence with respect to the meaning of a teardrop tattoo from a sociologist who was an acknowledged expert in gang culture. Here, the (descriptive) scientific hypothesis under consideration is that the presence of a teardrop tattoo increases the likelihood that the wearer had murdered a rival gang member. (Such a hypothesis might be tested in any number of ways. An obvious one – though not a particularly good one – would be to compare the prevalence of self-reported murders of rival gang members in a sample of young male gang members wearing the tattoo, with a (matched) sample of young male gang members without a teardrop tattoo. The prediction is that the self-reported prevalence of murders would be higher in the former group than the latter.) In the present case, the sociologist's opinion evidence was ruled inadmissible by the trial judge because it was considered to depend upon a novel scientific theory and hence, be subject to a higher standard of admissibility. This decision was subsequently overturned on appeal on the basis that the evidence was not scientific, but rather based on specialized knowledge and therefore need not have been subject to a threshold of scientific validity.

- In *Daubert,* Jason Daubert and Eric Schuller, both of whom had been born with serious birth defects, and their parents sued Merrell Dow Pharmaceuticals Inc., claiming that the drug Bendectin had caused the birth defects. In this case, the (causal) scientific hypothesis under consideration was that prolonged exposure to Bendectin increased the risk of the type of birth defects afflicting the plaintiffs. Daubert and Schuller's evidence was based on in vitro and in vivo animal studies, pharmacological studies, and reanalysis of other published studies – methods that were not, at the time, accepted within the general scientific community. For this reason, the trial judge granted summary judgment to Merrell Dow, who had argued that no published scientific study had documented any relationship between Bendectin exposure and birth defects. Clearly, the studies in question provided evidence

[8] *R v Abbey*, [2007] OJ No 277 (Sup Ct).

bearing on the hypothesis. However, the probative value of such evidence is a different question.

2.5. What Distinguishes Scientific and Technical Evidence?

Although the categories of scientific and technical evidence overlap (*e.g.*, technical knowledge often depends upon science and *vice versa*), it can be important for trial judges to determine whether the evidence falls in one category or the other, as the gatekeeping function may need to be exercised differently for each.

The distinction is, inevitably, not clear-cut. All scientists employ technology in attempting to understand Nature's workings. Many are skilled technicians. In the courtroom then, scientists may be called upon to present technical evidence, scientific evidence, or both – so, for that matter, might technicians. In this context, technical evidence pertains to procedures, practices or tools and their associated operating standards and outcomes. The distinction is thus not in who gives the evidence, but rather in the nature of the evidence given (Table 1).

Table 1
Distinguishing Features of Science and Technology

Feature / Attribute	Science	Technology
Question(s)	How (of Nature)? Why? What? Where? When?	How (procedural)?
Object of knowledge	The inscrutable workings of Nature (including humans)	Procedures, practices, tools
Operating standards, operator's manuals	Not likely!	Yes
Inherent certainty of knowledge	Often low, rarely high	Usually high
Means or end?	End (N.B. in the eyes of scientists!)	Means only

Unsurprisingly, there are many definitions and characterizations of technology, but four themes recur.

1) Technological knowledge almost always refers to procedures, practices, or the tools (broadly construed) employed therein. Thus, it is knowledge about how to *do* something.
2) Implicit to technological knowledge are comparatively standardized yard-sticks ("operating standards") for assessing the quality or extent of technical knowledge, *e.g.*, effectiveness, efficiency, accuracy, precision, etc.
3) Technological knowledge is invariably considered a means to some end (*i.e.*, the thing one wants to accomplish), rather than an end in itself.
4) Technical knowledge is usually characterized by high replicability, high predictability and low uncertainty: given a set of procedures/tools and operating standards, the outcome has comparatively low uncertainty. Indeed, operating standards – known colloquially as "Quality Assurance/Quality Control (QA/QC)" – are designed specifically to minimize outcome variability.

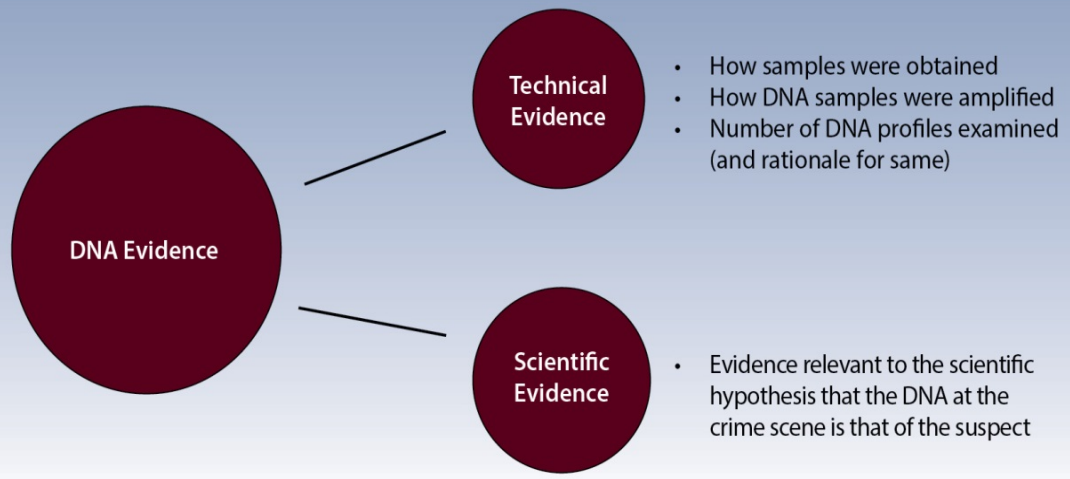
Scientific knowledge is quite different. First, as noted earlier, its scope of inquiry extends far beyond How? to Why? Moreover, How? questions in science relate to the functioning of the world around us, *viz.* how Nature works, not how to carry out a specified procedure. Second, although there is a (more or less) universally acknowledged set of general principles for the practice of science, there are no set operating standards and no owner's manuals. As such, the outcome of a scientific inquiry is often highly uncertain at the outset. Indeed, if one can infallibly predict the outcome of an experiment, it is not, by definition, an experiment. Third, most scientists regard scientific knowledge as an end unto itself, and consider (perhaps erroneously) that human welfare is enhanced by such knowledge, almost by definition. In the courtroom then, scientific evidence pertains to the inferred workings of an often inscrutable – if not downright devious – Nature, specifically her How's and Why's.

In the context of evidence adduced in the courtroom, a simple litmus test can be applied: scientific evidence is that which speaks explicitly to the truth or falsity of one or more scientific hypotheses.

For example, in criminal cases where DNA profiling is admitted as evidence, technological evidence would pertain to how samples were obtained, how small amounts of DNA are amplified, how many loci were involved in profiling, how many DNA profiles were examined for possible matches, and the like. Scientific evidence, on the other hand, is that pertaining directly or indirectly to the (scientific) hypothesis at hand, for example, that DNA samples obtained from the crime scene are indeed from the accused. Here the scientific expert witness is being asked to draw an inference regarding the likelihood that the scientific hypothesis is indeed true, based not only on his or her (technical) knowledge of sample preparation, DNA amplification, etc., but also on his or her knowledge of, for example, alternate explanations (hypotheses) for matches or mismatches,

factors that might contribute to errors in inference based on DNA profiling, and the like (Figure 2).

Figure 2
Technical and Scientific Elements of DNA Evidence



IN THE COURTROOM

R. v. Impaired Driver

Consider the example of a witness who has been called to testify as an expert about how breathalyzers work in a criminal case of impaired driving causing death. When asked whether her testimony provides evidence that speaks directly to the truth or falsity of a particular hypothesis, she is likely to answer no. Asked whether she relies upon operating standards or manuals, she is likely to refer to an instruction or operating manual related to the device. This kind of response would indicate the evidence is of a technical rather than a scientific nature.

IN BRIEF: Scientific *Versus* Other Types of Expert Opinion Evidence

For judges, the following questions may be particularly helpful for distinguishing scientific from other types of expert opinion evidence.

Is the opinion evidence scientific or non-scientific?

- Does the opinion present evidence that speaks directly to the truth or falsity of a hypothesis?
- If so, is the hypothesis scientific, *i.e.*, capable (at least in principle) of refutation?
- If unsure, judges may wish to request that the expert explicitly state the hypothesis in question, and provide examples of evidence (*i.e.*, experimental or study results) that he or she would consider inconsistent with the hypothesis.

What is the nature of the facts relied upon?

- Does the expert clearly distinguish between facts as empirical observations, versus facts as theories for which there is, apparently, such overwhelming support they have come to be regarded as provisionally "true"?
- If not, judges may contemplate having the expert make this distinction clear.

Is the evidence scientific or technical?

- Is it comparatively easy to distinguish between evidence pertaining either implicitly or explicitly, to (a) procedures, practices or tools for which there are existing precedents of use, established protocols, user's manuals or the like (technical evidence), versus (b) inferences drawn from the results of application of such procedures, practices or tools to the truth or falsity of some scientific hypothesis (scientific evidence)?
- If not, judges may contemplate having the expert make this distinction clear.

Is the expert opinion logically relevant?

- Does the proffered opinion change the probability that the scientific hypothesis under consideration is true?
- If, in the judge's opinion, this probability would be the same in the absence of the opinion, then the opinion is irrelevant to the hypothesis under consideration.

3. WHAT IS THE SCIENTIFIC METHOD?

Daubert established that an important criterion for the admissibility of expert scientific opinion is that it be based on scientific knowledge inferred from the application of the scientific method. In other words, when determining threshold reliability, judges must not simply accept proffered scientific conclusions, but rather are obliged to understand how and why the scientific conclusions were made.[9]

Although not explicitly stated, this determination of threshold reliability is integral to the *Mohan* test for admissibility and should be considered by judges when exercising their gatekeeping role.[10] Assessing reliability therefore requires judges to possess a clear understanding of the scientific method; that is, how scientists proceed from a set of experimental or study results to inferences about the truth or falsity of scientific hypotheses.

Science is distinguished from other modes of inquiry principally by a more or less universal method (the scientific method) that practising scientists attempt to apply with varying degrees of success. Proper application of the scientific method results in the accumulation over time of observations that are either consistent with (support), inconsistent with (refute), or irrelevant to particular scientific hypotheses. Also over time, observations can move from one class to another; that is, observations once considered irrelevant may become pertinent, and *vice versa*. Application of the scientific method results in knowledge that is purported to differ from other types of knowledge, arguably qualitatively, in its greater objectivity (but see *Normative Issues in Science – The Myth of Scientific Objectivity* at 105).

9 Erica Beecher-Monas, *Evaluating Scientific Evidence: An Interdisciplinary Framework for Intellectual Due Process* (Cambridge: Cambridge University Press, 2007) at 11.

10 Ontario, *Inquiry into Pediatric Forensic Pathology in Ontario*, Report, vol. 3, chap. 18 (Toronto: Queen's Printer, 2008) at 479 [*The Goudge Report*].

Unsurprisingly, philosophers of science, as well as scientists themselves, disagree on whether there is one scientific method or several.[11] But most concur that to be considered scientific, a method must include (Figure 3):

a) scientific (that is, testable or refutable) hypotheses;
b) systematic observations in the context of a study or experiment; and
c) inferences from b) to a).

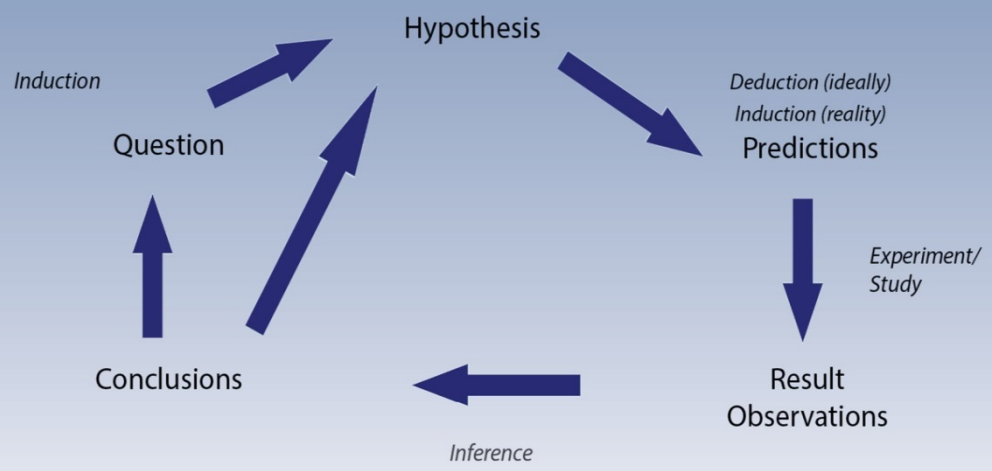
3.1. Hypotheses, Experiments and Predictions

As shown in Figure 3, the hypothetico-deductive method comprises several steps. One begins with a hypothesis, and designs a study to test the hypothesis. A candidate study design is appropriate only if, given the design, the hypothesis in question generates one or more predictions. These predictions are the experimental results one expects to see if the hypothesis is indeed true. If the obtained results are sufficiently similar to those predicted, the hypothesis is corroborated – *i.e.*, such results constitute evidence (although not necessarily strong evidence) that the hypothesis is true, at least in the study context. Insufficiently similar results are considered to be inconsistent with the hypothesis – *i.e.*, such results constitute evidence (though again, not necessarily strong evidence) that, in the study context under investigation, the hypothesis is false.

Several points are worth mentioning here. First, causal hypotheses and predictions are not the same thing, even though there is a tendency (even among scientists, who ought to know better) to use these terms interchangeably. Causal hypotheses are statements about causes; predictions are statements about experimental or study outcomes that should obtain if the hypothesis is true. Second, there are usually many different experiments that one might employ to test the same hypothesis, each of which will result in different predictions.

[11] For a general and eminently readable account of the scientific method and its philosophical foundations, see Alan Francis Chalmers, *What Is This Thing Called Science?*, 3rd ed. (Queensland, Australia: University of Queensland Press, 1999).

Figure 3
The Hypothetico-Deductive Method



Predictions are simply the results one expects to see if the hypothesis under consideration is true. A scientific study yields a set of results or observations, from which one draws an inference about the truth or falsity of the hypothesis under investigation. As George Bernard Shaw observed, answers to one scientific question invariably give birth to the next generation of questions, to which in turn, possible answers are advanced in the form of testable hypotheses. Shaw and others have interpreted this endless cycle as evidence that the scientific enterprise amounts to little more than chasing an ever-receding truth. Unsurprisingly, practising scientists are more sanguine.

Suppose, for example, that one is awakened at night and discovers that the wall switch does not produce illumination from an overhead light fixture. One might be inclined to wonder why the light does not work. Several different hypotheses may be advanced, which can be tested using any number of different experimental designs.

Table 2
Candidate Causal Hypotheses, Alternate Experiments and Associated Predictions for the Observation that a Light Does Not Work

Hypothesis	Experiment	Prediction
Light bulb burnt out	Replace bulb with new bulb	Light will work
Power off to house	Try other electrical switches	No other switch will work
Short in circuit	Replace bulb with new bulb	New bulb will blow and/ or breaker will trip

Table 2
Candidate Causal Hypotheses, Alternate Experiments and Associated Predictions for the Observation that a Light Does Not Work

Hypothesis	Experiment	Prediction
Light bulb burnt out	Replace bulb with new bulb	Light will work
Power off to house	Try other electrical switches	No other switch will work
Short in circuit	Replace bulb with new bulb	New bulb will blow and/ or breaker will trip

All this is fine in principle. But in practice, at least two problems arise. First, Popper's falsifiability criterion technically assumes a deductive relationship between hypothesis and prediction (Figure 3): that is, if in the experiment or study under consideration the hypothesis is indeed true, the predicted result must obtain. If, therefore, it does not obtain, then one can legitimately conclude the hypothesis is false. But the assumed deductive relationship between hypothesis and prediction rarely, if ever, obtains for the simple reason that it invariably depends on the validity of other ancillary assumptions (often themselves scientific hypotheses), which may or may not themselves be true (see *The Logical Structure of Popper's Criterion of Falsifiability* at 126). The result is that, in reality, the relationship is never truly deductive; it is always inductive (Figure 3) and hence, susceptible to error.

Consider again a malfunctioning light. One hypothesis is that the bulb is burnt out. A simple experiment would be to replace the bulb with a new one from an unopened package. The prediction under the hypothesis would be that when the light switch is activated, the light will work. But this prediction follows deductively from the hypothesis only if other assumptions (premises) are indeed true. One such assumption is that the new bulb itself works. Suppose, for example, that the hypothesis that the bulb is burnt out is true. Suppose further that the new bulb does not work (a not uncommon occurrence). Because the predicted result (*i.e.*, that when the old bulb is replaced with the new bulb, the light should work) is not obtained, we conclude that the hypothesis is false, and perhaps embark on an interminable series of ultimately doomed – yet expensive – domestic rewiring experiments in a vain attempt to find the cause of the problem.

It is precisely in this manner that science can – and often does – get involved in fruitless attempts to hunt down scientific Snarks. Yet all of this could have been avoided by the simple expedient of first testing to see whether the auxiliary assumption(s) required to render the relationship between hypothesis and prediction deductive (rather than inductive) was valid – in this example, by checking to make sure that the new bulb in fact worked.

A second problem is also posed, as in practice an operational question arises as to how close the resemblance must be between the observed results and those predicted, in order to conclude that the predictions are (or are not) upheld. Answering this question almost always involves statistical testing of the hypothesis that the observed results are the same as those predicted (see *Statistics* at 76). But because this too is a hypothesis – and a scientific one at that – one can only assign a finite probability that the results obtained are (or are not) those predicted. The consequence is that when one tests a hypothesis, the evidentiary support – *i.e.*, the probative value – for the conclusion that the hypothesis is either true or false can vary dramatically, depending both on the experimental design itself and the experimental outcomes (see also section on *Inferential Strength* at 64).

In the case of the malfunctioning light, two possible hypotheses could be considered: (1) power to the house is off; (2) there is a short circuit and the breaker has been tripped. To test these hypotheses, one might test other lights or outlets on the same breaker. In this experiment, both hypotheses lead to the same prediction, namely, that other lights or outlets will not work. If indeed this is the obtained result, it is consistent with both hypotheses. Thus, we have evidence – but not very strong evidence – supporting either one of the hypotheses. As such, the inference that one or the other is the true explanation is comparatively weak (Table 3A).

By contrast, one could easily design another experiment for which the two hypotheses generate different predictions (Table 3B). In this design, results consistent with one hypothesis are inconsistent with the other. As such, the inference that the hypothesis for which consistent results are obtained is in fact the true explanation is stronger than in the previous case. Hence, the probative value of the second experiment is considerably greater than the first, irrespective of the actual results.

3.2. Inferential Strength

Judges and juries are often faced with trying to make sense of what appears to be conflicting scientific evidence. Often this conflict arises because one scientific witness adduces evidence which, on the face of it, is consistent with some hypothesis (for example, that a breast implant was the cause of the plaintiff's breast cancer) versus another witness who adduces evidence apparently inconsistent with the hypothesis. Thus, judging between apparently conflicting studies requires an evaluation of their differences in inferential strength.

Table 3
Additional Candidate Causal Hypotheses, Alternate Experiments and Associated Predictions for the Observation that a Light Does Not Work

Hypothesis	Experiment	Prediction
Power off to house	Try other lights/ outlets	No other lights/ outlets work
Short circuit and thrown breaker	Try other lights/ outlets	No other lights/ outlets work

(A)

Hypothesis	Experiment	Prediction
Power off to house	Try outlets on *different* breaker	No other lights will work
Short circuit and thrown breaker	Try outlets on *different* breaker	Other outlets will work

(B)

Two different candidate hypotheses for a malfunctioning light, for which a candidate experiment yields the same (A) or different (B) predictions. Insofar as the second experiment (B) allows one to distinguish between two alternate hypotheses, it is a better experiment, *i.e.*, yields stronger inference.

Note that if in (A), no other lights/outlets work, this result is consistent with the hypothesis that the problem is a short circuit, and thus can be construed as having some probative value. But this value is low: the experiment only permits weak inference because the result is equally consistent with other hypotheses. The experimental set-up in (B) provides for stronger inference: if indeed other outlets work, not only is the short circuit hypothesis supported, but other alternate hypotheses (*e.g.*, that the power is off to the house) are refuted.

As we have seen, application of the hypothetico-deductive (or, in practice, the hypothetico-inductive) method leads the investigator to draw an inference as to the truth or falsity of the hypothesis under consideration. Inferential strength is the probability that this inference is indeed correct. A study that permits strong inference is one that, in drawing a conclusion (*viz.*, either that the hypothesis is true, or alternatively, that it is false) based on the study results, the investigator is very unlikely to have made an error. By contrast, some studies only permit weak inference; in such cases, the inferred conclusion is quite likely to be wrong. For judges then, the probative value of scientific evidence is equivalent to its inferential strength.

3.2.1. Inferential Strength Determined by Experimental Design

What determines the inferential strength of a study? As we have seen above, experimental design is clearly important. All else being equal, experiments or studies with comparatively high probative value are those for which, if the hypothesis is true, the prediction (in the context of the experiment) almost surely follows. This means minimizing the number of ancillary assumptions that must be true in order that the relationship between hypothesis and prediction be deductive. So in the case of a malfunctioning light, testing to make sure a new bulb from an unopened package actually does work immediately increases the inferential strength of the study by increasing the strength of the inductive relationship between hypothesis and prediction (see *Hypotheses, Experiments and Predictions* at 50). Probative value is also greater for experiments or studies wherein different hypotheses give rise to different predictions (Figure 3), as such experiments can simultaneously provide evidence consistent with one or more hypotheses as well as inconsistent with others.

For judges, there are two simple questions that might be employed to assess the probative value of scientific results that are adduced as evidence (supporting or inconsistent with the hypothesis under consideration):

1) Under what conditions might one expect the hypothesis under consideration to be true, but the "predicted" pattern nonetheless not obtain?
2) Under what conditions might the hypothesis be false, but the "predicted" pattern nonetheless be observed?

The larger the set of conditions for either question, the lower the inferential strength of the study, and thus the lower the probative value of the evidence. Note that the answers to both questions have nothing to do with the actual results of the study: they relate only to experimental design. As such, they can only be employed in assessing the *a priori* inferential strength of the study (see also *Inferential Strength, Redux* at p. 101).

3.2.2. Inferential Strength Determined by Experimental Outcomes

The overall, or *a posteriori,* inferential strength of a study depends not only on the experimental design, but on the actual results. If the observed results match very closely those predicted, the inference that they are consistent with the hypothesis in question is stronger, given a certain *a priori* inferential strength. The inference that they are not consistent with the hypothesis is also stronger if they are wildly discrepant from those predicted. Problems arise with more equivocal results – *i.e.*, results that do not match those expected especially well, but on the other hand are not wildly discrepant either.

Once again, to test our hypothesis that a light bulb is burnt out, we replace the bulb with a new one that we have determined actually does work. But the room has several large east facing windows which, early in the morning, let in considerable light and the replacement bulb is only 40 watts. In bright early morning sunshine, with the blinds wide open, we must look carefully indeed to distinguish the light from the replacement bulb from background light. Thus it is comparatively difficult to know whether the result is that predicted. By contrast, if we use a 100 watt replacement bulb, and conduct the experiment with the blinds drawn, we can be much more certain the result is, or is not, that predicted (see also *Inferential Strength, Redux* at p. 101).

IN BRIEF: The Scientific Method

One of the most important outcomes of *Daubert* is the recognition that in determining threshold reliability, judges must not simply accept proffered scientific conclusions. Reliability depends, in large part, on the scientific basis for the conclusions. Invariably, conclusions will be based on one or more studies, the results of which are adduced either as supporting or as being inconsistent with some scientific hypothesis, the truth of which bears on the ultimate issue. In essence, the question boils down to: Given the proffered evidence, how likely is it that in the case under consideration, the hypothesis is true (or alternatively, false)?

In attempting to get a feel for this likelihood, judges may consider the following questions:

(1) For the study in question, what is the hypothesis being tested, and what are the associated predictions?

Expert scientific witnesses ought to be able to state explicitly, and with little ambiguity, both the scientific hypothesis and associated predictions.

(2) Given the stated hypothesis and predictions from (1), is it the case that predictions are derived deductively from the hypothesis? If not, what other ancillary assumptions must be true so that this relationship is minimally inductive?

As in (1) above, expert witnesses ought to be able to state explicitly not only what other ancillary assumptions are required, but also which of them (if any) have actually been tested. All else being equal, the greater the number of assumptions that have not been tested and found to be valid, the lower the inferential strength of the study, and the lower the probative value of the evidence.

(3) Under what circumstances might (a) the hypothesis be true in the study under consideration, yet the predicted results still not be observed; and (b) the hypothesis be false, yet the predicted results still be observed?

This question is not only an excellent means of determining the extent to which the witness has truly considered the fallibility of his or her opinion and/or the adduced science, but it also allows the trier of fact to get a first impression of the probative value of the evidence: the greater the set of circumstances for which (a) or (b) apply, the lower the inferential strength of the study, and the lower the probative value of the evidence.

(4) How closely do the observed results match those predicted?

The closer the match, the greater the inferential strength of the study, and the more probative the evidence – all else being equal.

IN THE COURTROOM

R. v. Thief

Consider a criminal case against a defendant who has been accused of armed robbery. The Crown is relying upon the testimony of two eyewitnesses – the store owner and a nearby customer – for identification.

The defence has called expert witness 'A' who is a university professor in criminology. 'A' has documented 15 criminal trials in the last decade in which the accused was identified by eyewitnesses but was later acquitted based on DNA evidence. The defense wishes to use this evidence to suggest that eyewitness testimony is often unreliable.

The Crown has called expert witness 'B' who is a university professor in psychology. His research evaluates the impact of stress on a person's ability to recall facial characteristics. He has conducted a study to test the accuracy of visual recall of facial features under circumstances with varying stress levels (induced by increasing time pressure, *i.e.*, the time allotted for recall, and stressing the need for accurate results). His study found that the ability to recall facial features is not significantly impacted by stress.

The following outlines how the questioning of witness 'B' would proceed along the lines of the model set out above (see 67).

(1) For the study in question, what is the hypothesis being tested, and what are the associated predictions?

Witness 'B' states that the hypothesis of his study was that the accuracy and reliability of recall of facial features is reduced under stress. The predicted result was that visual recall accuracy would decrease as the time allotted for recall (as an index of stress) decreased.

(2) Given the stated hypothesis and predictions from (1), is it the case that predictions are derived deductively from the hypothesis? If not, what other ancillary assumptions must be true so that this relationship is minimally inductive?

Witness 'B' should explicitly state that there are at least three such assumptions:

a) that the number of, and variability among, images used to evaluate recall is sufficient to provide an accurate estimate (for example, if only a couple of images are used, and these images are not sufficiently variable to span the range of patterns in shape, colour, size, etc., that one might expect to see in facial features, then an accurate assessment of facial feature recall is unlikely to obtain);
b) in the sample of subjects studied, the range of recall times employed (as an index of stress) actually does induce stress (it is possible, for example, that even the shortest permitted recall times are comparatively unstressful for most subjects); and
c) that subjects allowed short recall times are actually more stressed than those given longer times. (It is possible that given individual variation in stress susceptibility, especially if the sample of subjects is small, those permitted short recall times experience no more stress, on average, than those allotted longer times. Both (2) and (3) could be validated by, for example, examining physiological measures of stress, such as the production of stress hormones in relation to the time allocated for recall).

(3) Under what circumstances might (a) the hypothesis be true in the study under consideration, yet the predicted results still not be observed; and (b) the hypothesis be false, yet the predicted results still be observed?

When asked in cross-examination under what circumstances the hypothesis might be true in the study under consideration, yet the predicted results still not be observed, witness 'B' responds that this could occur if any one of assumptions 2(a)-(c) are invalid. If, for example, stress levels do not actually increase as the time allotted for recall declines, then the predicted pattern will not obtain because the index of stress employed in the study is invalid.

When asked under what circumstances the hypothesis might be false even if the predicted results are observed, witness B offers a similar list of study design features. For example, if the number of subjects in each recall time "treatment" was small, and the experimental design involved subjects being allocated to one and only one treatment, it is entirely possible that by chance, the average baseline ability to recall facial features was lower in the short recall time treatment group compared to the longer recall time treatment group. If so, the predicted pattern would be observed, but has nothing to do with stress levels.

(*N.B.* This is an example of an experimental design with comparatively low *a priori* inferential strength. A much better design would involve testing the same subject under different recall time constraints, so that each subject experiences the full range of stress treatments. In this way, variation among subjects in baseline recall ability is much more effectively controlled.)

At this point, one could pursue a line of questioning about the specifics of the study design (*i.e.*, sample size, number of images used, types of images used, etc.) to get a better sense of the quality of the study design. One might also ask witness 'B' to evaluate the quality of his own study design.

(4) How closely do the observed results match those predicted?

Witness 'B' should be able to provide a quantitative estimate of how discrepant the results are from those predicted. One such estimate is provided by the Type I error rate, that is, the probability of obtaining results as discrepant from those predicted by the statistical null hypothesis, given that it is true. In this example, the appropriate statistical null hypothesis is that average recall accuracy does not decrease as recall time is reduced. The associated Type I error in this case would be large, say 0.5 or greater. The closer it is to 1.0, the greater the probability that the observed results are consistent with the statistical null hypothesis, and inconsistent with the associated scientific hypothesis (for a detailed discussion on null hypotheses and Type I errors, see *Statistics* at 76.)

4. PROBABILITY[12]

As noted in *The Logical Relevance of Expert Scientific Opinion* (at 54), the *Mohan* and *Abbey* criterion of logical relevance requires judges to evaluate scientific relevance with respect to a particular scientific hypothesis. The scientific relevance of expert (notionally scientific) evidence is determined by the extent to which, in the opinion of the trier of fact, the proffered evidence changes the probability that the scientific hypothesis under consideration is true (or, conversely, false). The greater the change in this probability, the greater the relevance and probative value that can be assigned to the adduced evidence.

Using the example of *R. v. Thief* (immediately above), one could imagine two different studies that might be introduced as evidence by the defence. In a first study, subjects who are assigned only to one recall time treatment are asked to recall general spatial patterns (not specifically facial features), and no attempt is

[12] For a lucid and concise description of a number of statistical misconceptions, several of which involve the concept of probability, see Jonathan J. Koehler, "Misconceptions about Statistics and Statistical Evidence" in Richard L. Wiener and Brian H. Bornstein (eds), *Handbook of Trial Consulting: A Psychological Perspective* (New York: Springer, 2010) at 121-136.

made to verify that subjects allotted shorter recall times are actually more stressed than those given longer times. By contrast, a second study provides each subject with a wide range of facial images designed specifically to evaluate different dimensions of facial recognition (skin colour, hair colour, ethnicity, facial shape, etc.), each subject is tested over the full range of recall time periods, and a blood sample is taken from which serum concentrations of several well-described stress hormones are measured.

Before each study is conducted (that is, *a priori)*, the hypothesis is as likely to be true as it is to be false. Once the results are in, we can ask: How likely is it now that the hypothesis is true? In the first study, the low *a priori* inferential strength of the study design means that this probability will not be much different from the *a priori* value of 0.5 because any result will be rather equivocal owing to limitations in the experimental design. By contrast, in the second study, any result will be less equivocal, *i.e.*, the estimated *a posteriori* probability will be closer to 0 (so, we are quite sure the hypothesis is false) or 1 (so we are reasonably convinced it is true). Thus, the (absolute) difference between the *a priori* and *a posteriori* probabilities will be greater in the latter case than the former. Or in other words, the second study has considerably higher scientific relevance, and hence, probative value.

The above discussion highlights the need for judges to have a reasonably firm grasp of the concept of probability, at least as it is employed in the testing of scientific hypotheses. Probability is a superficially simple concept for which two principal interpretations are employed in science: frequentist and Bayesian. In elementary statistics, one invariably is taught the frequentist interpretation of probability. Under this interpretation, the probability of some "event" or "outcome" or "result" is the long-run frequency of that event relative to other possible outcomes. "Long run" means the repetition of the same experiment under the same conditions a large number of times.

For example, in an experiment in which we roll a (fair) die, there are six possible outcomes, and the long run probability of rolling, say, a six, is one in six, or 1/6. Note that for a small number of experiments, the frequency of rolling a six, even if the die is indeed fair, need not be 1/6. For example, if we roll a fair die six times, it is entirely possible that none of the rolls will produce a six, in which case the observed frequency is 0/6 = 0, not 1/6. This is why under the frequentist view, probability is considered the "long run" frequency of an outcome or result based on a large number of independent experiments.

Under the Bayesian interpretation, the probability of an outcome is a measure of our belief that, in the experiment in question, a particular outcome will result. Thus, Bayesian probability is interpreted as a measure of the current state of knowledge. So, in the context of a dice-rolling experiment, the Bayesian probability associated with the event of rolling a six is, in essence, my belief that on this roll, the outcome will be a six.

Judges are ultimately concerned with the probability that a scientific hypothesis is true (or false) in the situation before the court. Expert evidence will be introduced in an attempt to convince the court that it is true (to some standard of proof) or, alternatively, false. To support one or the other contention, probability estimates of various sorts will be presented. What are the implications of different approaches (frequentists versus Bayesian) to the interpretation of these estimates?

For frequentists, a hypothesis is either true or false – there is no probability about it. Rather, the probability that is estimated under a frequentist interpretation is the probability of obtaining the observed data, given the specified hypothesis. We might hypothesize that the die is fair, and ask: What is the probability of getting four sixes in 10 rolls if the die is fair? This probability (0.054) is in fact rather small, which would lead us to question the validity of the hypothesis.

By contrast, under a Bayesian interpretation we ask: What is our belief (as quantified by probability) that the die is fair, given the observed results? To calculate this probability, we must first specify our belief that the die is fair *before* we conducted the experiment. That is, the Bayesian probability that the hypothesis (*viz.*, that the die is fair) is true, given the experimental results, is estimated with reference to the prior probability, *i.e.*, the probability of the hypothesis being true before the study was undertaken (so, before any results are known). Under a Bayesian approach, as more experimental tests of a hypothesis are conducted, the prior changes (*i.e.*, is updated) to reflect our changing belief.

To return to the dice-rolling experiment, under a Bayesian approach, for the first experiment to test whether the die is fair (the hypothesis), it is reasonable to set the prior at 0.5, *i.e.*, in the absence of any information whatsoever, there is an equal chance that my hypothesis is true or false. Suppose that in the first 10 rolls, I roll four sixes. For the 11^{th} roll, now the prior is substantially less than 0.5 because the chance of rolling four sixes in 10 tries, given the die is fair, is rather small. Thus, the estimated probability of the hypothesis being true, given the results of the 11^{th} experiment is very different because of constant updating of the prior based on the results of previous experiments. In other words, while I might initially have believed that the die was fair, after 10 rolls of which four turned up a six, I am now rather skeptical.

Given the same results (*viz.*, rolling four sixes in 10 throws), we have two probability estimates: one (Bayesian) based on an initial prior, which gives the probability that the die is fair, given the results, and another (frequentist) that gives the probability of the results, given the die is fair.

These estimates are not the same, for two reasons. First, in the Bayesian case, the estimated probability depends on the prior initially chosen; change the prior, and the estimated probability changes. Second, while both probabilities are conditional (probability of the die being fair, given the results (Bayesian);

probability of results, given the die is fair (frequentist)), they are nonetheless generally different. This fact alone means that they may differ dramatically.

To see this, consider swans, most species (and most individuals) of which are white. Therefore, if I know a bird is a swan, there is a very high probability that it is white, *i.e.*, the probability that the bird is white, given it is a swan, is high. But the probability that it is a swan, given it is white, is in fact quite low, as there are many species of birds that are white but are not swans.

The issue of which interpretation (Bayesian or frequentist) is more appropriate in a given situation is not primarily a scientific question. Both interpretations are reasonable and have their underlying logic. But there are least two points of which judges should be aware.

1) The estimated probability of an event or hypothesis under a frequentist interpretation of probability may differ dramatically from that estimated under a Bayesian approach. For example, in forensic DNA cases, frequentist and Bayesian estimates of the so-called "random match probability" may differ a million-fold or more (see the courtroom example *People v. Puckett* at 74).
2) Scientific witnesses are often asked to give probability estimates for certain events, outcomes, or hypotheses, such as the probability of a DNA match or the probability that a convicted offender will reoffend if granted parole. Given the potentially large differences in estimated probabilities based on frequentist versus Bayesian interpretations, witnesses should be explicit not only about the set of results from which the estimate is derived, but also about which interpretation is being used to generate the probability estimate. When Bayesian estimates are given, the witness should provide a clear statement of the prior probability employed in the estimate, and its justification.

Read more… (Please refer to Appendix 2 at 114)

- *Contrasting Frequentist and Bayesian Probabilities*
- *Frequentist and Bayesian Probabilities in Forensic DNA Profiling*

IN THE COURTROOM

People v. Puckett

Differences in estimated probability based on frequentist versus Bayesian interpretations may be enormous. Consider the American case of *People v. Puckett.*[13] John Puckett was convicted in 2005 for the 1972 rape and murder of a San Francisco nurse. Although there were no witnesses and no physical evidence such as footprints or fingerprints, linking the accused to the crime, California's convicted-offender database matched Puckett's DNA to biological evidence found at the crime scene. Based on this random "cold hit," many years after the crime, the accused was convicted. The prosecution's expert estimated that the Bayesian probability of a coincidental match between Puckett's DNA and the biological evidence found at the crime scene were 1 in 1.1 million (p = 0.000001). A frequentist estimate of probability (which was not admitted into evidence) would have placed the odds of a coincidental match in Puckett's case at 1 in 3 (p = 0.33).

[13] No A121368, (Cal Ct App, 1st Dist, May 1, 2008). The accused appealed the conviction, but died before the matter went to court. A motion was brought for an abatement of the appeal given the death of the appellant. Counsel for the appellant asked the court to issue an opinion on the DNA evidence issues even though the issue was technically moot. The Court of Appeal denied this request (*The People v John Puckett*, No A121368, (Cal Ct App, 1st Dist, 10 June 2010)).

IN BRIEF: Probability

The concept of probability lies at the heart of any decision on the relevance or probative value of scientific expert testimony. As we have seen, all expert scientific testimony ought properly to concern specific scientific hypotheses. For the trier of fact, the question is then: How does the probability that the hypothesis is true differ in the presence, versus the absence, of the evidence? If this difference in probability is near zero, the proffered evidence is scientifically irrelevant. The greater the difference, the greater the scientific relevance, and hence, the probative value of the evidence.

In attempting to get a feel for this likelihood, judges may consider the following questions (N.B. the questions that follow assume that the trier of fact has already established the set of hypotheses to which, at least notionally, the proffered evidence pertains).

- When the witness refers to the concept of probability, in what sense (frequentist or Bayesian) is he or she employing the term?
  Any credible scientific witness who uses the term should be able not only to state explicitly in which sense he or she is using it, but also to provide the rationale for his or her interpretation.

- In the witness's opinion, how much does the proffered evidence change the probability that the hypothesis under consideration is true, relative to the corresponding probability in the absence of the evidence, and on what is this estimate based?

  As noted above, this is the fundamental question in determining scientific relevance and probative value. Any credible scientific witness should be able to state explicitly:
  a) what he or she considers to be the *a priori* probability (*i.e.*, in the absence of the evidence being adduced by the witness) of the hypothesis being true;
  b) on what this probability is based (N.B. If, in response to (a), the witness replies that this probability is 0.5, the implication is that he or she is estimating this probability solely with respect to the evidence he or she is adducing, and is not taking into account any other evidence that might be available. By contrast, if this *a priori* probability is not 0.5, then he or she is necessarily taking into account other evidence, in which case judges may wish to inquire as to the nature of this evidence); and
  c) what he or she estimates to be the *a posteriori* probability of the hypothesis being true, and on what basis. Judges should be aware that estimates close to zero or one are only reasonable if indeed the study or studies in question have very high *a posteriori* inferential strength (see section on

Inferential Strength at 52) which, in fact, few individual studies have. Such questioning can, therefore, assist the trier of fact in establishing the extent to which the witness appreciates the fallibility of his opinion and the evidence being presented.

5. STATISTICS[14]

Statistical analysis is a critical element of virtually all modern science. It is also increasingly important in judicial decision-making, as judges and juries are frequently required to evaluate evidence with a strong – sometimes overwhelming – statistical flavour. As was the case with probability (see *Probability* at 58), a working knowledge of basic statistical concepts is critical for assessing the reliability of scientific evidence simply because in many cases, some sort of statistical method or procedure is employed to estimate the probability that the scientific hypothesis under consideration is true. As it is on the basis of this probability that the scientist often concludes, provisionally, that a hypothesis is true (or alternatively, false), assessing the reliability of this conclusion requires that judges understand not only the nature of the probability estimate itself (*i.e.*, frequentist or Bayesian), but also how it was determined.

There are two types of statistics: descriptive and inferential. Descriptive statistics, as the name implies, are simply statistical descriptors of a set of observations. Well-known quantities such as the mean (average) and the variance of a sample are examples of simple descriptive statistics. Other examples are measures of the association between two variables such as correlation (*e.g.*, the correlation between height and weight in a sample of students). No attempt is made to infer any other proposition from these descriptors.

With inferential statistics, an inference is drawn, based on a sample, to some other proposition such as, for example, the value of a parameter in a population (*e.g.*, the mean of the population, or the correlation between two variables in the population). This proposition may take many different forms, but the two most likely to be encountered in a courtroom are: a) an estimate of some population parameter based on a sample (*e.g.*, the prevalence of particular genes in a population in DNA profiling studies), or b) the truth or falsity of some hypothesis (*e.g.*, that the accused is indeed guilty, given a forensic DNA profiling match).

[14] For a considerably more detailed review of a number of statistical concepts and statistical approaches that might be encountered in the courtroom, see David H. Kaye & David A. Freedman, "Reference Guide to Statistics" [*Reference Guide to Statistics*].

Drawing an inference from a sample to some proposition requires a set of assumptions. A commonly encountered assumption is that the sample is a random sample of the population whose attributes one is attempting to estimate. For example, if an estimate of the average height of citizens of Toronto is desired, a sample of first generation male Scandinavian immigrants does not permit valid inference because this is not a random sample of Torontonians. As a result, the estimate of the average height is biased – in this case, overestimated – due in part to non-random sampling.

The potential importance of these underlying assumptions cannot be overstated. All statistical inferences depend, to a greater or lesser degree, on their validity. Some assumptions are critical: if they are invalid, then any inference will be highly suspect. Others are less critical, so that minor – or sometimes even quite substantial – violations have comparatively little effect on the validity of the inference. For example, if one is attempting to draw an inference about a parameter (*e.g.*, the frequency of a particular DNA profile in the Canadian population) from a finite sample (*e.g.*, a collection of DNA samples in a database), the assumption that the sample is a random sample is critical. By contrast, in statistical hypothesis testing, the estimate of the Type I error rate *p* (see *Errors in Statistical Inference* at 73) often depends on the assumption that the data conform to a certain distribution, often a normal distribution. This distributional assumption is required so that the derived estimate of the Type I error is an unbiased estimate of the true value of However, when sample sizes are large, even large departures from normality still provide estimates of *p* that are very close to the true value. In such circumstances, the assumption of normality – while technically a requirement – is not very important, *i.e.*, under conditions of large sample size, the estimate of *p* is robust with respect to the normality assumption.

In the courtroom, most statistical evidence involves inferential statistics: based on some sample, an inference is drawn concerning the truth or falsity of one or more propositions. The relevant questions are then:

a) What assumptions underlie the statistical inference?
b) How sensitive/robust is the inference to violations of these assumptions?
c) What is the evidence that the assumptions are met?

Scientific witnesses should be able to give explicit answers to each of these questions. These answers are crucial for correct interpretation of the validity of the statistical inference.

IN THE COURTROOM

Employee v. Employer [15]

Consider a civil suit alleging employment discrimination against the federal government. When the plaintiff – a First Nations citizen – applied for a job in the federal government in 2012, he was asked to pass a test. He would only proceed to the next stage of being granted an interview if he achieved a certain score on the test. He was informed that he had not achieved the score, thus would not be interviewed.

The plaintiff alleges the practice of hiring based on this pre-employment test eliminates a disproportionate number of minorities from further consideration for the job, and is thus discriminatory. To determine whether the practice of having candidates write the test has a disparate impact on First Nations, one might compare the passing rates of First Nations applicants versus applicants who are not from First Nation communities.

In discovery, the government was asked to reveal the results of its pre-employment tests according to race. Over a two-year period (2006-2008), seven out of 18 racial minority candidates had passed the exam (a passing rate of 38%), versus 45 of 59 (a passing rate of 76%) racial majority candidates. Applying a statistical test for equality of proportions to these data, a Type I error probability (see *Errors in Statistical Inference* at 86) of $p = 0.003$ is obtained; *i.e.*, the probability of getting the observed results, given that the null hypothesis (of equality) is true, is estimated to be 3/1000. Thus, one might be led to conclude that it is highly likely that the test discriminates against First Nations members.

The conclusion that the test discriminates against First Nations members is based on at least two principal assumptions.

1) The performance of First Nations members on the test is, on average, the same as that of the sample of 18 minority applicants in the sample. If the 18 minority applicants were all first-generation Cuban immigrants, for example, their average performance might be very different than First Nations members.

[15] This example was inspired by the following article: Joseph L. Gastwirth *et al.,* "Statistical Issues Arising in Disparate Impact Cases and the Use of the Expectancy Curve in Assessing the Validity of Pre-Employment Tests" (2003) 71:3 International Statistical Review 565.

2) The performance of applicants (both minority and majority) in 2012 is, on average, the same as it was in 2006-2008. If, for example, a new test was introduced in 2011, it is entirely possible that failure rates have changed for minority applicants, majority applicants, or both.

If either of these assumptions is invalid, the inferred likelihood that the 2012 test is non-discriminatory for First Nations members based on the estimated Type I error (*viz.*, $p = 0.003$) may be wildly inaccurate.

5.1. Statistical Hypothesis Testing

Virtually all statistical inferences can be represented as scientific hypotheses. One class of simple statistical hypotheses concerns the value of a population parameter. For example, a statistical hypothesis may posit that the frequency of some DNA profile *G* in a population is 0.003. A researcher might then collect profiles of a sample of individuals, calculate the prevalence of *G* in the sample, and use these data to test the hypothesis that the population prevalence is 0.003. Intuitively, the smaller the absolute difference between the sample estimate and 0.003, the greater the chance that the hypothesis is true.

Statisticians and scientists often refer to statistical null hypotheses. These are (statistical) hypotheses that are assumed to be true, unless demonstrated otherwise. Statistical null hypotheses specify patterns that are the opposite of what is predicted under the scientific hypothesis, such that rejection of the statistical null hypothesis corresponds to support for the scientific hypothesis (see *Statistical Null Hypotheses Versus Scientific Hypotheses* at 70). Statistical hypothesis testing involves answering the question: if indeed the null hypothesis is true, how likely is it that we would obtain results that are at least as discrepant (from those predicted by the null) as the observed results?

Returning to the above example, suppose our (null) hypothesis is that the population prevalence of genotype *G* is 0.003 (in the population, three out of a thousand have that profile), but we find the prevalence of *G* in our sample is 0.031 (31 out of a thousand have the profile *G*). This is ten times larger than the value specified by the null hypothesis. The probability (denoted *p*) of getting a value at least as discrepant as 0.031, if the true value is indeed 0.003, is rather small. By contrast, the chances of getting a sample value as discrepant as 0.002 under the same null hypothesis is much larger. So *p* in the first case is much smaller than in the second, which we interpret as meaning that the null hypothesis is less likely to be true in the former case than in the latter.

Inferential statistics adopt the frequentist view of probability whereby a proposition is either true or false, and the task at hand is to estimate the probability of getting results as discrepant as, or more discrepant than, those observed, given the null hypothesis. Thus, in statistical hypothesis testing, the

usual inferred conclusion is either that the null is true (or rather, that we have insufficient evidence to reject it) or it is false (in which case we reject it).[16] The decision to reject or not is based on the value of p: if the estimated value of p is below some threshold value α, we reject the null; otherwise we accept it. By convention (and by convention only), scientists tend to set $\alpha = 0.05$; this corresponds to the collective – and, one assumes, consensual – scientific attitude that unless we are 95% sure the null hypothesis is false, we provisionally accept it. It is partly because of this that scientists have the reputation of being a notoriously conservative lot, given that a 95% threshold constitutes a rather high standard of proof (see *Statistical Significance* at 76).

There is a subtle but important point to be made here, about which the trier of fact should be particularly vigilant. As noted above, p is the probability of obtaining results at least as discrepant as those observed if the null is true. This is not the same as the probability of the null hypothesis being true, given the results. To see this, again consider the example of swans, most species of which are white. Therefore, if I know a bird is a swan, there is a very high probability that it is white, *i.e.*, the probability that the bird is white, given it is a swan, is high. But the probability that it is a swan, given it is white, is in fact quite low, as there are many species of birds that are white but are not swans.[17]

5.2. Data Reliability

All empirical scientific studies require data collection. In statistical analysis, these data are then used to estimate something: the value of a parameter, its precision, the probability of obtaining results at least as discrepant as those observed given some null hypothesis, etc. The validity of an estimate and any inference drawn therefrom thus depends on the data themselves, and more specifically, on their reliability.

What, precisely, is meant by data reliability? Any data set is made up of a collection of observations of individual objects, so-called "sampling units." These observations may be actual measurements (*e.g.*, size, shape, weight, etc.) or they may be more qualitative (*e.g.*, whether the object has a certain attribute such as

[16] Technically, the decision – based on the estimated value of p – is to either accept or reject the null hypothesis. Suppose that $\alpha = 0.05$. Then if the estimated p is less than 0.05, and one rejects the null, one is reasonably assured that, at least 95% of the time, no error has been committed. But if the null is accepted, there is no reliable way to estimate the probability that, in fact, it is true. Thus, as pointed out, the correct inference is not that the null is true, but rather that it cannot be rejected with the evidence at hand. For a lucid discussion of this point in the context of legal implications of the *U.S. Endangered Species Act*, see Berry J. Brosi & Eric G. Bibber, "Statistical Inference, Type II Error, and Decision-Making Under the *US Endangered Species Act*" (2009) 7:9 Front. Ecol. Environ. 487.

[17] This equating of conditional probabilities is referred to as the transpositional fallacy, which has been committed by the Court on a number of occasions. See *e.g.*, *Reference Guide to Statistics,* in *Reference Manual for Scientific Evidence*, 3rd ed. (Washington D.C.: National Academies Press, 2011) at 211 at 250 note 99.

blue eyes). As such, there are two aspects of data reliability: the reliability of individual observations, and the reliability of the collection of observations.

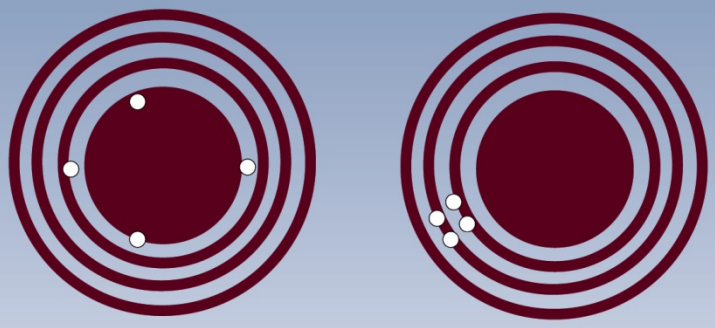
Observations have two attributes: accuracy and precision. Accuracy is a measure of the extent to which the measured value of the object reflects the true value. Precision, on the other hand, is a measure of the variability among repeated measurements of the same object under identical conditions (Figure 4).

Accuracy can be determined with reference to an object whose value is known; that is, a reference standard (*e.g.*, a sample with a known concentration of some chemical, or an object of known mass). One can then measure the reference standard, using the measuring system in question, and determine:

(a) the difference between the average of the set of repeated measurements and the reference value (known as the bias); and
(b) the variation among repeated measurements, *i.e.*, the precision.

Clearly, the greater the accuracy and precision of a measurement system, the greater the reliability of the data it generates.

Figure 4
Two Attributes of Data Reliability: Accuracy and Precision



In this illustration, repeated measurements of the same object are represented as shots at a target. The closer an individual observation (dot) is to the bull's-eye, the greater its accuracy, and the closer on average a measuring system's dots are to the bull's-eye, the greater its accuracy. Precision, on the other hand, measures how tightly repeated observations are grouped. Thus the left hand panel shows comparatively high system accuracy (on average, repeated measurements are close to the bull's-eye) but comparatively low precisions (dots are widely spaced), while the right hand panel illustrates a system for which precision is higher, but accuracy lower. Note, however, that accuracy in individual measurements cannot be achieved unless the measuring system has high precision – if the dots are not tightly grouped, they cannot all be close to the bull's-eye.

Why do accuracy and precision matter? Suppose we want to test the null hypothesis that the average weight of Canadian males is 75 kg. Suppose our scale consistently overestimates a person's weight by 1 kg. Suppose, moreover, that it is highly sensitive to ambient conditions (*e.g.*, temperature, vibration, etc.) such

that repeated measurements of the same person can differ by up to 1 kg. Thus, the estimated population mean based on a sample will be biased low, and the variance in weight in the sample will be greater than it would otherwise be because, in addition to registering the true variation in weight among males in the sample, an additional source of variation arising from the imprecision of our scale will also be included. On the one hand, bias in the sample estimate makes it more likely that the null will be rejected, because the sample estimate of the mean is "further away" from that specified by the null than it should be. On the other hand, the additional source of variation due to the scale's sensitivity to ambient conditions makes it less likely that the null will be rejected, as it increases the likelihood of achieving a result of a given discrepancy (*i.e.*, difference between the sample mean and the value specified by the null hypothesis). In either case, the estimate of *p* will be less accurate than if the scale was unbiased and insensitive to ambient conditions.

Judges must also be concerned with the reliability of the collection of observations, that is, the sample. Here the critical question is whether the method of sampling is appropriate for the question(s) being asked. By "sampling method," we mean how the objects in the collection were chosen.

An important attribute of a collection (sample) of objects is that they are independent. If the measured value of one object depends on, or is affected by, the value of another, they are not independent. Suppose one wanted to determine whether, in a particular high school, clothing brand X was more preferred than brand Y. One approach would be to affix two sheets of paper to the wall of a classroom (one for X and one for Y), give each student in the classroom a sticker, and ask them to place their sticker on the sheet of the brand they prefer. In this way, one could estimate the proportion that prefers X over Y, or *vice versa*. But where students place their sticker may well be influenced by where other classmates put their sticker, so that the observations may not be independent. (This is why all reliable voting systems strive to achieve anonymity of the voter's choice). Notwithstanding the availability of various statistical methods for detecting lack of independence of observations, reliable sampling methods strive to ensure this independence.

A second attribute of reliable sampling methods is that they aim to reduce bias. Sample bias occurs when the sample of objects does not, in some sense, "represent" the population of objects about which one is drawing an inference. If the sample is not representative, then any inference is, to some degree at least, unreliable. For example, an estimate of the average height of males in the city of Toronto, based on a sample of male college basketball players, will not be very reliable.

However, a representative sample need not be a random sample. Indeed, there may be good reasons for sampling non-randomly, depending on the intent. For example, if I wanted to obtain the most reliable estimate of the relationship between weight and height of male students at the University of Ottawa, would I

sample randomly? No. Why? Because in a finite random sample, I am very likely to miss students who are very tall or very short simply because they are rare. Thus, we will not know if the relationship derived on the basis of our random sample applies over the full range of possible heights of male students. This problem can be resolved by sampling non-randomly, that is, by ensuring that our sample includes roughly equal numbers of students from all possible height classes, from the very short to the very tall.

In summary, in assessing data reliability, the trier of fact should be concerned with:

(a) the accuracy and precision of observations;
(b) the independence of observations; and
(c) the extent to which the sample of observations represents the population from which the sample is drawn.

In particular, judges should not be reticent to question expert witnesses about potential biases in sampling, and how these potential biases were addressed.

5.3. Accuracy and Precision

Expert witnesses often provide estimates of quantities based on a sample drawn from a population; for example, the frequency of a particular DNA profile. Such estimates will always be subject to a certain amount of random error. The standard error of an estimate is a measure of the magnitude of this error: the smaller the standard error, the more precise the estimate. That is, the smaller the standard error of an estimate, the closer any estimate based on a finite sample is likely to be to the true value of the parameter in the population.

Closely related to the standard error is the confidence interval of an estimate. A confidence interval is usually centered at the sample estimate and extends in either direction in (usually) multiples of the standard error, with two standard errors corresponding to a 96% confidence interval, three standard errors a 99.7% confidence interval. If the associated standard error for a given estimate is large, with a correspondingly large confidence interval, the estimate itself may be quite inaccurate due to random error.

In the courtroom, the standard error or confidence interval is one measure of the reliability of the estimate. The trier of fact should, however, be aware that if there are problems with how the data used to derive the estimate itself were collected (*e.g.*, systematic sample bias; see *Data Reliability* at 80), these problems will not be reflected in estimated standard errors or confidence intervals. For example, if one wanted to estimate the average height of men living in Ottawa but sampled only first-generation Chinese male immigrants, the derived estimate of average height would be inaccurate, even if the sample was large and the resulting standard error small.

5.4. Statistical Null Hypotheses *Versus* Scientific Hypotheses

Few scientists are interested in statistics *per se*. Rather, they are interested in what statistical inference allows them to conclude about scientific hypotheses. But, as noted above, statistical inference concerns statistical null hypotheses. Hence, to infer anything about the scientific hypothesis under scrutiny, one must somehow relate it to the statistical null hypothesis – and not just to any statistical null hypothesis, but to the appropriate null hypothesis.

What is the appropriate null hypothesis? Recall that associated with a scientific hypothesis is one or more predictions. These predictions are simply the results one expects to see in the study or experiment under consideration if the hypothesis is indeed true. The corresponding appropriate null hypothesis is the *opposite* of this predicted (from the scientific hypothesis) pattern, so that rejection (refutation) of the statistical null hypothesis implies support for the scientific hypothesis. Hence, the more closely observed patterns and those predicted by the scientific hypothesis match, the smaller the probability of obtaining these results, given that the statistical null hypothesis is true. Or, equivalently, the closer the match, the greater the probability of achieving these results, given that the scientific hypothesis is true. The difference between an appropriate and inappropriate null hypothesis, given the scientific hypothesis in question, can be rather subtle (Table 4). For example, consider a toxic tort case where the plaintiff alleges that exposure to chemical X in the workplace caused her breast cancer. The scientific hypothesis is then that exposure to X increases breast cancer risk. The corresponding appropriate null hypothesis is that exposure to X does not increase breast cancer risk. An epidemiological study that finds a positive relationship between prevalence of breast cancer and levels of X in blood would then be inconsistent with the null hypothesis but consistent with the scientific hypothesis. All other patterns – including no relationship – would be consistent with the null, and inconsistent with the scientific hypothesis.

By contrast, the null hypothesis that there is no relationship between breast cancer prevalence and blood levels of X is inappropriate. The reason is straightforward: this null would be rejected if, for example, the prevalence of breast cancer decreased (rather than increased) with increasing levels of X in the blood; that is, the relationship between breast cancer prevalence and X is negative rather than positive. Yet a negative relationship is not that predicted by the scientific hypothesis. So there is a discrepancy: the null (of no relationship) is rejected, but the observed pattern is not that predicted by the scientific hypothesis.

Table 4
Exposure to Chemical X and Breast Cancer: Examples of Appropriate and Inappropriate Null Hypotheses

	Case 1: Exposure to X increases breast cancer risk	Case 2: Exposure to X decreases breast cancer risk	Case 3: Breast cancer risk is affected by exposure to X
Appropriate Null Hypothesis	Breast cancer risk does not increase with increasing exposure to X (one-tailed)	Breast cancer risk does not decrease with increasing exposure to X (one-tailed)	Breast cancer risk is independent of exposure to X (two-tailed)
Inappropriate Null Hypothesis	Breast cancer risk is independent of exposure to X (two-tailed)	Breast cancer risk is independent of exposure to X (two-tailed)	Exposure to X increases breast cancer risk (one-tailed)

In cases 1 and 2, the scientific hypothesis specifies a directional pattern. Hence, the appropriate null hypothesis is one-tailed (for case 1, rejection of the null would occur only if the relationship is found to be positive, that is, breast cancer prevalence increases with increasing exposure), not two-tailed (whereby rejection of the null would occur if the relationship is found to be either positive or negative). For example, in case 1, the inappropriate null hypothesis would be rejected if it were found that breast cancer prevalence decreased with exposure. Yet this is not the pattern predicted by the scientific hypothesis. So there is a contradiction: the null is rejected, but the observed pattern is not consistent with that predicted by the scientific hypothesis. For case 3, the scientific hypothesis is non-directional, so the appropriate null hypothesis is two-tailed: in essence, any relationship between breast cancer prevalence and exposure to X (increasing, decreasing, non-monotonic, etc.) is consistent with the scientific hypothesis and hence, should lead to rejection of the null. If prevalence increases with exposure, the inappropriate null would be accepted. Yet this is a pattern that is clearly consistent with the scientific hypothesis. Once again there is a contradiction: the (inappropriate) null is accepted, but the observed pattern is inconsistent with the scientific hypothesis.

In inferential statistics, patterns (or non-patterns) consistent with the statistical null hypothesis are invariably represented as mathematical models. "Testing" the null hypothesis then involves determining whether the associated mathematical model provides an adequate description of the observed pattern. If it does, the null is accepted. If it does not, the null is rejected – in which case the observed pattern resembles more that predicted by the scientific hypothesis than that corresponding to the null (see *Appendix 1: The Logical Structure of Popper's Criterion of Falsifiability* at 126).

Read more… (Please refer to Appendix 3 at 132)

- *Statistical Inference and Mathematical Models*

5.5. Errors in Statistical Inference

Given a statistical null hypothesis and a set of data with which to test it, there are two types of errors in inference: a) a true null hypothesis could be rejected (Type I or α error); or b) a false null hypothesis could be accepted (Type II or β error) (Table 5A). Both matter.

Consider, for example, a blood test to detect the HIV virus in male subject. Application of the test results in one of two conclusions: either the subject is, or is not, an HIV carrier, based on whether he tests seropositive or seronegative respectively. Because carriers are less prevalent, the null hypothesis would usually be that the subject does not carry HIV, and two types of errors are possible: (1) a subject who is not a carrier tests positive (Type I error: we infer he is a carrier when he is not, so we have rejected the true null hypothesis); or (2) a subject who is a carrier tests negative (Type II error: we infer he is not a carrier when indeed he is, *i.e.*, we have accepted a false null) (Table 5B).

Table 5

(A) Errors in Inference in Statistical Hypothesis Testing

	Reality	
Conclusion	H_0 is true	H_0 is false
Accept H_0	No error	β
Reject H_0	α	No error

(B) Errors in Inference in a Hypothetical Blood Test for HIV

	Reality	
Test	No HIV	HIV
Seronegative	99%	5%
Seropositive	1%	95%

(A) H_0 denotes the null hypothesis, which may be either true or false. Based on study results, one may infer either that it is true, or that it is false. There are then two types of errors: Type I or α, whereby a true null hypothesis is rejected (*i.e.*, one concludes the null is false when it is in fact true); and Type II or β, whereby one accepts a false null (*i.e.*, one concludes the null is true, when it is in fact false).

(B) Here, the null hypothesis is that the subject does not carry the HIV virus. If a patient who is not an HIV carrier tests positive, one rejects the null, and incorrectly concludes that he is a carrier (Type I error). If, on the other hand, a patient who is a carrier tests negative, one accepts the null, and incorrectly concludes that he is not a carrier (Type II error). In this hypothetical example, the Type I and II error rates are estimated at 0.01 and 0.05 respectively.

Closely associated with Type I and II errors are the concepts of specificity and sensitivity, often encountered in the context of diagnostic testing. Suppose a test (usually referred to as an assay) is required to determine whether a subject is carrying the H1N1 influenza virus. A highly sensitive assay is one for which true positives are very unlikely to be missed; that is, very few people who are carriers are undiagnosed. If the null is that the subject is not carrying H1N1, then this is an assay with a small Type II error rate – in very few cases is the false null accepted, so the likelihood of a false negative is low. By contrast, a highly specific assay is one for which the true negatives, *i.e.*, subjects who are not carrying H1N1, are

unlikely to be mistakenly identified as carriers. Here the rate of false positives (rejection of a true null, or Type I error) is low.

Type I and II error rates are determined by attributes of both the study design and the study results. But there is always a trade-off: designs and results that give comparatively smaller Type I errors give comparatively larger Type II errors, and *vice versa*. The question then becomes: Which is worse?

Scientists, collectively, have decided that Type I errors are more important, simply because they want to be very sure that the null is false before rejecting it (or alternatively, because the null is the opposite of the scientific hypothesis, they want to be very sure the scientific hypothesis is true before provisionally accepting it). But the determination of importance is not a statistical question, nor even a scientific question. Which is the lesser of the two evils: to inform a patient that he is an HIV carrier when he is not, or to tell him he isn't a carrier when in fact he is? The moral, ethical or economic issues involved in such a decision transcend completely the more pedestrian statistical or scientific issues.[18]

Consideration of both Type I and Type II errors is important in the weighing of scientific evidence. When evidence notionally *in support* of some scientific hypothesis is being presented, judges may wish to pay particular attention to Type I errors: evidence purporting to support a scientific hypothesis is, as noted above, evidence inconsistent with the associated statistical null hypothesis – assuming the null hypothesis has been correctly specified. The smaller the Type I error, the greater the support for the scientific hypothesis in question, all else being equal (which it never is, of course).

There are two circumstances that will lead to the null being accepted: (1) it is true; (2) it is false, but by chance, the observed data more or less conform to expectations under the null, corresponding to a Type II error. So when evidence *inconsistent* with some scientific hypothesis (and hence, consistent with the null) is being presented, judges may wish to pay particular attention to the Type II

[18] The issue of what critical Type I error rate is appropriate and in what context, has exercised scientists and statisticians for more than a century. The fact that (a) critical Type I error rates (*e.g.*, $\alpha = 0.05$) are arbitrary; and (b) once set, they implicitly determine the Type II error rate, is problematic enough. More problematic still is that depending on the context, the "cost" of a Type I versus Type II error may vary dramatically. These problems, among others, have led to some scientists suggesting that statistical null hypothesis testing ought to be abandoned altogether (see *e.g.*, Ronald Carver, "The Case Against Statistical Significance Testing, Revisited" (2003) 61 Journal of Experimental Education 287; Jeff Gill, "The Insignificance of Null Hypothesis Significance Testing" (1999) 52 Political Research Quarterly 647; and especially David R. Anderson *et al.*, "Null Hypothesis Testing: Problems, Prevalence, and an Alternative" (2000) 64 Journal of Wildlife Management 912). A possibly less heretical alternative is to set a combined (that is, both Type I and Type II) critical error rate or to determine the critical Type I rate based on the relative costs of errors (Joseph F. Mudge *et al.*, "Setting an Optimal α that Minimizes Errors in Null Hypothesis Significance Tests" (2012) 7 PLOS One e32734).

error rate. Again, all else being equal, the lower the Type II error, the stronger the evidence against the scientific hypothesis in question.[19]

IN THE COURTROOM

HIV Victim v. Dentist

Consider the example of an HIV positive plaintiff who is suing a dentist for battery and negligence further to the allegation that he had transmitted the virus to her. Two critical facts to be established are: whether the defendant dentist was indeed HIV positive, and whether he was aware of this. The defendant who is in fact an HIV carrier, was tested and found to be seronegative (a Type II error – a false null was accepted). He relied upon this test and inadvertently infected several others, including the plaintiff. An expert witness is called to explain how the tests work, and the likelihood of such an error. The court inquires about the null hypothesis (that test subjects are HIV negative) and the design of the study (Was it designed to minimize Type I or Type II errors? Was it very sensitive or specific?) to better understand the likelihood of a false negative.

The expert explains that a common screening test known as enzyme-linked immunosorbent assay (ELISA) was employed. Positive results are followed up by a second confirmatory test (western blotting or direct immunofluorescence assay) to detect antibodies to HIV in serum or plasma. The trial judge asks the expert to describe the null hypothesis for both tests. The expert replies that the null in both cases is that the patient is not an HIV carrier. When the expert is asked about the likelihood of Type I error for the first test, she responds that because ELISA is highly sensitive, it suffers from a higher than desirable Type I error rate of about 0.01, *i.e.*, that there will be one false positive result for every 100 tests.

[19] If the null hypothesis is accepted, then by definition the Type II error rate (β) must be comparatively large, usually greater than 0.5, simply because if power ($1-\beta$) were large, a significant deviation from the null would have been detected, and the null rejected. A more precise estimate of Type II error is only possible if one first specifies a minimal effect size (see *Statistical Error Rates, Sample Size and Effect Size* at 90) that one wishes to detect. For example, in a randomized clinical trial comparing the effectiveness of two different drugs, A and B, in the treatment of metastatic breast cancer, one might ask whether the response rate (the proportion of patients in a trial arm that show an objective shrinkage of their tumour mass) differs between the two drugs. Given a specified difference in response rate that would be considered clinically significant, one could then estimate quite reliably the power to detect this difference, given the number of patients enrolled in the trial. For a discussion of this and related issues, see Philip C. Kendall & William M. Grove, "Normative Comparisons in Therapy Outcome" (1988) 10 Behavioral Assessment 147.

(Recall that a Type I error means that one concludes the patient is an HIV carrier when they in fact are not – what would, in this instance, be called a "false positive." A Type II error would correspond to the situation where the null (that the patient is not a carrier) is accepted, when in fact he is a carrier, *i.e.*, a "false negative").

Because of this comparatively high sensitivity, and corresponding high false positive rate, a positive ELISA result is followed by a second confirmatory test for which the false positive rate is much lower, in low prevalence settings on the order of 1 in 250,000. Because Type II errors (false negatives) are comparatively rare due to the high sensitivity of the ELISA, the judge may conclude that it was reasonable for the defendant dentist to rely upon this test in determining his status as an HIV carrier. Of course, the court may consider other evidence in determining the reasonableness of the defendant's reliance upon the test results, such as the defendant's exposure to HIV and any symptoms he may have experienced.

Read more… (Please refer to

Appendix** 4 **at 134)

- *Statistical Errors and Inferential Strength: A Cautionary Tale*
- *Why Can We Not Minimize Both Type I and Type II Errors in Hypothesis Testing?*

5.6. Statistical Significance

The scientific literature is awash with phrases like "statistically significant difference," "no statistically significant difference," "highly significant difference," etc. What precisely do these phrases mean? They certainly sound sophisticated, exalted even. But are they?

No. As noted above, by convention scientists have adopted a threshold probability of $\alpha = 0.05$ for what is referred to as the "nominal" (or "critical") Type I error, usually denoted α. That is, by convention, scientists have decided that if the probability of getting results as discrepant as those observed, given the null hypothesis, is more than 1 in 20, they (provisionally) accept it. When scientists speak of "statistical significance," what they really mean is that for their study, the estimated probability of a Type I error (p) is less than $\alpha = 0.05$. In a similar vein, "highly statistically significant" and "very highly significant" usually correspond to $p < 0.01$ and $p < 0.001$ respectively (Table 6).

Table 6
Common Scientific Phrases and Their Meaning

Common Scientific Phrases	Meaning	Type I Error Rate
Statistically significant	The chances of getting results at least as discrepant as those observed if the null hypothesis is true, are **less than 5 in 100**.	$p < 0.05$
Highly significant	The chances of getting results at least as discrepant as those observed if the null hypothesis is true, are **less than 1 in 100**.	$p < 0.01$
Very highly significant	The chances of getting results at least as discrepant as those observed if the null hypothesis is true, are **less than 1 in 1000**.	$p < 0.001$

Phrases likely to crop up in expert witness testimony reporting on study results, their meaning, and associated Type I error rates.

Note that the nominal thresholds for rejection (α = 0.05, 0.01, etc.) of the null hypothesis are low. Because scientists are, by convention, only willing to accept low Type I error rates, the implication is that they are willing to accept comparatively high Type II error rates. This in turn implies that they care more about Type I than Type II errors.

In science this makes a certain amount of sense. Because it is a cumulative pursuit, a weak foundation would threaten to turn the house of science into a house of cards. All elements of the structure must be strong, so science is unwilling to accept any particular scientific hypothesis as even provisionally true unless the supporting evidence is relatively unequivocal.

But making good sense in the context of the *practice* of science does not imply it necessarily makes good sense in the *application* of science in the courtroom. To return to HIV testing, relying on any assay with low nominal Type I error rate (corresponding to a highly specific assay) implies that: 1) we think that telling someone they are a carrier when in fact they are not is worse than telling them they are not a carrier when in fact they are; and 2) we are willing to accept that the latter will happen more frequently than if we employed a less specific assay. This would be the consequence of the usual scientific practice of setting comparatively low nominal Type I error rates.

5.7. Compounding Statistical Errors and Statistical Significance

Suppose we set $\alpha = 0.05$ for our first experimental test of a null hypothesis that is in fact true. Suppose further that, based on the results of this experiment, we reject the null, *i.e.*, we estimate that $p = 0.04$. Because $p < \alpha$, we conclude (erroneously) that the results support our (scientific) hypothesis (as noted above in *Statistical Null Hypotheses Versus Scientific Hypotheses* at 84, statistical null hypotheses are set up so that the associated prediction is opposite to that predicted by the corresponding scientific hypothesis. Hence, if one rejects the null, one has supported the scientific hypothesis in question).

But because one swallow – even a scientific swallow – does not a summer make, we decide to do a second experiment, just to be sure. Perhaps it is a better experimental design, as we have realized – *post hoc*, inevitably – that the first design was less than ideal. We forge ahead, do the second experiment, for which we again obtain $p = 0.04 < \alpha$, and again reject the (true) null hypothesis. Let's further assume the two experiments are independent. The probability that we have drawn the wrong conclusion in Experiment 1 is 0.04, as it is for Experiment 2. So, the probability that we have drawn the right conclusion (that is, that the null is true) in both experiments is (1-0.04)(1-0.04) = 0.9216. Thus, the probability that we have drawn the wrong conclusion in at least one experiment is 1-0.9216 = 0.0784. In general, for N different tests of the same (true) null hypothesis, there is a $1-(1-0.05)^N$ probability of finding at least one experimental result that is nonetheless inconsistent with the null based on a nominal Type I error of $\alpha = 0.05$. If $N = 10$, this works out to about 0.6. We would then expect to reject the (true) null hypothesis for at least one experiment about 60% of the time. The message here is clear: if one does enough tests, one can almost always find a result that supports the scientific hypothesis under consideration, even if it is false.

A classic example of this sort of problem arises in Genome Wide Association Studies (GWAS), where scientists try to associate specific diseases with mutations in certain genes. To do this, the prevalence of the disease in question is compared in two samples: those persons who have a mutation in a certain gene, and those who do not. Usually, this is done for thousands of genes, as the objective is to find genetic mutations associated with an increased (or possibly decreased) risk of developing a particular disease. For each gene, the same null hypothesis is tested: that disease prevalence is the same in the two groups of people, those carrying and not carrying the mutation. This results in thousands of tests of the same null hypothesis. If we set $\alpha = 0.05$ for each test of, say, 2000 genes, then the overall (so called experiment-wise) Type I error rate is $p = 1-(1-0.05)^{2000}$, which is essentially 1. So, in such a study, if we set $\alpha = 0.05$ for each test, we expect to find many dozens of genes for which we reject the null and conclude there is an association with the disease in question, yet there is not. It is precisely for this reason that in such studies, researchers usually set the nominal Type I error rate for individual tests to be very small indeed (*e.g.*, $\alpha = 10^{-15}$).

The point in all this is straightforward: given a sufficiently large number of tests or experiments, one can almost always find a set of "statistically significant" results that appears to support the scientific hypothesis in question. This is particularly true in studies where large numbers of variables (such as different genes in GWAS) are measured in an attempt to uncover empirical associations with certain outcomes of interest (*e.g.*, disease presence).

Judges should be wary of individual "statistically significant" results that are mined from comparatively large numbers of trials or experiments, as the results may be "cherry picked" from a larger set of experiments or studies that yielded mostly negative results. The court might ask the expert how many other trials or experiments testing the same hypothesis he or she is aware of, and to describe the outcome of those studies.

5.8. Statistical Error Rates, Sample Size and Effect Size

As noted above, Type I and II errors are determined by attributes of both the study design and the study outcomes. One study design element that is crucially important is the sample size, *i.e.*, the number of independent observations from which one is attempting to infer whether the observed pattern is consistent with the scientific hypothesis under consideration.

Consider again the die-rolling experiment. Our null hypothesis is that the die is fair. The prediction under the null is that in a large number of rolls, the probability of getting a one, a two,…, a six, is 1/6. So in N experiments, we expect (1/6) N ones, (1/6) N twos, etc. The greater the observed deviation from these ratios, the smaller the probability of obtaining the observed deviation if the null is true.

Suppose we roll the die six times, and get one six, one five, three fours, no threes or twos, and one one. Clearly, the expected ratios for four, three and two are quite different than what we expected under the null hypothesis. For throws of four, the difference is 3/6 – 1/6 = 2/6 = 0.33. But in only six rolls, there is a reasonable chance of getting this pattern even if the die is fair, because the predicted pattern applies to the long-run probability of different outcomes. As a consequence, the null hypothesis is likely to be true even if the observed pattern deviates substantially from expectations. So if we reject the null, it is quite likely that we have committed an error, *i.e.*, the probability of a Type I error is comparatively large, in this case because the number of independent trials of die rolls (the sample size) was very small.

Now suppose we have rolled the die 600 times, of which in 300 cases, the result is a four. Again, the observed probability of a four is 300/600 = 0.5, the same as in the first experiment. So too is the deviation from expectation under the null: 300/600 – 100/600 = 0.33. But if indeed the die is fair, it is very unlikely that in 600 trials, half would turn up a four. So if we reject the null, the Type I error is far

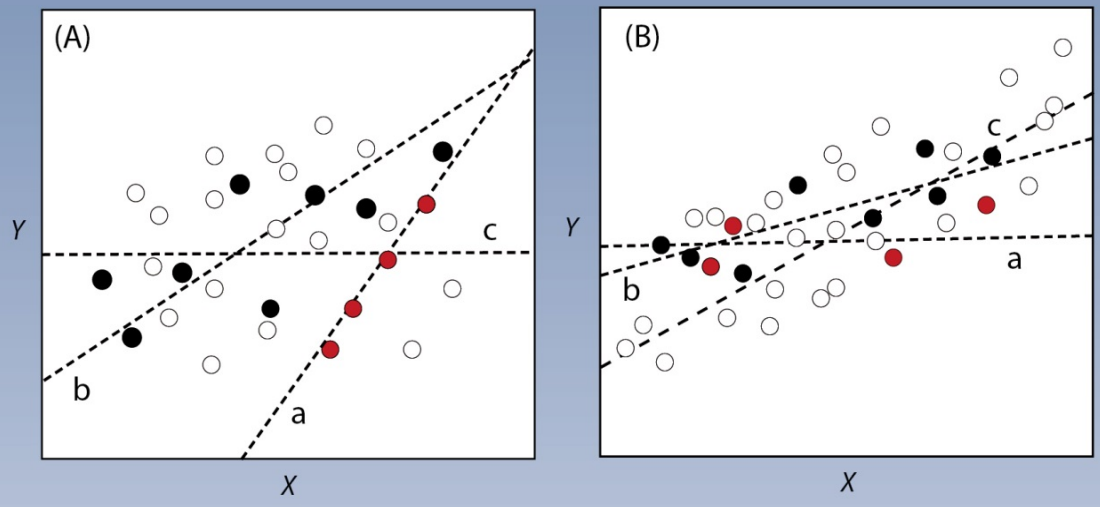
smaller than it was in the first experiment simply because of the far larger number of (independent) trials.

Why is this important? Let us return to the notion of threshold Type I error rates. As noted above, by convention scientists often set $\alpha = 0.05$ as the so-called "nominal" Type I error rate. In the first study with $N = 6$ observations, because there is quite a good chance of getting three fours even if the die is fair, p is quite large, well above the threshold $\alpha = 0.05$. So, we would accept the null, and conclude there is no evidence the die is loaded. By contrast, in the second case, the chances of getting 300 throws of four in 600 rolls if the die is indeed fair are very small. Hence, $p < \alpha = 0.05$. We therefore reject the null, and conclude the die is loaded. Notice that even though the deviation from the ratios expected under the null hypothesis is identical in both experiments, we are nonetheless led to opposite conclusions ( ).

In summary, given a certain deviation from expectations under the null, both the Type I and Type II error decrease with increasing sample size. And given a certain sample size, both the Type I and Type II error decrease with increasing deviation from the null. As explained below in more detail, this deviation from expectations under the null hypothesis is called the effect size.

Why is all this important for judges? Because a finding of "statistically significant" is meaningless unless one knows more details about both the study design and study results. If the sample size is small, the null hypothesis will be rejected only when the effect size is large, because it is only with large effect sizes that the estimated Type I error will be smaller than the nominal Type I error threshold (*e.g.*, $\alpha = 0.05$) for rejection. By contrast, when sample sizes are very large, even small effect sizes will be sufficient to reject the null. In the former case, it is entirely possible that even though the hypothesis is true, and the results are consistent with the prediction, nonetheless the null is accepted because the sample size is small. In the latter case, patterns consistent with the scientific hypothesis are detected (*i.e.*, the null is rejected), but the pattern is nonetheless very weak and may, consequently, be of little significance – that is, have little probative value.

Figure 5
The Relationship Between Statistical Errors and Sample Size



Suppose that the scientific hypothesis in question predicts a positive relationship between *Y* (*e.g.,* the prevalence of breast cancer) and *X* (*e.g.,* exposure to pesticide *X*). Here, each datum (white, black or red circles) represents a sample population for which breast cancer prevalence and average blood concentration of *X* have been estimated.

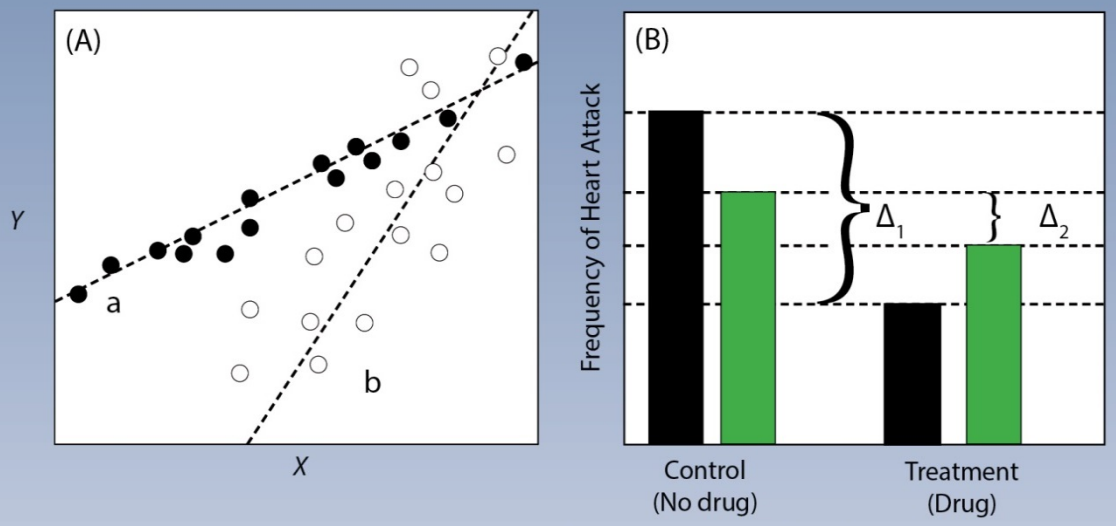
(A) Suppose that in fact, there is no relationship between *Y* and *X*. In the case of only a few observations (red circles), it is entirely possible that a strong positive relationship (*line a*) will be obtained purely by chance. As the sample size increases (red and black circles), the chances of getting a strong positive relationship purely by chance declines; if a positive relationship is observed, it is likely to be weaker (*line b*; more scatter around the fitted line). In the case of a large sample (red, black and white circles combined), a positive relationship is unlikely to obtain purely by chance, and the fitted relationship (*line c*) is much more likely to accurately represent the true relationship. In this case, the (appropriate) null hypothesis, given the scientific hypothesis, is that the relationship between *Y* and *X* will be zero or negative. Hence, as sample size increases, the chances of a Type I error decreases.

(B) Suppose now that there is in fact a positive relationship between *Y* and *X*. Given a sample with only a few observations (red circles), it is in fact quite likely that the fitted line (*line a*) will show no relationship. As sample size increases (red and black circles, and all circles combined), the chance of the fitted relationship accurately representing the true relationship increases (*lines b* and *c* respectively increase). Thus, as sample size increases, the chances of a Type II error decrease.

The above example illustrates another attribute of studies – or rather, study outcomes – that influences the "statistical significance" of a result, and already referred to above. This is the so-called effect size, which we can consider the difference between the results expected under the null hypothesis and the observed results. Suppose, for example, that in the second experiment, we observed 105 throws of four instead of 300. In this case, the difference between observed and expected ratios is small (105/600 – 100/600 = 5/600 = 0.008). It is

quite likely that this small difference might well obtain even if the die is fair. So in this case, *p* would be comparatively large, greater than the nominal level required for rejection, and we would accept the null. By contrast, (300/600 – 100/600) = 0.33 is a much larger effect size, *p* is small, and we would therefore reject the null. Note that the effect size is not a property of the experimental design, it is a property of the experimental outcome (Figure 6).

Figure 6
Effect Size and Type I Error



Because in (A), the linear relationship shown by *line a* based on the sample denoted by black circles is stronger (less scatter around the fitted line) than that based on the sample denoted by white circles (*line b*), the effect size is larger. If in fact there was no relationship between *Y* and *X*, the probability of getting the data denoted by the black circles, purely by chance, is much less than the probability of getting the data denoted by the white circles. Thus, the probability that in rejecting the null hypothesis (that there is no relationship between *Y* and *X*), one has made an error, is less for *line a* than for *line b*. Hence, Type I error decreases with increasing effect size, given a fixed sample size (in this hypothetical example, $N = 15$ in both samples); for a given sample size, the larger the effect size, the smaller the value of *p*.

(B) shows a hypothetical clinical trial in which patients are randomized into two arms: one (treatment) in which an anti-diabetic drug is given, the other (control) where no drug is administered. For both groups the frequency of heart attacks is estimated. Two possible experimental results are shown. In one (denoted by black bars), the difference between the control and treatment groups (Δ_1) is comparatively large, indicating a comparatively large effect size; in the other (denoted by green bars), the effect size (Δ_2) is much smaller. Hence the probability that in rejecting the null hypothesis (that there is no effect of the drug on risk of heart attack), one has made an error, is smaller for the first case than for the second.

IN THE COURTROOM

Lung Cancer Victim v. Polluter

Judges must be particularly vigilant when considering both statistically "significant" and "non-significant" findings. Imagine a common toxic tort proceeding where a plaintiff suffering from lung cancer claims that chronic exposure to chemical X has caused her disease. A toxicologist called by the defence might report the results of a toxicological study where sets of rats were exposed to differing chronic levels (treatments) of X and the researcher assessed the prevalence of changes in the cellular morphology of the lung (changes known to be a precursor of lung cancer) in each treatment group.

In this experiment, the hypothesis is that chronic exposure to X increases the risk of lung cancer in the exposed group. The prediction is that at some level of exposure (treatment), prevalence of these changes will be greater than in the control (unexposed) group. But the changes in lung cellular morphology associated with later progression to full blown lung cancer are rare in healthy rats, and unless chemical X is a very powerful carcinogen (in rats), these changes will also be rare (although perhaps not as rare) in exposed groups. Thus, unless the sample size in each treatment group is very large indeed, the difference (*i.e.*, effect size) in prevalence of pre-cancerous changes in the control and exposed groups may be insufficient to reject the null hypothesis. The result is that unless X is a very powerful carcinogen, the statistical null will be accepted, leading to the conclusion that X does not increase lung cancer risk, at least in rats, even if X is a moderately strong carcinogen. It is for precisely this reason that the detection of causal factors with small – or even moderate – effects on the outcome of interest (prevalence of disease, say) usually requires large or very large sample sizes.

The statistical shoe may be on the other foot. If the sample size is very large, even small effect sizes may be sufficient to reject the null hypothesis. The question then becomes: Are these effects meaningful? Consider, for example, a toxic tort case in which the plaintiff alleges that chronic exposure to asbestos in the workplace has caused her lung cancer. One piece of scientific evidence pertains to the relationship between lung cancer prevalence and asbestos exposure. A common type of epidemiological study would address this question by comparing the attributes of a large number of persons with lung cancer ("cases") with those of matched subjects ("controls") who do not have lung cancer, and compute what is called the relative risk (RR).

From these data, the relative risk of developing lung cancer if chronically exposed to low level asbestos might be estimated as, say, 1.2, meaning that persons in the low

exposure group are 1.2 times as likely to develop lung cancer as those who are not so exposed. With a large sample of cases and controls, this comparatively small effect size will be sufficient to reject the null hypothesis that prevalence of lung cancer is the same in both groups (*i.e.*, that RR = 1). In other words, there is a "statistically significant" positive association between lung cancer and asbestos exposure, at least at the population level. But if for the same sample, the relative risk for heavy smokers is 4.0, heavy smoking is much more closely associated with the development of lung cancer than chronic low-level asbestos exposure, even though both RRs are "statistically significant." Suppose that a plaintiff in a toxic tort case is a heavy smoker. This difference in effect sizes for heavy smoking versus asbestos exposure is important in evaluating the plaintiff's allegation that her lung cancer was caused by low-level asbestos exposure in the workplace. In particular, the large discrepancy in the two estimated RRs implies that the effect of asbestos exposure will, in general, be overshadowed by the smoking effect.

Whether a particular finding is statistically significant or not, the evidentiary weight of the finding must be considered carefully. As noted above, if sample size is small, large effect sizes are required to achieve statistical significance. Judges should, therefore, consider carefully the size of the estimated effect, statistical significance (or lack thereof) notwithstanding. On the other hand, when sample sizes are large, even small effect sizes will be statistically significant – but the importance of this effect relative to that of other contributing factors may be small.

The probative value of statistical evidence is, in general, large when sample size is large and effect sizes are small (corresponding to stronger evidence that the [statistical] null is true, or, equivalently, that the scientific hypothesis in question is false in the experiment under consideration) or when sample size is large and effect size is large (corresponding to stronger evidence that the statistical null is false, or equivalently, that the scientific hypothesis is true in the experiment under consideration, and the importance of the effect is comparatively large).

IN THE COURTROOM

In ***Re: Avandia*** [20]

In a recent American decision, GlaxoSmithKline (GSK) sought a motion to exclude the testimony of three of the plaintiffs' expert witnesses to support the claim that Avandia, a drug used to treat diabetes, causes heart attacks. In pursuing the motion, GSK argued that randomized control trials (RCTs) are the "gold standard" for clinical research, and that the plaintiffs' experts had no grounds for their assertion based on such trials because the association between Avandia and myocardial infarction did not reach statistical significance in any of the RCTs entered as evidence. For their part, the plaintiffs' experts argued that the RCTs upon which GSK relied were all underpowered; given the comparatively high background risk of heart attack associated with diabetes, a large sample of patients would have needed to be enrolled in the trials to have any real chance of detecting an additional effect of Avandia, and that to overcome the power problem, results from several studies would need to be combined in a statistical procedure called metanalysis. As the court found that the plaintiffs' experts' methods were the product of reliable principles and methods, and the experts had good grounds to reach their conclusions, the motion was rejected.

[20] *In Re: Avandia Marketing, Sales Practices and Products Liability Litigation* (2011), MDL No 1871 (Dist Ct ED Pa).

IN BRIEF: Statistics

As statistical analysis is an increasingly important element of expert evidence, judges find themselves in the position of having to evaluate the probative value of statistical evidence adduced notionally to support or refute some scientific hypothesis. In evaluating evidence of a statistical flavour (or aroma), judges may wish to consider the following questions.

What is the scientific hypothesis under consideration and what are the associated predictions?

As with any evaluation of scientific evidence, this is the starting point: witnesses ought to be able to clearly state both the hypothesis itself, and the associated predictions in the context of the experiment or study under consideration.

What is the associated null hypothesis?

For each scientific hypothesis and prediction, witnesses should be able to state explicitly:

a) whether there is a corresponding statistical null hypothesis that was evaluated; and

b) what, precisely, the null hypothesis was. In evaluating answers, judges should pay particular attention to the appropriateness of the stated nulls; remember that appropriate statistical null hypotheses are the logical opposite of the predictions derived from the corresponding scientific hypothesis.

How reliable are the data?

In assessing data reliability, the trier of fact should be concerned with: (a) the accuracy and precision of observations; (b) the independence of observations; and (c) the extent to which the sample of observations from which an inference is drawn truly represents the population in question. In particular, judges should not be reticent to question expert witnesses about potential biases in sampling, and how these potential biases were addressed.[21]

[21] For a more comprehensive treatment of the issue of data collection and reliability, see *Reference Guide to Statistics*, in *Reference Manual for Scientific Evidence*, 3rd ed. (Washington D.C.: National Academies Press, 2011) at 211 at 216.

How likely is it that in inferring a statistical conclusion (viz., that the null hypothesis has been accepted or, alternatively, rejected) an error has been made?

Witnesses should be able to state explicitly what the Type I and II error rates are, or, in the case of diagnostic tests, the sensitivity and specificity of the test.[22]

What size of an effect does the expert consider to be important (and why?), and how does this compare to the estimated effect size?

Unceasing vigilance on the part of judges is required here. Remember that non-zero effect sizes, even if very small, can nonetheless be "statistically significant" if the sample size is very large, whereas quite large estimated effect sizes might nonetheless be statistically non-significant if the sample size is small. Judges may wish, therefore, to consider carefully how estimated effect sizes compare to those considered by the witness to be "important." If, for example, an estimated effect size is statistically significant but much smaller than that considered important, the probative value of the evidence may be lower than in cases where the estimated effect is larger, *i.e.*, exceeding that considered important, but which is still (statistically) "non-significant" owing to small sample size.

Is the study appropriately powered?

If sample size is small, then large effect sizes will be required to reject the statistical null. It is not uncommon for the lack of statistical significance to be adduced as evidence of the absence of an effect in, among others, cases of personal injury.

[22] Type II error rates can only be calculated conditional on a specific alternate hypothesis different from the null. That is, β is the estimated probability of (incorrectly) accepting the null hypothesis when in fact a specific alternative is true. Thus, as one changes the alternative, β will change. Judges must, therefore, be very cautious when considering proffered estimates of Type II errors: it is, for example, possible to choose alternative hypotheses that are so similar to the null (*e.g.*, in a personal injury case, that the relative risk of having a heart attack while on drug X is 1.02 (the alternative) versus 1.0 (the standard null, equivalent to no elevated risk)), that the likelihood of distinguishing between the two in any study is essentially zero (*i.e.*, the Type II error rate is very high). As such, in considering estimates of Type II errors, judges must not only be clear about what the specific alternative hypothesis is, but whether in the study under consideration, it makes any sense. For more details on this issue, see Stephen E. Fienberg *et al.*, "Understanding and Evaluating Statistical Evidence in Litigation" (1995) 36:1 Jurimetrics Journal at 22-23.

6. INFERENTIAL STRENGTH, REDUX

The end result of the application of the scientific method is a conclusion, namely, that the hypothesis under consideration is supported or not. Judges are charged first with determining the admissibility of this conclusion offered as expert scientific evidence and, assuming it is admissible, its weight and probative value. Assuming the proffered evidence satisfies the *Mohan* admissibility criteria, the question before the trier of fact is: What is the likelihood that, given the adduced evidence, the scientific hypothesis is true or, alternatively, false? As we have seen, answering this question requires judges to evaluate the inferential strength of the studies adduced as evidence (see *Inferential Strength* at 64). A study that permits strong inference is one in which the investigator is very unlikely to have made an error in drawing a conclusion (*viz.*, either that the hypothesis is supported, or alternatively, that it is refuted) based on the study results. By contrast, some studies only permit weak inference: in such cases, the inferred conclusion is quite likely to be wrong. For judges then, the probative value of a scientific study is equivalent to its inferential strength.

Some factors influencing inferential strength have already been discussed (see *Inferential Strength* at 64) including:

1) the number of different ancillary assumptions required to make the relationship between hypothesis and experimental prediction deductive (or at least strongly inductive); and
2) the proportion of these assumptions that have been evaluated and found to be, at least provisionally, true.

The greater the number of ancillary assumptions, and the smaller the proportion that have been tested and found valid, the lower the inferential strength of the study.[23] The best studies thus adopt a precautionary approach to heed Murphy's Law: their authors invest considerable effort in determining the things that could go wrong in deducing predictions from a hypothesis, and design the experiment so that they are reasonably convinced that these pitfalls are avoided.

[23] The Quine-Duhem postulate of theory underdeterminism asserts that a deductive relationship between hypothesis and prediction requires that a number of ancillary assumptions (which are themselves hypotheses) must be true. So, for example, the hypothesis that the light fixture is burnt out leads to the prediction that if I replace it with a new bulb, the light will work. But there is an additional assumption here, namely, that indeed the new bulb works. How do I know this? I could, for example, test it in another fixture that I know works. Suppose that the new bulb works in the test fixture. I now take the new bulb and install it and the light still does not work. Can I now confidently refute the hypothesis that the problem with the light was a burnt out bulb? No. Why? Because it is possible that in transporting the new, apparently functional bulb, the bulb filament broke so that in fact it no longer works. According to this argument then, all hypotheses have an indeterminate number of ancillary assumptions that must be true for the hypothesis to be reliably refuted.

But there are other factors that determine inferential strength. According to Popper, science proceeds by the elimination of candidate hypotheses advanced to explain observed effects, as perhaps best evoked in Sherlock Holmes: "Whenever you have eliminated the impossible, Watson, whatever remains – however improbable – is the truth."

We can consider each test of a hypothesis as a trial by fire: the more tests it survives, the more likely it is to be true. For an individual study, each independent prediction is a test of the hypothesis in question. So, the more independent predictions, the more opportunity for refutation. Hence, inferential strength increases with the number of independent predictions that are testable within the experimental design.

Moreover, given that there are always multiple possible hypotheses, efficient winnowing of the candidate set means that studies which simultaneously test multiple hypotheses yield stronger inference. It is one thing for an experiment to provide empirical support for hypothesis A, and quite another to both provide evidence in support of hypothesis A as well as evidence against competing hypothesis B. Hence, inferential strength increases with: (1) the number of independent predictions of a given hypothesis that are tested; and (2) the number of different hypotheses tested.

Inferential strength is influenced by other attributes of study design. A particularly important attribute is whether the study is observational or manipulative (experimental). Consider two different studies designed to test the hypothesis that mercury exposure increases diabetes risk in a linear dose-dependent manner (*i.e.*, the relationship between diabetes prevalence and exposure is linear and positive). In a classic epidemiological observational study, one might measure mercury in the hair of individuals from a number of aboriginal communities across northern Canada. In each of these communities, the prevalence of Type II diabetes is also estimated. Plotting diabetes prevalence against average mercury concentration of each community might then yield a clear positive linear relationship consistent with the hypothesis.

Now consider a different experiment, where a large number of rat pups are randomized into treatment groups corresponding to differing concentrations of mercury added to their drinking water. The pups are monitored over the course of a year for signs of diabetes (rats can develop diabetes), after which the prevalence of diabetes in relation to mercury exposure is plotted. The second study obtains a relationship identical to that obtained in the first experiment. So the strength of statistical inference is very similar in the two studies.

But is the overall inferential strength the same? No. In the first case, it is quite possible to observe the predicted pattern even though the causal hypothesis is false. For example, communities with higher levels of mercury may also be those which, because of wildlife consumption advisories, have diets that are low in wild

game, and high in foods with refined sugar. By contrast, communities with low average levels of mercury in hair may have a much higher proportion of their diet being wild game. If foods high in refined sugar increase diabetes risk (and there is plenty of evidence they do), then a strongly positive relationship between average hair mercury levels and diabetes prevalence may be evident even though there is not a direct causal link between mercury exposure and diabetes. Because the second study experimentally manipulated the putative causal factor, and observed the predicted outcome, the inference that the effect is due directly to the factor in question is stronger, all else being equal.

But inevitably, all else is never equal. One problem with the second study is that the hypothesis before the court is not about the causes of diabetes in rats, but in people. It is, of course, possible that because of physiological differences that affect rates of mercury clearance from the body, rats may well develop diabetes on chronic exposure to mercury, whereas humans do not. Hence, inferential strength is reduced in the first study because of the nature of the experimental design (observational *versus* manipulative), and in the second because of the additional assumption that at least with respect to the physiology of mercury metabolism (and possibly other characteristics), people are just big rats.

The extrapolation issue requires particular vigilance on the part of the trier of fact. In the rat pup example, the critical extrapolation is from one species to another. But extrapolation can take many different forms. It is, for example, very common to encounter extrapolations in time, space or both. In studies of drug efficacy and potential side effects, even the best designed randomized controlled trials have fixed time horizons (usually two to five years). There is clearly no guarantee that estimates of, say, lifetime risk based on these data are reliable: it is entirely possible that serious side effects might manifest themselves only after 10 or 20 years of treatment. In environmental science, the effect of pollutants is often investigated in simplified laboratory settings, where, for example, routes of exposure may be very different than in real ecosystems.[24] Such studies are always susceptible to the charge that what is true in the lab need not be true in the field.

[24] The extrapolation problem has especially bedeviled the environmental sciences. Almost invariably, manipulative studies are conducted at small spatial and temporal scales (*e.g.*, experimental manipulation of "model" laboratory systems, with effects monitored over a period of weeks or months), and the results extrapolated to larger/longer scales. Observational studies are often conducted at the right spatial and temporal scale (thereby reducing the extrapolation problem and increasing inferential strength) but are not manipulative. The solution is to conduct manipulative studies at the appropriate scale – easy in principle, but often very difficult in practice. A notable exception is the Experimental Lakes Area in northern Ontario, for which controlled whole lake manipulation experiments on the impacts of pollutants, climate change, forestry and invasive species have been conducted for decades (see *e.g.*, David W. Schindler *et al.*, "Eutrophication of Lakes Cannot Be Controlled by Reducing Nitrogen Input: Results of a 37-Year Whole-Ecosystem Experiment" (2008) 105 Proc. Nat. Acad. Sci. 11254).

IN BRIEF: Inferential Strength

For judges, the probative value of a scientific study that is adduced as evidence relevant to a specific scientific hypothesis is determined by its inferential strength. A study that permits strong inference is one in which the investigator is very unlikely to have made an error in drawing a conclusion (*viz.*, either that the hypothesis is true, or alternatively, that it is false) based on the study results. By contrast, some studies only permit weak inference: in such cases, there is a greater chance that the inferred conclusion is wrong.

In attempting to assess the inferential strength of a study, judges may wish to consider the questions and issues set out in Table 7.

Table 7
Questions and Interpretation: Study Design and Experimental Results

Questions	Interpretation*
Study Design	
What assumptions must be made to make the relationship between hypothesis and prediction deductive? How many of these were actually tested?	The greater the number of assumptions that have not been tested, the lower the inferential strength.
How many different hypotheses were tested?	The greater the number of tested hypotheses, the greater the inferential strength.
For a given hypothesis, how many independent predictions were tested?	The greater the number of independent predictions tested, the greater the inferential strength.
Are putative causal variables manipulated?	Manipulative studies have, in general, greater inferential strength than observational studies.
How much extrapolation (in time, space, among species, etc.) is required?	The greater the extrapolation, the lower the inferential strength.
What is the sample size?	The larger the sample size, the lower the Type I and Type II error rate, and hence, the greater the inferential strength.
Experimental Results	
How closely do experimental results match predictions?	The closer the match, the greater the strength of the inference that the scientific hypothesis is true – at least in the study in question. The more discrepant the results from those predicted, the greater the strength of the inference that the hypothesis is false. For studies involving multiple predictions, the strength of the inference that the (scientific) hypothesis is true increases with the proportion that are borne out (that is, for which observed and predicted results match). For studies where multiple hypotheses are tested, the strength of the inference that a particular hypothesis (say, A) is true increases with the extent to which results match predictions from A, but do not match predictions from other hypotheses B, C, etc.
What is the effect size?	The larger the effect size, the greater the strength of the inference that the scientific hypothesis is true.**

* All listed interpretations should be understood as being predicated by "all else being equal" which, of course, it never is.
** This presumes that the statistical null hypothesis has been appropriately specified.

7. Normative Issues in Science – The Myth of Scientific Objectivity

Although no human undertaking can be completely objective, one redeeming value of science is that it strives for maximum objectivity. Maximum objectivity is, however, qualitatively different from total objectivity. Despite the relentless scientific pursuit of objectivity, no science is immune from subjective bias.

Consider the issue of who bears the burden of (scientific) proof. In toxic torts, there are two possibilities:

1) substance X could be assumed to be toxic (and therefore, "unsafe") unless demonstrated otherwise; or
2) substance X could be assumed to be non-toxic (and therefore, "safe") unless demonstrated otherwise.

Both represent perfectly sound scientific hypotheses as both are, at least in principle, testable, as long as "toxic" and/or "safe" are empirically well-characterized, *a priori.*

Can science tell us which one to choose? No. The choice is principally a normative one, influenced by a wide range of factors. Historically in toxic torts the convention was to, at least implicitly, consider (2) as the presumption, *i.e.*, the hypothesis was that X is non-toxic (or safe) unless demonstrated otherwise. However, some argue that the presumption should rather be (1), *i.e.*, that decision-making institutions ought properly to adopt a "precautionary" approach, or one that advocates a reverse onus (reversing the dominant presumption). In either case, the choice is a normative, not a scientific one. So we see that a critical element in the scientific process – arguably, the *most* critical element (*viz.,* the selection of the scientific hypothesis to test) – is determined largely by normative considerations.

Consider further the case of statistical errors in hypothesis testing. Being a notoriously conservative (epistemologically speaking) lot, scientists conventionally use $\alpha = 0.05$ as the nominal Type I error threshold. Thus, by convention scientists set the evidentiary standard for (null) hypothesis rejection very high indeed. But this choice of standard is arbitrary: there is no scientific principle that tells science it ought to set $\alpha = 0.05$ *versus*, say, $\alpha = 0.10$ or even $\alpha = 0.20$. In the classic Humean dichotomy, this is an "ought" question, not an "is" question, and therefore, many would argue, lies outside science.

Arbitrariness notwithstanding, the conventional setting of a low Type I error threshold may have important implications both to science itself, and in the courtroom. It means that in any study, the scientific hypothesis under investigation is presumed to be false unless we are quite sure it is not. The result is that Type II error rates are comparatively high: in other words, we will often

conclude that the scientific hypothesis is false when in fact it is true. It is clear that in some judicial contexts, the consequences of a Type I *versus* Type II error may be very different, in which case relentless cleaving to scientific convention may be inadvisable.

For example, consider a personal injury case where the plaintiff alleges that medication X, designed to treat another condition, has caused his heart attack. One potential piece of evidence adduced by the plaintiff is a randomized control trial with results showing that the estimated prevalence of heart attacks was greater in those patients who received X compared to those who did not, although the estimated Type I error was $p = 0.06$. As this is above the conventional Type I error rate ($\alpha = 0.05$) for rejection of the null that there was no difference in prevalence between the two groups, the conclusion is that the study provides no evidence in support of the scientific hypothesis. From a strict statistical hypothesis-testing perspective, such a result has no probative value with respect to the plaintiff's claim. Yet most would agree that a Type I error of 0.06 represents a higher standard of proof than, for example, the considerably lower "balance of probabilities" employed in civil proceedings. It is partly for this reason that some have argued that statistical significance may be irrelevant[25] in some judicial contexts, as strict interpretation would imply that *any* evidence for which the estimated p is greater than the critical threshold has no probative value whatsoever. On the other hand, it is generally agreed that equating Type I error with legal standards of proof is an error.[26]

Finally, consider what scientists actually measure or estimate, in the field, in the laboratory, or in the clinic. How is the "what" that they measure or estimate determined? Suppose, for example, an assessment is to be made of whether a new candidate drug for the treatment of breast cancer is better than the current gold (far from gold, actually) standard. How, operationally, do we characterize "better" – better in what sense, precisely? In cancer therapy, regulatory institutions and clinical oncologists have traditionally been concerned with specific endpoints such as response rate (the proportion of patients on the drug who show an objective reduction of their tumour mass) or duration of response (how long the patient is under therapy before the tumour begins to grow again). Thus, "better" therapies were those that had higher response rates and/or a longer duration of response.

But many cancer therapies are toxic to normal cells, so side effects can be debilitating – indeed, lethal in some patients. Especially in the palliative treatment of advanced cancer, patients are often less concerned with how long they will live, and more concerned with enjoying as high a quality of life as possible while alive. For patients who value quality of life, the severe morbidity associated with some

[25] See *e.g.*, Terence Ison, "Statistical Significance and the Distraction of 'Scientific Proof '" (2008) 27:1 Canadian Bar Review 119.

[26] See e.g., *Reference Guide to Statistics*, in *Reference Manual for Scientific Evidence*, 3rd ed. (Washington D.C.: National Academies Press, 2011) at 211 at 577.

therapies is too large a price to pay for a few additional weeks or months before death. The increased prevalence of this attitude in advanced cancer patients has led to expansion of the set of endpoints now regularly considered in clinical studies to include not only response rates and overall survival, but also specific quality of life indicators. So the *locus* of scientific investigation – what is investigated – has shifted in response to changing patient values. In science, as in all human undertakings, complete fact-value decoherence is unachievable.

Read more… (Please refer to Appendix 5 at 136)

- *Appreciating Normative and Factual Elements in Scientific Evidence*

8. WEIGHING SCIENTIFIC EVIDENCE

In both science and law, weight of evidence affects probative value: in criminal legal proceedings, for example, judges are confronted with a range of evidence, each element of which bears, in some fashion, on the guilt or innocence of the accused. In science, we are confronted with a set of scientific studies, each of which provides evidence bearing on the truth or falsity of the hypothesis under consideration. Thus in science as in law, there is the probative value of individual pieces of evidence (*e.g.*, evidence provided by different witnesses/evidence provided by individual studies) as well as the collection of evidence (the collection of witness testimonies/the collection of studies).

In science, the probative value of a study is simply its inferential strength: studies that permit strong inference have comparatively higher probative value, whereas those permitting only weak inference have comparatively lower probative value. We have already discussed some of the attributes of a study that influence its inferential strength (see *Inferential Strength* at 64 and *Statistical Null Hypotheses Versus Scientific Hypotheses* at 84). At this level then, the issue is, at least conceptually, reasonably clear-cut.

It is at the level of the collection of evidence that the issue becomes much more obscure. Just as in the legal context, scientific standards of proof based on weight of evidence pertain to the collection of evidence, not simply that provided by individual studies. Thus in adjudging scientific weight of evidence, one needs somehow to get from the attributes of a collection of studies to a conclusion – albeit provisional – that the hypothesis under consideration is indeed true or false, in precisely the same manner that a judge in a criminal proceeding must somehow get from a collection of witness testimonies to a determination of whether, given the collection of evidence, the accused is innocent or guilty.

At present, there exists no established prescriptive methodology for weight of evidence assessment in science. What we have is a set of general guidelines flowing from three general principles.

1) **Inferential strength of individual studies.** All else being equal, a collection of studies, each of which yields strong inference in itself, has an overall greater weight of evidence than a collection of studies with lower average inferential strength.
2) **Consistency.** Greater weight of evidence is associated with a collection of studies that, in some sense, all say the same thing. That is, if all or most studies in the sample provide evidence supporting the hypothesis in question, the weight of evidence that the hypothesis is true is greater than if only some of the studies are consistent with, while others are inconsistent with, the hypothesis under consideration.
   Consistency is usually evaluated within a fairly prescribed universe of experimental designs, experimental contexts, and endpoints. For example, clinical trials are often conducted in several different hospitals in several different regions using the same patient recruitment and experimental regimens. If results in each of these settings are similar, the collection would be considered to show high consistency. To return to the light fixture example (see 62), one could replicate the experiment of replacing the original bulb with a new bulb from an unopened package multiple times. The finding that in all experiments the light now works would provide strong weight of evidence for the hypothesis that the cause of the problem was a burnt-out bulb. By contrast, the result that in some cases the light works, and in others it does not, provides much more equivocal evidence.
3) **Complementarity.** This criterion is just the scientist's rendition of the old adage that if it has a bill, quacks, and waddles, it's more likely to be a duck than if it only has a bill (so do platypuses), quacks (so do coots and moorhens, among other species), and waddles (so do geese, penguins and auks, among others). In the context of chemical risk assessment, for example, the hypothesis that chemical X is toxic might be tested using a wide range of experimental designs from controlled laboratory exposure studies to large-scale epidemiological studies, using a wide range of toxicity endpoints. Convergence of multiple but complementary lines of evidence (different study designs, different endpoints, etc.) would, all else being equal, be regarded as conferring a greater weight of evidence that X is indeed toxic.

To return again to the malfunctioning light, one might conduct two quite different experiments:

a) replace the original bulb in the light with a new bulb; and
b) test the original bulb in a different fixture that is known to work.

If in a), the light works, but in b) it does not, the weight of evidence in favour of the hypothesis that the cause of the malfunction was a burnt-out bulb is greater than if the light does not work in both a) or b). In the former case, there are multiple lines of evidence from different experiments that support the hypothesis in question, whereas in the latter case, the results are conflicting.

In summary, given a collection of study results of differing inferential strength, the larger the inferential strength of studies in the collection, the greater the consistency in findings among studies, and the greater the convergence of multiple lines of evidence, the greater the overall weight of evidence.

Read more… (Please refer to Appendix 6 at 138)

- *Biases in Weight of Evidence Assessment*

9. LEGAL *VERSUS* SCIENTIFIC TERMINOLOGY

Both a comparative lexicon and a general lexicon have been prepared. The comparative lexicon serves to contrast terminology which may have different meanings in scientific or legal settings. The general lexicon covers terminology which is frequently used in the scientific community and invoked throughout this manual.

9.1. Comparative Lexicon

Term		Definition
Burden of Proof	*Legal*	There are two meanings to the phrase burden of proof. 1. The burden of production or the evidentiary burden of proof is the obligation of adducing sufficient evidence to raise an issue in the case (Rollin M. Perkins & Ronald N. Boyce, 78). 2. The burden of persuasion or the "legal burden" is a party's duty to persuade the trier of fact of the merits of their case (Rollin M. Perkins & Ronald N. Boyce, 78).
	Scientific	A scientist's responsibility to gather and adduce sufficient scientific evidence to justify the inference that the (scientific) hypothesis in question is either (a) supported; or (b) unsupported.
Evidence	*Legal*	Data such as testimony, documents or material objects presented to the trier of fact which confirm or refute alleged facts (Black's Law Dictionary, 2004). "Evidence of a fact is information that tends to prove it" (David M. Paciocco & Lee Stuesser, 1).

Fact	*Scientific*	Data gathered from scientific studies in a manner which can be reproduced by others, and serves to either support or refute a scientific theory or hypothesis. Scientific evidence pertains to one or more scientific hypotheses: data only constitute scientific evidence for (or against) a hypothesis if the probability that the hypothesis is true, given the data, is different than the probability of it being true in the absence of the data.
	Legal	A finding of fact refers to what a judge, jury, or tribunal has determined to be a fact based on evidence in the record which is often presented at the trial or hearing (Black's Law Dictionary, 2004).
Hypothesis	*Scientific*	1. An observation that is considered valid by virtue of being repeatable by independent, notionally objective, observers. Scientific facts either (1) support; (2) contradict; or (3) are irrelevant to, a given hypothesis. 2. A scientific hypothesis or theory that has been so rigorously tested as to be accepted by the overwhelming majority of scientists as true.
	Legal	A belief, idea or prediction based on evidence but not proven, which serves as a starting point of any investigation (Black's Law Dictionary, 2004).
Null Hypothesis	*Scientific*	A proposition – often causal – supposition that is the starting point of scientific inquiry and the *locus* of experimental testing. In the Popperian view, a scientific hypothesis is a proposition that can be empirically falsified; that is, could at least in principle be shown to be false.
	Legal	The null hypothesis is the presumption of innocence in the criminal justice system.
	Scientific	In statistics, null hypotheses are propositions that are assumed to be true unless demonstrated otherwise. This demonstration involves answering the question: if the null hypothesis is true, how likely is it that we would obtain the observed results?

Standard of Proof	*Legal*	The amount/weight of evidence required for the trier of fact to decide upon the ultimate issue in a given case. In the criminal justice system, the standard for rejecting the null hypothesis of "the presumption of innocence" is "beyond a reasonable doubt." In civil proceedings, the plaintiff must prove the elements of the case using the standard of the preponderance of probabilities.
	Scientific	The amount/weight of evidence required to infer, at least provisionally, that the hypothesis under investigation is supported or refuted. In statistics, it is the amount/weight of evidence required to overturn the null hypothesis; that is, to infer that the null hypothesis is false. Usually such standards are high: conventionally, unless the chance of getting the observed results, given the null is true, is very small (less than 5/100), the null hypothesis is (provisionally) accepted.
Type I Error	*Legal*	In the justice system, if the null hypothesis of the presumption of innocence is falsely rejected, the defendant is wrongly pronounced guilty, resulting in an innocent defendant being found guilty (Lynn A. Stout, 711).
	Scientific	A Type I error, also known as a false positive, is the error committed by rejecting the null hypothesis when it is in fact true (Christopher Clapham & James Nicholson, 2009).
Type II Error	*Legal*	In the justice system, if the null hypothesis of the presumption of innocence is falsely accepted (or not rejected), the defendant is wrongly pronounced innocent, resulting in a guilty defendant being found innocent (Lynn A. Stout, 711).
	Scientific	A Type II error, also known as a false negative, is the error of accepting the null hypothesis (or failing to reject a null hypothesis), when it is actually false (Christopher Clapham & James Nicholson, 2009).
Weight of Evidence	*Legal*	"What must be considered in respect of 'the weight of the evidence' is the whole that has been heard during the hearing in the light of what has gone before during the trial" (***R v Moulton***, para 66). "The weight of an individual item of evidence describes the importance that is to be attached to it. When a trier of fact weighs evidence, it considers its credibility, its reliability, and the strength of the inferences it gives rise to." (David M. Paciocco & Lee Stuesser, 41.)

	Scientific	The degree to which the available data, results or, more generally, scientific evidence, supports or refutes the hypothesis in question. In science, the weight of an individual study is determined by its inferential strength: the greater the inferential strength, the greater the weight.

9.2. General Lexicon

Term	Definition
Accuracy	The difference between the true value of an observation or parameter, and that obtained through measurement or sampling: the smaller the difference, the more accurate the measurement or estimate.
Assay	A test or procedure designed to measure or evaluate a specific chemical, physical or biological response.
Bayesian Probability	See *Probability*.
Bias	In scientific studies, bias refers to the case where, owing to problems in methodology of one form or another, the results one obtains are systematically different than reality. Bias may arise from technical problems with measurement (instrument bias), from how the set of sample units were selected (sampling bias), or from the experimental design itself. In statistics, parameter estimates are biased if the method of estimation gives rise to an estimate that is systematically different from the true value of the parameter of interest. Biased measurements or estimates are, by definition, inaccurate.
Causal Hypothesis	See *Hypothesis*.
Confidence Interval	Usually associated with a parameter estimate, and usually considered a measure of the reliability of the estimate. Confidence intervals are ranges of values that include a good estimate of the (unknown) population parameter. The correct interpretation of, say, a 95% confidence interval is that if, from a sample, one estimated a population parameter and a confidence interval, and did this many times (each time generating an estimate and confidence interval), in 95% of the cases the confidence interval would include the true value of the population parameter.
Controls	In science, controls are used to reduce the unintended influence of other factors that could potentially influence the outcome of the

	experiment; a control group or subject is (in principle, at least – if not in practice) identical to the experimental group, except the variable of interest which is being tested in the experiment is eliminated (Paula D. Johnson & David G. Besselsen, 2004).
Deductive Reasoning	See *Reasoning*.
Descriptive Hypothesis	See *Hypothesis.*
Descriptive Science	See *Science*.
Descriptive Statistics	Statistical descriptors of empirical patterns. Quantities such as the mean (average) and the variance of a sample are examples of simple descriptive statistical parameters. Other examples include measures of the association between two variables (*e.g.*, between height and weight in a sample of people) such as correlation. No attempt is made to infer any other proposition from these descriptors.
Effect Size	A measure of the magnitude of the difference between the pattern expected under the null hypothesis and the observed pattern.
Experiment	An experiment is a scientific study in which one or more variables are deliberately and (one hopes) systematically manipulated by the investigator. Scientific studies are experimental or observational – the latter being those where no variables are deliberately manipulated, but hypotheses are tested with respect to existing patterns in nature.
Frequentist Probability	See *Probability.*
Hypothesis	
Causal Hypothesis	A hypothesis that postulates a cause-and-effect relationship between attributes or variables of interest. So, for example, the causal hypothesis that carbon dioxide levels regulate photosynthetic rates in plants leads to the prediction that if CO_2 levels are increased in an experimental growth chamber, rates of photosynthesis should increase.
Descriptive Hypothesis	A descriptive hypothesis is a statement about patterns or associations between attributes of variables of interest. For example, one might hypothesize a positive linear relationship between height and weight. Such a hypothesis is not causal: the proposition is not, for example, that if one lost weight, one's height would decrease.

Scientific Hypothesis	According to Sir Karl Popper, scientific hypotheses are those capable of refutation. This means that, at least in principle, there must exist at least one result which, if observed, would lead one to conclude that the hypothesis is false. This in turn means that hypotheses may be refutable in principle, but not in practice: for example, current technological limitations may preclude measurements with sufficient accuracy or precision. By this definition, unscientific theories are propositions which are irrefutable. Non-scientific hypotheses may be true – but science provides no means by which to establish their truth or falsity.
Strong Hypothesis	A hypothesis for which, in the context of a specific study, the associated predictions are very specific; that is, there are few alternate hypotheses that make the same prediction. Thus, if one observed the predicted results, the inference that the hypothesis is true is strong.
Weak Hypothesis	A hypothesis for which, in the context of a specific study, the associated predictions are not very diagnostic; that is, there are a number of alternate hypotheses that make the same prediction. Thus, even if one observed the predicted results, the inference that the hypothesis is true is weak.
Hypothetico-Deductive Method	The "ideal" scientific method in which (a) hypotheses are advanced; (b) predictions are derived deductively therefrom; (c) observed results are compared with those predicted; and (d) an inference is drawn concerning the truth or falsity of the hypothesis under investigation. In practice, predictions are rarely – if ever – deduced from hypotheses, simply because a true deductive relationship between the two always implies the validity of other assumptions (premises) which are themselves (scientific) hypotheses and, as such, can never be known to be absolutely true.
Inductive Reasoning	See *Reasoning*.
Inferential Statistics	With inferential statistics, an inference is drawn, on the basis of a sample, to some other proposition, often about the population from which the sample is drawn. This proposition may take many different forms, but the two most likely to be encountered in a courtroom are: a) an estimate of some population parameter (and perhaps the uncertainty associated with same) based on a sample (*e.g.*, the prevalence of particular genes in a population in DNA profiling studies), or b) the truth or falsity of some statistical hypothesis (*e.g.*, that the accused is indeed guilty, given a forensic DNA profiling match).

Inferential Strength	An attribute of individual studies that depends both on the study design and the study results. The inferential strength of a study is a measure of confidence in the inference (based on the study results) that the hypothesis in question is true, or, alternatively, that it is false. Studies with high inferential strength are those for which, if one infers based on the study results that indeed the hypothesis is true (or, equivalently, false), an error is very unlikely to have been committed.
Knowledge	
Scientific Knowledge	Knowledge bearing on the truth or falsity of scientific hypotheses, rather than the procedure or methods employed to generate such knowledge (see *Technological Knowledge*). Although there are a (more or less) universally acknowledged set of general principles for the practice of science, there are no set operating standards and no owner's manuals.
Technological Knowledge	Knowledge about procedures, practices, or tools and their associated operating standards and outcomes. Usually characterized by high replicability, high predictability and low uncertainty.
Natural Sciences	See *Science*.
Precision	The variability among different measurements of the same sampling unit carried out under identical conditions, often characterized by sample statistics such as the standard error. The precision of a parameter estimate is a measure of the uncertainty associated with the estimate: the larger the uncertainty, the smaller the precision of the estimate.
Prediction	The pattern one expects to see in a particular experiment or study if the hypothesis under investigation is true.
Probability	
Bayesian Probability	The likelihood of a specified outcome or result. Under the Bayesian interpretation, the probability of an outcome is a measure of one's belief that in the experiment in question, the outcome will obtain. Thus, Bayesian probability is interpreted as measure *of the current state of knowledge*.

Frequentist Probability	The likelihood of a specified outcome or result. Under the frequentist interpretation, the probability of some "event" or "outcome" or "result" in an experiment, is the long-run frequency of that event relative to other possible outcomes.
Physical Sciences	See *Science.*
Population	In statistics, the group about which an inference based on a sample is to be drawn.
Reasoning	
Deductive Reasoning	In deductive reasoning, if all premises of the argument are true, then the conclusion must be true, assuming that a deductive fallacy has not been committed.
Inductive Reasoning	In inductive reasoning, all of the premises of the argument may be true, but the conclusion is nonetheless false.
Relative Risk	The likelihood of an outcome of interest (*e.g.*, lung cancer) occurring in sample units that have some attribute (*e.g.*, a history of smoking) compared to the likelihood in the absence of the attribute (*e.g.*, non-smokers). Relative risk greater than one means that the outcome of interest is more prevalent among those with the attribute in question than those without, whereas a relative risk less than one means that the outcome of interest is less prevalent among those with the attribute in question.
Sample	The set of sample units considered in a study, and on which observations (*e.g.*, measurements) are made. This collection of observations defines the study results.
Sample Size	The number of sample units on which information is collected during the course of a study.
Scientific Hypothesis	See *Hypothesis.*
Scientific Knowledge	See *Knowledge.*
Scientific Method	A scientific method includes a) scientific hypotheses, *i.e.*, refutable propositions; b) systematic observations in the context of a study or experiment; and c) inferences from b) to a). Predictions are the patterns one expects to see in a particular experiment or study if indeed the hypothesis is true. The study yields a set of results or observations from which one draws an inference about the truth or falsity of the hypothesis in question.

Sensitivity	Usually applied to an assay or test, in which context it relates to the probability that subjects with the outcome of interest (*e.g.*, a particular disease) are likely to be missed. A highly sensitive test or assay is one for which this probability is low.
Science	
Natural Sciences	The branches of science (*e.g.*, biology, physics and chemistry) that deal with the study of the natural world, including both the physical sciences (whose focus is the non-living elements of the natural world) and the life sciences (whose focus is the living world).
Physical Sciences	The branches of natural science (including physics and chemistry) concerned with the nature and properties of energy and non-living matter.
Social Sciences	The study of society and the relationships between individuals in society such as political science and sociology (The Oxford Dictionary of English, 2009).
Descriptive Science	Science that aims to describe patterns in the natural world, for which descriptive hypotheses can be advanced.
Specificity	Usually applied to an assay or test, in which context it relates to the probability that subjects who do not have the outcome of interest (*e.g.*, a particular disease) are identified as having it. A highly specific test or assay is one for which this probability is very low.
Statistical Significance	A Type I error rate estimated to be less than five in one hundred ("statistically significant"), less than one in one hundred ("highly significant") or less than one in one thousand ("very highly significant").
Strong Hypothesis	See *Hypothesis.*
Technological Knowledge	See *Knowledge.*
Type I error	Incorrect rejection of a true null hypothesis. The Type I error rate is the associated probability, *i.e.*, the probability that one has made an error in rejecting the null hypothesis based on the study results.
Type II error	Incorrect acceptance of a false null hypothesis. The Type II error rate is the associated probability, *i.e.*, the probability that one has made an error in accepting the null hypothesis based on the study results.

Weak Hypothesis	See *Hypothesis*.

9.3. Lexicon Bibliography

Black's Law Dictionary, 8th ed., *s.v.* "evidence" (WL).

Black's Law Dictionary, 8th ed., s.v. "fact" (WL).

Black's Law Dictionary, 8th ed., *s.v.* "hypothesis" (WL).

Black's Law Dictionary, 8th ed., *s.v.* "standard of proof" (WL).

Christopher Clapham & James Nicholson, *The Concise Oxford Dictionary of Mathematics*, (Oxford University Press, 2009) *s.v.* "probability" (Oxford Reference Online).

Christopher Clapham & James Nicholson, *The Concise Oxford Dictionary of Mathematics*, (Oxford University Press, 2009) *s.v.* "Type I error" (Oxford Reference Online).

Christopher Clapham & James Nicholson, *The Concise Oxford Dictionary of Mathematics*, (Oxford University Press, 2009) *s.v.* "Type II error" (Oxford Reference Online).

Chris Park, *A Dictionary of Environment and Conservation*, (Oxford University Press, 2007) *s.v.* "natural sciences" (Oxford Reference Online).

David M. Paciocco & Lee Stuesser, *The Law of Evidence* 5th ed. (Toronto: Irwin Law Inc., 2008).

Lynn A. Stout, "Type I Error, Type II, and the Private Securities Litigation Reform Act" (1996) 38 Ariz. L. Rev. 711.

Paula D. Johnson & David G. Besselsen, "Practical Aspects of Experimental Design in Animal Research" (2002) 43 ILAR Journal 202.

Panu Raatikainen, "The scope and limits of value-freedom in science" in Heikki J. Koskinen & Sami Pihlström, eds., *Science – A Challenge to philosophy?* (New York: Peter Lang Inc., 2006).

R.I. Levin, *Statistics for Management* 4th ed. (Englewood Cliffs: Prentice-Hall Inc., 1987).

Robert Ireland, *A Dictionary of Dentistry*, (Oxford University Press, 2007) *s.v.* "hypothesis" (Oxford Reference Online).

Rollin M. Perkins & Ronald N. Boyce, *Criminal Law and Procedure Cases and Materials* 3d ed. (New York: Foundation Press, 1982).

R v Moulton (1979), 19 AR 286 (CA).

The Dictionary of Environment and Conservation, s.v. "assay" (Oxford Reference Online).

The Oxford Dictionary of English, 2d ed., *s.v.* "fact" (Oxford Reference Online).

The Oxford Dictionary of English, 2d ed., *s.v.* "null hypothesis" (Oxford Reference Online).

The Oxford Dictionary of English, 2d ed., *s.v.* "physical sciences" (Oxford Reference Online).

The Oxford Dictionary of English, 2d ed., *s.v.* "prediction" (Oxford Reference Online).

10. APPENDIX 1

10.1. Descriptive *Versus* Causal Scientific Hypotheses

For *Daubert*, an important criterion for admissibility of scientific expert opinion is that the opinion be derived from scientific knowledge inferred from application of the scientific method.[27] A key element of the scientific method is the formulation of testable hypotheses; *i.e.*, hypotheses that can, at least in principle, be refuted (see *What is the Scientific Method?* at 60).

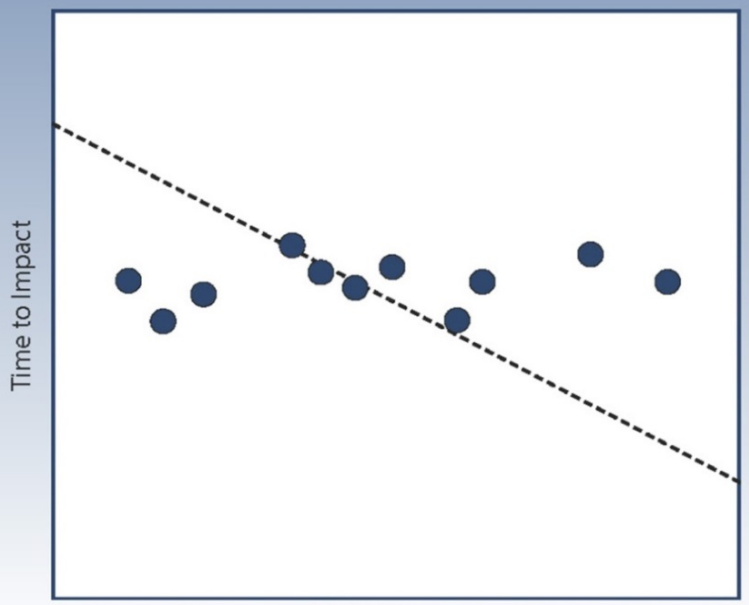
There are two types of scientific hypotheses: descriptive and causal. Descriptive hypotheses are propositions about patterns. Science that is concerned with these sorts of hypotheses is referred to – unsurprisingly – as descriptive science. As with all science, its lifeblood is systematic observation, through which scientists begin to discern patterns in the natural world. Documented patterns take the form of empirical associations between attributes of the natural world (Who? What?); in space (Where?); or time (When?). As children, many of us experimented with dropping objects off bridges into the water below. In so doing, we soon realized that heavier objects (such as stones) made far more satisfying splashes than lighter objects (such as spitballs). This pattern, which we came to recognize through a set of largely unplanned experiments, can be represented as an empirical relationship between two variables: impact force (as measured by splash size) and object mass.

A scientifically precocious child might elaborate a number of different hypotheses about the specific form of the pattern that relates impact force to mass. For example, one hypothesis might be that force of impact is directly proportional to mass. If this is indeed the true relationship, then a plot of splash size *versus* object mass should be a straight line. One could conduct a simple experiment: using a camera to record the size of the splash of objects of varying masses but approximately the same shape and volume that would be dropped from the bridge. The resulting set of observations relating splash size to the mass of the dropped object would then be analyzed to determine if the observed pattern matches the predicted relationship, namely, that the two show a linear association. If the match is good, the hypothesis is supported; if the match is poor, the hypothesis is not supported. Here, the hypothesis being tested is not a causal hypothesis (an answer to the question of *why* impact force increases with mass is not sought), it is a hypothesis about the form of the relationship; that is, a hypothesis about pattern, or a descriptive hypothesis.

[27] *Daubert*, 509 US 579 at 590.

Causal hypotheses, on the other hand, are hypotheses about how or why observed patterns are the way they are. To return to the above example, suppose we have described the pattern; that is, we now know that splash size is indeed linearly related to the mass of the object. An obvious question is: Why? One potential causal hypothesis is that the rate at which an object falls increases in proportion to its mass; that is, it is differences in the velocity at which objects of different mass fall that cause heavier objects to have larger impact force. This might be tested by recording the length of time objects of varying masses (but the same volume and shape – for example, balls of the same size but made of different materials) take to hit the water when dropped from a fixed height on a bridge. The prediction from this hypothesis is that the time to impact will decrease with the mass of the object – that is, that heavier objects will fall faster. As it turns out, the causal hypothesis is not supported (Figure 7). In contrast to the predicted pattern, the time to impact for objects of different mass is the same: as adults, we know (or ought to know) that in a fixed gravitational field, acceleration is constant and independent of mass.

Figure 7
Hypothetical Data Relating Time to Impact to Object Mass



Observations (where objects of different masses are weighed, dropped from a bridge, and the time to impact recorded) are denoted by blue dots. The predicted relationship is shown by the dashed line: this is the pattern one would predict if indeed, as hypothesized, objects fall faster in proportion to their mass. In this case, the results are inconsistent with the hypothesis, as objects of different mass fall at the same (within experimental error) rate.

From this example, it is clear that descriptive hypotheses (*i.e.*, the hypothesis that the time to impact will decrease with the mass of the object) may be predictions derived from causal hypotheses (*e.g.*, that objects of greater mass fall faster in the Earth's gravitational field), or they may simply be postulated as empirical relationships (*e.g.*, that the relationship between splash size and mass is linear). As the probative value of expert scientific evidence with respect to the ultimate

issue may well depend on whether the relevant scientific hypotheses are descriptive or causal, it is important that judges know when hypotheses about patterns are simply descriptions of empirical patterns or *bona fide* predictions derived from causal hypotheses.

10.2. Descriptive Hypotheses as Description of Patterns *Versus* Tests of Scientific Hypotheses

Descriptive hypotheses can serve two functions. On the one hand, we may advance and test descriptive hypotheses simply to achieve a more rigorous representation of patterns in nature. Here the question is of the what/where/when variety. One might ask: What is the empirical relationship between weight and height? In almost all such cases, the descriptive hypothesis being tested usually takes the form of a (sometimes implicit, often explicit) mathematical model. For example, one hypothesis might be that in people, the relationship between weight (W) and height (H) is linear; that is, a person's weight increases in proportion to his or her height. Testing the hypothesis that this linear model accurately describes the relationship between W and H involves determining whether the model provides a good representation of (technically referred to as a good "fit" to) a collection of observations, where both an individual's weight and height are recorded (Figure 8).

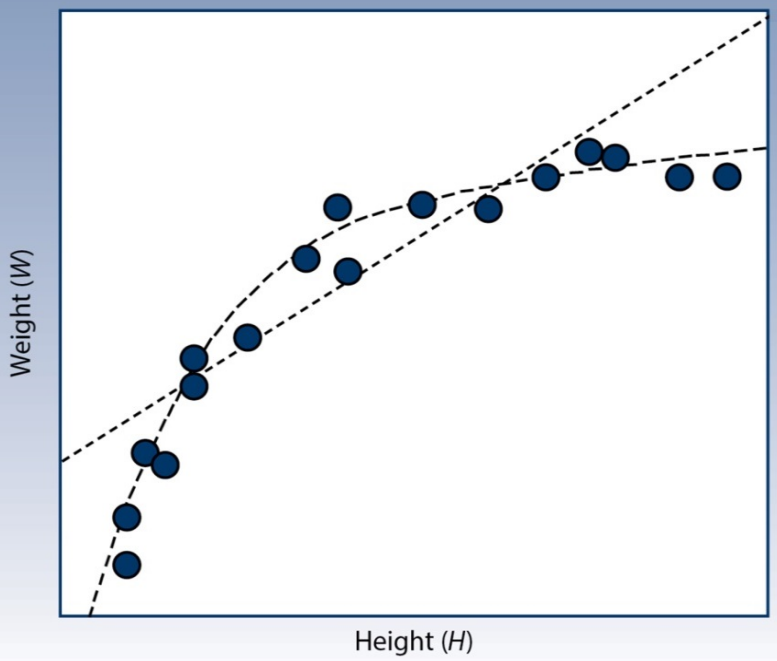
On the other hand, descriptive hypotheses may represent predictions derived from causal hypotheses. In the case of objects dropped from a bridge, one possible causal hypothesis for the observation that heavier objects produce a more satisfying splash, is that heavier objects fall faster. The prediction from this causal hypothesis is: the heavier the object, the sooner it will hit the water when dropped from the bridge; this is a statement about the empirical pattern one would expect to observe in the experiment if the causal hypothesis is true.

Thus, the difference between the two types of hypotheses is really a difference in intent. In the case of a descriptive hypothesis, the testing is to determine whether an empirical relationship (pattern) of a particular form (say, a linear relationship) exists, but there is no particular reason for evaluating a specific form. So, for example, the relationship between weight and height might be linear, but it might also be curvilinear so that the increase in weight associated with a unit change in height becomes smaller as height increases (Figure 8). These represent two different patterns, hence two different descriptive hypotheses. We can then compare the results against the two different models and evaluate which provides a better fit.

In the case of a causal hypothesis, however, we have an explicit *a priori* expectation of the form the relationship will take. This expectation is, in fact, the prediction that one derives from the causal hypothesis under consideration. To return to the example of objects being dropped from a bridge, the causal

hypothesis makes a specific prediction, namely that the relationship between impact force and object mass will be linear (Figure 7). If this model does not provide a reasonably good fit, then the causal hypothesis is not supported; if it does, the hypothesis is supported.

Figure 8
Hypothetical Data Relating to Weight and Height



Observations (where people are weighed and measured) are denoted by blue dots. Two possible relationships, each corresponding to a different descriptive hypothesis, are denoted by dashed lines. In this case, the data strongly suggest that weight and height are not linearly proportional, but rather show a nonlinear relationship.

Why does this matter for judges? It matters because the weight and/or probative value attached to the results of a study may depend on whether the study is descriptive or hypothesis-driven. Suppose one is asked to pass judgment on the hypothesis that fish respond primarily to visual cues in the yellow/orange part of the visual spectrum. A set of results is presented clearly showing that in a controlled laboratory experiment, fish attack bait coloured orange or yellow much more frequently than bait of the same size and shape but of a different colour.

There are at least two ways such a set of results could have been obtained.

1) Causal hypothesis: The experiment was designed specifically to test the hypothesis in question, *i.e.*, fish are visual predators that respond to the yellow/orange part of the visible light spectrum.
2) Descriptive hypothesis: The experiment was designed to determine whether there is any relationship between bait colour and attack rates; *i.e.*, which bait colours, if any, are most successful in attracting fish.

Suppose that the results are the same under both approaches: attack rates are found to be higher for yellow/orange bait than for other colours. Is there a difference in the weight or probative value of the study with respect to the specified hypothesis?

Yes. In the first case, only one pattern (*i.e.*, that attack rates for yellow/orange bait are greater than for bait of other colours) is consistent with the hypothesis. All other patterns – including the "non-pattern" whereby bait of different colours all have the same attack rates – would lead to rejection of the hypothesis; that is, the conclusion that the hypothesis is false.

In the second case, the implicit hypothesis being tested is that there is some relationship between bait colour and attack rates. In this case, *any* pattern is consistent with the hypothesis. Indeed, the only pattern that would lead one to reject the hypothesis is the lack of any pattern; that is, when attack rates are the same for bait of all different colours of the rainbow.

In the experiment working with the causal hypothesis, then, all outcomes but one would lead to hypothesis rejection. In the experiment working with the descriptive hypothesis, all outcomes but one lead to hypothesis acceptance. If any possible pattern but one is consistent with the hypothesis, an experimental outcome consistent with the hypothesis is, *a priori*, very likely. In such cases, the hypothesis, in the context of the chosen experimental design, is weak. Strong hypotheses are those for which the number of experimental outcomes consistent with the hypothesis is small compared to those that are inconsistent: in such cases, a finding consistent with the hypothesis yields comparatively strong evidentiary support – that is, has comparatively high probative value.

In case (1), the experimental results are indeed evidence supporting the hypothesis that fish respond preferentially to visual cues in the yellow/orange part of the spectrum. But in case (2), the results do not provide the basis for making a judgement – rather, they are evidence only that some relationship exists between bait colour and attack rates, as this was the (weak and implicit) hypothesis being tested. To present these results as a test of the specific hypothesis under consideration is not only to misrepresent the intent of the experiment, but more importantly, will lead the trier of fact to assign a greater weight and/or probative value to the evidence than it deserves. For judges then, it is critical that when an empirical pattern is presented as evidence, the nature of the study that generated it be clear.

There is a tendency among some scientists to regard descriptive science as somehow inferior (one often sees, for example, allusions to "mere" descriptive science). But this ignores the fact that the two types of science simply have very different objectives. There is also an unfortunate tendency for scientists to, *a posteriori*, present descriptive science as if it were causal hypothesis-testing science: here, as elsewhere, *caveat emptor* applies.

10.3. How Does One Distinguish Causal Hypothesis-Driven Science from Descriptive Science?

The following are some questions that may help judges make the distinction between these two sorts of scientific approaches.

- Is there a clearly stated, scientific, causal hypothesis? If not, the study is descriptive by definition. Descriptive science focuses explicitly on questions of the what/where/when variety.
- If a causal hypothesis is explicitly stated, do the variables measured, or the parameters experimentally controlled in the study, logically relate to the stated hypothesis? Are there measured quantities that were not used in the analysis, or were "added" to "see if they had any effect"? If so, this strongly suggests a descriptive study.
- Are there measured variables for which the hypotheses considered make no real predictions, *i.e.*, which appear irrelevant? If so, this strongly suggests a descriptive study.
- Are there a large number of variables measured or estimated in the study? Most tests of causal hypotheses focus on only a few experimental outcome variables or involve only a few parameters. Descriptive studies often involve measurements/estimates of dozens, hundreds or even thousands of variables (*e.g.*, expression levels of thousands of genes).
- Are stated hypotheses simply "expectations" based on what other researchers have found? Are causal mechanisms explicitly stated? "Yes" and "No" to the respective questions suggest a descriptive study.
- If the hypothesis is as stated, does the experimental design have reasonable *a priori* inferential strength? While it is certainly possible for true hypothesis-driven studies to have low *a priori* inferential strength, in a surprising number of cases this is a consequence of the fact that the study was not hypothesis-driven to begin with. (For a discussion on *Inferential Strength*, see 57.)
- In sum, the scientific weight attached to an empirical pattern that is adduced to support or refute a scientific hypothesis depends on the type of study. All else being equal, patterns derived from descriptive studies carry less weight than studies where causal hypotheses are tested explicitly. In other words, inferences as to the truth or falsity of some causal hypothesis based on patterns discerned in a descriptive study will be weaker than inferences based on patterns observed in a study explicitly designed to test the hypothesis in question. Judges should be especially vigilant in situations where results from descriptive studies may well be presented – through the glass darkly of posterity – as if they were "tests" of particular causal hypotheses, when in fact they were not.

10.4. The Logical Structure of Popper's Criterion of Falsifiability

Ideally, predictions are derived deductively from the hypothesis. That is, if in the chosen experimental set-up the hypothesis is indeed true, then the prediction follows with absolute certainty. In fact, this criterion is the foundation of Popper's notion of scientific hypotheses being refutable. The underlying logic is the classic *modus tollens* syllogism of deductive logic: If H(ypothesis) then P(rediction); not P(rediction); therefore not H(ypothesis). From this, it is also immediately apparent why, in Popper's view, scientific hypotheses can never be proven: doing so would involve the logical fallacy of affirming the consequence (if H(ypothesis) then P(rediction); P(rediction) therefore H(ypothesis)). Thus in Popper's view, even when a hypothesis has been tested and corroborated many times, there must remain, by definition, some small residual probability that it is nonetheless false.

As noted above, the logical structure of Popper's view of science as hypothesis refutation technically requires that predictions be derived deductively from hypotheses. This is rarely the case because a deductive relationship between hypothesis and prediction usually requires that other assumptions be true. Invariably these premises are either themselves hypotheses or existential statements. If the former, then in the Popperian scheme, they cannot be shown to be absolutely true – there is always some residual doubt. And if the latter, there is always some chance that the statement is in error. The consequence is that invariably, predictions are derived inductively, not deductively, from hypotheses. Hence, the proposition "If H(ypothesis) then P(rediction)" always has some finite probability of being false.[28]

[28] The dependence of the deductive relationship between hypothesis and prediction on other ancillary assumptions forms the basis for the Quine-Duhem thesis of conformational holism. This thesis asserts that all theories are underdetermined, in the sense that every theory contains an indeterminate number of underlying ancillary assumptions. Thus, empirical evidence apparently inconsistent with the theory in question does not necessarily mean the theory is false; it may be simply that one or more of the ancillary assumptions are (see *e.g.*, Pierre Duhem, *The Aim and Structure of Physical Theory* (Princeton, New Jersey: Princeton University Press, 1954); Willard Van Quine, "Two Dogmas of Empiricism" (1951) 60 The Philosophical Review 20; Willard Van Quine, *Word and Object* (Cambridge, Massachusetts: MIT Press, 1960).

Table 8
The Logic of Science as Originally Conceived by Popper, and as it is Applied in Practice

	Popperian Ideal	Reality Bites
Premise 1 (Experimental design)	If (Hypothesis is true) then (Prediction will be observed)	If (Hypothesis is true) then (Prediction will be observed with probability $q < 1$)
Premise 2 (Experimental results)	Prediction not observed	Prediction not observed with probability $p < 1$
Conclusion	Therefore, Hypothesis is false	Therefore, Hypothesis is false with probability $pq + (1-p)(1-q)$

In the table, q denotes the probability that the prediction will be observed, given the hypothesis is true. For a given hypothesis, this depends only on the experimental design. It can, therefore, be determined before any experimental results are known. Values close to 1 denote the case where predictions will almost always follow from the hypothesis, indicating a stronger experimental design. Values close to 0.5 indicate a weak experimental design: for $q = 0.5$, there is an equal chance that the prediction would or would not be observed even if the hypothesis were true. Hence, there is no real prediction, indicating a very weak experimental design.

By contrast, p values refer to the probability in the context of experimental results. When p is close to 1, it means that the results are almost surely inconsistent with predictions. When p is close to 0, it means that the results are almost surely consistent with predictions. The worst situation is when p is close to 0.5: this means that one is about equally sure that the results are, or are not, consistent with the prediction. Note that when $p = q = 0.5$, the probability of the hypothesis being false, given the results, is the same as the probability of it being true, that is, 0.5, which is the same as the probability of it being false (or true) in the absence of any results whatsoever. In other words, the experiment conveys no scientific information, *i.e.,* is scientifically irrelevant to the hypothesis under consideration.

IN THE COURTROOM

Breast Cancer Victim v. Breast Implant Manufacturer

In their role as gatekeepers, judges may need to determine whether an expert's methodology was scientifically valid. Does it matter whether the expert is reporting about results from a descriptive *versus* a causal hypothesis? The distinction can be important.

Consider the example of a negligence lawsuit in which the cause of the plaintiff's breast cancer is at issue. The plaintiff is suing the manufacturer of breast implants, alleging that the implants caused her cancer. Part of the company's defence is that the plaintiff was genetically predisposed to breast cancer, and that the

implants were not responsible. The defence has managed to enter into evidence the fact that the plaintiff has overexpression of gene A, which they argue substantially increased her risk of breast cancer. The defence has the choice of calling scientist X or scientist Y as an expert.

Scientist X's research involves trying to identify genes involved in breast cancer. Her causal hypothesis is that overexpression of gene A (which means that gene A produces much more than the normal amount of its associated protein, which results in aberrations in the signals that control cell proliferation) is a causal determinant of breast cancer. She tests this hypothesis by selecting a sample of women with diagnosed breast cancer and a set of control subjects matched with respect to age, diet, socioeconomic status, etc., who do not have breast cancer, so far as can be determined. If the hypothesis is true, the prediction is that average expression levels of gene A in women with diagnosed breast cancer should be greater than levels in the matched controls. Finding this pattern would therefore be considered to support the hypothesis. On the other hand, the finding that average expression levels of A are the same in the two groups would be inconsistent with the hypothesis.

Scientist Y's research involves describing patterns of association between levels of gene expression and breast cancer risk. She assays expression levels of thousands of genes in the same two groups described above – with current technology, this is quite easy to do. Suppose further that in doing so, she discovers that, as in the first case, average expression of A is greater in breast cancer patients than in matched controls. Note that in this case, there was no *a priori* specification of a causal hypothesis that leads to a specific prediction about average expression levels of gene A in cases *versus* controls.

Suppose that in both cases, the same result is presented, namely, that average expression of A is greater in breast cancer patients than matched controls. The presented pattern is the same – but the strength of the evidence is not the same. Because if one considers (as in the case of scientist Y's research) thousands of genes, and a comparatively small sample of women (say, several hundred in each group), it is very likely that one will find some genes for which average expression levels are greater in women with breast cancer, purely by chance. On the other hand, if one specifically considers only one gene (*i.e.*, gene A) *a priori*, out of the tens of thousands that could be assessed, it is much less likely that the observed pattern arises purely by chance. As such, scientist X's results allow for stronger inference that overexpression of A is indeed a causal determinant of breast cancer risk. In the latter case, it is entirely possible that the result has been "cherry picked" from a large number of results, which would clearly give it less probative value. In the former case, there is no cherry tree from which to pick.

Thus the defence would be better served by calling scientist X as an expert, assuming that the outcome of the research supported the causal hypothesis. Should scientist X's research be inconsistent with the hypothesis, this witness would be much more useful to the plaintiff.

11. APPENDIX 2

11.1. Contrasting Frequentist and Bayesian Probabilities

Under a Bayesian interpretation, the probability of a study outcome is a measure of our belief that for the study in question, a particular result will obtain. Thus, Bayesian probability is interpreted as a measure of the current state of knowledge; that is, as more experimental tests of a hypothesis are conducted, the prior $p(H)$ changes (*i.e.*, is updated) for the next experiment.

In the context of a scientific hypothesis designated H, the Bayesian probability that H is true given the results D of the study is given by Bayes theorem:

$$p(H|D)_B = \frac{p(D|H)p(H)}{p(D)}$$

where:

- $p(H|D)_B$ is the *posterior* probability, the probability that H is true given the study results D;
- $p(H)$ is the *prior* probability; that is, the probability that H is true before the study in question was undertaken;
- $p(D|H)$ is the probability of obtaining the results D given that the hypothesis is true; and
- $p(D)$ is the marginal probability of D, *i.e.*, the probability of obtaining the results D under any hypothesis, not just H.

In the context of the dice-rolling experiment, under a Bayesian approach, for the first experiment to test the hypothesis that the die is fair, it is reasonable to set $p(H) = 0.5$; that is, in the absence of any information whatsoever, there is an equal chance that the hypothesis is true or false. Suppose that over the first 10 rolls (experiments), we roll 5 sixes. For the 11th roll, now $p(H)$ is substantially less than 0.5, because the chances of rolling 5 sixes in 10 tries if the die is fair are rather small. Thus although neither $p(D)$ nor $p(D|H)$ change, $p(H|D)$ (the estimated probability of the hypothesis being true, given the results of the 11th experiment) is very different because of constant updating of the prior $p(H)$ based on the results of the 10 previous experiments.

What does the corresponding frequentist probability look like? Recall that under the frequentist interpretation, each "experiment" (in a long run of such experiments) is considered independent of all the previous experiments. This means that whatever $p(H)$ is for the first roll of the die, it does not change for the 10th, or 100th, or 1,000th roll.

So we can rewrite Bayes theorem as:

$$p(H|D)_F = \frac{p(D|H)K}{p(D)}$$

where K is the (unchanging) "prior" probability. From these two equations, we can compute the difference between frequentist and Bayesian probabilities:

$$p(H|D)_F = \frac{p(H|D)_F - p(H|D)_B}{p(D)} = \frac{p(D|H)(K - p(\mathrm{H}))}{p(D)}$$

From this expression, it is clear that, for any given study or experiment, the difference between the two probabilities may be substantial, depending on the values of K and $p(H)$. Note that this difference has nothing to do with the results of the experiment under consideration, which are identical in the two cases. The difference arises solely from the difference in interpretation of the probability concept and, in particular, the extent to which pre-existing information is used to define the prior probability.

11.2. Frequentist and Bayesian Probabilities in Forensic DNA Profiling

The difference between frequentist and Bayesian probability estimates can be profound. A case in point relates to forensic DNA profiling. Here two different situations can arise. In "confirmatory cases," evidence other than DNA suggests the suspect committed the crime – for example, an eye witness account. A DNA sample is taken from the suspect and when compared with DNA obtained from the crime scene is determined to be a match. This match is taken as evidence supporting the hypothesis that the DNA recovered at the scene comes from the suspect.

The question at hand is then: What is the probability of a match between the two DNA samples if they are not from the same person? This is equivalent to the probability of finding a match "at random," *i.e.*, between the DNA sample from the crime scene and a person selected at random from the population. This probability is referred to as the Random Match Probability (RMP).

In so-called "cold hit cases," DNA from the crime scene is compared with a DNA database of known offenders. If a match is found, the person is then deemed to be a suspect, and a new DNA sample is obtained. If this profile matches that found at the scene, the suspect is charged. The issue is the same as in a confirmatory case: What is the probability of a match between the new DNA sample and that recovered from the crime scene if they are not from the same person?

The difference between the two situations relates to how the suspect was first identified – through independent non-DNA evidence in confirmatory cases, or via a hit through a DNA database trawl in the second. The question is: Should this make a difference in the reported RMP? Most experts agree that the fact that the suspect was first identified via DNA match in cold hit cases introduces some ascertainment bias. Where they disagree is whether this makes the link between suspect and crime scene more or less probative, and to what degree.

This disagreement illustrates the difference between frequentist and Bayesian interpretations of probability. Under the frequentist view, if the probability of a random database entry matching a DNA sample collected at the crime scene (*i.e.*, the RMP) is p, then in a database of N entries, the probability of the first entry in the database being a match is p, as is the probability of the second entry being a match, and so on. Because each entry is considered an independent test of the hypothesis that the entry in question does match the DNA sample from the crime scene, under the frequentist interpretation, the probability of at least one match is then $p + p + p + \ldots = Np$. Thus, ascertainment bias makes for a less probative link, and the larger the database (N), the less probative the evidence.

This accords well with our intuition: if you look for a match to your surname in the telephone directory for, say, the village of Kaladar, Ontario, you might be surprised to find an unrelated match. It would be substantially less surprising to find a match in the telephone directory of greater Toronto.

By contrast, under the Bayesian interpretation, one begins with the idea that the DNA at the scene had to come from somebody. Thus, if you know that the first N entries considered in the trawl are not matches, this increases the likelihood that if there is a match, it will be from the remaining entries. And the larger the number of entries (suspects) that have been eliminated by trawling, the greater the probability that the true perpetrator is in the group that remain. The larger the number of suspects that have been eliminated (*i.e.*, the larger the size of the database), the more probative the evidence.

To return to the telephone directory example, if you consider all those Canadians who could match your name by chance, examination of the directory of greater Toronto will eliminate a larger number that do not than will examination of the Kaladar directory. Thus if you do get a match in the former, it is more likely that

the match is not by chance, thus more probative, at least from a Bayesian point of view.[29]

12. APPENDIX 3

12.1. Statistical Inference and Mathematical Models

Let us return to the childhood hypothesis that heavier objects accelerate faster in the Earth's gravitational field in proportion to their mass (see *Descriptive Versus Causal Scientific Hypotheses* at 120). The prediction from this hypothesis is that objects launched from the same height will hit the water at different times, with impact time decreasing linearly with mass. Corresponding to this prediction is a linear mathematical model of the form $T = a + bM$, where T is the time to impact, M is mass, and a and b are constants, with $b < 0$. This is just the equation of a line with a negative slope $-b$. The prediction from this hypothesis is that if objects of varying mass are dropped from the same height with the time to impact recorded, the relationship between the two will be a line with negative slope (Figure 9A).

A scientist would then proceed by "fitting" this mathematical model to the data. Fitting means, in essence, using various procedures to estimate values for the model parameters a and b. Statistical null hypotheses often then pertain to the parameters of the fitted model. These null hypotheses must always be set up so that they are the opposite of what is predicted under the scientific hypothesis, such that rejection of the statistical null hypothesis corresponds to support for the scientific hypothesis. Hence, in the present example, because the scientific hypothesis predicts a line of negative slope (*i.e.*, a value of $b < 0$), the corresponding statistical null hypothesis is $b \geq 0$ (Figure 9B).

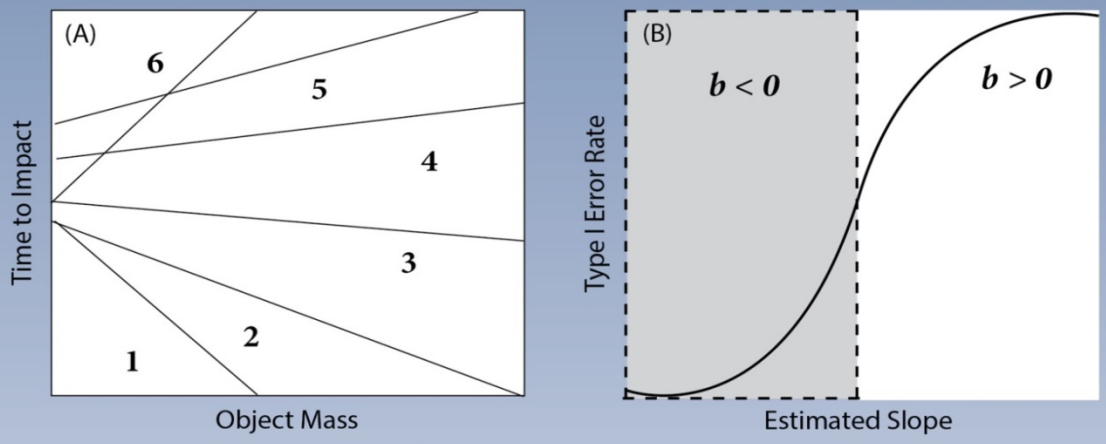
Suppose from fitting a linear model to the data we get an estimate of $b = -0.01$. Suppose further that the true value of the slope of this relationship really is zero (which in fact it is: in a uniform gravitational field, acceleration is constant independent of mass). So, even though we might be tempted to reject the null, because the estimated value is negative, it is very close to the range of values specified by the statistical null ($b \geq 0$). Given that there is always some level of uncertainty associated with any parameter estimate, it is in fact quite likely that we would get an estimated value of $b = -0.01$ even if the true value were zero. Hence, if we reject the null, it is more likely that we have committed an error – specifically, a Type I error (Figure 9B). So the Type I error rate is large, because the difference between the slope predicted under the null ($b \geq 0$) and the observed

[29] See David H. Kaye and George Sensabaugh, "Reference Guide on DNA Identification Evidence" in *Reference Guide to Statistics,* in *Reference Manual for Scientific Evidence*, 3rd ed. (Washington D.C.: National Academies Press, 2011) at 211 at 163.

slope (b = -0.01) is very small. This difference is referred to as the effect size (see *Statistical Error Rates, Sample Size and Effect Size* at 92).

Suppose that on the basis of the observed data, we obtained an estimate b = -10. Now the estimated value is very divergent from the range specified by the null – the effect size is much larger. Is it very likely that we would get an estimated value of this size if the true value were greater than or equal to zero? No. Thus, if we reject the null hypothesis in this case, the Type I error is much smaller (Figure 9B).

Figure 9
Hypothetical Predictions of the Acceleration of an Object in the Earth's Gravitational Field in Proportion to its Mass



(A) Under the hypothesis that the acceleration of an object in the Earth's gravitational field is proportional to its mass, the prediction is that the time to impact should decrease linearly with the mass of the dropped object. Thus patterns such as 1, 2 or 3 (that is, lines with a negative slope) are all consistent with the hypothesis (because a specific value of the slope is not specified; the only prediction is that it will be less than zero), whereas patterns 4 to 6 (lines with a positive slope) are inconsistent with the hypothesis.

(B) As the scientific hypothesis predicts $b < 0$, the corresponding statistical null hypothesis has $b \geq 0$, so that rejection of the statistical null hypothesis constitutes support for the scientific hypothesis. Thus, if the estimate of b based on experimental results is large and negative (*e.g., line 1* in panel A), in rejecting the null hypothesis, it is highly unlikely that an error has been made because it is very unlikely that one would observe a large negative slope if the true value is greater than or equal to zero. So, the probability of a Type I error is small. By contrast, if the estimate of b is small and negative (*e.g., line 3* in panel A) and the null is rejected, then the probability of a Type I error is substantially larger; it is, for example, entirely possible that even if the true value were zero, experimental error might result in a small negative slope estimate. On the other hand, if the estimate of b is large and positive, it is very unlikely that one would get such an estimate if indeed the true value were negative. Thus, if one were to reject the null, it is almost certain that one has made an error, so that the Type I error rate is large.

13. APPENDIX 4

13.1. Statistical Errors and Inferential Strength: A Cautionary Tale

A statistical error has occurred when, on the basis of the experimental or study results, we infer that the predicted pattern under the null hypothesis does not obtain when in fact it does (Type I error), or we infer that the pattern (or lack of pattern) predicted by the null hypothesis does obtain when in fact it does not (Type II error). There is a common tendency – even among scientists – to think that studies with a low Type I error rate yield strong inference. This is incorrect. They may yield strong statistical inference (so for premise 2 in Table 8 in **10.4: The Logical Structure of Popper's Criterion of Falsifiability**, p is either close to zero or close to one), but this says nothing about premise 1 (q in Table 8). Thus it is entirely possible for a study to permit strong statistical inference (that is, we are very sure that the results are/are not consistent with the hypothesis) yet nonetheless permit only weak overall inference because the experimental design itself has low *a priori* inferential strength.

To better understand this, consider the example of the light that doesn't work. One hypothesis is that the bulb is burnt out. Consider two different experimental designs: one (Experiment 1) in which I replace the bulb with a new one taken from an unopened package; a second (Experiment 2) where I replace the bulb from an unopened package, but I also test the new bulb in another fixture, which I have ascertained is working. For both experiments, my prediction is the same: when the old bulb is replaced with the new one, the light should work. The experimental result in both cases is the same: the light still does not work, which is inconsistent with the prediction. Whereas for these two experiments, p is the same, q is larger in Experiment 2; q for Experiment 1 is reduced by the probability that a new bulb from an unopened package does not in fact work.

13.2. Why Can We Not Minimize Both Type I and Type II Errors in Hypothesis Testing?

As noted above, statistical null hypothesis testing is liable to two different types of errors: either a true null hypothesis might be rejected (Type I error) or a false null hypothesis might be accepted (Type II error).[30] Ideally, of course, one would like to minimize both Type I and II error rates. Alas, in general we cannot: as Type I error increases, Type II error decreases, and *vice versa*.

To see this, consider again the example of designing a test for the presence of the H1N1 flu virus. As the test is designed to detect the presence of the virus, the scientific hypothesis to be tested is that the subject is indeed a carrier. The appropriate null hypothesis is then that the subject is not an H1N1 carrier, so that rejection of the null is consistent with the scientific hypothesis that the subject is a carrier. Thus, a Type II error has been committed if the subject tests negative (*i.e.*, the null is accepted), but he is in fact a carrier.

How might we go about minimizing this Type II error? Standard tests for H1N1 make use of a procedure called reverse-transcription polymerase chain reaction (RT-PCR) for amplifying (making many copies of, thereby allowing quantitation) the DNA from the RNA of cells collected with a nose swab. Diagnosis is based on the rate of amplification of specific DNA sequences (called probes or probesets) that are considered to be diagnostic of the H1N1 virus. Thus, if these sequences can be amplified rapidly, it means they were present at comparatively high abundance in the original sample, which means the source was likely an H1N1 virus. On the other hand, if they cannot be amplified, or can be amplified only very slowly, this suggests that the original source was not an H1N1 virus.

Unfortunately genetic variation exists in DNA sequences within H1N1 viral populations. Thus, if the probeset is highly specific, one risks missing *bona fide* H1N1 variants that have sequences slightly different than those included in the probeset. So, if one wants to capture these variants, the obvious solution is to expand the probeset. In this way, we ensure that subjects with variant forms are detected. This means that we will catch virtually all H1N1 carriers (so Type II error rate is low – the test is very sensitive).

The problem, however, is that as we expand the size of the probeset, we increase the likelihood of amplifying DNA that comes from a different virus, and not H1N1, as many different viruses have a high degree of genetic similarity. So by making our assay more sensitive, we reduce its specificity: while we catch almost all variants of H1N1, we also catch everything else (so that Type I error rates are high).

[30] Closely related to Type II error in statistical hypothesis testing is what is known as statistical power, defined as 1 - β. "Power" is simply the probability of correctly rejecting a false null hypothesis.

Alternatively, to minimize Type I error, we might choose a very narrow probeset. By doing so, we certainly reduce Type I error, because virtually all positive tests (*i.e.*, subjects for whom the null is rejected) are indeed H1N1 carriers. The problem is that we will also miss a lot: any variant form of H1N1 with a sequence different from those included in the (narrow) probeset returns a negative test – that is, the (false) null is accepted – so that Type II error increases.

This is the eternal conundrum faced by scientists: designing an experiment to reduce Type I error rates necessarily increases Type II error rates, and *vice versa*. As we have seen, which is more desirable is not a scientific question: this will depend on the comparative costs of Type I and Type II errors. The critical point for judges, however, is that the costs of Type I and II errors, viewed from a strictly scientific perspective – that is, as impediment to the inexorable march of science – may be very different from the extent to which they retard the (perhaps) inexorable march of justice in the courtroom.

14. APPENDIX 5

14.1. Appreciating Normative and Factual Elements in Scientific Evidence

The failure to appreciate the intermingling of the normative and non-normative in science has resulted in some interesting legal challenges. In April 2009, Dow Agrosciences LLC filed a Notice of Arbitration (NoA) under NAFTA Chapter 11, alleging that Québec failed to apply a strict science-based test to its ban of the lawn pesticide 2,4-D, as required by NAFTA's rules protecting foreign investors. The NoA asserts:

> All of these documents make clear that Québec recognized the absence of a scientific basis for its ban of 2,4-D. Moreover, even its stated reliance on an interpretation of the precautionary approach was motivated by political considerations, rather than any legitimate scientific concerns.[31]

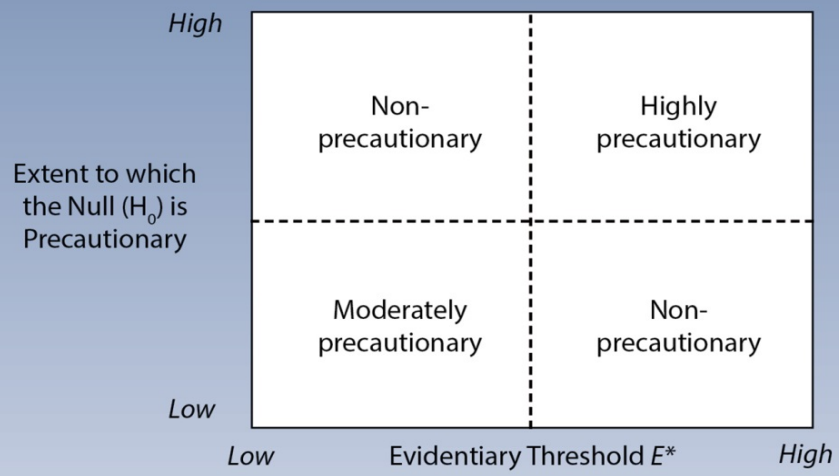
This argument attempts to represent so-called "science-based" and "precautionary" decision-making as horses of wildly different colours: the coldly analytic, precise, impersonal and objective on the one hand, *versus* the warm, fuzzy, vague and subjective on the other. But in fact these two approaches differ primarily with respect to two elements: 1) the null hypothesis and the *locus* of burden of proof (2,4-D is presumed toxic unless demonstrated otherwise, *versus*

[31] *Dow Agrosciences LLC v Government of Canada*, Notice of Arbitration (March 31, 2009) at para 25, online: < http://www.naftaclaims.com/disputes/canada/dow/dow-02.pdf >.

2,4-D is presumed non-toxic unless demonstrated otherwise); and 2) the evidentiary threshold required for rejection of the null hypothesis (Figure 10). As discussed in *Errors in Statistical Inference* at 86, neither of these issues are scientific issues – they are purely normative.

It is, therefore, entirely possible that completely opposite conclusions may be reached based on the same scientific evidence depending on where the burden of proof is determined to lie, and what standard of proof is required for rejection of the null hypothesis. It is for precisely this reason that, for example, different regulatory agencies can easily reach opposite decisions (*e.g.*, substance X is banned; substance X is registered): it is not because the science is different, but because differing normative elements in decision-making are applied.

Figure 10
Precautionary *Versus* Non-Precautionary Decision-Making



Highly precautionary decision-making in practice means: (a) choosing a highly precautionary null (reverse onus, *e.g.,* substance X is presumed to be toxic unless demonstrated otherwise); and (b) choosing a high evidentiary threshold (high standard of proof, low nominal Type I error; *e.g.,* unless there is very strong evidence X is not toxic, the null is accepted).

By contrast, non-precautionary decision-making obtains when a non-precautionary null is adopted (*e.g.,* that substance X is non-toxic) and when the evidentiary threshold for rejection of the null is high (*i.e.*, when there must be strong evidence that X is toxic before it is determined as such). Both the selection of the null (or, in legal parlance, the determination of where the onus of proof lies), and the choice of evidentiary threshold (the standard of proof) are normative, not scientific issues. Yet both have major consequences for the manner not only in which scientific evidence is presented, but in how it is interpreted.

15. APPENDIX 6

15.1. Biases in Weight of Evidence Assessment

In science, there are two dimensions to scientific evidence:

1) its mass,[32] *i.e.*, the total amount of evidence, including evidence that contradicts, as well as evidence that supports, the hypothesis under consideration; and
2) the comparative weight of evidence for and against the hypothesis in question, given the evidence pool in (1).

The mass dimension is particularly problematic in both legal and scientific settings. First, a novel hypothesis will, by definition, have little evidence one way or another simply because it has not yet been subjected to much vigorous testing. This was true initially even of today's most well-established theories. In comparison to hypotheses that have been around for some time then, novel hypotheses will necessarily score poorly on the mass dimension, at least in their infancy. In a real sense then, any weight of evidence assessment is biased against novel hypotheses or theories, simply because there isn't much mass to weigh.

Second, the amount of scientific evidence at any given time is determined, at least in part, by the number of scientific studies in which the hypothesis under consideration has been investigated. This is at least partially determined by factors that lie outside science, having more to do with the sociopolitical and economic contexts in which science is prosecuted. Moreover, scientists are as prone to clambering on passing bandwagons as anyone else. The result of these sorts of biases is that at any given time, differences between competing hypotheses in the mass of evidence – and hence, the amount of evidence that could in principle be adduced – should be interpreted carefully.

[32] In science, mass and weight are different concepts, even though the terms are often used interchangeably. The mass of an object is a measure of the amount of material present, and is context-independent. The weight of an object is a measure of the force experienced by an object due to gravity. Thus, as one changes the gravitational field, the weight of an object changes, but its mass is constant. So objects weigh less on the moon because of the weaker lunar gravitational field.

The problem of bias is not restricted to the mass dimension. Scientists have also identified several clear sources of bias relevant to the comparative weight of evidence dimension. One such bias is the so-called "file drawer" problem arising from institutional biases against the publication of "negative" results. That is, studies that provide support for a given hypothesis are more likely to be published than those that provide contradictory evidence.[33]

One potential source of the file drawer problem is the premature termination of clinical drug trials, especially those sponsored by industry.[34] Various reasons for premature termination have been given, usually falling under the general rubric of "commercial reason." One concern raised about such practices is that if interim results provided no evidence that the candidate drug was performing better than the comparator, premature termination would ensure that these negative results would not be reported, and hopefully not influence investor behaviour. Despite concerted attempts to overcome this selective publication bias through the creation of clinical trials registries,[35] major problems still remain. The consequence is that negative results tend to be left in file drawers, accumulating (possibly electronic) dust.

A second documented bias concerns the relationship between who is funding the study, and the results obtained therefrom. In drug clinical trials, positive findings of efficacy are more likely to be reported in studies funded in whole or in part by the drug manufacturer, compared to studies funded solely by public institutions. One hypothesis for this finding is that drug manufacturers are likely to invest in expensive clinical trials only for lead products for which pre-clinical work itself provides strong evidence of clinical efficacy, whereas public institutions – largely in response to public pressure – are more likely to fund trials of drugs for which the pre-clinical evidence of efficacy is weaker. Unsurprisingly, other hypotheses have been advanced.[36]

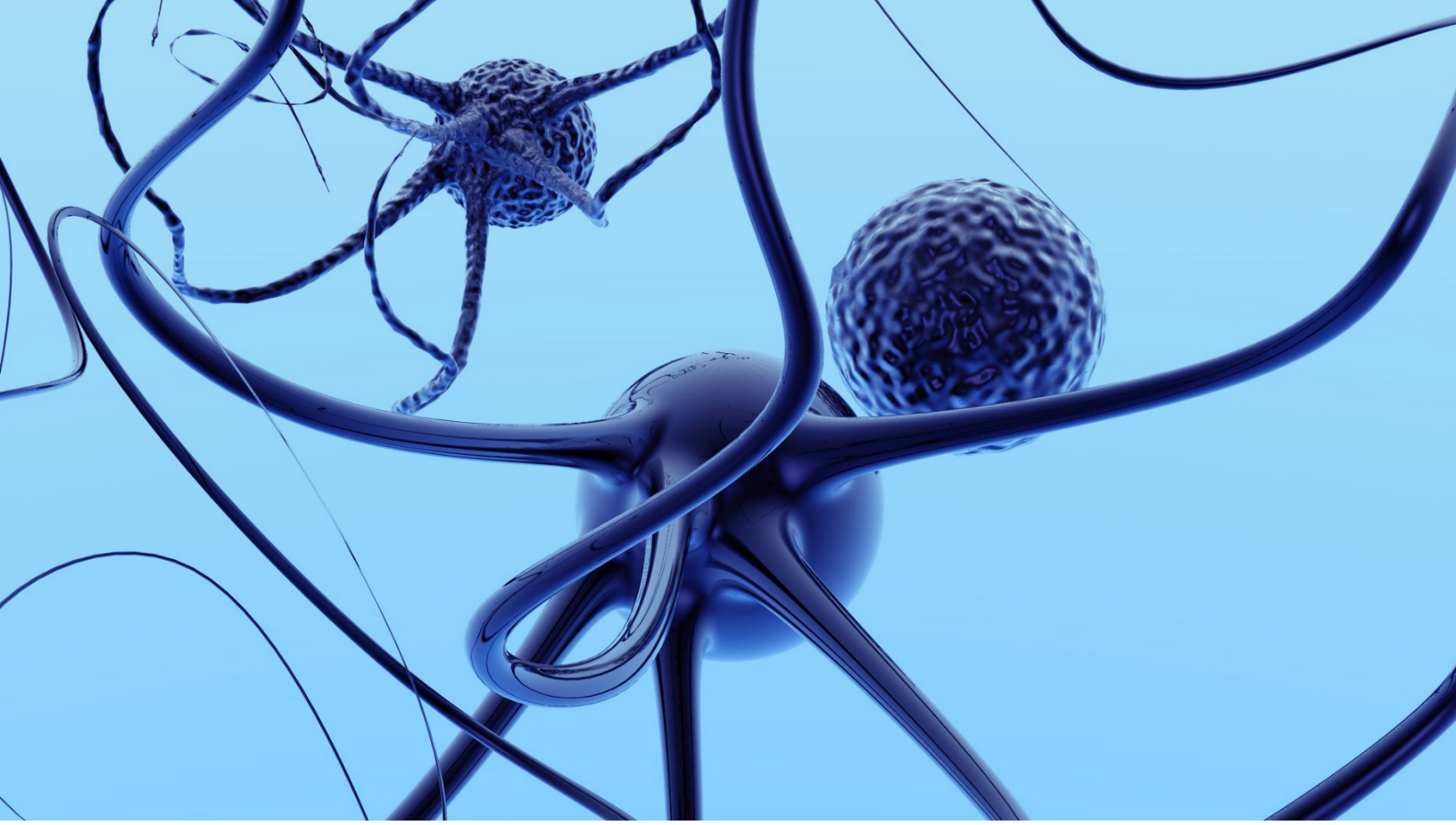
---

[33] There has been extensive documentation of publication bias, especially in the biomedical sciences. See *e.g.*, Kay Dickersin, "The Existence of Publication Bias and Risk Factors for its Occurrence" (1990) 263:10 JAMA 1385; John A. Ioannidis, "Why Most Published Research Findings Are False" (2005) 2:8 PLOS Med e124; Philippa J. Easterbrook *et al.*, "Publication Bias in Clinical Research" (1991) 337: 8746 Lancet 867.

[34] See *e.g.*, Michel Lièvre *et al*., "Premature Discontinuation of Clinical Trial for Reasons Not Related to Efficacy, Safety, or Feasibility" (2001) 322 BMJ 603. See also Anna S. Iltis, "Stopping Trials Early for Commercial Reasons: The Risk-Benefit Relationship as a Moral Compass" (2005) 31:7 Journal of Medical Ethics 410.

[35] See *e.g.*, Catherine De Angelis *et al.*, "Clinical Trial Registration: A Statement from the International Committee of Medical Journal Editors" (2004) 351 New England Journal of Medicine 1250.

[36] The problems with industry-sponsored research in biomedicine, especially in the reporting of results from clinical trials (see *e.g.*, Joel Lexchin *et al*., "Pharmaceutical Industry Sponsorship and Research Outcome and Quality: Systematic Review," (2006) 326 BMJ 1167; and Justin E. Bekelman *et al*., "Scope and Impact of Financial Conflicts of Interest in Biomedical Research," (2003) 289:4 JAMA 454) have led to persistent calls that industry-sponsored research be better regulated (see *e.g.*, Matthew Wynia & David Boren, "Better Regulation of Industry-Sponsored Clinical Trials Is Long Overdue" (2009) 37:3 Journal of Law, Medicine, & Ethics 410). These calls have recently resulted in a joint industry-academy attempt to restore credibility (Bernadette A. Mansi *et al*., "Ten Recommendations for Closing the Credibility Gap in Reporting Industry-Sponsored Clinical Research: A Joint Journal and Pharmaceutical Industry Perspective" (2012) 87 Mayo Clinic Proceedings 424).



Chapter 3
Managing and Evaluating Expert Evidence

***Note**: Special thanks to the Honourable W. Ian C. C. Binnie for the introductory remarks, the Honourable Justice Frederick Ferguson from the Court of Queen's Bench of New Brunswick; the Honourable Justice Elizabeth A. Hughes, the Honourable Justice C. Adèle Kent and the Honourable Justice Earl C. Wilson from the Court of Queen's Bench of Alberta; the Honourable Justice Anthony J. Saunders from the Supreme Court of British Columbia; the Honourable Justice André Wery and the Honourable Justice J. Brian Riordan from the Québec Superior Court; and the Honourable Justice Roger T. Hughes and the Honourable Justice Simon Noël from the Federal Court for their contribution to Part II: Managing Expert Evidence: Practical Considerations; to the Honourable Justice Stephen T. Goudge for his contribution to Part III: Evaluating Expert Evidence: The Hallmarks of Reliability; and to Sarah Berger Richardson and Flora Lê from the National Judicial Institute.*

1. FOREWORD

The management and evaluation of expert evidence is becoming dispositive of an increasing number of civil and criminal cases, and the credibility of the outcome of these cases in the eyes of the litigants, lawyers, and the broader community will often depend on how the expert evidence is handled. The *Goudge Inquiry* into the multiple failures of the legal system to expose the frailties of Dr. Charles Smith in a timely way, and the *Kaufman Inquiry* into the wrongful conviction of Guy Paul Morin, are notable examinations of a prosecutorial process gone wrong. But equally, on the civil side, a judicial misunderstanding or mishandling of expert evidence can attract serious criticism. A well-known example is a British product liability trial involving an oral contraceptive, about which the British Medical Journal (*The Lancet*) editorialized that "despite millions of pounds spent, numerous intelligent minds locked in combat, the judge failed to get to the heart of the matter."[1] The editorial concluded, in a rather dismissive comment on the adversarial system in general, that "trying science in a court of law is doomed to failure."

Similarly, the American judge who tried ***Wells v Ortho Pharmaceutical Corp***,[2] which concerned a claim that serious birth defects had been caused by a spermicide, was greeted by an editorial in the *New York Times* declaring his reasons to be "an intellectual embarrassment."[3]

In the recent U.S. litigation over the patenting of fragments of the human genome (in this case the "breast cancer gene"), James Watson, a co-winner of the Nobel Prize for the discovery of the structure of DNA, filed an *amicus* brief with the Supreme Court of the United States stating unequivocally that the courts below had totally misunderstood what DNA is all about.[4] Yet a significant element of the multibillion dollar biotechnology industry hung on the outcome. Yikes!

The heart of the debate is reliability. The court's focus has to be on what is said, not just the credentials and demeanour of who says it. Moreover, a lot of court time may be wasted unless the trial judge properly exercises a "gatekeeper" function to exclude expert evidence that is unreliable, or beyond the expertise of the witness, or is wholly unnecessary to the disposition of the case.

---

1 Editorial, "Oral contraceptives, venous thromboembolism and the courts" (2002) 325 *British Medical Journal* 504.

2 *Wells v Ortho Pharmaceutical Corp* (1986), 788 F (2d) 741 (11th Cir).

3 Editorial, "Federal judges vs. science," *New York Times,* 27 December 1986, p A22.

4 *Association for Molecular Pathology v Myriad Genetics, Inc.* (2013), 133 S Ct 2107.

There is also increasing recognition that perhaps the traditional adversarial system approach to expert evidence in some respects is inadequate. Both the *Kaufman Inquiry* and the *Goudge Inquiry* gave numerous helpful recommendations in this respect. Some reforms can be accomplished within the existing rules. Some reforms might require legislation. However, given the state of the law as it exists, the National Judicial Institute plays a key role in helping trial and appellate judges to deal appropriately with expert evidence. The present manual, together with the numerous NJI seminars and publications on related topics, continues the NJI's enviable tradition of practicality and common sense.

The Honourable W. Ian C. Binnie, Q.C.
Supreme Court of Canada (Retired)

2. INTRODUCTION

The first two chapters of this manual detail the law and science relevant to the admission of expert scientific evidence. **Chapter 3**, in turn, provides practical advice on the process of receiving and weighing such evidence. It is divided into two main sections. The first section highlights the importance of assessing the necessity of expert evidence from the outset; establishing what should be included in expert reports; pre-trial and trial management of experts and their evidence; and the innovative process of hot-tubbing. Several experienced judges were consulted and asked to share their practical advice as well as raise specific issues they consider important.

The second section addresses what is arguably the central task of a trial judge – evaluating the admissibility of expert scientific evidence. Drawing largely on the recommendations of the *Goudge Inquiry into Pediatric Forensic Pathology in Ontario*, the judge's gatekeeper function is reviewed and consideration is given to the variety of tools that may assist judges in discharging this challenging task.

While each of the chapters in this manual will change and grow as the law and science develop, this chapter has not been drafted to serve as an exhaustive guide to managing expert evidence. It is one thing to arm trial judges with the tools to manage expert evidence, but quite another to prescribe how these tools should be employed in particular cases. Different circumstances and different courtrooms will mean that approaches to the effective and efficient management of scientific evidence will vary. With time, the evolving experiences of judges called upon to admit and consider science in their courtrooms will complement this practical portion of the manual and enhance understanding of the issues.

3. MANAGING EXPERT EVIDENCE: PRACTICAL CONSIDERATIONS

3.1. General Considerations

3.1.1. The Need to Intervene Varies According to the Circumstances of the Case

Managing expert evidence is an ongoing exercise, and judges must be sensitive to the need to intervene whenever necessary. The distinction between an expert providing evidence in criminal and civil proceedings is not, in itself, an important factor the judge should consider. Indeed, the role of the judge in managing expert evidence is not significantly different whether dealing with two engineers in a construction case, or two pathologists in a murder trial. Rather, unique challenges arise when comparing the roles of the judge and the expert in the context of a trial by jury or a trial by judge alone. Similarly, judges must be alive to the issues that arise when only one party has an expert, or power imbalances result in parties retaining experts with unequal expertise and qualifications. In all of these circumstances, the gatekeeper function of the judge may be heightened.

Assessment of the need to intervene is addressed in greater detail in the *Trial Management* section below.

3.1.2. The Necessity Requirement

In ***R v D.D.***, the Supreme Court explained that the necessity requirement of the *Mohan* analysis exists to ensure that the dangers associated with expert evidence are not lightly tolerated.[5] In particular, mere helpfulness or a finding that the evidence might reasonably assist the jury is not sufficient to admit an expert's opinion. Rather, expert opinion is admissible if exceptional issues require special knowledge outside the experience of the trier of fact. The necessity requirement suggests that judges, when exercising their gatekeeper function, must intervene from the outset to ask lawyers why the expert opinion they are seeking to introduce is necessary. Early on, judges should ask counsel: "Why is this expert being called?" "Does this issue require an expert?" "Why is this evidence necessary?"

In some cases, widespread agreement can be found about what does or does not constitute necessary expert evidence. For example, there is substantial agreement that expert evidence is unnecessary when explaining the potential weaknesses of eyewitness testimony to juries because instruction from the judge, their own

[5] *R v D.D.*, 2000 SCC 43, [2000] 2 SCR 275.

common sense, examination, and cross-examination will be sufficient to guide them. In ***R v McIntosh***, expert evidence of a psychologist on the frailties of eyewitness identification was refused on the basis that it was not outside the normal experience of the trier of fact and therefore not an appropriate area for opinion evidence.[6] In reaching his decision, Finlayson J.A. cautioned against the court abdicating too easily its fact-finding responsibilities to so-called experts in "soft" behavioral sciences.

However, the reality is that different judges may come to different conclusions in similar circumstances on the question of necessity. One only has to look at the Supreme Court's split decision in *R. v. D.D.* on whether or not to admit a psychologist's evidence relating to the timing of disclosure of sexual assault by children to appreciate the difficulty that can arise in evaluating the necessity of expert evidence.

QUESTIONS TO ASK:

"Why is this expert being called?"
"Does this issue require an expert?"
"Why is this evidence necessary?"

Another example of courts inquiring into the necessity of expert opinion is ***Freyberg v Fletcher Challenge Oil and Gas Inc.***[7] In this case, Ritter J.A. found that the trial judge made palpable and overriding errors by considering expert evidence to determine the existence of an economical and profitable market for gas, because there was a sufficient factual basis to reach a decision without the need for expert evidence.

3.1.3. Proportionality

In addition to necessity, a complementary issue is the principle of proportionality. As the use of expert evidence grows, so too has the practice of using multiple experts to bolster one's case and to intimidate the opposing party. When one party retains multiple experts, the other party is often compelled to retain just as many, if not more, to compete on an equal playing field. Where both parties have the financial means to cover these additional costs, the result can be lengthy delays and more complex disputes between experts that the judge or jury will eventually be required to resolve. Where a party does not have the financial means to retain as many experts, they may find themselves at a significant disadvantage.

Judges should evaluate the proportionality of parties' intended use of expert evidence, question the necessity of multiple experts, and invoke their trial management powers to ensure the equitable and judicious use of experts so that all parties are on the same playing field.

---

6 *R v McIntosh* (1997), 35 OR (3d) 97 (CA).
7 *Freyberg v Fletcher Challenge Oil and Gas Inc.*, 2005 ABCA 46, 363 AR 35.

In Québec, the ***Code of Civil Procedure*** provides that "parties cannot seek more than one expert opinion, whether joint or not, per area or matter, unless the court authorizes otherwise given the complexity or importance of the case or the state of knowledge in the area or matter concerned."[8] The Code of Civil Procedure also requires parties to provide the court with reasons if they do not intend to jointly seek expert opinion.[9] Québec judges have the discretion to assess the usefulness of expert opinion at any stage of a proceeding. Where necessary to uphold the principles of proportionality, the court may set a time limit for submission of the expert report, or impose joint expert evidence.[10]

3.1.4. Novel Disciplines and Junk Science

The judge should always be vigilant for junk science trying to slip into the courtroom disguised as expert opinion evidence. The gatekeeper function is broader than simply assessing expert qualifications; often it involves an assessment of the expertise itself. The judge can refuse to hear the evidence of an expert whose area of expertise is not a recognized, that is, a legally reliable scientific or social scientific discipline. For example, in ***R v Dimitrov***, the analysis of barefoot insole impressions to establish that the accused wore a particular shoe was not deemed to be admissible expert opinion evidence because the expert's research had not reached the stage where he could make a categorical identification from barefoot impressions.[11]

3.1.5. The Importance of the Gatekeeper Function

Judges must exercise their gatekeeping function even in seemingly routine situations or in proceedings where the rules of evidence may be relaxed. Judges should carefully scrutinize the qualifications of experts and the adequacy and reliability of their methodology before admitting expert reports. This is equally important in Criminal, Family, and Civil Law cases. Miscarriages of justice can occur in any of these areas if judges admit flawed expert evidence.

The ***Report of the Motherisk Hair Analysis Independent Review***[12] suggests that the trial judge should remain vigilant even where expert evidence is tendered on consent. For approximately 15 years, the Motherisk Drug Testing Laboratory at Toronto's Hospital for Sick Children (MDTL) was a leader for hair strand drug and alcohol testing. Child protection agencies frequently relied on the results of these

---

8 *Code of Civil Procedure*, CQLR c C-25.01, art 232 (*CCP*).
9 *CCP*, CQLR c C-25.01, art 148.
10 *CCP*, CQLR c C-25.01, art 158(2).
11 *R v Dimitrov* (2003), 68 OR (3d) 641 (CA), leave to appeal refused [2004] SCCA No 59.
12 Ontario Ministry of the Attorney General, *Report of the Motherisk Hair Analysis Independent Review*, by Susan Lang (Toronto: Ontario Ministry of the Attorney General, 2015) [Lang].

tests, and courts routinely admitted them. As the Independent Reviewer, the Honourable Susan Lang, noted:

Although these results and interpretations were expert opinion evidence, they were infrequently recognized or treated as such by child protection counsel, parents' and children's counsel, or the court. They were often introduced into evidence on consent. As a result, the MDTL evidence was rarely tested against the admissibility requirements for expert opinion evidence that apply in civil proceedings, including child protection proceedings. Accordingly, courts frequently accepted the test results and the concentration ranges as a reliable measure of use – with little, if any, examination of the forensic qualifications of the MDTL representative who communicated those results to the child protection agency, and without any analysis of the adequacy and reliability of the test methodology employed by the Laboratory.[13]

The Independent Review identified a number of flaws in the MDTL's analytical practices and concluded that the "hair-strand drug and alcohol testing used by MDTL between 2005 and 2015 was inadequate and unreliable for use in child protection and criminal proceedings."[14] Additionally, none of the staff were trained forensic toxicologists and they lacked the expertise to provide a forensic interpretation of the results.

Positive results from MDTK were used in over 9,000 cases between 2005 and 2015.[15] The Government of Ontario established the Motherisk Commission in January 2016 to review child protection cases where individuals have been affected by the MDTL's flawed hair strand testing. This is but one example of the potential impact of admitting expert evidence without proper scrutiny.

3.2. The Report

3.2.1. Preliminary Matters

Language: The judge should remind counsel that expert reports must be written so as to be understandable to a judge and jury. The report should anticipate the language that the expert will use when testifying, and define all terminology used.

Time limits: While expert reports must be filed as evidence within the established time limits, judges should be wary of sanctioning delays. Sanctions ultimately prejudice the client, who may not be the person responsible for the delay.

[13] Lang, *Report of the Motherisk Hair Analysis Independent Review* at 16.
[14] Lang, *Report of the Motherisk Hair Analysis Independent Review* at 5.
[15] Lang, *Report of the Motherisk Hair Analysis Independent Review* at 222.

Report in lieu of testimony: In British Columbia, when an expert report is filed in a civil case, the report is the evidence, and an examination in chief may only be used to explain ambiguities or technical terms. In Québec, the ***Code of Civil Procedure*** provides that the expert report stands in lieu of the expert's testimony.[16] Québec legislation restricts examination in chief to seeking clarification or obtaining an opinion on new evidence introduced during the trial.[17]

Admissibility of the expert report: In criminal matters, the common practice is to admit the expert report into evidence regardless of whether the witness will be giving *viva voce* evidence. This practice is not explicitly set out in s 657.3(1) of the *Criminal Code*, but this apparent inconsistency does not appear to raise any issue in practice.

3.2.2. Proving All Factual Assumptions

As a matter of common sense, there must be a foundation in the evidence for all opinions stated in the expert report, and if facts are being assumed, they must be proven. The report should explain, in detail, the evidence that the expert has reviewed. It should do more than summarize the evidence that has been consulted. Practically, however, this does not always happen. Increasingly, the practice of experts has become to review volumes of records and reports, but only list them in the report without discussing their significance.

In British Columbia, notwithstanding that medical records are an exception to the hearsay rule, the Court of Appeal has discouraged the practice of filing volumes of medical records that are not referred to in the evidence.[18] As a result, only a limited number of records are filed as evidence, leaving judges with an incomplete picture of the records that were reviewed by the experts. This can lead to problems if, during cross-examination (or any other time), experts are asked to clarify a statement or assumption they have made, and they consult their own notes or refer to documents that were not filed as evidence in order to respond to the question.

QUESTION TO ASK:

"It has become evident to me that the records just referred to by the expert should be included in the evidence. Could you please discuss amongst yourself and reach an agreement as to documents?"

[16] *CCP*, CQLR c C-25.01, art 293.
[17] *CCP*, CQLR c C-25.01, art 294.
[18] *Samuel v Chrysler Credit Canada Ltd.*, 2007 BCCA 431; 70 BCLR (4th) 247.

Where cross-examinations uncover assumptions by experts that were not made explicit in their reports, the judge may intervene to say: "It has become evident to me that the records just referred to by the expert should be included in the evidence. Could you please discuss amongst yourselves and reach an agreement as to documents?" To avoid this problem altogether, the judge should, from the outset, ask counsel to ensure that all evidence that will be referred to by the expert (including during cross-examination) is properly filed as evidence. The judge can also ask counsel to confirm when they file a joint book document that the contents represent the entirety of the documentary record upon which the expert is relying.

3.2.3. Rebuttal Reports

Rebuttal reports are an exception to the requirement for advance written notice of the expert's view. Judges should remain vigilant to ensure that rebuttal reports are not used as a device to introduce new evidence or a new opinion. For example, in ***Canadian National Railway Co. v Canada***, Henderson J. found that an expert report was inadmissible as reply or rebuttal evidence because it made no effort to respond directly to the defence's experts nor criticize their assumptions and methodology.[19] In the same case, portions of a separate report were admitted as reply evidence because they provided a critical review of the analysis of the defence experts' reports.

While the judge must be wary of fresh opinion evidence "masquerading"[20] as rebuttal evidence, the expert should be allowed to comment on the theories of the primary report and provide competing theories to explain the phenomena in issue. Indeed, in ***Wade v Baxter***, Slatter J. cautioned that: "The concept of a 'rebuttal' report should not be so narrowly construed that the rebutting expert must accept the way the original expert has defined the question."[21] The judge must give a certain degree of leeway to the expert to further explain his methodology, or why the opposing expert is incorrect by giving additional details on the expert's own analysis.

---

[19] *Canadian National Railway Co. v Canada*, 2002 BCSC 1669, 8 BCLR (4th) 323 at paras 25-26.

[20] See *Kroll v Eli Lilly Canada Inc.* (1995), 5 BCLR (3d) 7 (SC) at para 7, Saunders J.: "I consider that the law as enunciated in *Pedersen v Degelder* is still applicable to response to expert reports, and note that this exception to the requirement for advance written notice of the expert's view, limited strictly to true response evidence, does not permit fresh opinion evidence to masquerade as answer to the other side's reports."

[21] *Wade v Baxter*, 2001 ABQB 812, 302 AR 1 at para 71.

If there is a concern that rebuttal reports are being improperly used to circumvent the requirement of advance written notice of the expert's view, judges should intervene and inquire further into the purpose of the report, even if the opposing counsel is silent on the matter. For reasons of fairness, judges must ensure that there is no abuse of the rule for their delivery. The judge may ask: "Could you please explain to me how this is rebuttal? This appears to me to be a fresh opinion." The decision as to whether the rebuttal report is in fact limited to a response and critique lies with the trial judge.[22]

QUESTION TO ASK:

"Could you please explain to me how this is rebuttal? This appears to me to be a fresh opinion."

3.3. Pre-Trial Management

Pre-trial management provides an important opportunity to assess the necessity of expert evidence, and to ensure best practices early on. At the outset, the judge should ask the parties if they intend to use expert evidence.

Where neither party has an expert, the judge might consider the appropriateness of suggesting they retain a joint expert. Similarly, in cases where there are multiple defendants, the judge should consider whether it would be appropriate to encourage the defendants to retain a joint expert to provide evidence for areas of shared interest.

Using a joint expert has the advantage of reducing costs and delays. It can also contribute to a greater perception of fairness in the outcome by the parties, especially among those who lose their case. Where both parties agree on a joint expert and acknowledge their impartiality and competence, the losing party will be less inclined to later critique that expert's qualifications or evidence. An additional advantage to using a joint expert is that, by resolving an important and often central issue separating the parties, this practice can often facilitate an out-of-court settlement.

A common misconception about joint experts is that it will be difficult for the parties to agree on an expert. One suggestion to overcome this challenge is to ask each party to draft a list of their top five independent experts and to compare the results. In most cases, there will be at least one common name on each list. This expert has the advantage of having been individually chosen by both parties, and is seen as independent, impartial, objective and competent in everyone's eyes.

[22] *Janis v Janis*, 2013 BCSC 116, [2013] BCJ No 130 at para 22 (Mast).

Another common misconception is that lawyers will resist the appointment of joint experts because it means losing control of their expert. Lawyers usually want to retain as much control as possible over what their experts can and cannot say in court. However, experience has shown that lawyers are warming to the idea of using of a joint expert provided that, in so doing, the interests of their client continue to be met. In some cases, the financial benefits or the associated reductions in delays may warrant the use of a joint expert. For example, there is a government-financed service in Québec in family matters whereby a joint expert is named by the administrators to evaluate parental capacity and the wishes of the children in files dealing with custody and access. Where the parties consent to using this program, and a large number do, there is no charge to them, although the party who wishes to file a reply report from another expert will bear that cost alone.

In addition to considering the use of joint experts, pre-trial management also provides an opportunity to pinpoint the focus of what the expert evidence will cover at trial. This allows all parties to anticipate either the issues about which the experts will testify, or what the areas of contention will likely be, and to prepare accordingly. In criminal matters, it may be more difficult to identify what the joined issue(s) and the focus of the expert evidence will be. In most cases, the defence will challenge the Crown's experts on their conclusions by calling their own evidence, but they are under no obligation to reveal their arguments until the trial. Nonetheless, the defence may be encouraged to at least advise the judge if they intend to challenge the expert's conclusions. In practice, they are usually prepared to do so. However, it may sometimes be the case that judges will have to accept a broader qualification of the expert than the defence is prepared to accept. Such a decision will, of course, be based upon an application of the ***Mohan***[23] criteria.

[23] *R v Mohan*, [1994] 2 SCR 9 [*Mohan*].

3.4. Trial Management

3.4.1. The Qualification Process

Qualification should be in all the areas in which the expert is to give opinion evidence: In ***R v Marquard***, the Supreme Court held that the proper practice for counsel presenting an expert witness is to qualify the expert in all the areas in which they are to give opinion evidence and to define the expert's field of expertise and its limits.[24] A proper qualification should highlight: (1) the important aspects of the witness's education with respect to the evidence about which they will testify; (2) the extent of their practical experience in that particular field; and (3) what they have done to disseminate their expertise to others. It is the lawyer's responsibility to qualify the expert. Usually, the judge should not interfere in the qualification process. However, if, at the end of the qualification stage, counsel has not addressed all of the important issues, the judge should exercise his gatekeeper function and ask counsel to explain how the expert's qualifications relate to the evidence they hope to establish at trial (*e.g.*, "You are seeking to qualify this expert to give evidence in what?").

QUESTION TO ASK:

"You are seeking to qualify this expert to give evidence in what?"

Even when precautions such as these are taken, the reality is that sometimes questions and answers venture into territory which counsel had not foreseen.[25] For instance, a lawyer may want to qualify an expert as a pediatrician, but later have them testify about pediatric oncology or shaken baby syndrome. However, not every pediatrician will be qualified to testify in these particular subcategories. In these circumstances, McLachlin J. (as she then was) noted as follows:

> [37] Important as the initial qualification of an expert witness may be, it would be overly technical to reject expert evidence simply because the witness ventures an opinion beyond the area of expertise in which he or she has been qualified. As a practical matter, it is for opposing counsel to object if the witness goes beyond the proper limits of his or her expertise. The objection to the witness's expertise may be made at the stage of initial qualification, or during the witness's evidence if it becomes apparent the

[24] *R v Marquard*, [1993] 4 SCR 223, [1993] SCJ No 119 at para 36.
[25] *R v Marquard*, [1993] 4 SCR 223.

> witness is going beyond the area in which he or she was qualified to give expert opinion. In the absence of objection, a technical failure to qualify a witness who clearly has expertise in the area will not mean that the witness's evidence should be struck. However, if the witness is not shown to have possessed expertise to testify in the area, his or her evidence must be disregarded and the jury so instructed.

Where an expert expresses an opinion outside of the specific area for which they were originally qualified by the court, the judge may choose to hold a second *voir dire* on their competencies in that new area.

Qualification should focus on the expertise more than the expert: The qualification process provides an important exposé of who the expert is and why the expert should be taken seriously. The qualification process should emphasize the expert's expertise (education and practical experience) as it relates to the opinion they are about to give, not their background. In other words, the emphasis should be less on the expert and more on the expertise itself. A question to ask counsel is: "What, specifically, is the focus of the issue about which the expert will testify?"

QUESTION TO ASK:

"What, specifically, is the focus of the issue about which the expert will testify?"

3.4.2. Managing the Overreaching Expert

The judge's gatekeeper function does not end at the qualification stage. Judges must remain vigilant throughout the proceedings to prevent an expert from testifying outside their area of qualification or on new subject matter not covered in the initial report unless the expert is requalified for that purpose. Where an expert begins to stray outside the qualified area of expertise, and prejudice may result, an adjournment can be granted. The adjournment should be granted if it is absolutely necessary to allow the expert to be requalified in a broader range of topics, or to allow a supplemental report to be filed, or to give time to the other side to prepare, including the possibility of filing a rebuttal report. Alternatively, an agreement may be reached that allows the expert to testify about these issues without adjourning and requalifying the expert. There is no clear rule in this type of situation. The judge should use his or her discretion whether to admit the testimony, provided it does not cause prejudice to the opposing party, and does not take them by surprise.[26] However, this type of decision should be ruled on

[26] *Cascades Conversion inc. c Yergeau*, 2006 QCCA 464, 63 Admin LR (4th) 1 at para 68.

immediately in light of the effect that it will have on the evidence to be adduced by the opposing party. It should not be taken under reserve until final judgment.

3.4.3. Impartiality and Expert Bias

Assessing impartiality: Experts owe a duty to the court to provide evidence that is fair, objective, and non-partisan. This common-law duty has been codified in the Rules of most of the provinces, which provide explicit guidance related to the duty of expert witnesses, typically requiring mandatory attestations by the experts that they are providing an objective opinion. This impartiality, however, should not always be assumed. Expert witnesses often tend to espouse the cause of those instructing them.[27] Questions can be put to the expert in order to assess their level of impartiality and their understanding of their role in the proceedings. If cross-examination does not cover this aspect, the judge could ask the expert: "How do you see your role here today?" "How many mandates have you executed for this client?" and other questions of this nature in order to assess the expert's impartiality.

QUESTIONS TO ASK:

"How do you see your role here today?"

"How many mandates have you executed for this client?"

Motions for exclusion for bias before the qualification stage: In the case of ***White Burgess Langille Inman v Abbott and Haliburton Co.*** (*White Burgess*),[28] Justice Cromwell, writing for a united Supreme Court of Canada, clarified that independence and impartiality are to be considered at the threshold stage. Prior to *White Burgess*, there had been conflicting lines of authority as to how this issue should be handled, with some decisions suggesting that a lack of independence and/or impartiality would go only to the weight given to an opinion, rather than to admissibility.

In *White Burgess*, the Court confirmed that "a proposed expert's independence and impartiality **goes to admissibility and not simply to weight**" (emphasis added) and that the analysis of a witness's independence and impartiality is properly to be undertaken under the "qualified expert" prong of the *Mohan* test.

Absent a challenge to an expert's independence and impartiality, an attestation or testimony recognizing and accepting the duty will generally be sufficient to establish that this threshold is met. The onus will be on the party opposed to admission of the expert's evidence to show that the expert is unable and/or

[27] *Abbey National v Key Surveyers*, [1996] 3 All ER 184.

[28] *White Burgess Langille Inman v Abbott and Haliburton Co.*, 2015 SCC 23, [2015] 2 SCR 182 [*White Burgess*].

unwilling to comply with the duty to the court to provide fair, objective and non-partisan evidence. Note that this is not to be assessed based on the *appearance of* bias: Exclusion at the threshold stage will only occur in cases where it is "very clear" that the proposed expert is unable or unwilling to fulfill their duty to the court. Where the inquiry concludes that there are concerns, but they fall below clear unwillingness or inability to comply with the duty to the court, the evidence should be admitted and the concerns weighed during the second stage of the test:

> "In my view, expert witnesses have a duty to the court to give fair, objective and non-partisan opinion evidence. They must be aware of this duty and able and willing to carry it out. If they do not meet this threshold requirement, their evidence should not be admitted. Once this threshold is met, however, concerns about an expert witness's independence or impartiality should be considered as part of the overall weighing of the costs and benefits of admitting the evidence. This common law approach is, of course, subject to statutory and related provisions which may establish different rules of admissibility." (*White Burgess* at para 10)

Although the Court predicts it will be rare that this threshold is not met, it does provide numerous examples of cases where an expert's evidence was ruled inadmissible due to the nature of the relationship between expert and party, while stressing that the analysis is contextual rather than precedent-driven.[29]

Even though the witness may meet the threshold test, judges must conduct the cost-benefit analysis as part of their gatekeeper role. Failure to conduct this analysis is a reviewable error.[30] Conducting this test can lead to the conclusion that the expert should be excluded from testifying on the basis of lack of independence, even though the threshold test was met.

Distinguishing bias from partisanship to a particular idea: There is a difference between an expert being an advocate for a particular idea, and an advocate for a party. While the former is not necessarily problematic, the latter is. In fields where there may be conflicts within the expert's discipline as to methodology (this is especially true in the social sciences, such as sociology or

[29] An expert's interest in the litigation or relationship to the parties has led to exclusion in a number of cases: see, e.g., *Fellowes, McNeil v Kansa General International Insurance Co.* (1998), 40 OR (3d) 456 (Gen. Div.) (proposed expert was the defendant's lawyer in related matters and had investigated from the outset of his retainer the matter of a potential negligence claim against the plaintiff); *Royal Trust Corp. of Canada v Fisherman* (2000), 49 OR (3d) 187 (Sup Ct) (expert was the party's lawyer in related U.S. proceedings); *R v Docherty*, 2010 ONSC 3628 (expert was the defence counsel's father); *Ocean v Economical Mutual Insurance Co.*, 2010 NSSC 315, 293 NSR (2d) 394 (expert was also a party to the litigation); *Handley v Punnett*, 2003 BCSC 294 (expert was also a party to the litigation); *Bank of Montreal v Citak*, [2001] OJ No 1096 (Sup Ct) (expert was effectively a "co-venturer" in the case due in part to the fact that 40 percent of his remuneration was contingent upon success at trial: para 7); *Dean Construction Co. v M.J. Dixon Construction Ltd.*, 2011 ONSC 4629, 5 CLR (4th) 240 (expert's retainer agreement was inappropriate); *Hutchingame v Johnstone*, 2006 BCSC 271 (expert stood to incur liability depending on the result of the trial). In other cases, the expert's stance or behaviour as an advocate has justified exclusion: see, e.g., *Alfano v Piersanti*, 2012 ONCA 297, 291 OAC 62; *Kirby Lowbed Services Ltd. v Bank of Nova Scotia*, 2003 BCSC 617; *Gould v Western Coal Corp.*, 2012 ONSC 5184, 7 BLR (5th) 19.

[30] *Bruff-Murphy v Gunawardena*, 2017 ONCA 502 at para 41.

cultural anthropology), an expert's expression of his strong preference for a certain way of doing things should not be confused with bias toward a party.

In ***Keefer Laundry Ltd. v Pellerin Milnor Corp.***, Smith J. said:

> [15] […] the statement that an expert should not be an advocate […] is sometimes misunderstood. There is a difference between an expert who advocates for a party and one who advocates for his or her opinion. By that I mean that an expert opinion should be confined to the expert's field of expertise and to the question within that field that is at issue. It should be the result of careful and objective consideration of all relevant facts and scientific principles and not based on extraneous considerations.
>
> [16] In short, the Court should be able to approach the opinion with some confidence that the expert would have rendered the same opinion if he or she had been consulted by the opposite party. However, once an expert has formed an opinion through that process, he or she may be firm, emphatic or even strident in the way he or she expresses the opinion or defends it against contrary opinions.[31]
>
> [Emphasis in original].

In ***Tsilhqot'in Nation v Canada (Attorney General)***, Vickers J. noted that the strength of an expert's convictions on matters within her field of expertise would not disqualify her from testifying.

> [28] In addition, counsel argue that Dr. Turner's evidence is more in the nature of advocacy and does not reflect the unbiased views of an expert. Dr. Turner has spent her entire career studying the First Nations people of British Columbia and their relationships with plants and their ecology. It is expected she will have firm opinions to express on such matters and the impact of the Europeans on the First Nations people and their environment. The fact that her opinions are sympathetic to the positions advanced by the plaintiff in this case does not transform her into an advocate for the plaintiff. The manner in which she expresses those opinions, in writing and orally, and her demeanour under cross examination all bear on whether the opinions she expresses can be relied upon.[32]

[31] *Keefer Laundry Ltd. v Pellerin Milnor Corp.*, 2007 BCSC 899, 72 BCLR (4th) 187.
[32] *Tsilhqot'in Nation v Canada (Attorney General)*, 2005 BCSC 131.

Bias in high profile cases: In cases where the stakes are elevated (*e.g.*, monetary, media presence, or cases involving liberty of the person or death of a child), experts are at a greater risk of becoming personally and emotionally involved, as well as more susceptible to bias. When an expert loses their objectivity, they are more likely to reach beyond their expertise to support the side that hired them. In these cases, the gatekeeper function of the judge is heightened.

3.4.4. Special Considerations for Jury Trials

The expert report: Where an expert report contains an executive summary, the latter should be deleted on a written exhibit. There are two reasons for this. First, juries may be tempted to simply adopt the executive summary without looking further into the expert's considerations and logic. Second, juries may be tempted to pay less attention to both the expert's testimony and cross-examination if they know they can rely on a summary in their deliberations.

Qualification of the expert: Should there be any issue with the qualifications, this should not be raised with the jury present.

Explaining expert bias: Juries have a tendency to assume the objectivity of experts, who can often make a considerable impression on them by virtue of their credentials alone. Juries should not assume that all experts are scientists who only conduct empirical research and provide objective answers. It is important for judges to explain to juries that experts may have their own biases, and may not be independent either because one party pays them, or they have developed an attachment for the side for which they are testifying.

Controlling quantity and quality of evidence: The judge must be careful to control both the volume and the quality of the evidence submitted to the jury, for whom the ability to absorb and understand the expert report may be more limited.

Mid-trial instructions to the jury: Where the judge is concerned that the jury has become overwhelmed by the scope and nature of the expert evidence, or this potential exists, mid-trial instructions on the meaning and value of expert evidence can be very helpful. These instructions provide an opportunity for the judge to remind the jury that it is up to them to accept some, all, or none of the expert's evidence, and to put the jury in the right frame of mind to approach the expert evidence, not skeptically, but very carefully. The Canadian Judicial Council's *Model Jury Instructions* offer this sample mid-trial instruction to the jury:

Canadian Judicial Council's *Model Jury Instructions*

7.18 Expert Opinion Evidence

[1] Normally, witnesses may only testify about what they have seen or heard, and may not testify about their opinions. In the case of [expert], however, because of his/her special training, education and experience, s/he will be permitted to give an opinion about (specify).

[2] Although [expert] will be permitted to give his/her opinion, it is up to you to decide the extent to which you will rely on it. Here are some things to consider as [expert] testifies:

- [expert]'s qualifications and experience;
- the reasons given for the opinion;
- the suitability of the methods used;
- whether [expert] is impartial; and
- the other evidence in the case.

It is up to you to decide how much or little to rely on [expert]'s opinion.

[3] [Expert] may be asked to assume or rely on certain facts in giving his/her opinion. Those facts may be the same or different from what you later find as facts on the basis of all the evidence in this case. The closer the facts assumed or relied on by [expert] are to the facts as you find them to be, the more helpful [expert]'s opinion may be to you. To the extent [expert] relies on facts that you do not find supported by the evidence, you may find [expert]'s opinion less helpful.

Final charge instructions to the jury on disagreements between experts: When there is a disagreement between experts, the final charge instruction on this topic is also helpful to remind the jury that it is their decision whether or not to accept the expert's evidence. The credentials or pedigree of the expert must not intimidate the jury. The Canadian Judicial Council's *Model Jury Instructions* offer this sample final jury instruction regarding conflicting expert evidence:

Canadian Judicial Council's *Model Jury Instructions*

10.4 Expert Opinion Evidence

[1] There is a disagreement between (among) the expert opinions of (identify witnesses by name) about (describe briefly subject-matter of dispute).

[2] The issue on which these experts differ is an essential element that the Crown must prove beyond a reasonable doubt. Before you accept the opinion of the Crown's expert on this issue you must be satisfied beyond a reasonable doubt that s/he is correct. If you are not sure that s/he is correct, then the Crown has failed to prove beyond a reasonable doubt that essential element of the offence charged.

When to permit or not permit examination in chief: In a jury trial where expert reports have been filed, there are risks with limiting the scope of direct examination and proceeding directly to cross-examination. While judges may have the expertise or the experience to review the expert report on their own, juries may not understand the opinion in the report and appreciate its nuances without the benefit of an examination in chief. If the jury's first exposure to the expert's report is during cross-examination, this gives an unfair advantage to the party seeking to discredit the expert's opinion.

Using evidence from a *voir dire* in a jury charge: Where, in the case of a *voir dire*, it is determined that the expert evidence is not necessary, pieces of the evidence (and perhaps more importantly, the framework used by the expert for organizing the concepts presented) may nevertheless still be used by the judge in the jury charge to assist the jury in making its determination. For example, in ***R v McIntosh*,** Finlayson J.A. found that the writings of an expert, whose evidence on the frailties of eyewitness identification was refused on the basis that it was not outside the normal experience of the trier of fact, could be used by the trial judge to draft the jury charge.

> [22] This is not to say that a reminder as to cross-racial identification is not appropriate in a case where it is an issue. However, the argument that impresses me is that such a reminder from the trial judge is more than adequate, especially when it is incorporated into the well established warnings in the standard jury charge on the frailties of identification evidence. Writings, such as those of Dr. Yarmey, are helpful in stimulating an ongoing evaluation of the problem of witness identification, but they should be used to update the judge's charge, not instruct the jury. I think that there is a very real danger that such evidence would "distort the fact-finding process."[33]

[33] *R v McIntosh*, (1997), 35 OR (3d) 97 (CA).

Dealing with prejudicial effects revealed during the course of testimony: The trial judge retains discretion to exclude prejudicial evidence throughout the trial. This includes expert evidence that is admitted and later revealed to have a strong prejudicial effect.[34] Where the expert witness provides evidence that is overly prejudicial, the trial judge has a number of options. For example, he or she can seek submissions from counsel on excluding the evidence and provide either mid-trial or final instructions to the jury; or he or she can seek submissions from counsel on a mistrial. Both of these options should be exercised in the absence of the jury.[35]

3.5. Concurrent Expert Evidence: Hot-Tubbing

Concurrent evidence, commonly known as hot-tubbing, has been described by the Chief Justice of the Supreme Court of New South Wales as:

> […] essentially a discussion chaired by the judge in which the various experts, the parties, the advocates and the judge engage in a cooperative endeavour to identify the issues and arrive, where possible, at a common resolution of them. Where resolution of issues is not possible, a structured discussion, with the judge as chairperson, allows the experts to give their opinions without the constraints of the adversarial process and in a forum which enables them to respond directly to each other. The judge is not confined to the opinion of one advisor but has the benefit of multiple advisors who are rigorously examined in public.[36]

Hot-tubbing originates from Australia, but has since been introduced to countries including Malaysia, Singapore, Hong Kong, Japan, the United States, England and Canada. In Canada, the ***Federal Court Rules***, SOR/98-106, were amended in 2010 to permit hot-tubbing of experts at pre-trial and at trial (see sections 52.6, 282.1 and 282.2 of the ***Federal Court Rules*** below).

[34] *Bruff-Murphy v Gunawardena*, 2017 ONCA 502 at paras 65-66.
[35] *Bruff-Murphy v Gunawardena*, 2017 ONCA 502 at para 67.
[36] The Hon. Justice Peter McClellan, "Concurrent Evidence," (2011) 14 Effectius Newsletter at 3.

***Federal Court Rules*, SOR/98-106.**

Pre-trial

Expert conference

52.6 (1) The Court may order expert witnesses to confer with one another in advance of the hearing of the proceeding in order to narrow the issues and identify the points on which their views differ.

Presence of parties and counsel

(2) Subsection (1) does not preclude the parties and their counsel from attending an expert conference but the conference may take place in their absence if the parties agree.

Trial

Expert witness panel

282.1 The Court may require that some or all of the expert witnesses testify as a panel after the completion of the testimony of the non-expert witnesses of each party or at any other time that the Court may determine.

Testimony of panel members

282.2 (1) Expert witnesses shall give their views and may be directed to comment on the views of other panel members and to make concluding statements. With leave of the Court, they may pose questions to other panel members.

Examination of panel members

(2) On completion of the testimony of the panel, the panel members may be cross-examined and re-examined in the sequence directed by Court.

The process for hearing concurrent evidence can vary. For example, at the pre-trial stage, the reported experience of Justice Steven Rares of the Federal Court of Australia in ***Australasian Performing Right Association ltd v Monster Communications Pty Ltd***[37] was that after the experts submitted their individual reports, they met in the absence of counsel and prepared a joint report on matters about which they were in agreement and disagreement.[38] This helped to narrow the range of difference between the experts.

[37] *Australasian Performing Right Association ltd v Monster Communications Pty Ltd*, [2006] FCA 1806 (Australia).

[38] The Hon. Justice Steven Rares, "Using the 'Hot Tub' – How Concurrent Expert Evidence Aids Understanding Issues" (Paper presented at the *New South Wales Bar Association Continuing Professional Development seminar: Views of the "Hot Tub" from the Bar and the Bench*, 23 August 2010) at para 22 [Rares].

Instead of, or in addition to, a pre-trial process, hot-tubbing at the trial stage is described by Rares J. as follows:

> First, each expert will be asked to identify and explain the principal issues, as they see them, in their own words. After that each can comment on the other's exposition. Each may ask then, or afterwards, questions of the other about what has been said or left unsaid. Next, counsel is invited to identify the topics upon which they will cross-examine. Each of the topics is then addressed in turn. Again, if need be, the experts comment on the issue and then counsel, in the order they choose, begin questioning the experts. If counsel's question receives an unfavourable answer, or one counsel does not fully understand it, he or she can turn to their expert and ask what that expert says about the other's answer.[39]

In Canada, the first reported case of hot-tubbing pursuant to the amended *Federal Court Rules* is ***Apotex Inc. v AstraZeneca Canada Inc.***[40] In this case, Hughes J. adopted a procedure similar to that described by Rares J. above. Following the testimony of two experts, both experts took the stand at the same time while remaining under oath. They were asked to explain where their opinions differed from each other, and Hughes J. moderated the exchange. At the end of the process, each counsel was invited to put follow-up questions to the witnesses.[41]

In Ontario, the ***Rules of Civil Procedure***, RRO 1990, Reg 194, were also amended in 2010 to allow either a pre-trial judge or a trial judge to order experts to meet on a without prejudice basis to identify areas of agreement or disagreement, attempt to resolve and clarify the latter, and prepare a joint statement about areas of agreement and disagreement (see sections 50.07 and 20.05 of the ***Rules of Civil Procedure*** below).[42] The express requirement that hot-tubbing is to be without prejudice in Ontario distinguishes Ontario's hot-tubbing provisions from those in the ***Federal Court Rules***.

---

[39] Rares, at para 27.

[40] *Apotex Inc. v AstraZeneca Canada Inc.*, 2012 FC 559; 410 FTR 168 [*Apotex*].

[41] *Apotex*, 2012 FC 559 at para 10.

[42] See also *Glass v 618717 Ontario Inc.*, 2011 ONSC 2926 at para 25 for a discussion on availability of hot-tubbing for both pre-trial and trial judges.

Ontario *Rules of Civil Procedure*, RRO 1990, Reg 194

Pre-trial

Powers

50.07 (1) If the proceeding is not settled at the pre-trial conference, the presiding judge or case management master may,

(c) make such order as the judge or case management master considers necessary or advisable with respect to the conduct of the proceeding, including any order under subrule 20.05 (1) or (2).

Trial

Powers of Court

20.05 (1) Where summary judgment is refused or is granted only in part, the court may make an order specifying what material facts are not in dispute and defining the issues to be tried, and order that the action proceed to trial expeditiously.

(2) If an action is ordered to proceed to trial under subrule (1), the court may give such directions or impose such terms as are just, including an order,

(k) that any experts engaged by or on behalf of the parties in relation to the action meet on a without prejudice basis in order to identify the issues on which the experts agree and the issues on which they do not agree, to attempt to clarify and resolve any issues that are the subject of disagreement and to prepare a joint statement setting out the areas of agreement and any areas of disagreement and the reasons for it if, in the opinion of the court, the cost or time savings or other benefits that may be achieved from the meeting are proportionate to the amounts at stake or the importance of the issues involved in the case and,

(i) there is a reasonable prospect for agreement on some or all of the issues, or

(ii) the rationale for opposing expert opinions is unknown and clarification on areas of disagreement would assist the parties or the court;

In Ontario, courts have identified pre-trial hot-tubbing as a way to mitigate the complexity and confusion generated by multiple opposing expert reports.[43] These provisions have been used in ***Glass v 618717 Ontario***, 2011 ONSC 2926 at para 26 and ***Karrys v Karrys***, 2014 ONSC 713 at para 19, to require opposing experts to prepare a joint statement setting out areas of agreement and disagreement and explaining in detail the reasons for the disagreements. In both cases, the judge made the order to receive non-partisan assistance in understanding why the experts had such divergent views.

[43] See e.g. *Livent Inc. v Deloitte & Touche LLP*, 2014 ONSC 2176 at para 276.

In Québec, art. 240 of the ***Code of Civil Procedure***, CQLR c C-25.01 (*CCP*) also provides that at any stage of the proceeding the judge may require that the experts meet in order to determine where, if at all, their differences of opinion can be reconciled. At least one decision has cited the usefulness of these expert conferences in helping the court resolve complex issues, such as retroactive child support, in a cost-effective manner.[44]

Québec *Code of Civil Procedure*, CQLR c C-25.01

240. After the report has been filed but before the trial begins, the joint or court-appointed expert must, if the court so requires or on the parties' request, provide clarifications on certain aspects of the report and meet the parties to discuss the expert's opinions ahead of the trial.

If conflicting expert reports are filed, the parties may call the experts to a meeting so that they may reconcile their opinions, identify the points on which they differ and, if necessary, prepare an additional report on those points. At any stage of the proceeding, the court, even on its own initiative, may order the experts to meet and file an additional report within a specified time.

Although lawyers may attend these meetings, experts should be encouraged to meet alone. Meeting individually provides experts with an opportunity to speak openly to their counterpart, often for the first time, and to clarify their opinions away from the influence of their clients. Following the meeting, the experts then draft a report outlining points of agreement, and areas that have not been resolved. The report helps to crystallize the key areas of dispute, which in turn, helps to reduce the length of hearings and associated costs.

In 2012, British Columbia introduced pre-trial hot-tubbing to the ***Supreme Court Civil Rules***, BC Reg 168/2009, providing that a judge may order, with or without an application by a party, "that the parties' experts must confer before the service of their respective reports" (Rule 5-3(1)(k)(iii)).

[44] *Droit de la famille – 162293*, 2016 QCCS 4410 at para 217

British Columbia *Supreme Court Civil Rules*, BC Reg 168/2009

Orders

5-3 (1) At a case planning conference, the case planning conference judge or master may make one or more of the following orders in respect of the action, whether or not on the application of a party:

[…]

k) respecting experts, including, without limitation, orders

[…]

(iii) that the parties' experts must confer before the service of their respective reports,

In provinces where the rules of the court do not expressly incorporate hot-tubbing provisions, broad powers given to case management judges may provide support for similar practices. Consider, for example, Rule 4.14(1) of the ***Alberta Rules of Court***, Alta Reg 124/2010.

Alberta *Alberta Rules of Court*, Alta Reg 124/2010

Authority of case management judge

4.14 (1) A case management judge, or if the circumstances require, any other judge, may

(a) order that steps be taken by the parties to identify, simplify or clarify the real issues in dispute,

(b) establish, substitute or amend a complex case litigation plan and order the parties to comply with it,

(c) make an order to facilitate an application, proceeding, questioning or pre-trial proceeding,

(d) make an order to promote the fair and efficient resolution of the action by trial,

(e) facilitate efforts the parties may be willing to take towards the efficient resolution of the action or any issue in the action through negotiation or a dispute resolution process other than trial, or

(f) make any procedural order that the judge considers necessary.

Advocates of hot-tubbing identify four advantages to the process: greater clarity and comprehension of the evidence, better communication and cooperation amongst experts, a reduction in the adversarial character of expert evidence, and efficiency.[45] However, hot-tubbing is not likely to work in highly-contested cases. Hot-tubbing requires the goodwill of the parties to find a timely and efficient answer to a technical question. For this process to be helpful, the parties themselves must be ready and willing to cooperate in the search for the "right" answer. Moreover, in criminal trials hot-tubbing is of limited application because it conflicts with the right of the accused to withhold all arguments until trial.

4. Evaluating Expert Evidence: The Hallmarks of Reliability

4.1. Applying the Standard of Threshold Reliability

It is one thing for jurisprudence to arm trial judges as gatekeepers, with threshold reliability as an admissibility screen for expert scientific evidence, and quite another to describe how the standard can be applied in particular cases. Judges need to exercise an element of judgement when evaluating expert opinion evidence. There are a variety of tools that may assist judges in discharging this challenging task. Some of these tools will undoubtedly be more useful than others, depending on the nature of the case and the particular evidence being scrutinized. The tools should, however, provide a reasonable basis for that judgment. It may therefore be helpful to outline a few of these tools and to provide some evaluation of their potential assistance to a trial judge in fulfilling the gatekeeper role.

4.2. *General Criteria for Admissibility*

The general criteria for the admissibility of expert opinion evidence are discussed in detail in Professor Hamish Stewart's chapter **The Legal Framework for Scientific Evidence**. In his chapter, Professor Stewart reviews the ***Mohan***[46] criteria for admission of expert evidence, and the two-stage test confirmed by the Supreme Court of Canada in ***White Burgess Langille Inman v Abbott and Haliburton Co.***[47] For the purposes of the present discussion on threshold reliability, a brief review of the case law is reproduced here.

---

[45] Freya Kristjanson,"'Hot-Tubs' and Concurrent Evidence: Improving Administrative Proceedings" (2011) 25 Can J Admin L & Prac 79.

[46] *Mohan*, [1994] 2 SCR 9.

[47] *White Burgess Langille Inman v Abbott and Haliburton Co.*, 2015 SCC 23, [2015] 2 SCR 182 [*White Burgess*].

This test was summarized by the Court of Appeal for Ontario in ***R v Abbey*** (2017):[48]

> Expert evidence is admissible when:
>
> (1) It meets the threshold requirements of admissibility, which are:
> a. The evidence must be logically relevant;
> b. The evidence must be necessary to assist the trier of fact;
> c. The evidence must not be subject to any other exclusionary rule;
> d. The expert must be properly qualified, which includes the requirement that the expert be willing and able to fulfil the expert's duty to the court to provide evidence that is:
> i. Impartial,
> ii. Independent, and
> iii. Unbiased.
> e. For opinions based on novel or contested science or science used for a novel purpose, the underlying science must be reliable for that purpose,
> and
> (2) The trial judge, in a gatekeeper role, determines that the benefits of admitting the evidence outweigh its potential risks, considering such factors as:
> a. Legal relevance,
> b. Necessity,
> c. Reliability, and
> d. Absence of bias.

Should the costs at this stage be found to outweigh the benefits, the evidence may be deemed inadmissible, notwithstanding that it met all the *Mohan* factors.

The "benefit" side of the cost-benefit evaluation requires a consideration of the probative potential of the evidence. This is where the trial judge, as gatekeeper, must consider whether the evidence is sufficiently reliable to get it over the threshold of admissibility. As Doherty J.A. said in ***Abbey*** (2009):

> [87] Reliability concerns reach not only the subject matter of the evidence, but also the methodology used by the proposed expert in arriving at his or her opinion, the expert's expertise and the extent to which the expert is shown to be impartial and objective.[49]

[48] *R v Abbey*, 2017 ONCA 640 at para 48.
[49] *Abbey*, 2009 ONCA 624 at para 87.

The "cost" side of the cost-benefit evaluation requires an assessment of the risks of consumption of time, prejudice, and confusion.[50]

This assessment of reliability requires the trial judge to intrude into territory often seen as the exclusive domain of the jury in a criminal jury trial.[51] To be clear, the trial judge's evaluation of threshold reliability is not the same as the jury's ultimate reliability assessment. The trial judge has the responsibility to determine the threshold reliability of expert evidence, but the jury has the responsibility to determine ultimate reliability and to resolve disagreements between competing experts.

This balancing act is described as follows by Justice Doherty in ***Abbey*** (2009):

> [142] In assessing threshold reliability, I think trial judges should be concerned with factors that are fundamental to the reliability of the opinion offered and responsive to the specific dangers posed by expert opinion evidence. Trial judges, in assessing threshold reliability, should not be concerned with those factors which, while relevant to the ultimate reliability of the evidence, are common with those relevant to the evaluation of evidence provided by witnesses other than experts. For example, I would not think that inconsistencies in an expert's testimony, save perhaps in extreme cases, would ever justify keeping the expert's opinion from the jury. Juries are perfectly able to consider the impact of inconsistencies on the reliability of a witness's testimony.[52]

4.3. Specific Factors for the Trial Judge to Consider

The following is a non-exhaustive list of specific factors that the trial judge may consider when evaluating threshold liability. Some factors will be relevant in all cases involving expert opinion evidence, while others may be more relevant in cases involving scientific evidence. The factors below are sorted according the following broad themes: (1) factors related to the discipline in which the expert operates; (2) factors related to the expert's qualifications or experience; (3) factors related directly to the actual opinion; and (4) factors relating to the methodology used in arriving at the opinion.[53]

---

[50] *Abbey*, 2009 ONCA 624 at para 90.

[51] *Abbey*, 2009 ONCA 624 at para 89.

[52] *Abbey*, 2009 ONCA 624 at para 142.

[53] For a poignant example of the injustices that can result from a failure to interrogate the underlying science and expert qualifications, see Susan E. Lang, *Report of the Motherisk Hair Analysis Independent Review* (15 December 2015) online: < http://www.m-hair.ca/docs/default-source/default-document-library/motherisk_enbfb30b45b7f266cc881aff0000960f99.pdf?sfvrsn=2>.

4.3.1. Factors Related to the Discipline in which the Expert Operates

A preliminary inquiry should be made to ensure that the discipline used by the proposed expert witness has sufficient threshold reliability. Factors the trial judge may wish to consider to assess the discipline in which the expert operates include the following:

- Does the expert evidence actually form part of a field or specialization? Is the technique or theory novel? Does it rely on established principles? Is it controversial?
- Has the technique or theory been described and endorsed in academic and/or specialized literature? If so, is the reference in the literature substantial or incidental?
- Has the technique or theory been generally accepted by experts in the field? Has it undergone peer review? In assessing the extent of acceptance, the judge should consider what evidence supports acceptance.[54]

4.3.2. Factors Related to the Expert's Qualifications or Experience

In assessing the expert's qualifications or experience, the trial judge will consider the reliability of the expert, including the expert's credentials, and evaluate the prospect of bias.[55] Although the trial judge may examine the expert's training, education and experience, including any allegations of serious mistakes in other investigations or prosecutions, depending on the nature of the opinion evidence, the expert's background may or may not be directly relevant.[56]

The trial judge will also consider the independence of the expert: "A proposed expert witness who is unable or unwilling to fulfill this duty to the court is not properly qualified to perform the role of an expert."[57] Some questions to ask are:

- Did the experts have close contact with the investigators, or were they formally and substantially independent?
- Does the expert have a financial or professional interest in the evidence of the technique?
- Does the expert invariably work for the prosecution (or defence)?

---

[54] See *R v Johnston* (1992), 69 CCC (3d) 395 (Ont Ct J (Gen Div)) at 415; *R v J.E.T.*, [1994] OJ No 3067 (Ct J (Gen Div)); Gary Edmond, "Pathological Science? Demonstrable Reliability and Expert Forensic Pathology Evidence" (Research paper commissioned by the Inquiry into Pediatric Forensic Pathology in Ontario, November 12, 2007) [Edmond].

[55] David Paciocco, "Evaluating Expert Opinion Evidence for the Purpose of Determining Admissibility: Lessons from the Law of Evidence" (1994) 27 CR (4th) 302 at 313-318 [Paciocco].

[56] Edmond, "Pathological Science? Demonstrable Reliability and Expert Forensic Pathology Evidence" (Research paper commissioned by the Inquiry into Pediatric Forensic Pathology in Ontario, November 12, 2007).

[57] *White Burgess*, 2015 SCC 23 at para 53.

- To what extent is the proffered opinion based on data and other information gathered independently of the specific case or, more broadly, the litigation process?[58]

4.3.3. Factors Related Directly to the Actual Opinion

In order to evaluate threshold reliability, the trial judge must delineate the scope of the expert's opinion, noted in *Abbey* (2009) as follows.

> [62] Before deciding admissibility, a trial judge must determine the nature and scope of the proposed expert evidence. In doing so, the trial judge sets not only the boundaries of the proposed expert evidence but also, if necessary, the language in which the expert's opinion may be proffered so as to minimize any potential harm to the trial process. A cautious delineation of the scope of the proposed expert evidence and strict adherence to those boundaries, if the evidence is admitted, are essential. The case law demonstrates that overreaching by expert witnesses is probably the most common fault leading to reversals on appeal: see, for example, ***R v Ranger*** (2003), 67 OR (3d) 1 (CA); ***R v Klymchuk*** (2005), 203 CCC (3d) 341 (Ont CA); ***R v K.(A.)*** (1999), 45 OR (3d) 641 (CA), at paras. 123-35; ***R v Llorenz*** (2000), 145 CCC (3d) 535 (Ont CA), at paras. 33-40.[59]

This duty does not end after the initial ruling. The Supreme Court in ***R v Sekhon***[60] notes at para 46 that "all trial judges – including those in judge-alone trials – have an ongoing duty to ensure that expert evidence remains within its proper scope…The trial judge must do his or her best to ensure that throughout the expert's testimony, the testimony remains within the proper boundaries of expert evidence." The trial judge retains discretion to exclude prejudicial evidence throughout the trial.[61]

Specific factors to consider include:

- To what extent has the witness honoured the boundaries and limits of his/her discipline?[62]
- Is the expert merely expressing a personal opinion? To what extent is the expert evidence speculation?

---

58 Edmond, "Pathological Science? Demonstrable Reliability and Expert Forensic Pathology Evidence" (Research paper commissioned by the Inquiry into Pediatric Forensic Pathology in Ontario, November 12, 2007); see also *Abbey*, 2009 ONCA 624.

59 *Abbey*, 2009 ONCA 624 at para 62.

60 *R v Sekhon*, 2014 SCC 15, [2014] 1 SCR 272.

61 *Bruff-Murphy v Gunawardena*, 2017 ONCA 502 at paras 65-66.

62 See *R v Johnston*, (1992), 69 CCC (3d) 395 (Ont Ct J (Gen Div)); Edmond, "Pathological Science? Demonstrable Reliability and Expert Forensic Pathology Evidence" (Research paper commissioned by the Inquiry into Pediatric Forensic Pathology in Ontario, November 12, 2007); *Abbey*, 2009 ONCA 624.

- 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?
- Does the opinion set out the facts relied on, the reasoning process, and the opinion reached in a logical and understandable way?
- Are the techniques or conclusions based on individual case studies, or more broadly based on statistical approaches such as epidemiology and meta-analysis?

4.3.4. Factors Relating to the Methodology Used in Arriving at the Opinion

Where, for example, the proffered expert opinion evidence is of a scientific nature, the trial judge should consider the following four factors in deciding whether the expert opinion is sufficiently reliable ("the ***Daubert*** factors" mandated by the United States Supreme Court):

1. Whether the underlying theory can be (and has been) tested;
2. Whether the theory or technique has been subjected to peer review and publication;
3. The known or potential rate of error associated with use of the technique; and
4. Whether the technique has been generally accepted in the relevant scientific community.[63]

While testing and error rates are optimal, it is important to reiterate that many kinds of expert opinion are not readily susceptible to empirical testing or repeatability. The inability to provide testing results does not necessarily render these kinds of expert evidence unreliable. However, it does call for vigilant use of other indicators of reliability which are more germane to the task. The trial judge must guard against inappropriate applications of scientific factors where the expert opinion is not the product of a scientific inquiry.

Indeed, this was made clear in ***Abbey*** (2009), where Justice Doherty noted:

> [109] Scientific validity is not a condition precedent to the admissibility of expert opinion evidence. Most expert evidence routinely heard and acted upon in the courts cannot be scientifically validated.[64]

[63] *Daubert v Merrell Dow Pharmaceuticals Inc.* (1993), 509 US 579 at 593-95; see also *R v J.-L.J.*, 2000 SCC 51, [2000] 2 SCR 600 at para 33.

[64] *Abbey*, 2009 ONCA 624 at para 109.

And again:

> [114] The same caution against the inappropriate use of the *Daubert* factors to assess the reliability of expert opinion evidence can be found in Canadian commentary. Professor Paciocco has observed:
>
> *Clearly it is inappropriate to consider all expertise as science, or to require all expertise to attain the scientific method. Some expert witnesses rely on science only in a loose sense. Actuaries apply probability theory and mathematics to produce decidedly unscientific results. Appraisers make subjective assessments of objective data, as do family assessment experts. Professionals testifying to standards of care within their profession are doing nothing scientific. Yet Daubert spawned a jurisprudence that was fixated for a time with science. This led lower courts to commit two kinds of error. First, it caused some lower courts to hold that the Daubert test and the gatekeeping role is confined to scientific expertise. Experts who were not scientists would not be subjected to the reliability inquiry prescribed by Daubert. Second, it caused other courts to apply the criteria listed in Daubert in a wooden fashion, even to non-scientific forms of expertise. Each of these two kinds of errors was caused by the failure to take context into account.*[65]

Despite these concerns, in cases where scientific evidence and scientific methodologies are under review, the following factors can be used to supplement and flesh out the *Daubert* criteria:

- What is the error rate – for the technique, as well as the equipment and practitioner? To what extent do standards exist? If they do exist, have they been maintained?
- Has the technique or theory been applied in circumstances that reflect its intended purpose or known accuracy? Has the technique been employed with care? 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?
- Is the evidence processed or interpreted by humans or machines? How often are the machines tested or calibrated?
- Does the evidence have a verification process? Was it applied? Were protocols followed?[66]

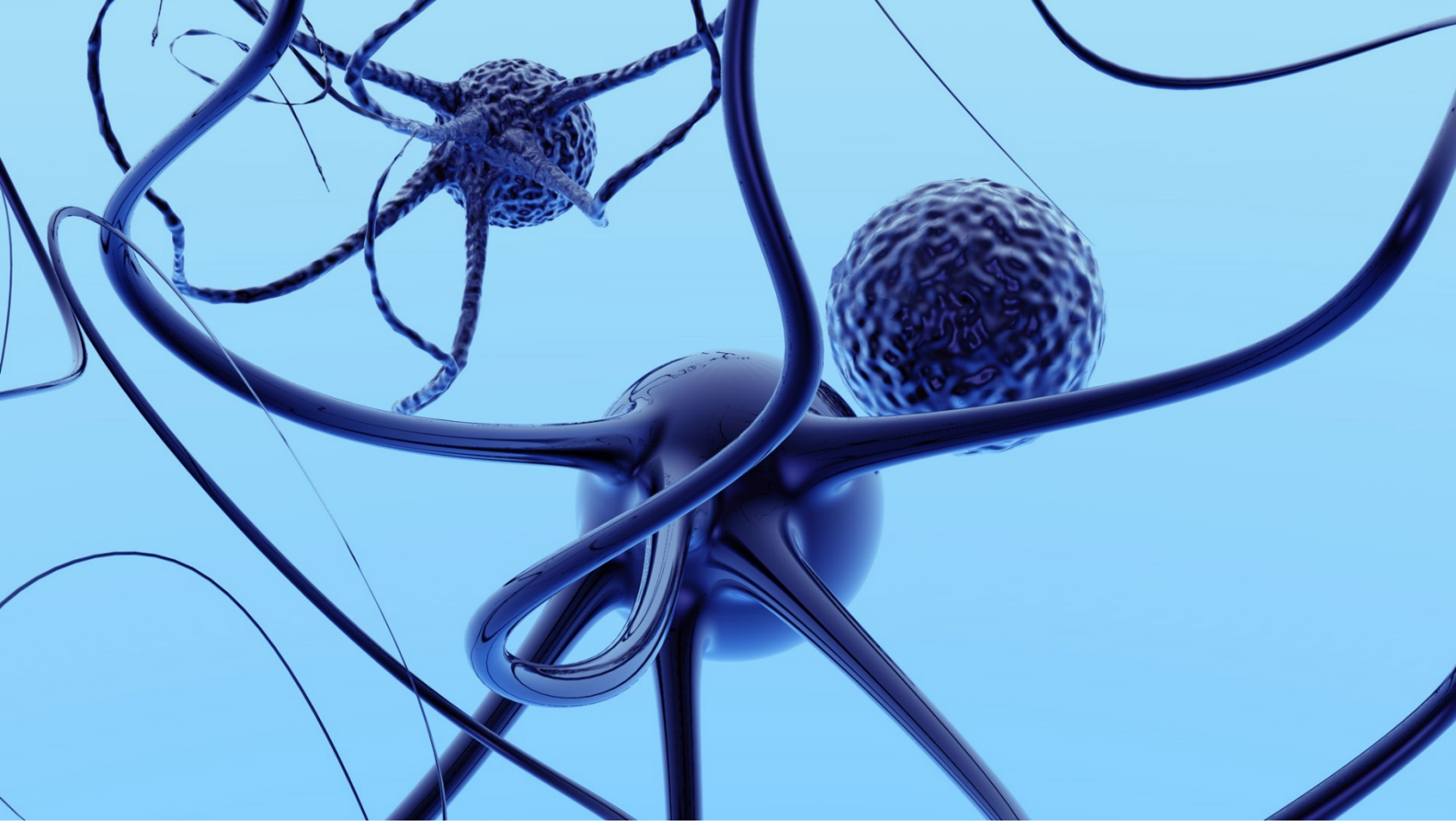
---

[65] *Abbey*, 2009 ONCA 624 at para 114.

[66] See *R v Johnston*, (1992), 69 CCC (3d) 395 (Ont Ct J (Gen Div)); Edmond, "Pathological Science? Demonstrable Reliability and Expert Forensic Pathology Evidence" (Research paper commissioned by the Inquiry into Pediatric Forensic Pathology in Ontario, November 12, 2007); *R v J.E.T.*, [1994] OJ No 3067 (Ct J (Gen Div)).

4.4. Thinking Logically in Assessing Evidence

Expert evidence has become increasingly important in the trial process. In many cases it is central to the outcome. As specialties proliferate and the fields from which opinions emerge become more complex, the task of the trial judge in evaluating threshold reliability becomes more daunting. The tools outlined in this chapter may therefore be of increasing utility. Ultimately though, judges have their own speciality, namely the training to think logically in assessing evidence – of which opinion evidence is simply a subset. The trial judge is both mandated and best equipped to apply these tools to successfully determine whether opinion evidence is sufficiently reliable to be considered by the ultimate fact finder.



Chapter 4
Ethics of the Expert Witness

Professor Adam Dodek
Faculty of Law
University of Ottawa
Ottawa, Ontario

This chapter benefitted from the excellent research assistance of four superb uOttawa law students: Katelyn Wiley JD 2012, Samantha Newton JD 2012, Dara Jospe JD 2012 and Carolyn Burkholder-James JD 2014.

1. INTRODUCTION

As set out in **Chapter 1**, the traditional common law rule, also applicable in Québec courts, is that witnesses are not supposed to offer opinions. They are called upon to testify only about their personal knowledge of the facts in dispute. The expert witness, however, is a special kind of witness. Provided certain admissibility criteria are met, the expert is the only type of witness who is generally expected to testify as to his or her opinion.

The expert's role is to assist the court by providing an independent and unbiased opinion about the matters coming within the witness's expertise. The expert witness is thus in a unique position, a witness upon whom the unqualified judge may rely in making a finding that may have significant import for the parties to the litigation.[1] For reasons set out elsewhere in this manual, the judge may be at a significant disadvantage in evaluating scientific evidence as compared to other evidence.

There is a fundamental tension between the independence required of the expert and the operation of the expert witness within the adversarial system. Specifically, the expert witness is usually retained, instructed and remunerated by one of the parties. This immediately raises questions about the expert's required objectivity because either consciously or subconsciously, it is natural that the expert will want to please the party that has hired him or her, a phenomenon known as *adversarial bias.*

Moreover, many experts themselves misunderstand their role,[2] as a leading Canadian text on expert evidence has explained:

> Attitudes and expectations inherent in the adversarial system foster certain beliefs about the role of the expert witness. Some expert witnesses genuinely view it as their proper role to assist persons employing them by whatever means is enabled by their specialized knowledge. These experts are biased, but not necessarily dishonest. They do, however, overlook their primary duty to assist judges and juries. These expert advocates assist their clients in several ways, including providing opinions outside specialized knowledge or skill; considering irrelevant facts; using scientific evidence with little probative value; failing to consider alternative explanations; failing to disclose facts, documents or errors; failing to communicate limitations; overstating probative value and using potentially misleading language.[3]

---

1 *Corner Brook Pulp & Paper Ltd. v Geocon*, [2000] NJ No 446 at para 4 (Nfld SCTD).

2 The Goudge Inquiry found that Dr. Charles Smith failed to understand his duty of impartiality. See Ontario, *Inquiry into Pediatric Forensic Pathology in Ontario, Report*, vol 1 (Toronto: Queen's Printer, 2008) at 16 [*Goudge Inquiry*].

3 Glenn R. Anderson, *Expert Evidence*, 2d (Toronto: Lexis Nexis, 2009) 341 quoted with approval in *Posthumous v Foubert*, 2009 MBQB 206, 243 Man R (2d) 143.

The expert's duties to the court are set out in the common law, professional codes of conduct, expert witness codes of conduct, and obligations that may be incorporated into court rules. These are generally consistent and tend to differ only in the extent of their detailing of the expert's duties.

In ***White Burgess Langille Inman v Abbott and Haliburton Co.*** (*White Burgess*), the Supreme Court confirmed that "expert witnesses have a duty to assist the court that overrides their obligation to the party calling them."[4] This duty is comprised of three related concepts:

> The expert's opinion must be **impartial** in the sense that it reflects an objective assessment of the questions at hand. It must be **independent** in the sense that it is the product of the expert's independent judgment, uninfluenced by who has retained him or her or the outcome of the litigation. It must be **unbiased** in the sense that it does not unfairly favour one party's position over another. The acid test is whether the expert's opinion would not change regardless of which party retained him or her.[5] [emphasis added]

To the extent that the conduct of an expert witness raises ethical issues, the key points for the judge are to determine: (1) what conduct by the expert witness is improper; and (2) whether any appropriate additional measures are required when an expert is found to have violated his or her duty to the court.

2. The Ethics of the Expert Witness and the Judge's Gatekeeper Role

Chapter 1 discusses "The Legal Framework for Scientific Evidence" and the trial judge's "gatekeeper" role.[6] As the authors of that chapter state, issues of the independence of the expert witness may be considered under the "qualified expert" prong of the ***Mohan*** test at the threshold stage.[7] The expert's independence and impartiality should also be considered at the gatekeeping stage: "the judge must be satisfied that the potential helpfulness of the evidence is not outweighed by the risk of the dangers materializing that are associated with expert evidence."[8] Finally, an expert's lack of independence and impartiality can influence the weight to be given to the evidence if admitted.[9]

---

4 *White Burgess Langille Inman v Abbott and Haliburton Co.*, 2015 SCC 23, [2015] 2 SCR 182 at para 46.
5 *White Burgess*, 2015 SCC 23 at para 32.
6 See generally *R v Mohan*, [1994] 2 SCR 9 [*Mohan*]; *R v Abbey*, 2009 ONCA 624, 97 OR (3d) 330.
7 *White Burgess*, 2015 SCC 23, at para 53.
8 *White Burgess*, 2015 SCC 23, at para 54.
9 *White Burgess*, 2015 SCC 23, at para 45.

3. THE DUTIES AND RESPONSIBILITIES OF EXPERT WITNESSES

This section sets out the general duties and responsibilities of experts. A strong consensus exists about the content of these duties at the level of general principles. However, much of this consensus breaks down when these general duties must be translated to specific fact scenarios. The following sections break down the duties identified in this section and attempt to apply them to common scenarios found in the courtroom.

The general duties and obligations of the expert witness may be established under the common law, set out in rules of court or a code of conduct for experts within such rules, or specified by a code of conduct for the expert's profession.

As a matter of the common law, the leading statement in the Commonwealth is provided by the court in ***The Ikarian Reefer*** case:[10]

1) Expert evidence presented to the Court should be and should be seen to be the independent product of the expert uninfluenced as to form or content by the exigencies of litigation…
2) An expert witness should provide independent assistance to the court by way of objective unbiased opinion in relation to matters within his expertise… An expert witness in the High Court should never assume the role of advocate.
3) An expert witness should state the facts or assumptions on which his opinion is based. He should not omit to consider material facts which detract from his concluded opinion…
4) An expert witness should make it clear when a particular question or issue falls outside his expertise.
5) If an expert's opinion is not properly researched because he considers that insufficient data is available then this must be stated with an indication that the opinion is no more than a provisional one…
6) If, after exchange of reports, an expert witness changes his view on a material matter… such change of view should be communicated… to the other side without delay and when appropriate to the Court.
7) Where expert evidence refers to photographs, plans, calculations, … survey reports or other similar documents these must be provided to the opposite party at the same time as the exchange of reports...

[10] *National Justice Compania Naviera S.A. v Prudential Assurance Co. Ltd.*, ("The Ikarian Reefer") [1993] 2 Lloyd's Rep 68 at 81-82 (Com Ct), rev'd on other grounds but aff'd on this point [1995] 1 Lloyd's Rep 455 at 496 (CA).

These statements from *The Ikarian Reefer* have been embraced by many Canadian courts and may be considered generally established under Canadian common law.[11] In *White Burgess*, the Supreme Court noted that the first two elements of this duty were particularly influential in the development of Canadian law.[12] As I have already noted, the Supreme Court held that experts have a duty to be impartial, independent, and unbiased. However, it is worth highlighting that the mere appearance of bias is not enough to render expert evidence inadmissible.[13]

This duty has also been incorporated into codes of conduct for experts, *e.g.*, the Federal Court Rules (Canada) (see Appendix A – Federal Court Rules and Expert Code of Conduct, at 213). Further, the *Ikarian Reefer* principles may also be included in *Rules of Procedure*. For example, the Ontario ***Rules of Civil Procedure***, RRO 1990, Reg 194, Rule 4.1, entitled "Duty of Expert" (see *Appendix B – Ontario Form 53* at 216) provides as follows:

> **4.1.01** (1) It is the duty of every expert engaged by or on behalf of a party to provide evidence in relation to a proceeding under these rules,
>
> (a) to provide opinion evidence that is fair, objective and non-partisan;
>
> (b) to provide opinion evidence that is related only to matters that are within the expert's area of expertise; and
>
> (c) to provide such additional assistance as the court may reasonably require to determine a matter in issue.
>
> *Duty Prevails*
>
> (2) The duty in subrule (1) prevails over any obligation owed by the expert to the party by whom or on whose behalf he or she is engaged.

See also ***Alberta Rules of Court***, Alta Reg 124/2010, Rules 5.34-5.40 and Form 25;[14] British Columbia ***Supreme Court Civil Rules***, BC Reg 168/2009, Rules 11-2 and 11-6;[15] Nova Scotia ***Civil Procedure Rules***, Rule 55.04;[16] Prince Edward Island ***Rules of Civil Procedure***, Rule 53.03 and Form 53E.[17]

---

[11] But see *United City Properties Ltd. v Tong*, 2010 BCSC 111 at para 44 (wherein Romilly J. considered the test insufficient where bias is not clearly apparent).

[12] *White Burgess*, 2015 SCC 23 at para 27.

[13] *White Burgess*, 2015 SCC 23 at para 36.

[14] *Alberta Rules of Court*, Alta Reg 124/2010.

[15] *Supreme Court Civil Rules*, BC Reg 168/2009.

[16] *Nova Scotia Civil Procedure Rules*, Nova Scotia Civil Procedure Rules, Royal Gaz, Nov 19, 2008.

[17] Prince Edward Island *Rules of Civil Procedure*, online: <http://www.gov.pe.ca/courts/supreme/rules/annotated/a-rule53.pdf>.

The Québec ***Code of Civil Procedure***[18] specifically sets out the duty of experts and requires them to swear to perform these duties. The mission of all experts, whether retained by a single party, by the parties jointly, or appointed by the court, "is to enlighten the court. This mission overrides the parties' interests." All experts "must fulfill their mission objectively, impartially and thoroughly."[19] Further, "experts must sign a declaration regarding the carrying out of their mission… and attach it to their report."[20]

The subsequent sections elaborate on the principles identified above and analyze how courts have applied them in the courtroom.

4. The Independence and Impartiality of Experts

4.1. General Principles

Despite being hired and paid in most cases by one of the parties before the court, experts are expected both to appear and to be independent of the party or counsel who retained their services. This requirement represents a significant challenge for the expert, the party, counsel, and for the judge, because the expert may have had considerable history or interaction with the party or counsel that retained the expert. The expert's perception, likely shared by many of the actors within the justice system and perhaps the public, is that the expert is testifying *for* the party that retained the expert. This is known as *adversarial bias*.

Independence is an instrumental concept. As with judicial independence, the independence of experts is not an end in itself. Rather, the independence of the expert is intended to foster the formation and delivery of an objective and impartial opinion by the expert. Independence refers to the status or relationship between the expert and others, whereas impartiality refers to a state of mind or attitude.[21]

In *White Burgess*, the Supreme Court noted that the trial judge must ask whether the expert witnesses "are impartial in the sense that they are expressing their own unbiased professional opinion and whether they are independent in the sense that their opinion is the product of their own, independent conclusions based on their own knowledge and judgement."[22]

---

18 *Code of Civil Procedure*, CQLR c C-25.01 (*CCP*).
19 *CCP*, CQLR c C-25.01, art 22.
20 *CCP*, CQLR c C-25.01, art 235
21 On this distinction between independence and impartiality, see *R v Valente*, [1985] 2 SCR 673 at 685.
22 *White Burgess*, 2015 SCC 23 at para 11.

The challenge is how to ensure such independence and impartiality.

More often than not, judges have chosen to rely on counsel, the parties retaining experts or the experts themselves to ensure the independence of experts.

Thus, policy statements or professional codes of ethics may recognize the independence of certain experts. *Appendix D – Selected Codes of Conduct for Various Experts* at 219 contains relevant provisions from codes of conduct for selected experts.

For example, the ***Rules of Professional Conduct*** of the Canadian Society of Forensic Science provide that a member "be impartial and independent in their analysis, reporting and testimony."[23] In 2006, the Ontario Ministry of the Attorney General issued a practice memorandum "to reinforce the necessity for a clear and impartial presentation of the evidence to the court" in light of the role of scientists in the Centre of Forensic Science being employed in government-sponsored forensic laboratories.[24] This practice memorandum was issued in order to "protect the integrity of the role of forensic scientists and ensure that their evidence is available with all its legitimate force in the criminal process."[25] This also necessitates that the forensic scientist include any information adverse to the Crown in their report.[26]

Other scientists may be governed by similar duties of impartiality in their professional codes.[27]

The next section analyzes how these principles are applied in common scenarios involving experts.

---

[23] Canadian Society of Forensic Science, *Code of Conduct: Rules of Professional Conduct* (adopted 5 November 1994), online: <http://www.csfs.ca/become-a-member/code-of-conduct/> [Canadian Society of Forensic Science, Code of Conduct].

[24] Memorandum from Ontario Ministry of the Attorney General, Criminal Law Division, to Counsel, Criminal Law Division (31 March 2006) "Practice Memorandum [2006] No. 7: Physical Scientific Evidence" at 3 [Memorandum from Ontario Ministry of the Attorney General].

[25] Memorandum from Ontario Ministry of the Attorney General.

[26] Memorandum from Ontario Ministry of the Attorney General.

[27] See Canadian Society of Forensic Science, *Code of Conduct*; Canadian Academy of Psychiatry and the Law, *Ethical Guidelines for the Practice of Forensic Psychiatry,* online: <http://www.capl-acpd.org/ethical-guidelines/>. A complete list of national specialty societies in medicine can be found on the website of the Royal College of Physicians and Surgeons of Canada, online: < http://www.royalcollege.ca/rcsite/resources/national-specialty-societies-e/ >

4.2. Examples: In the Courtroom

The scenarios below identify situations where the independence or impartiality of experts has been called into question and explain how Canadian courts have addressed these circumstances. In all cases, the admissibility of evidence is subject to some degree of discretion by the trial judge. I have suggested Best Practices for the reader's consideration.

In considering the following situations, remember that it is the nature of the interest or connection that matters, not the mere fact that it exists. The Supreme Court in ***Mouvement laïque québécois v Saguenay (City)*** held that the mere appearance of bias is not enough to justify excluding expert evidence:

> For expert testimony to be inadmissible, more than a simple appearance of bias is necessary. The question is not whether a reasonable person would consider that the expert is not independent. Rather, what must be determined is whether the expert's lack of independence renders him or her incapable of giving an impartial opinion in the specific circumstances of the case.[28]

Once an expert attests to his or her independence and impartiality, it is up to the challenging party to raise a realistic concern that the expert is unable or unwilling to comply with his or her duties. If such a concern exists, the party calling the evidence will have to establish on a balance of probabilities that the expert is willing and able to meet his or her duty to the court.[29]

4.2.1. Party / Spouse of Party as Expert

Courts generally do not allow parties or their spouses to offer expert evidence.[30] Courts have refused to qualify a party or party's spouse as an expert in the party's own case. For example, in one case the court stated: "…it is unreasonable to believe that the Appellant… could have offered an entirely objective opinion uninfluenced by his personal interest. Given his interest in the matter, he definitely could not have provided the objectivity essential to expert status."[31] The courts are close to adopting a categorical rule against parties or their spouses offering expert evidence.

---

[28] *Mouvement laïque québécois v Saguenay (City)*, 2015 SCC 16, [2015] 2 SCR 3, at para 106.

[29] *White Burgess*, 2015 SCC 23 at para 48.

[30] See *Handley v Punnett*, 2003 BCSC 294; *White Burgess*, 2015 SCC 23 at para 49.

[31] *Diotte v Canada*, 2008 TCC 244, 2008 DTC 4558 at paras 25-28. See also *Handley v Punnett*, 2003 BCSC 294 at para 19; *Ocean v Economical Mutual Insurance Co.*, 2010 NSSC 315, 293 NSR (2d) 394 at para 22.

Best Practice:	I believe that the following statement by Burnyeat J. of the British Columbia Supreme Court best encapsulates the approach that courts have taken and should take. Burnyeat J. expressed serious reservations that a party could provide the objectivity required of an expert witness, leading him to conclude that "it would be an almost insurmountable barrier for a party to be in a position to express an expert opinion."[32]

4.2.2. Employee of Party as Expert

While an employment relationship can raise the spectre of bias, the Supreme Court has recently stated that "a mere employment relationship with the party calling the evidence will be insufficient" to render the proposed expert evidence inadmissible.[33] It might well be the case that the person best qualified (in terms of scientific knowledge) to give evidence is, indeed, the research scientist employed by one of the parties to a lawsuit. In many areas of science, some of the most respected researchers are employed by private companies who may be parties to litigation. Automatically excluding such persons because of their pecuniary connections to one of the parties risks depriving the court of the benefit of top quality scientific expertise. As in many areas, the issue for the court is one of weighing competing risks and benefits.

In one case, the court accepted as an expert witness the applicant corporation's Vice-President, Pharmaceutical Technology, despite objections by the respondent, who questioned the expert's financial interest in the outcome of the litigation.[34] The court stated: "[t]he weight to be given to his evidence could have been tempered by that fact but I am satisfied his evidence was objective and helpful. I qualify him as an expert."[35] Similarly, in a 2010 decision the Alberta Court of Appeal allowed the expert testimony of the Chief Accountant of the Alberta Securities Commission. It found that although the Chief Accountant worked for the Commission, there was no reasonable apprehension of bias created by his testimony.[36]

In a 2010 decision, the Federal Court of Appeal took a very strong position in support of allowing the testimony of a chemist who was employed by one of the parties. The court stated: "Counsel for Hospira was unable to cite any authority for the proposition that an employee cannot give opinion evidence

---

32 *Handley v Punnett*, 2003 BCSC 294 at para 19.
33 *White Burgess*, 2015 SCC 23 at para 49.
34 *Biovail Pharmaceuticals Inc. v Canada (Minister of National Health & Welfare)*, 2005 FC 9, 267 FTR 243 at para 18.
35 *Biovail Pharmaceuticals Inc. v Canada (Minister of National Health & Welfare)*, 2005 FC 9 at para 18.
36 *Alberta (Securities Commission) v Workum*, 2010 ABCA 405, 493 AR 1 .

on behalf of his or her employer merely because of the employee's lack of independence from the employer. I am unaware of any basis for such a sweeping proposition which would have wide ranging consequences."[37]

Other cases have refused to recognize an employee of a party as having the independence required of an expert. In one case, the court concluded that a senior executive of the defendant was not an independent expert witness. The court stated that his testimony: "simply amounted to advocacy for his company. He is too connected to one side of this litigation for his opinions to have much value in this context."[38]

In another case, the trial court excluded an employee of the prosecuting Ontario Ministry of the Environment. However, the decision of the trial judge was reversed on appeal.[39] The trial court had reasoned as follows:

> Mr. Mak is not only employed by the Ministry of the Environment, but is attached to and intimately concerned with the day-to-day operations involving investigations and enforcement by instructions to and education of other members of the Branch and including experts. I have no doubt that Mr. Mak is an honourable person. I have no doubt that he would attempt to be honest and fair in his testimony, and in giving his opinions, but he is not being proffered in the same light as those government expert/employees such as, first instance, work in the Centre of Forensic Sciences, the Ministry of the Solicitor General, with which I am probably aware of more than other experts that are preferred by the government in prosecutions. These experts are used not only extensively in prosecutions, but also to a very large degree by the defence bar of Ontario and I dare say outside of the province and even the country. They do not have that connection as does Mr. Mak. They do not gather, direct or instruct as does he. His position, in my view, could only be perceived by the public as capable of lacking independence.
>
> There is nothing in the evidence to suggest actual bias, but his position does not lend to the appearance of professional objectivity. In my view, who pays him, who assesses him, is no more relevant than who pays and who assesses experts from the Centre of Forensic Sciences. That is not in consideration in this case.

---

37 *Eli Lilly Canada Inc. v Hospira Healthcare Corp.*, 2010 FCA 282, 409 NR 167 at para 7.

38 *Prairie Well Servicing Ltd. v Tundra Oil & Gas Ltd.*, 2000 MBQB 52, 146 Man R (2d) 284 at para 24.

39 *R v Inco Ltd.* (2006), 80 OR (3d) 594 (Sup Ct).

> Basically, the bottom line here is that there is not the separation between Mr. Mak and the Crown/Prosecution that ensures the vital appearance of impartiality. He will not, therefore, be permitted to testify as an expert.[40]

On appeal, the court held: "the mere fact that the witness in this case was employed in the Investigations & Enforcement Branch as a 'technical enforcement specialist' is not a sufficient basis on which to find him incapable of providing an independent opinion."[41]

In another case, the experts were directly employed by a holding company that owned a significant interest in the plaintiff. The court held that the witnesses' objective was parallel to that of the individual, his company and that of the plaintiff and therefore they were found not to be independent witnesses.[42]

Best Practice:	Employees of parties should be treated with caution as potential expert witnesses. They should not be excluded *per se*, but because of their links to the party they should be subjected to heightened scrutiny to ensure objectivity in their analysis. This practice should also apply to witnesses who may not technically be employed by the party but have a strong connection to the party (such as pathologists, police officers, etc., in cases involving the Crown/government).

4.2.3. Current or Past Business Relationship Between Expert and Party

Courts have taken contrasting approaches to current or past business relationships between an expert and the party proffering the expert's testimony.

In one case, two engineers had collaborated on at least two studies and articles highlighting the benefit of the type of construction used by the party that retained them in the litigation. The court found that there had been "a close professional collaboration" between these two experts and one of the parties to the litigation, at a time when the party was designing buildings such as the one at issue in the lawsuit. Together with "their past explanation to the rest of the world of the merits of this commonly used form of construction in Ottawa,"

---

40 *R v Inco Ltd.* (2006), 80 OR (3d) 594 (Sup Ct) (citing 10-11 of trial decision of Mahaffy J.).

41 *R v Inco Ltd.* (2006), 80 OR (3d) 594 at para 44 (Sup Ct).

42 *1159465 Alberta Ltd. v Adwood Manufacturing Ltd.*, 2010 ABQB 133, 25 Alta LR (5th) 237 at Sch. 2 (The Admissibility of Expert Evidence), para 2.3.

the court concluded this suggested that the two experts were not "totally disinterested bystanders offering a completely independent perspective."[43]

In another case, the court refused to draw any adverse inferences from the fact that an expert had testified for Apotex more than 20 times in a career spanning over 30 years. The court explained: "[t]he implication is that if he is not a man for all seasons, he is certainly a man for all patents. However, I was not aware that there was a limit to a number of times a witness could appear."[44]

In another case, an expert's independence was called into question because of his close association with the party that retained him. The expert had known the party for 25 to 30 years and the party was one of the expert's first customers. The expert had constructed and serviced three radio towers for party.[45] Similarly, in another case the proposed expert had a long-standing relationship with the individual plaintiffs and their company for some 25 years, which called into question his objectivity.[46] In another case, the court rejected the expert evidence of the founder and CEO of a shareholder services company on the relationship between shareholder turnout at meetings and whether the meetings resulted from a requisition or a disputed proxy simulation. The judge ruled that the CEO's testimony was inadmissible because he had a previous and ongoing business relationship with the respondents by providing services for shareholder meetings. He therefore lacked the independence and objectivity required of an expert.[47]

Conversely, in ***More v Bauer Nike Hockey Inc.***,[48] the plaintiffs objected to the evidence given by three expert witnesses, contending that they were not independent or impartial because of past associations with the defendants. One of the witnesses had worked for Bauer, another had been a voluntary member of the defendant Canadian Standards Association (CSA), and the third had served on committees for the CSA. The judge found that despite their connections with the defendants, none of the experts showed bias. Their opinions, the judge found, were "academically sound and were carefully presented."[49] Similarly, in ***Bursey v St. John's (City)***,[50] the plaintiffs challenged the testimony of the defendant's expert witness on the grounds that the professional engineer and his company had done consulting work for the defendant, and that he was therefore biased. The judge refused to find any bias on this basis.

---

43 *Carleton Condominium Corp. No. 21 v Minto Construction Ltd.* (2001), 15 CLR (3d) 23 at para 29 (Ont Sup Ct) [*Carleton Condominium v Minto Construction*].

44 *Lundbeck Canada Inc. v Canada (Minister of Health)*, 2009 FC 146, 343 FTR 53 at para 74 [*Lundbeck Canada Inc. v Canada*].

45 *Widelitz v Robertson*, 2009 PESCTD 21, 289 Nfld & PEIR 111 at para 37.

46 *Kirby Lowbed Services Ltd. v Bank of Nova Scotia*, 2003 BCSC 617, [2003] B.C.J. No. 917.

47 *Ebrahim v Continental Precious Metals Inc.*, 2012 ONSC 2918, 111 OR (3d) 110.

48 *More v Bauer Nike Hockey Inc.*, 2010 BCSC 1395, 78 CCLT (3d) 89.

49 *More v Bauer Nike Hockey Inc.*, 2010 BCSC 1395 at para 184.

50 *Bursey v St. John's (City)*, 2011 NLTD(G) 130, 314 Nfld & PEIR 75.

Best Practice:	Similar concerns arise here as for employees of a party. A distinction may be drawn between a past association and a current one. If the expert has a current business association with a party, the expert may have a direct or indirect stake in the litigation and it is difficult to see how he or she could be considered objective or independent; he or she should probably not be allowed to qualify as an expert.

4.2.4. Lawyer for Party in Related Proceedings

Courts are ambivalent about whether a current or past lawyer-client relationship between the party and the proposed expert disqualifies the proposed expert as a witness.

One case held that current involvement as an advocate for a party disqualifies one as an expert.[51] Similarly, in another case, the court refused to qualify the party's American counsel as an expert on American law because of his connection to the party.[52] However, another case held exactly to the contrary and allowed as an expert a lawyer who had represented the party in a foreign proceeding, but held that the expert's connection to the party may affect the weight given to his evidence.[53]

Another case refused to allow the testimony of lawyers who had been instructed and paid by the testator because they had personal involvement in the preparation and execution of the very documents that were contested in the lawsuit, and were therefore not independent.[54]

Best Practice:	The lawyer for a party in a related proceeding is in the same category as a proposed expert with an ongoing business relationship with the party above. It is difficult to see how the lawyer/proposed expert could be considered objective or independent; he or she should probably not be allowed to qualify as an expert.

51 *Fellowes, McNeil v Kansa General International Insurance Co.* (1998), 40 OR (3d) 456 (Ont Ct (GD)) [*Fellowes, McNeil v Kansa Insurance*].

52 *Royal Trust Corp. of Canada v Fisherman* (2000), 49 OR (3d) 187 (Sup Ct).

53 *Bil'in (Village Council) c Green Park International*, 2009 QCCS 4151, [2009] RJQ 2579 at paras 82-84.

54 *Dansereau Estate v Vallee*, 1999 ABQB 557, 247 AR 342 at paras 134-41.

4.2.5. Participants in the Background to the Lawsuit

An expert's independence or objectivity may be tainted by his or her involvement in matters connected to the lawsuit. In one case, the court concluded that an expert who was involved in the construction process at issue in the lawsuit lacked the requisite objectivity. There had been a close association between the expert and the party that retained him as they had worked on various construction projects together.[55] However, the more serious problem arose because the expert had been involved in advising the party during the construction problem that was at issue in the lawsuit. Thus, the expert was essentially providing an opinion on his own work.[56]

In another case, the court refused to allow a consultant who had advised the defendant about repairs on the roof of its building to testify as an expert because he was directly involved in the case and therefore was not a "disinterested party."[57] However, the court did allow the consultant to testify as to "both factual evidence and opinions."[58] Similarly, a court rejected the attempt by a party to introduce her own home inspector as an expert witness.[59] And in another case, the court refused to qualify two Canada Revenue Agency officers as experts because they were investigators in the tax fraud case against the defendants.[60]

It is not uncommon for a party to retain an expert as a consultant during litigation and then also to provide an expert opinion. While such retainers are problematic in terms of an expert's independence and objectivity, they are not prohibited nor are they particularly frowned upon. In one of the strongest cases setting out the responsibilities of experts, Justice Farley stated as follows:

> Whatever role the expert may have undertaken during the course of the litigation in assisting counsel to a fuller appreciation of the facts in dispute and the inferences that might be drawn from them, the expert must set aside that role upon entering the witness box at trial. From the witness box the expert speaks only to assist the court.[61]

More consideration is needed regarding the compatibility of the roles of litigation consultant and expert witness combined in a single person.

---

55 *Carleton Condominium v Minto Construction* (2001), 15 CLR (3d) 23 at para 29.
56 *Carleton Condominium v Minto Construction* (2001), 15 CLR (3d) 23 at para 26.
57 *Continental Roofing Ltd v J.J.'s Hospitality Ltd.*, 2012 ONSC 1751, 12 CLR (4th) 90 at para 40.
58 *Continental Roofing Ltd v J.J.'s Hospitality Ltd.*, 2012 ONSC 1751 at para 42.
59 *Ricchio v Rota*, 2011 ONSC 7500, [2011] OJ No 6111.
60 *R v He*, 2010 BCPC 457, [2010] BCJ No 2938.
61 *Frazer v Haukioja* (2008), 58 CCLT (3d) 259 at para 138 (Ont Sup Ct).

Best Practice:	It is extremely problematic to permit experts to testify as experts when they have been involved in matters relating to the lawsuit. This confuses the roles of factual witnesses and expert witnesses and generally should not be permitted. While the quote above from Justice Farley accurately sets out the law and the duties of the expert, it is unrealistic to expect the expert who has been involved in matters relating to the lawsuit to set aside that involvement and be objective in assisting the court as an expert. Accordingly, persons who have been involved in the underlying factual matters of the lawsuit should not be permitted to testify as experts on the grounds that they lack the requisite independence and impartiality.

4.2.6. Expert's Interactions with Counsel

An expert's interaction with counsel may lead the trier of fact to question the expert's independence because of the perception that the expert's opinion may have been influenced by counsel.

Thus, an expert's independence was called into question by his agreement to provide a formal opinion to counsel in writing only after discussing his views with counsel orally.[62] The court explained that it questioned the expert's impartiality because the expert had several hours of telephone conversations and a meeting with counsel before preparing his reports, and read transcripts of evidence containing facts clearly different from the facts upon which his opinions were based without stating that he had done so, let alone explaining why.[63]

In another case, the court criticized the communication and collaboration between the expert and the client, the Financial Services Commission of Ontario, which was also the prosecuting authority in the case.[64] This communication and collaboration occurred in all aspects of the expert's work, including the various stages and draft preparation of what became the expert's final report, and considerations as to the bases and foundational material and sources for his report. The expert took these suggestions and used the words that the client gave him in his e-mail, including them in his final report. He also took out completely and entirely what he had written in an earlier draft about testing the sufficiency and reliability of data. The court found it troubling that the expert kept the client completely in the loop for input, suggestions and review of his draft reports. The court stated that the expert lost sight of his role

62 *Frazer v Haukioja* (2008), 58 CCLT (3d) 259 at para 142.
63 *Frazer v Haukioja* (2008), 58 CCLT (3d) 259.
64 *R v Norton*, 2007 ONCJ 414, 59 CCPB 27.

and the distance he should maintain from the party. Moreover, the expert had made a conscious decision not to set out the sources with whom he spoke.[65] Another court came to a similar conclusion based on its finding that e-mails between the defendants and the expert showed that the expert based his position on the defendants' theory and therefore "assumed the role of advocate."[66] The judge disqualified the expert in the case.

An unusual case occurred in Prince Edward Island in 2009, in which the party prepared the report for the expert who reviewed it and signed it. The court held that expert's report was "a sham. It is not an independent report. It is not a report uninfluenced by the exigencies of litigation. It is not [the expert's] unbiased, objective opinion. In fact, it is not [the expert's] report at all. Not one word of that report belongs to [the expert]. It is all [the party's] evidence. As a piece of evidence it is utterly worthless, and I place zero weight on it."[67]

While this is an extreme example, it demonstrates the general concerns with experts' interactions with counsel or parties. As the Federal Court explained in one case where the parties played a very significant role in the drafting of expert reports: "the more the lawyers are involved the more careful an expert must be in reviewing the text proposed to ensure that it truly reflects his or her views."[68]

The BC Supreme Court provided some words of caution on the role of counsel in preparation of the expert's report: "Often counsel is able to state a set of facts to an expert witness and have the witness express an opinion based upon those assumed facts. In this case that option did not appear open to counsel because Dr. Turner's opinions are based, in part at least, on her own field research."[69] In this case, the exchange of correspondence between counsel and the expert contained "unfortunate" language which left open the argument that counsel dictated the opinion required of the expert and that the expert complied with the dictates of counsel. Counsel's editing of the expert's report left open a similar argument. The court stated that counsel "should strive, at all times, not to place themselves in the position where their conduct becomes a focal point of the court's concerns."[70] The court concluded that the expert's report was admissible.[71]

---

65 *R v Norton*, 2007 ONCJ 414 at para 89.
66 *Alfano (Trustee of) v Piersanti*, [2009] OJ No 1224 (Sup Ct) at para 2.
67 *Widelitz v Robertson*, 2009 PESCTD 21 at para 40.
68 *Eli Lilly & Co. v Apotex Inc.*, 2009 FC 991, 351 FTR 1 at para 62.
69 *Tsilhqot'in Nation v Canada (Attorney General)*, 2005 BCSC 131 at para 33.
70 *Tsilhqot'in Nation v Canada (Attorney General)*, 2005 BCSC 131 at para 34.
71 *Tsilhqot'in Nation v Canada (Attorney General)*, 2005 BCSC 131.

In other cases, courts have been quite accepting of the expert's collaboration with counsel. In one case, the court explained that it had become apparent during the expert's cross-examination that in preparing his affidavit, the expert had copied or paraphrased some 29 passages from a brief submitted by the party's counsel in the U.S. proceedings, including the dictionary definitions of eight terms. The court stated that it was "regrettable" that the expert "chose to rely upon the terms and phrases employed in the U.S. brief to convey his views but I am not convinced that in so doing he lost the objectivity and impartiality required to assist the court with his expertise."[72] The court further concluded:

> In effect, he adopted the language in that brief as his own in writing his report as it accorded with his understanding of the technology at issue and used it to describe his opinion. While I do not condone this practice, I doubt that there are many expert reports that are not, to some extent, the product of collaboration between counsel and the expert if only to conform to varying legal requirements in different jurisdictions or to focus the report on the issues. I am unable to conclude that Mr. Phillips' evidence lacked such objectivity that I should ignore it or give it little weight. I found his evidence, generally, to be helpful, clear and precise.[73]

The problem identified above seems to be the failure of the expert, and by extension the lawyer who drafted the report, to acknowledge the source of the information relied upon or paraphrased. There may be more benign reasons for counsel to assist in the drafting of a report. Time or resource pressures may prohibit an expert from drafting a report. An expert may have limited English or French language skills. In such circumstances, the expert may explain his or her analysis to counsel who may then draft the report for the expert's review and signature. In such circumstances, what is critical is not that counsel drafted the report but the full context of the preparation of the report, including the facts provided by counsel to the expert that served as the basis for the expert's findings and conclusions, and the process by which the expert came to review and adopt the report.

In a leading Ontario case, ***Moore v Getahun***,[74] the Court of Appeal directly addressed the issue of consultation between counsel and experts. The court observed that independence and objectivity of an expert witness is fostered by the ethical and professional standards of the legal profession, as well as those of the professional bodies representing potential expert witnesses, and can be tested through cross-examination.[75] These safeguards help ensure that the expert performs his or her duty to the court, without requiring disclosure of all communications between counsel and expert witness.

---

[72] *Dimplex North America Ltd. v CFM Corp.*, 2006 FC 586, 292 FTR 38 at paras 43-44.
[73] *Dimplex North America Ltd. v CFM Corp.*, 2006 FC 586.
[74] *Moore v Getahun*, 2015 ONCA 55, 124 OR (3d) 321.
[75] *Moore v Getahun*, 2015 ONCA 55 at paras 56-61.

The Court held that consultation and collaboration between experts and counsel is essential:

> Counsel need to ensure that the expert witness understands matters such as the difference between the legal burden of proof and scientific certainty, the need to clarify the facts and assumptions underlying the expert's opinion, the need to confine the report to matters within the expert witness's area of expertise and the need to avoid usurping the court's function as the ultimate arbiter of the issues.[76]

Not only is this consultation necessary, but the draft reports, notes and records of these consultations are protected from disclosure under litigation privilege.[77] However, this protection is not absolute:

> Where the party seeking production of draft reports or notes of discussions between counsel and an expert can show reasonable grounds to suspect that counsel communicated with an expert witness in a manner likely to interfere with the expert witness's duties of independence and objectivity, the court can order disclosure of such discussions.[78]

Best Practice:	Judges should be cautious of undue collaboration between experts and lawyers. Where the judge believes that the expert's report is actually the product of the lawyer's work and not the expert's, the report should not be admitted.

4.2.7. The Expert for Hire

There is a frequent perception that experts may be "hired guns." Certain activities by experts may contribute to this perception and call into question their independence.

When an expert testifies only for one side (*e.g.*, defendants or plaintiffs), this may raise a red flag about his or her independence and signal to the trier of fact the need to be vigilant in assessing the expert's impartiality.

[76] *Moore v Getahun*, 2015 ONCA 55 at para 63.
[77] *Moore v Getahun*, 2015 ONCA 55 at paras 69-70.
[78] *Moore v Getahun*, 2015 ONCA 55 at para 77.

For example, in one case the practice profile of the expert indicated that 80% of his medical legal work was undertaken on behalf of defendants, with approximately 25% of his professional time being devoted to medical legal matters from which he earned "probably twice as much income" as he did from his clinical practice.[79] The court in this case commented as follows:

> That an expert is paid for services rendered in a case is not, of itself, a concern but the profile elicited from Dr. Reznek is a red flag, the sight of which focuses the court's attention upon the need for impartiality to be demonstrated in the evidence the proposed to give. The demonstration of that impartiality was found wanting.[80]

In the same case, the independence of the expert was called into question because the doctor chose to include a reference in his C.V. to the fact that he was a medical expert for the law firm that retained him in the case.[81]

The frequency of testimony alone will not usually suffice to taint an expert's independence. For example, in one case an expert had been called as a witness by the same party more than 20 times in 30 years. The court commented that it was not aware that "that there was a limit to a number of times a witness could appear."[82]

Best Practice:	The predictability of an expert's testimony may be a function of the expert's experience in a field or of the expert's lack of objectivity. Where the judge believes it is the latter, this consideration may properly go to the admissibility of the testimony.

[79] *Frazer v Haukioja* (2008), 58 CCLT (3d) 259 at para 145.
[80] *Frazer v Haukioja* (2008), 58 CCLT (3d) 259 at para 146.
[81] *Frazer v Haukioja* (2008), 58 CCLT (3d) 259 at para 142.
[82] *Lundbeck Canada Inc. v Canada*, 2009 FC 146.

4.2.8. The Expert's Stake in the Litigation

Courts have recognized that all experts have a stake in the outcome of litigation in the sense that their future employment and remuneration or professional status may depend on the success of their testimony. This is an accepted part of the litigation process. Canadian courts have been concerned with more direct stakes that the expert may have in the litigation.

Somewhat surprisingly, there are no rules of court barring contingency fees for expert witnesses in Canada.[83] The issue has arisen infrequently in Canada and it does not appear to be a common practice. When the issue has arisen, Canadian courts have reacted strongly against such practices.

In one of the few reported instances of an expert being a true co-venturer in the litigation, the party had guaranteed the expert 60% of his fee, and the other 40% would be contingent on the outcome of the case. After delivery of the expert's report, this was changed to a 100% guarantee. This contingency fee made the expert a co-venturer in the litigation. While the expert may not have appreciated that result, according to the judge, "the fact is that [the expert] had lost his neutrality and objectivity."[84]

In another case, the court refused to qualify an expert when he had entered into a retainer agreement entitling him to $1000 plus $100 per hour (plus expenses), or 10% of the value of the award or settlement. The judge found that the proffered expert had "lost his neutrality and objectivity" through the retainer with the plaintiff since the more favourable his evidence, the more money he would be paid.[85] The judge refused to allow the witness to testify as an expert, but did let him give evidence as a fact witness, assigning it little weight due to errors which he admitted.

In such circumstances, it is probably more accurate to describe the situation as being that, in the court's view, the contingency fee increased the risk of undermining the neutrality or objectivity of the expert beyond what was acceptable.

Courts may also reject expert witnesses who have other personal interests in the outcome of the case. Thus, in one case the court held that the investigator/expert witness had a personal interest in establishing that his manager's suspicions were well-founded in order to help secure a promotion.

---

[83] The issue has been studied in other jurisdictions. See, *e.g.*, Law Reform Commission New South Wales, "The 'No Win, No Pay' Expert Witness (Contingency Fee Expert Witnesses)," in *Report 109: Expert Witnesses* (Sydney: NSW Law Reform Commission, 2005) 143 et seq. [NSW Law Reform Commission, *Expert Witness Report*].

[84] *Bank of Montréal v Citak*, [2001] OJ No 1096 at para 5 (Sup Ct).

[85] *Dean Construction Co. v M.J. Dixon Construction Ltd.*, 2011 ONSC 4629, 5 CLR (4th) 240 at para 59 (Mast).

The investigator/expert personally met with the Crown attorney to convince him that charges were warranted. The court held that this suggested that the investigator/expert had a personal interest in the outcome of the case. In addition, the court held that the investigator's employer also had an interest in the successful prosecution of the charge.[86]

Similarly, where expert witnesses themselves are potentially liable, courts have stated that their evidence should be rejected as not being impartial.[87]

Best Practice:	Where an expert has a direct financial stake in the litigation through a contingency fee or other pay-for-performance arrangement, he or she has clearly crossed the line from being a neutral and objective witness into becoming a participant in the litigation, and his or her testimony should not be admitted on the grounds that it lacks the requisite impartiality.

4.2.9. The Expert Assumes the Role of Advocate

The general prohibition has been stated as follows: "Experts must not be permitted to become advocates. To do so would change or tamper with the essence of the role of the expert, which was developed to assist the Court in matters which require a special knowledge or expertise beyond the knowledge of the Court."[88]

The caselaw is replete with examples of courts chastising experts for assuming the role of advocates. Rarely are the examples as explicit as in ***United City Properties Ltd. v Tong***, where cross-examination of the expert proceeded as follows:

> *Q. You see yourself as your client's advocate; isn't that correct?*
>
> *A. Of course I do.*
>
> *Q. And you see yourself as your client's advocate here today, sir?*
>
> *A. Yes, I do.*[89]

[86] *R v Payette*, 2010 MBQB 73, 253 Man R (2d) 181 at para 30.
[87] *Hutchingame v Johnstone*, 2006 BCSC 271 at paras 6-8.
[88] *Fellowes, McNeil v Kansa Insurance*, (1998), 40 OR (3d) 456 (Ont Ct (GD)).
[89] *United City Properties Ltd. v Tong*, 2010 BCSC 111.

In such cases, the judge should inquire what the expert believes his or her role and responsibilities to be more precisely. It is possible that the term "advocate" may mean something different to an expert witness not steeped in the law, in contrast to its meaning for those in the legal system.

This was clearly not the case in another situation where the expert stated that "in every matter of litigation, I always take the position of advocate for my client and [for] that I'm paid a good fee" while asserting his independence and objectivity.[90] The court did not accept that the witness was independent and objective.

Rarely are witnesses so explicit or candid. More often courts conclude that an expert has become an advocate based upon the conduct of the expert.

Thus, the court concluded in one case that the expert had assumed the role of advocate because he rejected evidence – an MRI prepared by another doctor – that detracted from his diagnosis. The expert disregarded evidence that did not support his findings – evidence of symptoms which could not be caused by that to which he had attributed the symptoms.[91]

In a case involving Ontario Hydro, the court found that the American expert, Dr. Rosenberg, had displayed a deep suspicion of Ontario Hydro and utilities in general. The court stated that the expert "simply did not display the characteristics that a court would expect of a truly impartial expert and that it would require in a complicated damages case such as this." The court stated that the expert's evidence "did not appear unbiased and he seemed far too influenced by the exigencies of litigation." The court concluded that Dr. Rosenberg "became an advocate for his client and lost any objectivity he might otherwise have had. It was only after extensive cross-examination…, for example, that he would concede a point on which he was obviously in error."[92]

In another case, the Federal Court of Appeal expressed "grave concerns" about the objectivity and independence of a proposed expert's opinion, in part because the expert had expressed "editorial comments about some of the decisions using language that is gratuitous, intemperate and ideological. Further, the opinion expresses dislike for some of the jurisprudence of the Federal Court and this Court. This colours the opinion's assessment of the judge's decisions, many of which follow this jurisprudence."[93]

---

90 *Bank of Montréal v Citak*, [2001] OJ No 1096 at para 6.

91 *Posthumous v Foubert*, 2009 MBQB 206 at para 63.

92 *Eastern Power Ltd. v Ontario Electricity Financial Corp.*, [2008] OJ No 3722 at para 306 (Sup Ct).

93 *Es-Sayyid v Canada (Minister of Public Safety and Emergency Preparedness)*, 2012 FCA 59, [2013] 4 FCR 3 at para 43.

Strongly held opinions do not necessarily equate with advocacy or a lack of independence. Thus, in one case, the expert had spent her entire career studying the First Nations people of British Columbia, their relationships with plants, and their ecology. The court held that this did not make her biased and did not convert her into an advocate. The court held that it was expected that she would have firm opinions to express on these issues and the impact of the Europeans on the First Nations people and their environments. The fact that the expert's opinions were sympathetic to the position advanced by the plaintiff in this case did not transform her into an advocate for the plaintiff. The court held that the manner in which the expert expressed those opinions, in writing and orally, and the expert's demeanour under cross-examination, all bore on whether the opinions she expressed could be relied upon.[94]

Similarly, in the **BC Polygamy Reference**,[95] the Attorneys General of BC and Canada objected to the evidence of a McGill law professor who was testifying on behalf of the *amicus curiae.* The professor had researched the practice of polygamy in the BC community of Bountiful and was explicit in stating that her research was intended to give a voice to the women she interviewed in that community. In rejecting the Attorney General's assertion that the expert lacked the requisite impartiality, the judge stated that experts with firmly held opinions are required in a reference case and that this does not impair their ability to provide objective evidence. Any concerns about impartiality could be dealt with through cross-examination or by assessing the weight of the evidence.

In two "honour killing" cases several years apart, the same expert was challenged on the grounds that she was a strong advocate for women's rights, causing her to be biased. In both cases, the judges allowed her testimony, noting that her strongly held opinions did not exclude her as an expert.[96]

Additionally, the Newfoundland and Labrador Court of Appeal found that the trial judge did not err in admitting a neurologist's expert evidence that was alleged to be biased because of his criticism of the chiropractic profession. The expert had stated – in a foreword of a book and in his report – that chiropractors were in denial about the risk of serious injury arising from neck manipulation. The Court of Appeal held that the trial judge did not err in admitting the expert's evidence despite the expert witness's strongly held opinions.[97]

---

94 *Tsilhqot'in Nation v Canada (Attorney General)*, 2005 BCSC 131.

95 *Reference re: Criminal Code of Canada (BC)*, 2011 BCSC 1588, 28 BCLR (5th) 96.

96 *R v Sadiqi* (2009), 68 CR (6th) 346 (Ont Sup Ct); *R v Shafia*, 2012 ONSC 1538, 285 CCC (3d) 283.

97 *Gallant v Brake-Patten*, 2012 NLCA 23, 321 Nfld & PEIR 77.

Problems arise when an expert has been involved as a counselor, therapist or treating physician of one of the parties in the case. It is not surprising that in such circumstances, the proposed expert has difficulties separating his or her duty of objectivity to the court from his or her duty of advocacy on behalf of his or her client. Thus, in a case involving a dispute over custody, access, and guardianship of the parties' two children, the court found that a psychologist who had been qualified as an expert showed "obvious advocacy" on behalf of one the parties for whom he had served as a counsellor and twice intervened in ministry investigations on behalf of the party.[98] Another court came to the same conclusion under different circumstances. In ***Children's Aid Society of London and Middlesex v A.L.N.***,[99] two experts, one of whom was the respondent's therapist, disagreed about the intelligence of the respondent. The respondent's therapist found that she was close to, or at an average intellectual level, while another expert witness found that she had only borderline intelligence. The other witness stated that the respondent's therapist/expert must have used "liberal scoring and interpretive methods" to come to his conclusion.[100] The judge agreed, finding that the therapist/expert could have unconsciously given the benefit of any doubt to the respondent and therefore inflated her score. The judge stated that the therapist had become an advocate for the respondent. Therefore, the judge gave more weight to the other expert's testimony finding it to be "relevant [...] and correct when all of the evidence is taken together."[101]

In ***Gutbir (Litigation guardian of) v University Health Network***,[102] the plaintiffs applied to qualify Dr. Perlman, a neonatologist, as an expert witness. The plaintiffs were suing the hospital for damages after their child was born brain damaged and disabled. Dr. Perlman had treated the child at a different hospital after she was born, and had written two reports about her condition. The plaintiffs argued that Dr. Perlman should be allowed to give expert testimony on the cause of the child's brain damage, her disability, and when they both occurred. The defendant argued that Dr. Perlman would be unable to give an objective opinion because he had treated the child. The judge agreed, finding that Dr. Perlman had an interest in agreeing with the conclusion of his previous reports. Therefore, his evidence could not be neutral or impartial.

In many cases, courts have allowed the expert testimony of treating physicians.[103]

---

98 *A.E.O. v T.K.O.*, 2012 BCSC 708, [2012] BCJ No 968 at para 36.

99 *Children's Aid Society of London and Middlesex v A.L.N.*, 2012 ONSC 585, [2012] OJ No 1434.

100 *Children's Aid Society of London and Middlesex v A.L.N.*, 2012 ONSC 585 at para 69.

101 *Children's Aid Society of London and Middlesex v A.L.N.*, 2012 ONSC 585 at para 73.

102 *Gutbir (Litigation guardian of) v University Health Network*, 2010 ONSC 6394, 7 CPC (7th) 208.

103 See *e.g.*, *Degennaro v Oakville Trafalgar Memorial Hospital*, 2011 ONCA 319, 81 CCLT (3d) 165; *Lynne v Taylor*, 2006 ABCA 12, 376 AR 353; *R v Ryszard*, 138 Man R (2d) 95 (CA); and *Farooq v Miceli*, 2012 ONSC 558.

There are numerous criminal cases that deal with what might be termed "alleged police officer bias." In these cases, it is alleged that current or former police officers lack the requisite objectivity or independence to testify as experts for the Crown because of their current or former associations with the police. Generally, courts have not been receptive to such claims. Thus, in an Alberta case, the defendants asked the judge to prevent a retired RCMP staff sergeant from testifying as an expert on the Hells Angels Motorcycle Club for the Crown on the grounds of "police officer bias." The judge found that although the retired RCMP staff sergeant was influenced by his career as a police officer, the defendants had not established that this constituted evidence of bias. Any issue of independence would go to the weight of the evidence.[104] A BC court had made a similar ruling several years earlier involving the same retired RCMP staff sergeant.[105]

In ***R v Baxter***,[106] the defendant argued that the Crown's expert witness should not be allowed to give evidence because he was not impartial. Detective Heroux, an expert in the area of computer forensics and data recovery, was set to testify for the Crown in the child pornography case. The defendant objected to the testimony, arguing that Detective Heroux was biased because of his past membership in a police-run organization that only testified for the Crown. The judge found that the detective's past membership in the organization did not disqualify him from giving testimony and that his evidence was properly researched, thoroughly investigated, and based on valid assumptions. In ***R v Gager***,[107] the Crown applied to have a detective declared an expert to testify about street gangs. The defendant argued that the detective was too biased to serve as an expert witness on gang culture in the murder trial because he was a police officer and therefore "naturally predisposed toward the prosecution."[108] The judge disagreed, stating that police officers are often called as expert witnesses. The judge also found that the detective had given evidence in a "fair minded fashion."[109] The judge found that the detective was qualified to serve as an expert witness and that any allegations of bias could be uncovered in cross-examination.

[104] *R v Alcantara*, 2012 ABQB 225, 538 AR 44.
[105] *R v Violette*, 2008 BCSC 920.
[106] *R v Baxter*, 2009 ONCJ 16.
[107] *R v Gager*, 2012 ONSC 388.
[108] *R v Gager*, 2012 ONSC 388 at para 189.
[109] *R v Gager*, 2012 ONSC 388 at para 194.

Best Practice:	There is a difference between an expert being an advocate for a position and an advocate for a party. Both raise concerns about the objectivity of the expert witness, but becoming an advocate for a party raises serious concerns about the impartiality of the expert. Experts with strongly held views should have those views challenged and tested. Experts who become advocates for a party have compromised their impartiality; their evidence should be excluded.

5. Stating Facts or Assumptions, and Considering all Material Facts

5.1. General Principles

It is not the province of the expert to find facts that will be 'suitable' to support a legal argument. The expert must simply set out the facts and the basis upon which these have been determined.[110] In an Ontario case, Justice Farley elaborated the nature of the problem:

> Cherry picking facts that support a diagnosis that just happens to support the cause of the client that retained the expert and failing to include the facts that hurt the cause, whether those latter facts are capable of explanation and elimination in the course of the development of the expert's analysis and opinion or not smacks of partiality.[111]

5.2. Examples / In the Courtroom

In one case, the expert based fair market value (FMV) calculations of certain assets on blatant guesses provided by the party who retained him. The expert acknowledged that for some adjustments there were no documents to support such adjustments. There thus seemed to be a lack of theoretical or empirical rationale for such adjustments. The court found that the expert had adopted the mindset that the opposing party was stealing from the company and that therefore the expert lost his objectivity. In approaching his task with the belief

[110] *Kirby Lowbed Services Ltd. v Bank of Nova Scotia*, 2003 BCSC 617.
[111] *Frazer v Haukioja* (2008), 58 CCLT (3d) 259 at para. 154

that the defendant had a plan to strip the plaintiff company of its assets, the expert merely went about verifying this belief. The court stated that the expert's "overall responses and approach were colored by his mind set and made his conclusions suspect."[112]

In another case, the expert had not seen a number of key documents before preparing his first report, and mischaracterized some that he did see. The expert based his opinion on incomplete information. When confronted with some of the relevant documents that he had not reviewed, the expert simply said that they would not have affected his opinion in any way.[113] While new data may legitimately have no impact on a particular scientific hypothesis, the question is whether new information ought to have reasonably affected an expert's opinion. In this case, the court found that it should have.

In a Manitoba case, the expert omitted consideration of material facts which would have affected his opinion. Specifically, the expert ignored the clear indications of the [Personality Assessment Inventory] and accepted without question the history provided by the plaintiff. The expert also ignored the narrative produced by the computer-generated program. The court held that the expert evidence was not sufficiently objective to assist the court in its pursuit of true facts; the expert failed to provide the court with an unbiased opinion. The court rejected the expert's evidence, stating: "In matters such as these it is vital that the Court have as much unbiased, objective assistance as possible from the expert witnesses."[114]

In a case involving conflicting medical testimony, the court expressly found that the expert had breached the specific *Ikarian Reefer* duty to consider all material facts. The court stated that the doctor had omitted consideration of material facts which should have affected his opinion. In the words of the court, the doctor "failed to provide this Court with an unbiased opinion. In matters such as these it is vital that the Court have as much unbiased, objective assistance as possible from the expert witnesses. In the circumstances, I reject Dr. Brodie's evidence."[115]

In a case referenced above involving a prosecution by the Financial Services Commission of Ontario, the expert completely removed what he had written in an earlier draft of his report about testing the sufficiency and reliability of the data. The expert also made a conscious decision not to set out the sources with whom he had spoken.[116]

---

[112] *Dun-Rite Plumbing & Heating Ltd. v Walbaum*, 2009 SKQB 174, 59 BLR (4th) 170 at para 563.
[113] *Eastern Power Ltd. v Ontario Electricity Financial Corp.*, [2008] OJ No 3722 at paras 307-8.
[114] *Posthumous v Foubert*, 2009 MBQB 206 at paras 86, 87, 89.
[115] *Teichgraber v Gallant*, 2003 ABQB 58 at para 89.
[116] *R v Norton*, 2007 ONCJ 414 at para 89.

6. Knowledge Outside Witness's Expertise

Justice Farley explained that the court "should not need to take the time to review the proposed evidence of any expert to determine whether the witness is qualified to offer the evidence. That is a function of the role of the expert. The court expects the expert to know his or her professional limitations and expects the expert to decline to speak to matters beyond them."[117] As with many ethical rules, this expectation is often found wanting.

7. Expert Candour, Need to Qualify Opinions and Change of Circumstances

Although he was in dissent in the case, the explanation provided by Justice Slatter of the Alberta Court of Appeal of the expert's general duty on this issue is nowhere challenged. He explained the general duty of expert candour as follows: "If the expert discovers an error in the report, or if the expert believes the report is being used for a purpose for which it is was never intended, or is unsuitable, the expert has an obligation to advise the court."[118] Justice Slatter explained that the expert cannot "simply sit back and run the risk that the court or other third parties may be misled by relying on the report. The expert is also entitled to protect its reputation by withdrawing reports that turn out to be flawed. The obligation and ability to withdraw a report, change an opinion, or correct an error in the report does not depend on the consent of the client; the client also has an obligation to ensure that the court is not misled by evidence it has filed."[119]

8. The Standard Expected of Various Experts

The standards of conduct expected of different types of experts do not vary substantially. Certainly, neither court rules nor the common law vary in the standards expected of different types of experts. Similarly, under the common law, judges do not apply different standards to experts in criminal as opposed to civil cases.

[117] *Frazer v Haukioja* (2008), 58 CCLT (3d) 259 at para 163.

[118] *Deloitte & Touche LLP v Institute of Chartered Accountants of Alberta*, 2008 ABCA 162, 433 AR 41 at para 96 (per Slatter J.A., dissenting).

[119] *Deloitte & Touche LLP v Institute of Chartered Accountants of Alberta*, 2008 ABCA 162.

However, an anomaly now exists in several Canadian jurisdictions where experts in civil cases are subject to explicit duties, whereas none exist in criminal proceedings. In the *Report of the Inquiry into Pediatric Forensic Pathology in Ontario*, the Honourable Stephen T. Goudge recommended that a code of conduct for experts giving evidence in criminal proceedings should be created, and stated that it would be anomalous if such a code applied for civil cases but not in criminal proceedings, given that the accused's liberty is at stake in the latter.[120] Commissioner Goudge recommended the creation of a particularized code of conduct for experts in criminal cases along the lines of that prepared by the Crown Prosecution Service in England and Wales in its *Guidance Booklet for Experts – Disclosure: Experts' Evidence, Case Management and Unused Material.*[121]

9. ACCOUNTABILITY

As set out above, the expert witness is subject to various ethical requirements. There are a number of possible means by which experts may be held accountable for their duties.

Some Rules may require an expert to complete an acknowledgment of the expert's duty. Ontario's Form 53, "Acknowledgment of Expert's Duty" (see *Appendix B – Ontario Form 53* at 216), for example, contains the following statements:

> *I acknowledge that it is my duty to provide evidence in relation to this proceeding as follows:*
>
> *(a) to provide opinion evidence that is fair, objective and non-partisan;*
>
> *(b) to provide opinion evidence that is related only to matters that are within my area of expertise; and*
>
> *(c) to provide such additional assistance as the court may reasonably require, to determine a matter in issue.*
>
> *I acknowledge that the duty referred to above prevails over any obligation which I may owe to any party by whom or on whose behalf I am engaged.*

[120] *Goudge Inquiry*, vol. II, (Toronto: Queen's Printer, 2008) at 505.

[121] Association of Chief Police Officers & Crown Prosecution Service, "Guidance Booklet for Experts Disclosure: Experts' Evidence, Case Management and Unused Material" (May 2010), online: <http://www.cps.gov.uk/legal/assets/uploads/files/Guidance_for_Experts_-_2010_edition.pdf>.

One expert explained the utility of the form as being "a mental jog to remind the expert to be objective. As we now have to put that down on paper, witnesses are starting to remind themselves and recognize that they need to be more objective."[122]

However, the value of such acknowledgement forms is limited. There is no requirement that the expert be advised of – or confirm – his or her duty *before* undertaking to provide an expert opinion. Concern has been expressed that, in Ontario, lawyers are not advising experts of the requirements of Form 53.[123] These forms may be signed after the expert completes his or her written report. In such cases, their utility may be reduced. There is no obligation that the court confirm the expert appreciated the nature of his or her duty. Further, the rules that impose such obligations on experts do not carry with them consequences if the expert is found to have been in non-compliance with the enumerated duties.[124]

Under the common law, expert witnesses have enjoyed absolute immunity from civil suit for any mistakes committed in the course of their testimony,[125] absent bad faith.[126] A 2011 decision of the U.K. Supreme Court[127] abolished this rule in that country, but it remains the common law of Canada.

Despite the fact that experts are said to owe their duties to the court, judges have been reticent to directly hold experts to account. While courts could, in theory, sanction experts in the same manner as they sanction counsel or parties, they have not done so.

A judge could initiate a complaint about an expert to the expert's professional body. Many judges might be hesitant to do so, although they could simply send a transcript of the judge's comments about the expert to the appropriate professional body. This seems to be shifting the focus from the expert's duty to the court to the expert's duty as a member of a particular profession – assuming the expert is a member of a regulated profession that imposes specific

---

[122] Dr. Michael Ford, "U.K. shows way forward on expert accountability" *Law Times* (22 August 2011) 7.

[123] See Michael McKiernan, "New rules on experts create headaches" *Law Times* (26 July 2010) 1, online: <http://www.lawtimesnews.com/201007261644/headline-news/new-rules-on-experts-create-headaches> [McKiernan].

[124] As a procedural matter, a court may reject an expert report that is not accompanied by the requisite form, but this does not address the underlying substantive issue of the failure of the expert to live up to their obligations.

[125] See, generally, Alan W. Mewett & Peter J. Sankoff, *Witnesses* (Toronto: Carswell, 1991), vol. 2, § 20.4. A very useful summary is found in *Elliott v Insurance Crime Prevention Bureau*, 2005 NSCA 115, 236 NSR (2d) 104. See, *e.g.*, *Pearlman v Critchley*, 2011 BCSC 1479; *Mund v Sovio*, 2010 BCSC 252, 82 CCLI (4th) 246 at paras 68-79; and *Carnahan v Coates* (1990), 47 BCLR (2d) 127 (SC).

[126] See *Carnahan v Coates* (1990), 47 BCLR (2d) 127 (SC) and *Lower v Stasiuk*, 2012 BCSC 1087, 36 BCLR (5th) 397 (a claim for abuse of process against an expert witness may be a viable claim, if the expert knowingly engages in conduct that could support such a claim).

[127] *Jones v Kaney*, [2011] UKSC 13, [2011] 2 All ER 671.

requirements on the professional acting as an expert witness and that those obligations are consistent with the statutory and common law requirements. For reasons explained below by one expert, professional sanctions may be less effective than direct sanctions by the court.

The court's statements about the expert may constitute a form of indirect accountability. It is widely recognized that negative statements by the court about an expert's testimony may impact the expert's reputation and his or her future employability as an expert. One expert explained as follows:

> When the court gives you a slap on the wrist, your livelihood will be going on a bit of sabbatical. It can take years to build back up from something like that. I've seen an expert boutique go from 15 to 20 people down to about four after a critical decision from a judge. That will not happen when it's just the professional body doing it.[128]

For these reasons, a judge should be temperate in his or her comments about an expert witness. The judge must distinguish between his or her own dislike of an expert witness and actions of the expert which warrant critical comments from the bench. However, there may be circumstances where strong statements from the bench are entirely warranted. This may have been the case in ***Jayetileke v Blake***,[129] wherein the trial judge said as follows of the expert witness:

> [35] Dr. Davis had a history before the courts where his evidence was rejected and his objectivity called into question: ***Grewal v Brar et al***, 2004 BCSC 1157, [2004] BCJ No 1819; ***Gosal v Singh***, 2009 BCSC 1471, [2009] BCJ No 2131; ***Kelly v Sanmugathas***, 2009 BCSC 958, [2009] BCJ No 1413; and ***Smusz v Wolfe Chevrolet***, 2010 BCSC 82, [2010] BCJ No 114.
>
> [36] A witness may have a poor day in court – that does not mean the witness was dishonest or forever unreliable. However, Dr. Davis had displayed an alarming inability to appreciate his role as an expert and the accompanying privilege to provide opinion evidence.
>
> [37] The defence was alive to his propensity to abuse the role of an expert. His reputation would have been known from the cited decisions.

Another indirect form of accountability would be for courts to begin to regulate experts indirectly, by including in the duties of lawyers as officers of the court – and the primary gatekeepers to experts – that they inform experts of their duties and ensure compliance is observed. Many cases chastise experts for

---

[128] *McKiernan*, "New rules on experts create headaches" *Law Times* (26 July 2010) 1, online: <http://www.lawtimesnews.com/201007261644/headline-news/new-rules-on-experts-create-headaches>.

[129] *Jayetileke v Blake*, 2010 BCSC 1478, 12 BCLR (5th) 388.

their inappropriate interactions with counsel without imposing any sanctions on the counsel or parties involved in the case.

The New South Wales Law Reform Commission suggests additional possible sanctions against an unprofessional expert witness:

- The court might make a cost order against the expert witness.
- The expert witness might be charged with contempt or perjury.[130]

10. Red Flags: Indicia of Lack of Objectivity / Impartiality

Based on the examples cited in the previous sections, the following should be considered "red flags" or indicia of a lack of objectivity or a lack of impartiality by the expert witness. They might not demonstrate bias or a lack of objectivity in and of themselves, but they should cause the judge to pause and inquire further into the impartiality of the expert:

- The expert is an employee of the party or otherwise has a direct stake in the outcome of the litigation (*e.g.*, contingency fee).
- The expert has testified for the party in the past.
- The expert has frequently testified for the same position.
- The expert is a close personal friend of the party that retained expert.
- There has been frequent interaction between the expert and counsel/the party, indicative of a collaborative process in preparation of the expert report.
- Counsel/the party has written portions of the expert's report or portions are taken from the party's own materials.
- The expert's report contains passages or paraphrases portions of a brief filed by the party in another proceeding.

[130] NSW Law Reform Commission, *Expert Witness Report*, "The 'No Win, No Pay' Expert Witness (Contingency Fee Expert Witnesses)," in *Report 109: Expert Witnesses* (Sydney: NSW Law Reform Commission, 2005) at 162.

11. CHECKLIST FOR JUDGES

The following are a list of questions the judge may pose to counsel or to the expert. Many judges may find the suggestion of "cross-examining" the expert witness foreign or inappropriate.[131] However, I would suggest that the situation of expert witnesses is different because the expert owes a duty to the court, and in many jurisdictions that duty is now set out in legislation. For these reasons, I believe it is entirely appropriate for the judge to ensure that the expert understands and has complied with the expert's duty to the court.

In reviewing a draft of this chapter, one judge suggested that this checklist could be used in a pre-trial case management conference with counsel, with the judge explaining that counsel would be expected to address these issues in counsel's examination of the expert. If counsel fails to do so, opposing counsel may pick up on the suggestion and address these questions in cross-examination. The following are suggested questions for the expert:

Are you aware of your duty as an expert?

This question seeks to ensure that the expert understands that the expert's duty is to the court above all else. The expert is responsible to the court and not to the party who retained the expert.

Do you understand the nature of this duty?

This question seeks to ensure that the expert understands the content of the expert's duty to the court; not merely the mere existence of a duty. This duty includes objectivity, impartiality, fairness and completeness.

How were you made aware of this duty?

It is critical that experts be aware of this duty at the time they are retained to produce an expert opinion. The expert becoming aware of this duty after he or she has completed his or her report, or during preparation for testimony in court, is too late.

[131] Generally, appellate courts do not look kindly on excessive intervention by trial judges in the questioning of witnesses. See, *e.g.*, *R v Huang*, 2013 ONCA 240, 115 OR (3d) 596 .

When did you receive and review the code of conduct / Form? [If Rules of Court contain elements of a code of conduct for experts or there is a form to attach to their report – *e.g.*, Ontario Form 53.]

It is problematic if an expert does not receive and review the relevant information prior to beginning his or her work on a matter.

What instructions did you receive from counsel at the time you were first approached on this file?

This question is critical in order to ascertain the basis upon which the expert conducted his or her analysis, as well as how the expert's mandate was 'framed'. Did the instructions from counsel include information that might lead the expert to misapprehend his or her role, specifically with regard to being independent and objective? Did the instructions encourage the expert to become an advocate for the party?

What information did you receive from counsel (or the party) in preparation for undertaking your expert report?

This is necessary in order to understand the factual foundation for the expert's report.

Did you provide counsel with updates of your work? Did counsel have the opportunity to comment or contribute to your work-in-progress?

This and subsequent questions address the independent nature of the expert's work and attempt to ascertain how much of the expert's report is his or her own work and how much has been contributed by counsel.

Did you provide counsel (or the party) with draft reports or portions of drafts for comment?

Per above.

Did you discuss your work with counsel (on the phone, in person, e-mail, etc.)?

Per above.

Did you need to seek clarification from counsel (or the party) on any matter prior to submitting your report?

Per above.

How many times have you testified for [party] in the past?

This and subsequent questions go to the notion of the expert as a 'hired gun' and the extent to which the expert considers and fairly evaluates all relevant information in reaching his or her conclusions.

How many times have you testified for [position] in the past?

Per above.

Have you ever testified [against this position]?

Per above.

How are you being remunerated for your participation in the case?

This question goes to the extent of the expert's stake in the case which may impugn his or her impartiality.

12. SUGGESTED BEST PRACTICES

- Courts that do not have codes of conduct for experts should consider adopting codes as Practice Directions or Rules of Court.
- Courts with codes of conduct for experts/specific rules for experts should consider issuing Practice Directions requiring counsel to provide and explain such codes/rules to potential experts at the time they are first contacted about the potential for being retained as an expert.
- In the absence of a Practice Direction, the responsible judge or master should require all counsel to provide potential experts with all relevant information regarding their duties as expert witnesses. Courts without a relevant rule, code of conduct or Practice Direction may fashion a draft order or information sheet based on the common law duties of the expert (*Ikarian Reefer*).

IN THE COURTROOM

Smith v. Jones, Civil File # 0001, Superior Court

Sam Smith is suing Jessica Jones for injuries sustained in a motor vehicle accident. The trial is a bench trial and was bifurcated into liability and damages. The liability portion of the trial has been completed and it has now moved into the damages phase. Smith has claims for medical costs, lost wages, physiotherapy, and for pain and suffering. By far the most significant part of Smith's damages claim is for pain and suffering related to alleged chronic pain. Smith also alleges that he has been unable to return to work due to chronic pain. The defendant Jones' insurer questions the veracity of Smith's chronic pain claims and has hired a private investigator in an attempt to ascertain whether Smith's chronic pain claims are exaggerated or even fraudulent. Counsel for the defendant Jones led evidence from the private investigator during the damages phase of the trial. The private investigator's evidence is mixed and consists of some social media postings from Smith's Facebook page showing him having a good time in an apparent pain-free environment, as well as photos and video from Smith attending a party and some images of Smith raking leaves. Cross-examination of the private investigator by Smith's counsel reveals that the private investigator spent 20 hours investigating Smith online and over 50 hours observing Smith at his home and in public, and that these three incidents were the best indicators the private investigator had that Smith might be exaggerating his pain claims. Under cross-examination, the private investigator admitted that he observed Smith in public on numerous occasions apparently experiencing severe pain.

Counsel for Smith has produced an expert report from Dr. Solange Sotomineur. Dr. Sotomineur is a scientist who is a partner in Veritas Forensic Sciences Inc. (VFS), a for-profit company that provides functional magnetic resonance imaging (fMRI) and other scientific evaluations for use in court and other legal proceedings. fMRI services are VFS' fastest growing business. Dr. Sotomineur is an MD-PhD-MBA with degrees in neurology and neuroscience. She explains in her report that fMRI measures the hemodynamic response (*i.e.*, changes in blood oxygen) related to neural activity. The fMRI is considered a lie-detector test superior to the polygraph. fMRI is based on the same technology as magnetic resonance imaging (MRI) – a non-invasive test that uses a strong magnetic field and radio waves to create detailed images of the body. But instead of creating images of organs and tissues, fMRI looks at blood flow in the brain to detect areas of activity.

Areas of the brain that are active use more blood and, as a result, show up brighter when imaged. As different tasks are performed, blood flows to different parts of the brain, similar to how blood flows to muscles in motion. fMRI reveals the different parts of the brain people use when performing simple tasks. It can isolate the brain regions responsible for motor activity, perceptual activity (*e.g.*, hearing or seeing), and cognitive activity resulting from complex thought processes. Computer software generates a map of neural activity and the region of activity is then associated with specific cognitive functions.

fMRI scans can detect if a person is lying because different brain regions are activated when lying as opposed to being truthful. Studies have demonstrated that there is a neuro-physiological difference between deception and truth that is detectable by an fMRI scan. Lying activates brain areas associated with high-level executive functions since lying requires the suppression of a truthful response as well as the creation of a fictitious narrative.

Counsel for Smith sent his client to Dr. Sotomineur to conduct an fMRI to determine whether he was telling the truth about suffering chronic pain. This evidence is considered critical to rebut the evidence of the private investigator and the general assertion that is being presented by counsel for Jones that Smith is exaggerating his damages. Dr. Sotomineur's report concludes that Smith believes he is experiencing chronic pain. She testifies that there is no indication from the fMRI that Smith is lying.

Under cross-examination, Dr. Sotomineur reveals that VFS is a relatively new company that has only been in existence for 18 months. She admits that the company is actually in the growth phase and in need of outside financing. Acceptance of fMRI evidence in court is critical to the company's business plan and to securing external financing. At this point, the VFS is not making a profit and Dr. Sotomineur is drawing only a modest salary from the company. She actually earns three times her salary in expert witness fees.

The trial judge has determined that the expert is qualified and that the expert's testimony is admissible. However, when the judge reads the expert's report she suspects that there has been significant collaboration between counsel and Dr. Sotomineur. However, opposing counsel fails to cross-examine on this point. Additionally, the judge is not sure that Dr. Sotomineur completely appreciated her

role as an expert. While her expert report contained an acknowledgement of her role as an expert along the lines of Ontario's Form 53 (see *Appendix B – Ontario Form 53* at 216), the judge has no idea when Dr. Sotomineur signed the form, let alone whether she understood and appreciated the contents of it.

Questions

Assume that the judge has already determined that the expert is qualified to give evidence about fMRI.

What should the presiding judge do?

Should the judge ask counsel anything?

Should the judge ask the expert any questions? If so, what should the judge ask the expert?

Review and apply the Red Flags: Indicia of Lack of Objectivity / Impartiality *contained in section 10 and the* Checklist for Judges *in section 11*.

13. Appendix A – Federal Court Rules and Expert Code of Conduct

***Federal Courts Rules*, SOR/98-106**

Enabling Statute: Federal Courts Act

Expert Witnesses

Right to name expert

52.1 (1) A party to a proceeding may name an expert witness whether or not an assessor has been called on under rule 52.

Expert named jointly

(2) Two or more of the parties may jointly name an expert witness.

Expert's affidavit or statement

52.2 (1) An affidavit or statement of an expert witness shall

(*a*) set out in full the proposed evidence of the expert;

(*b*) set out the expert's qualifications and the areas in respect of which it is proposed that he or she be qualified as an expert;

(*c*) be accompanied by a certificate in Form 52.2 signed by the expert acknowledging that the expert has read the Code of Conduct for Expert Witnesses set out in the schedule and agrees to be bound by it; and

(*d*) in the case of a statement, be in writing, signed by the expert and accompanied by a solicitor's certificate.

Failure to comply

(2) If an expert fails to comply with the Code of Conduct for Expert Witnesses, the Court may exclude some or all of the expert's affidavit or statement.

SCHEDULE

(Rule 52.2)

CODE OF CONDUCT FOR EXPERT WITNESSES

General Duty to the Court

1. An expert witness named to provide a report for use as evidence, or to testify in a proceeding, has an overriding duty to assist the Court impartially on matters relevant to his or her area of expertise.
2. This duty overrides any duty to a party to the proceeding, including the person retaining the expert witness. An expert is to be independent and objective. An expert is not an advocate for a party.

Experts' Reports

3. An expert's report submitted as an affidavit or statement referred to in rule 52.2 of the *Federal Courts Rules* shall include:
    a. a statement of the issues addressed in the report;
    b. a description of the qualifications of the expert on the issues addressed in the report;
    c. the expert's current *curriculum vitae* attached to the report as a schedule;
    d. the facts and assumptions on which the opinions in the report are based; in that regard, a letter of instructions, if any, may be attached to the report as a schedule;
    e. a summary of the opinions expressed;
    f. in the case of a report that is provided in response to another expert's report, an indication of the points of agreement and of disagreement with the other expert's opinions;
    g. the reasons for each opinion expressed;
    h. any literature or other materials specifically relied on in support of the opinions;
    i. a summary of the methodology used, including any examinations, tests or other investigations on which the expert has relied, including details of the qualifications of the person who carried them out, and whether a representative of any other party was present;
    j. any caveats or qualifications necessary to render the report complete and accurate, including those relating to any insufficiency of data or research and an indication of any matters that fall outside the expert's field of expertise; and
    k. particulars of any aspect of the expert's relationship with a party to the proceeding or the subject matter of his or her proposed evidence that might affect his or her duty to the Court.

4. An expert witness must report without delay to persons in receipt of the report any material changes affecting the expert's qualifications or the opinions expressed or the data contained in the report.

Expert Conferences

5. An expert witness who is ordered by the Court to confer with another expert witness:
    a. must exercise independent, impartial and objective judgment on the issues addressed; and
    b. must endeavour to clarify with the other expert witness the points on which they agree and the points on which their views differ.

14. APPENDIX B – ONTARIO FORM 53

FORM 53

Courts of Justice Act

ACKNOWLEDGMENT OF EXPERT'S DUTY

(General heading)

ACKNOWLEDGMENT OF EXPERT'S DUTY

1. My name is ..... *(name)*. I live at ..... *(city)*, in the ..... *(province/state)* of ..... *(name of province/state)*.
2. I have been engaged by or on behalf of ..... *(name of party/parties)* to provide evidence in relation to the above-noted court proceeding.
3. I acknowledge that it is my duty to provide evidence in relation to this proceeding as follows:
    a. to provide opinion evidence that is fair, objective and non-partisan;
    b. to provide opinion evidence that is related only to matters that are within my area of expertise; and
    c. to provide such additional assistance as the court may reasonably require, to determine a matter in issue.
4. I acknowledge that the duty referred to above prevails over any obligation which I may owe to any party by whom or on whose behalf I am engaged.

Date .....

(*Signature*)

NOTE: This form must be attached to any report signed by the expert and provided for the purposes of subrule 53.03(1) or (2) of the *Rules of Civil Procedure*.

RCP-E 53 (November 1, 2008)

15. APPENDIX C – CODE OF PRACTICE FOR EXPERTS (ENGLAND AND WALES)

Preamble

This Code of Practice shows minimum standards of practice that should be maintained by all Experts.

It is recognised that there are different systems of law and many jurisdictions in Europe, any of which may impose additional duties and responsibilities which must be complied with by the Expert.

There are in addition to the Code of Practice, General Professional Principles with which an Expert should comply.

These include the Expert:

- Being a " fit and proper" person
- Having and maintaining a high standard of technical knowledge and practical experience in their professional field
- Keeping their knowledge up to date both in their expertise and as Experts and undertaking appropriate continuing professional developments and training.

The Code

1. Experts shall not do anything in the course of practising as an Expert, in any manner which compromises or impairs or is likely to compromise or impair any of the following:
    a. the Expert's independence, impartiality, objectivity and integrity,
    b. the Expert's duty to the Court or Tribunal,
    c. the good repute of the Expert or of Experts generally,
    d. the Expert's proper standard of work,
    e. the Expert's duty to maintain confidentiality.
2. An Expert who is retained or employed in any contentious proceeding shall not enter into any arrangement which could compromise his impartiality nor make his fee dependent on the outcome of the case nor should he accept any benefits other than his fee and expenses.
3. An Expert should not accept instructions in any matter where there is an actual or potential conflict of interests. Notwithstanding this rule, if full disclosure is made to the judge or to those appointing him, the Expert may in appropriate cases accept instructions when those concerned specifically acknowledge the disclosure. Should an actual or potential conflict occur after instructions have been accepted, the

Expert shall immediately notify all concerned and in appropriate cases resign his appointment.

4. An Expert shall for the protection of his client maintain with a reputable insurer proper insurance for an adequate indemnity
5. Experts shall not publicise their practices in any manner which may reasonably be regarded as being in bad taste. Publicity must not be inaccurate or misleading in any way.
6. An Expert shall comply with all appropriate *Codes of Practice and Guidelines*.

16. APPENDIX D – SELECTED CODES OF CONDUCT FOR VARIOUS EXPERTS

16.1.1. Canadian Medical Association, *Code of Ethics*, (2004):

http://policybase.cma.ca/dbtw-wpd/PolicyPDF/PD04-06.pdf

(*Note that the CMA's *Code of Ethics* has been adopted by the Colleges of Physicians and Surgeons of British Columbia, Alberta, Nova Scotia, Newfoundland & Labrador and Prince Edward Island).

> 45. Recognize a responsibility to give generally held opinions of the profession when interpreting scientific knowledge to the public; when presenting an opinion that is contrary to the generally held opinion of the profession, so indicate.

16.1.2. College of Physicians and Surgeons of Saskatchewan, *Code of Ethics*

http://www.cps.sk.ca/imis/CPSS/Legislation__ByLaws__Policies_and_Guidelines/Legislation_and_Bylaws.aspx?Legislation_BylawsCCO=3

(analogous provision to paragraph 45 of the CMA *Code of Ethics*).

16.1.3. College of Physicians and Surgeons of Manitoba, *Code of Conduct*

http://cpsm.mb.ca/cjj39alckF30a/wp-content/uploads/Code-of-Conduct-2010.pdf

(analogous provision found at paragraph 37).

16.1.4. College of Physicians and Surgeons of New Brunswick, *Code of Ethics*

http://www.cpsnb.org/english/code-of-ethics.html

(analogous provision found at paragraph 45).

16.1.5. College of Physicians and Surgeons of Ontario, Policy Statement #7-12: Medical Expert: Reports and Testimony (2012)

http://www.cpso.on.ca/uploadedFiles/policies/policies/policyitems/Medical-Records.pdf

16.1.6. Collège des médecins de Québec, *Code of Ethics of Physicians :*

http://www.cmq.org/publications-pdf/p-6-2015-01-07-en-code-de-deontologie-des-medecins.pdf

s. 67 (duty of impartiality and objectivity in acting on behalf of a third party),

s. 89 (communicating medical opinions)

16.1.7. Canadian Psychological Association, *Code of Ethics for Psychologists*, 3rd ed. (2000), Accuracy/Honesty (III.1-9), Objectivity/Lack of Bias (III.10-13)

http://www.cpa.ca/docs/File/Ethics/cpa_code_2000_eng_jp_jan2014.pdf

16.1.8. Canadian Society of Forensic Science, *Rules of Professional Conduct* (Ottawa, Ontario: 1994) (rules 7-12 re methods, disclosure, use of notes, rendering opinions, testing and testifying)

http://www.csfs.ca/become-a-member/code-of-conduct/

> Members of the Canadian Society of Forensic Science, with respect to their responsibilities to the C.S.F.S., shall:
>
> 6. take reasonable steps to ensure that all items in a case receive appropriate technical analysis;
>
> 7.a. utilise methods, techniques, standards and controls, provided that they exist, that they are generally accepted and that they are current and;
>
> b. utilise methods and techniques with standards and controls to conduct examinations and analysis such that they could be reproduced by another qualified and competent person;
>
> 8. make full and complete disclosure as required by law of the findings to the submitting agency or client;

9. make and keep worknotes on all items received, the examinations done, the results obtained and the findings and conclusions made in a timely fashion;

10. render opinions and conclusions strictly in accordance with the results and findings in the case and only to the extent justified by those results and findings;

11. make all efforts to testify in a clear, straightforward manner and refuse to extend themselves beyond their field of expertise or level of competence;

12. not exaggerate, embellish or otherwise misrepresent qualifications when testifying;

13. be impartial and independent in their analysis, reporting and testimony;

16.1.9. American College of Cardiology, *Code of Ethics*

http://www.acc.org/about-acc/our-bylaws-and-code-of-ethics/code-of-ethics

(specific rules re expert witness testimony in Rule 6)

16.1.10. American Academy of Neurology, *Code of Professional Conduct*

https://www.aan.com/uploadedFiles/Website_Library_Assets/Documents/8.Membership/5.Ethics/1.Code_of_Conduct/Membership-Ethics-American%20Academy%20of%20Neurology%20Code%20of%20Professional%20Conduct%20%282%29.pdf

(rules re expert witness 6.4)

16.1.11. Professional Engineers of Ontario Guideline: The Professional Engineer as an Expert Witness (2011)

http://www.peo.on.ca/index.php/ci_id/22088/la_id/1.htm

16.1.12. Association of Professional Engineers and Geoscientists of Alberta *Guideline for Professional Member as a Witness*, V1.0 (October 2003)

https://www.apega.ca/assets/PDFs/witness.pdf

NATIONAL JUDICIAL INSTITUTE
INSTITUT NATIONAL DE LA MAGISTRATURE

275 Slater Street, Suite 2000
Ottawa, Ontario K1P 5H9
TEL.: 613-237-1118 FAX.: 613-237-6155
www.nji-inm.ca