

GAITKEEPING IN CANADA: MIS-STEPS IN ASSESSING THE RELIABILITY OF EXPERT TESTIMONY

Emma Cunliffe* and Gary Edmond†

Incriminating expert testimony is a leading cause of wrongful convictions. Academic commentators and authoritative scientific research bodies agree that forensic comparison techniques such as fingerprint identification, toolmark comparison and bite mark analysis should be subjected to validation studies and experts subjected to proficiency testing prior to expert evidence being admitted in criminal trials. Canadian case law on the admissibility of expert testimony increasingly emphasizes demonstrable reliability as a condition of admission. In this article, we critically assess the BC courts' approach to reliability in R v Aitken. In R v Aitken, "forensic gait analysis" was offered for the first time in a Canadian courtroom. We suggest that the growing judicial attention to reliability is heartening, but that Canadian judges and lawyers have not yet developed the tools necessary to conduct a sound assessment of the reliability of incriminating expert testimony. The authors draw on authoritative research and policy reports to offer suggestions about how to improve the Canadian judicial approach to assessing reliability.

Les témoignages incriminants des experts sont l'une des principales causes d'erreurs judiciaires. Les auteurs et les organismes scientifiques faisant autorité sont unanimes pour souligner que les techniques judiciaires de comparaison, telles que l'identification des empreintes digitales, la comparaison des traces d'outils et les analyses des empreintes dentaires devraient faire l'objet d'études de validation et que les experts devraient être soumis à une vérification de leurs compétences avant que leurs témoignages ou leurs rapports ne soient admis dans les procès criminels. La jurisprudence canadienne sur l'admissibilité de la preuve d'expert insiste d'ailleurs de plus en plus sur la démonstration de leur

* Faculty of Law, University of British Columbia. An earlier version was presented in the UBC Faculty of Law faculty seminar series. We are grateful to the BC Court of Appeal registry for facilitating access to the court records on which much of this research is based, and to the UBC Faculty of Law and Peter Wall Institute for Advanced Studies for supplying funds in support. Thanks also to the anonymous reviewers.

† Professor and Director, Expertise, Evidence & Law Program, School of Law, The University of New South Wales; Professor (fractional), School of Law, Northumbria University. This research was supported by the ARC (FT0992041 and LP120100063). This paper was written while Edmond was a visitor at the Peter Wall Institute for Advanced Studies.

fiabilité comme condition d'admissibilité. Dans le présent article, les auteurs à un examen critique de l'approche adoptée par la Cour d'appel de la Colombie-Britannique pour évaluer la fiabilité dans l'arrêt R v Aitken dans laquelle l'« analyse judiciaire de la démarche » a été utilisée pour la première fois par un tribunal canadien. À cet égard, les auteurs sont d'avis que l'attention grandissante portée par les tribunaux à l'égard de la fiabilité s'avère un signe encourageant, mais que les juges et les avocats canadiens n'ont pas encore développé les outils permettant de procéder à une juste évaluation de la fiabilité des témoignages incriminants des experts. Pour étayer leur position, ils s'inspirent de travaux de recherche et de rapports de politique et formulent des suggestions quant à la manière d'améliorer l'approche appliquée par les tribunaux canadiens en matière d'évaluation de la fiabilité.

1. Introduction

Recent years have seen a growing recognition that incriminating expert testimony is a leading cause of wrongful convictions¹ and the realization that some forensic sciences – including identification by fingerprints, toolmark comparison and bite mark analysis – have not been subjected to validation studies, proficiency testing, or other forms of reliability assessment.² For the purposes of this article, we adopt a definition of validation proposed by analysts working at the Federal Bureau of Investigation (FBI) laboratory in Quantico:

Validation is defined as the process to assess the ability of defined procedures to reliably obtain results, to define conditions that are required to obtain the result, to determine the limitations of the analytical procedure, and to identify aspects that must be monitored and controlled. Validated methods are essential to the forensic sciences, are inherent in providing quality, and provide stability to continuously evolving scientific fields.³

¹ See e.g. Stephen T Goudge, *Final Report of the Inquiry into Ontario Pediatric Forensic Pathology Services* (Toronto: Queen's Printer, 2008); Samuel Gross and Michael Schaeffer, *Exonerations in the United States 1989 to 2012: Report by the National Registry of Exonerations*, online: <http://www.law.umich.edu/special/exoneration/Documents/exonerations_us_1989_2012_full_report.pdf>.

² National Research Council, *Strengthening Forensic Science in the United States* (Washington, DC: National Academies Press, 2009) [NRC Report]. Since the publication of this report, some fields have begun to conduct validation studies and proficiency assessments – for example, in relation to fingerprinting, see Matthew B Thompson, Jason M Tangen and Duncan J McCarthy, “Human Matching Performance of Genuine Crime Scene Latent Fingerprints” (2013) 38 *Law & Human Behavior* 84.

³ Bruce Budowle *et al*, “A Perspective on Errors, Bias and Interpretation in the Forensic Sciences and Direction for Continuing Advancement” (2009) 54 *J Forensic Sci* 798 at 807.

Whereas validation refers to the accuracy of a forensic method, proficiency testing focuses on the capacity of a given analyst or laboratory to perform that method.

The blunt conclusion drawn by the National Research Council (NRC) of the US National Academies of Science (NAS) is that:

In a number of forensic science disciplines, forensic science professionals have yet to establish the validity of their approach or the accuracy of their conclusions, and the courts have been utterly ineffective in addressing this problem.⁴

The failings of forensic science are methodological, in the sense that the procedures and research basis for much forensic science do not support the claims made by experts in criminal cases. Importantly, the adversarial process has failed to identify the vulnerabilities of expert testimony or to prevent wrongful convictions. One leading commentator has written, in a passage cited in the NRC Report:

For years in the forensic science community, the dominant argument against regulating experts was that every time a forensic scientist steps into a courtroom, his work is vigorously peer reviewed and scrutinized by opposing counsel. A forensic scientist might occasionally make an error in the crime laboratory, but the crucible of courtroom cross-examination would expose it at trial. This “crucible,” however, turned out to be utterly ineffective.⁵

Neufeld’s observations regarding the failures of cross-examination to ensure the reliability of expert testimony have been echoed in many common law jurisdictions.⁶

In response to incidents such as a large number of wrongful convictions for child homicide offences in Ontario, Canadian judges have

⁴ NRC Report, *supra* note 2 at 53.

⁵ Peter J Neufeld, “The (Near) Irrelevance of *Daubert* to Criminal Justice and Some Suggestions for Reform” (2005) 95 Am J Pub Health S107 at S109, cited in the NRC Report, *ibid* at 106-77. See also Gary Edmond and Mehera San Roque, “The Cool Crucible: Forensic Science and the Frailty of the Criminal Trial” (2012) 24 Curr Issues in Crim J 51.

⁶ See e.g. The Law Commission, *Expert Evidence in Criminal Proceedings in England and Wales* (London: Report No 235, March 2011), online: <<http://www.justice.gov.uk/lawcommission/expert-evidence-in-%20criminal-trials.html>>; Edmond and San Roque, *ibid*; Emma Cunliffe, “Independence, Reliability and Expert Testimony in Criminal Trials” (2013) 45 Austl J Forensic Sci 284; Gary Edmond, Simon Cole, Emma Cunliffe and Andrew Roberts, “Admissibility Compared: The Reception of Incriminating Expert Evidence (i.e., Forensic Science) in Four Adversarial Jurisdictions” (2013) 3 U Denver Crim L Rev 31.

also begun to emphasize the importance of conducting a thorough reliability assessment when considering the admissibility of inculpatory expert testimony. In 2007, the Supreme Court of Canada held in *R v Trochym* that reliability is a principle relevant to admissibility, even in relation to relatively well-established expert testimony.

... even if it has received judicial recognition in the past, a technique or science whose underlying assumptions are challenged should not be admitted in evidence without first confirming the validity of those assumptions.⁷

In *Trochym*, the Court excluded evidence of post-hypnosis memories on the basis that it is not possible to tell whether the information elicited by hypnosis is accurate or inaccurate. In reaching this conclusion, Charron J applied the *Daubert* criteria for assessing the reliability of scientific evidence.⁸ These criteria, discussed further in Part 2 of this paper, permit a court to consider whether a given technique has been appropriately validated, as well as directing attention to the state of knowledge within the field from which the technique originates. Accordingly, they are consistent with the recommendations of the NRC Report.

In his 2008 *Report of the Inquiry into Pediatric Forensic Pathology in Ontario*, Goudge JA described reliability as “a fundamental organizing principle in the law of evidence.” After reviewing the leading Supreme Court of Canada cases regarding the admissibility of expert testimony,⁹ Goudge JA quoted with approval Finlayson JA’s observation that:

it is important that the trial judge serve as a gatekeeper and allow into evidence opinion evidence that is reliable and furthers the goal of accurate fact-finding while at the same time refusing to admit evidence that is irrelevant or prejudicial or not based on an adequate scientific foundation.¹⁰

Goudge JA suggested that the reliability of expert evidence should be a “constant concern” of judges in their gatekeeping role when managing a criminal trial.¹¹ His report concludes that existing case law requires trial judges to engage in a critical methodological analysis of expert testimony when the reliability of that testimony is challenged, regardless of whether similar testimony has been admitted in previous cases.¹² In relation to

⁷ 2007 SCC 6, 1 SCR 239 at para 32 [*Trochym*].

⁸ *Ibid* at paras 36-54.

⁹ Goudge, *supra* note 1 at 475-84.

¹⁰ Finlayson, as cited in Goudge, *supra* note 1 at 478. Edmond was an adviser to the Goudge Inquiry and appeared before the Commissioner.

¹¹ *Ibid* at 479.

¹² *Ibid* at 486.

emerging fields or techniques, Goudge JA suggested that it was especially important for a trial judge to have regard to “whether sufficient material exists, either from the proffered expert or from competing experts, to understand the relevant controversies and frailties that may surround the scientific evidence.”¹³

Most recently, in the 2009 Ontario Court of Appeal decision *R v Abbey*,¹⁴ Doherty JA offered a reformulated version of the test for admissibility of expert evidence that was originally set out by the Supreme Court of Canada in *R v Mohan*.¹⁵ Doherty JA’s reformulated test offers a clear place for a reliability analysis. *Abbey* also warns trial judges against adopting a boilerplate approach to the costs and benefits of expert testimony. Rather, the reliability and dangers associated with a specific type of expertise, and the actual opinion offered, should guide the assessment of admissibility in the specific case.¹⁶ *Trochym* and *Abbey* require trial judges to engage critically with the methodology adopted by an expert witness when assessing threshold reliability. Such engagement requires trial judges and lawyers to actively consider how best to assess the validity of a given expert technique, rather than simply acceding to an expert’s testimony about reliability. The NRC Report supplies an authoritative set of guidelines for considering the validity of forensic comparison techniques. Accordingly, it is consistent with *Trochym* and *Abbey* for lawyers and trial judges to have regard to these guidelines whenever the reliability of forensic comparison evidence is fairly raised in a Canadian courtroom.

Abbey has proven to be an important decision, and the approach suggested by Doherty JA was quickly adopted, or cited with approval, by courts of appeal in other Canadian provinces.¹⁷ In this article, we will focus on the discussion of reliability, including the use of *Abbey*, in one such case, the BC Supreme Court trial of Daniel Aitken and his subsequent appeal to the BC Court of Appeal.¹⁸ Drawing on an analysis of the court records, including trial transcripts and stills of the CCTV video at the heart of the case, we suggest that while the growing judicial understanding of the importance of reliability is heartening, Canadian judges have not yet

¹³ *Ibid* at 489.

¹⁴ 2009 ONCA 624, 97 OR (3d) 330 [*Abbey*].

¹⁵ [1994] 2 SCR 239 [*Mohan*].

¹⁶ *Abbey*, *supra* note 14 at paras 108, 119.

¹⁷ For example, *R v Lee*, 2010 ABCA 1, 23 Alta LR (5th) 76; *R v Perrault*, 2013 QCCA 834, [2013] JQ No 4536; *R v Aitken*, 2012 BCCA 134, BCAC 125 [*Aitken CA*]; *Brake-Pattern v Gallant*, 2012 NLCA 23, 321 Nfld & PEIR 77; *LMM v Nova Scotia (Attorney General)*, 2011 NSCA 48, 303 NSR (2d) 243.

¹⁸ *Aitken CA*, *ibid*.

developed the tools necessary to conduct a sound assessment of the reliability of inculpatory expert evidence. In particular, while Doherty JA's observation that differing types of expert evidence should be assessed according to different reliability metrics is basically sound, it is our contention that the BC Court of Appeal's reliance on the flexibility potentially afforded by this principle in *Aitken* led the Court into error. *R v Aitken* addressed a new form of expert testimony – “forensic gait analysis” offered by a clinical podiatrist.¹⁹ The judicial willingness to classify “forensic gait analysis” as non-scientific and non-novel permitted an overly accommodating approach to admissibility. The characterization of forensic gait analysis as a matter of clinical experience rather than scientific practice simultaneously facilitated departure from indicia of reliability that were appropriate for this (and many other) species of forensic comparison evidence.²⁰

The challenges encountered by the BC courts in the course of this case engage more general principles about the relationship between science and experience, the capacity of criminal courts to judge the quality of claims that originate from non-legal sources, the role of precedent in determining admissibility, and the extent to which the traditional tools of adversarial justice permit defence counsel to effectively challenge the claims of seemingly well-credentialed prosecution experts. While we are necessarily somewhat critical of the expert testimony admitted in *Aitken*, this article is written with the constructive intention to demonstrate that having regard to the guidelines set out in the NRC Report may assist courts in future cases to critically engage with expert methodologies and thereby discharge their responsibility to ensure the threshold reliability of expert testimony.

2. *R v Aitken in the BC Courts*

Daniel Aitken was convicted of first-degree murder in January 2009 in relation to the death of Adan Merino. Aitken's trial lasted 41 days before a jury, and was preceded by 40 days of pre-trial argument.²¹ The case against Aitken was circumstantial. Evidence adduced by the Crown included wiretap evidence, CCTV video of the shooting, information about weapons and ammunition to which Aitken had access, information about

¹⁹ Haydn Kelly testified: “... most of my practice is biomechanically oriented. And a lesser percentage of it is surgically oriented.” See *R v Aitken* (17 November 2008), Victoria 135927 (BCSC) [*Aitken* transcript] Haydn Kelly, examination in chief at 3409.

²⁰ These comparison techniques are sometimes described as identification sciences because of their role in linking an individual or object with a trace.

²¹ The length and cost of proceedings might suggest the desirability of actually resolving some of the fundamental validation and reliability issues before untested and/or controversial techniques are drawn into criminal proceedings – often in trial after trial.

motive, and expert opinion evidence. Several experts were called at trial to testify about DNA, ballistics analysis, photogrammetry and forensic gait analysis. In this article, we focus on the forensic gait analysis proffered by Haydn Kelly, a Harley Street podiatrist who had previously offered expert testimony in a number of English cases. *Aitken* represents the first occasion on which forensic gait analysis seems to have been offered in a North American courtroom.

Kelly defined forensic gait analysis as “the analysis of the style or manner in which a person walks”²² as applied “to legal problems.”²³ Kelly asserted that “each person has an individual gait which is unique to them.”²⁴ Other aspects of his testimony suggested that this unique gait would be stable over time. He did not claim, however, that the features identified from CCTV images in this case were unique. Rather, and in compliance with an order made by Satanove J prior to trial,²⁵ Kelly pointed to what he claimed were similarities between Aitken’s gait (as captured in a number of covert recordings) and the stance or lower body position of the shooter at certain moments in the CCTV images. Overall, Kelly testified that the likeness between the shooter and Aitken was “very strong.”

The admissibility, and specifically the reliability, of Kelly’s testimony was vigorously contested by an unusually well-informed defence counsel before and throughout the trial.²⁶ On appeal, the defence renewed its argument that Kelly’s work did not meet the standard of reliability required of expert testimony in Canada. More particularly, the defence argued that Kelly had failed to follow scientific methods, and that the application of the clinical skills of podiatry to gait analysis was novel. On this basis, the defence challenged both Kelly’s qualifications and the “legal relevance” of his testimony.²⁷

²² *Aitken* transcript, *supra* note 19 (8 September 2008) Haydn Kelly, evidence in chief on *voir dire* at 1310, (17 November 2008); Haydn Kelly, examination in chief at 3410.

²³ *Ibid* (8 September 2008) Haydn Kelly, evidence in chief on *voir dire* at 1314, (17 November 2008) Haydn Kelly, examination in chief at 3412. Though there does not appear to be a settled definition of “gait analysis” among forensic podiatrists, see e.g. John DiMaggio and Wesley Vernon, *Forensic Podiatry: Principles and Methods* (New York: Humana Press, 2011) at 103-106.

²⁴ *Ibid* (9 September 2008) Haydn Kelly, evidence in chief on *voir dire* at 1360.

²⁵ *R v Aitken*, 2008 BCSC 1423, [2008] BCJ No 2858 [*Aitken* Expert Evidence *Voir Dire*].

²⁶ This level of effort and sophistication is not a particularly common feature of many adversarial trials.

²⁷ *R v Aitken*, Appellant’s Factum (BCCA Registry number CA36854, Victoria) at 35-45.

Even without the gait evidence there was a strong, perhaps compelling, circumstantial case against Aitken.²⁸ Our concern in this article is with the admission of the gait evidence and its presentation as independent corroboration of other evidence of guilt by an expert with seemingly impressive credentials. Having reviewed the court record, we consider that the nature of the gait evidence seems to have been misunderstood by the trial and appellate judges, might well have been unnecessary to obtain a conviction and, as we suggest, its admission rendered the process substantially unfair.²⁹ Placing the judicial reasoning in *Aitken* within the context of current trends within forensic science more generally, we argue that the BC Court of Appeal decision interpreted “reliability” in a way that is neither in keeping with other appellate decisions regarding the admissibility of expert evidence nor suited to evaluating expert comparison evidence in criminal proceedings.

A) Haydn Kelly’s Testimony on Voir Dire and at Trial

Adan Merino’s murder was captured by a medium-quality security camera positioned at the door of the apartment building where Merino lived. The shooter, whose face was obscured, was also captured in that video for approximately eighteen seconds. Six seconds of the CCTV video of the shooting provided the frames that were the focus of Kelly’s comparison (see figure 2).³⁰ Kelly’s testimony was offered by the Crown to strengthen the link between the shooter and Aitken.

On *voir dire* and at trial, Kelly testified that he has developed a methodology for forensic gait analysis. His description varied slightly over the course of the trial, but ultimately the methodology “for comparison of gait and features of gait”³¹ was comprised of five steps:

1. determining whether the images are of sufficient quality to be analyzed;³²

²⁸ The (perceived) strength of the overall case, however, should not mediate the admission of weak, speculative or unreliable expert evidence. With respect to admissibility, the techniques and derivative opinion evidence should stand on their own.

²⁹ It also added to the length, complexity and cost of the proceedings.

³⁰ *Aitken* transcript, *supra* note 19 (19 November 2008) Haydn Kelly, cross-examination at 3553, 3568.

³¹ *Ibid* (18 November 2008) Haydn Kelly, cross-examination at 3464.

³² There are no references in *ibid* to standards that might enable an analyst to determine whether the video is of sufficient quality for analysis, or for any of the other steps.

2. watching the images of the event involving an unknown person first, and identifying whether unusual or unique features of gait are present which may assist with identification;
3. only then watching known video of the person of interest and seeking to identify points of similarity and difference in gait;
4. watching all video numerous times, always beginning with the images of the unknown person; and
5. using a standard scale of strength of evidence to express a view about the degree of similarity or dissimilarity between the gait of the unknown person and that of the known person.³³

These processes are not entirely sequential, but Kelly did testify that he would not look closely at the images of the known person until he had documented any unusual gait features discernible in the images of the unknown person.³⁴ In response to questions asked by the defence, Kelly testified that he has not published his methodology or research in peer-reviewed journals because it was not yet at a point that warranted publication. In a passage that illustrates the acerbity directed towards the defence throughout the *voir dire* and trial, Kelly testified: “[W]hen there is something to be published that’s of real worth in terms of scientific community, I will be sure that ... you have a copy of the book, Mr Firestone, with the publisher’s discount.”³⁵

Despite the fact that the methodology was not ready for publication, and had not been subjected to formal evaluation or peer review, Kelly expressly claimed throughout the proceeding that his approach was scientific.³⁶ This claim emerged most clearly when Kelly was challenged by the defence on the subjectivity of his “method” of comparing the gait

³³ Generally, *ibid* (17 November 2008) Haydn Kelly, examination in chief at 3412 – 3415.

³⁴ On *voir dire*, Kelly testified that, when first considering the appropriateness of the video for forensic gait analysis, he also reviewed the images of the known person; see *ibid* (8 September 2008) at 1307. The ordering of the analysis is an attempt to avoid displacement. It does not necessarily do that, and does not address other threats from contextual bias and cross-contamination. See Part 3 below.

³⁵ *Ibid* (9 September 2008) Haydn Kelly, cross-examination on *voir dire* at 1408. Compare (17 November 2008) Haydn Kelly, examination in chief at 3422.

³⁶ *Ibid* (9 September 2008) Haydn Kelly, cross-examination on *voir dire* at 1366, (17 November 2008) Haydn Kelly, examination in chief at 3413, (18 November 2008) Haydn Kelly, cross-examination at 3470, 3473, 3504, (19 November 2008) Haydn Kelly, cross-examination at 3544.

of a person of interest with that of the incident video. Kelly responded to questions put by Firestone in the following terms:

it's a scientific approach to analyzing the material ...

it's being analyzed in a standardized, accepted scientific manner. The comparisons are then being performed in the same scientifically accepted manner. And they are being referred to, and compared by, an accepted scientific scale ...

they are methods that have been created by others in science for many, many, many, many years.³⁷

Notwithstanding Kelly's insistence that his methods were scientific, he strenuously resisted defence counsel's suggestion that processes and tests conventionally associated with the evaluation and refinement of forensic comparison techniques were relevant to assessing the reliability of forensic gait analysis. Most famously, in *Daubert v Merrell Dow Pharmaceuticals, Inc*, the US Supreme Court held that "in order to qualify as 'scientific knowledge,' an inference or assertion must be derived by the scientific method. Proposed testimony must be supported by appropriate validation—*i.e.*, 'good grounds,' based on what is known."³⁸ The US Supreme Court held that a court faced with determining the admissibility of scientific evidence must undertake

a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue.³⁹

While the US Supreme Court declined to establish a definitive checklist to help trial judges assess the reliability of scientific evidence, it suggested four criteria that would ordinarily be appropriate to this task. These criteria are: (1) whether the theory or technique 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 the technique and the application of standards; and (4) the degree to which the technique has gained general acceptance. In *R v J(J-L)*, the Supreme Court of Canada held that these criteria could assist a trial judge faced with assessing the reliability of novel scientific evidence.⁴⁰

³⁷ *Ibid* (18 November 2008) Haydn Kelly, cross-examination at 3473-74.

³⁸ (1993) 509 US 579 [*Daubert*].

³⁹ *Ibid* at 592-3.

⁴⁰ 2000 SCC 51, 2 SCR 600 [*J(J-L)*].

The appropriate use of criteria for determining threshold reliability also arose in *Abbey*.⁴¹ Readers may recall that the evidence at stake in that case was offered by a sociologist who had conducted research among prisoners with a history of gang involvement. It was proposed that the sociologist give evidence about the possible cultural meanings of a teardrop-shaped tattoo on the face of a gang member. Warren Abbey, the accused in that case, admitted to being gang-involved, and had obtained a teardrop tattoo sometime after the homicide with which he was charged. The sociologist was to explain the possible implications of such a tattoo – one of which was to record a gang-related killing. The trial judge excluded the expert’s testimony on the basis that his work had no error rates, had not been subjected to testing, and was not peer-reviewed. On appeal, Doherty JA held that the trial judge had erred by judging the reliability of the expert’s qualitative research methods by standards that were inappropriate to the expert’s field and the nature of the claims being made. In particular, Doherty JA observed, “Dr Totten did not pretend to employ the scientific method and did not depend on adherence to that methodology for the validity of his conclusions.”⁴² While reiterating that the list of criteria that may be relevant to assessing reliability was not closed, Doherty JA suggested a list of criteria that were more suited to assessing the reliability of qualitative evidence.⁴³

Returning to *Aitken*, Kelly plainly claimed that his evidence was scientific. He characterized the evidence as such before the jury and in *voir dire*. However, he rejected the relevance of several of the criteria set out by the US Supreme Court in *Daubert*⁴⁴ and cited with approval by the Supreme Court of Canada in *J(J-L)*.⁴⁵ The evidence set out above demonstrates that Kelly had not yet published his research – indeed he seems to imply in the passage cited above (“when there is something to be published that’s of real worth in terms of scientific community ...”)⁴⁶ that while his work is sufficiently reliable to be used in a criminal trial, it does not yet meet the higher standards required of *real* science.

The first *Daubert* criterion highlights the fact that courts seeking to assess the reliability of expert testimony should consider whether the evidence is susceptible to testing, and whether testing has been done. The third *Daubert* criterion draws attention to the relevance of error rates to the

⁴¹ *Abbey*, *supra* note 14.

⁴² *Ibid* at para 108.

⁴³ *Ibid* at para 119.

⁴⁴ *Supra* note 38.

⁴⁵ *Supra* note 40.

⁴⁶ *Aitken* transcript, *supra* note 19 (9 September 2008) Haydn Kelly, cross-examination on *voir dire* at 1408. See also transcript extract cited *infra* at note 47.

reliability of scientific evidence. In most cases formal testing and/or rigorous proficiency studies will provide an indication of the limits of a technique and the indicative range of errors. Regarding such blind tests and error rates, Kelly testified in response to defence questions:

Q: The word blind test suggests where, after you come to your conclusions, where you're shown others to see whether or not your identification is correct. Do you understand what I mean by blind test?

A: Yes.

Q: Have you ever done a blind test in any of your reports for the court?

A: No.

Q: I'm going to suggest to you, sir, that unless you do a blind test on any of your cases, you have no idea what, in fact, your error rate is, right?

A: I'd suggest that is a ridiculous suggestion.

Q: Why is that ridiculous?

A: Because you're not taking into account the points we discussed yesterday in relation to the 20-odd years of examining people's gait and what is unusual and what is not, and to infer that after 20 years of experience I can't tell one person from another is quite interesting.

...

Q: Are you claiming for the purposes of these court proceedings infallibility?

A: How do you interpret infallibility?

Q: Never being wrong.

A: Everything's possible that something could be wrong.

Q: Let me suggest this to you, sir, in terms of a scientific method. Unless you subject your work to blind testing, you can never determine whether or not you are in error or not, right?

A: Let me just explain it for you. At the meeting that took place with Dr Everett [sic] ... it was made very clear that the scale of comparison and the method of comparison being used was statistically satisfactory to be presented before courts.⁴⁷

Kelly suggested that blind testing is only relevant for those who do not have biomechanical training.⁴⁸ He asserted, however, that his clinical experience offered sufficient protection against error to render blind testing irrelevant to his work.⁴⁹ In this passage, a second important thread to Kelly's evidence emerges: namely, his reliance on his clinical experience in podiatrics when performing the forensic task of identifying similarities in gait.⁵⁰ This reliance on clinical experience became an important theme within both Satanove J's decision on the admissibility of Kelly's evidence and the BC Court of Appeal decision that upheld Satanove J's ruling.

The second *Daubert* criterion is peer review. The importance of peer review in all fields was also emphasized by Doherty JA in *Abbey*.⁵¹ In relation to peer review Kelly testified, again in response to defence questioning:

Q: Sir, let me just ask you directly, in terms of a methodology that you have adopted in this case, why have you not written an article to set out the science and method which you apply for a review by other podiatrists in the United Kingdom?

A: Well, there's — there's two parts to that. The first part is, as I mentioned yesterday, that there is research which is ongoing in relation to people's gait and features of gait. And in addition to the compilation of databases regarding people's gait and features of gait.

And there is also ongoing work in relation to the publication of a book which I'm the author of, and when those things are complete, when we have something which is not

⁴⁷ *Ibid* (9 September 2008) Haydn Kelly, cross-examination on *voir dire* at 1366-67.

⁴⁸ *Ibid* (18 November 2008) Haydn Kelly, cross-examination at 3471.

⁴⁹ Contrast the discussion of these issues in relation to latent fingerprint comparison in Gary Edmond, Matthew Thompson and Jason Tangen, "A Guide to Interpreting Forensic Testimony: Scientific Approaches to Fingerprint Evidence" (2013) 12 *Law, Probability & Risk* 1.

⁵⁰ Note also, the implied ability to be able to identify and discriminate – to be able to "tell one person from another."

⁵¹ The value of peer review as a form of verification is probably more constrained than most lawyers and judges seem to appreciate. For a review of recent literature on medical peer review, see Gary Edmond, "Judging the Scientific and Medical Literature: Some Legal Implications of Changes to Biomedical Research and Publication" (2008) 28 *Oxford J Legal Stud* 523.

an article which could be regarded as semi-scientific but when we can produce something which is in process that is very significant, it will be published. ...

And when there is a — when we, what I regard as being — having extensive data from a scientific perspective, then that work will be published. And that is in process. I'm not here or anywhere else to provide an article that others may assume from a scientific perspective to be of interest, make reference to it and then when it comes under proper scientific scrutiny is found to be wanting.⁵²

He later clarified:

A: ... the important thing about such an identifying aspect of forensic science is that when something is published, that there is sufficient data for it not to be misleading or misinterpreted by others.

And therefore the reason why journ [sic] — sorry, articles have not been published, in that aspect, at the moment, is precisely that reason.⁵³

In relation to verification, Kelly testified that he had never submitted a report for verification by anyone “other than myself.”⁵⁴

Kelly testified that he had concluded that there was a “very strong likeness” between the gait of the unknown shooter and that of Aitken. Kelly based his conclusion largely on the magnitude of abduction⁵⁵ he identified as common to the two gaits and on the eversion⁵⁶ of the shooter's left foot compared with the eversion of Aitken's feet.⁵⁷ He explained that “very strong likeness” was the second highest level of similarity on his scale, and that the highest level, “extremely strong likeness,” would be as close as a forensic podiatrist could come to

⁵² *Aitken* transcript, *supra* note 19 (18 November 2008) Haydn Kelly, cross-examination at 3466-67. DiMaggio and Vernon, *supra* note 23 at 171: “Whether working in or outside a large laboratory, there is a clear obligation to provide an acceptable level of quality assurance in the approach to and outcomes of the work provided. To this effect, forensic podiatrists are advised to work to strict protocols and to ensure where possible that their work is checked and verified by a peer practitioner.”

⁵³ *Aitken* transcript, *supra* note 19 (18 November 2008) Haydn Kelly, cross-examination at 3471.

⁵⁴ *Ibid* (18 November 2008) Haydn Kelly, cross-examination at 3472-73, 3466, 3481. Contrast *Re Truscott*, [2007] ONCA 575 at paras 166, 169, 183, 313, 349, 371-2, 226 OAC 200.

⁵⁵ “Abduction” means that feet turn out at the toes during the gait cycle.

⁵⁶ “Eversion” translates, roughly, to flat footedness and means that the foot rolls during the gait.

⁵⁷ Kelly testified that the video did not allow him to assess whether the shooter's right foot was everted.

identification.⁵⁸ However, he also agreed with Crown counsel's suggestions before the jury that the video images demonstrated some "unique features of gait."⁵⁹ The ordinary meaning of unique is, of course, that the object thereby described is the only one of its kind. Accordingly, and despite claims to the contrary, Kelly came very close to asserting that he could identify Daniel Aitken as the shooter portrayed in the incident video.

B) The Trial Judge's Conclusions Regarding Admissibility

Satanove J ruled on *voir dire* that Kelly was not qualified to testify to the frequency with which the unusual gait features he identified as common to the shooter and to known images of Aitken could be found in the general population.⁶⁰ Kelly had been willing to testify that they would be found in one per cent of the population, but could not point to a database or study that supported this claim.⁶¹ Satanove J otherwise ruled that Kelly's evidence was admissible, however, and expressed the view that "[t]here is no doubt that the evidence of Mr. Kelly has high probative value, that is, it is so related to a fact in issue that it tends to establish it."⁶² Probative value is the extent to which the evidence is able to rationally influence the assessment of facts in issue. Satanove J regarded Kelly's evidence as reliable and sufficiently probative to be heard by the jury. She concluded that Kelly's testimony was necessary, because "[i]t takes his trained eye to be able to isolate features of the stance and gait of the person of interest, the shooter and the accused to appreciate the magnitude and similarity of the unusual features."⁶³ Satanove J also noted that Kelly had offered an

⁵⁸ *Aitken* transcript, *supra* note 19 (9 September 2008) Haydn Kelly, cross-examination on *voir dire* at 1432 and (18 November 2008) Haydn Kelly, examination in chief at 3442-43.

⁵⁹ *Ibid* (18 November 2008) Haydn Kelly, examination in chief at 3442. Forensic analysts do not always use the word "unique" according to its conventional meaning. Unfortunately, given their role in identification and presenting evidence in reports and courtrooms, the word is often used to refer to features that are said to be rare or highly discriminating but not necessarily unlike anything else. See also DiMaggio and Vernon, *supra* note 23 at 19: "It is ... fundamentally important to note that there is currently no evidence considered and utilized by forensic podiatrists that has been demonstrated at the individual (unique) identification level. ... change is not anticipated in the foreseeable future." See also *ibid* at 78, 105.

⁶⁰ *Aitken* Expert Evidence *Voir Dire*, *supra* note 25 at para 35.

⁶¹ Interestingly, if this was based on Kelly's experience with patients – that is, those seeking the services of a podiatrist – then it could hardly be considered to be a representative sample of the (English) population.

⁶² *Aitken* Expert Evidence *Voir Dire*, *supra* note 25 at para 21.

⁶³ *Ibid* at para 25.

expert opinion in approximately 150 cases in England, and testified on 20 occasions in that jurisdiction.

The defence had argued that Kelly's evidence constituted novel science and should therefore be subjected to the heightened scrutiny required by *Mohan*.⁶⁴ In particular, the defence suggested that Kelly was untrained in forensics and unaware of the many dangers attending comparison and identification evidence. Satanove J concluded:

I am not satisfied that Mr. Kelly's evidence is in fact "novel" science. Podiatry has been in existence for a thousand years and the expertise of a podiatrist to analyze an individual's gait has long been accepted and practiced in a clinical setting. ... I do not think there is a danger of the jury being mesmerized by what is quite simply an exercise of expertise in observation and diagnosis.⁶⁵

In this passage, Satanove J positions Kelly's evidence as clinical – based on professional experience, rather than being properly characterized as scientific. However, Satanove J did not expressly consider whether the training and clinical experience of a podiatrist translates well to the forensic task of identifying similarities from CCTV images for the purpose of identification. This is an important, perhaps the crucial question. A podiatrist focuses on diagnosing biomechanical irregularities of the lower limb *in vivo* in patients who have presumably already self-selected as requiring clinical intervention.⁶⁶ In his or her office the podiatrist can control the angle from which to view (or record) the gait, the gait speed, the number of gait cycles necessary to make a diagnosis, the type of gait (such as walking or running), the clothing and footwear worn by the patient, the lighting, and so on. Podiatrists do a range of things with patients, but it is not a profession that is steeped in, nor familiar with, identification and its innumerable complexities.⁶⁷

In the criminal context, identification (whether individualization or the recognition and enumeration of similarities) is fraught with difficulties. To focus on a relevant example, CCTV video of an incident is frequently of relatively low quality and the expert must work with whatever images are

⁶⁴ *Supra*, note 15.

⁶⁵ *Aitken Expert Evidence Voir Dire*, *supra* note 25 at para 34. See *Aitken* transcript, *supra* note 19 (18 November 2008) Haydn Kelly, examination in chief at 3408.

⁶⁶ Kelly described podiatry as the "study ... and treatment of all conditions of the feet or conditions pertaining to the feet;" see *Aitken* transcript, *supra* note 19 (18 November 2008) Haydn Kelly, examination in chief at 3401.

⁶⁷ The newness of podiatric involvement with criminal investigations is suggested by Kelly's entry in the *Guinness Book of Records*; see below, text accompanying note 102.

available. In *Aitken*, the video was taken at night, the shooter was wearing loose clothing including track pants, socks and slip-on sandals and the frame of the apartment building's front door obscured part of the shooter's lower body for much of the six seconds of the video identified by Kelly as "usable." Figure 1 is a sample still from the images used at trial.⁶⁸

Figure 1: Relatively unobstructed sample CCTV image of shooter's lower limbs



This image is taken from the court record in *R v Aitken* (BCCA Registry number CA36854, Victoria). It corresponds to Exhibit 17, Tab 2 at 102. CCTV footage date and time stamp 28.12.2004/23:11:25:026.

Reproduced as figure 2 is an illustrative set of still images taken from the video of the incident. This sequence of images was crucial to Kelly's opinion at trial.⁶⁹

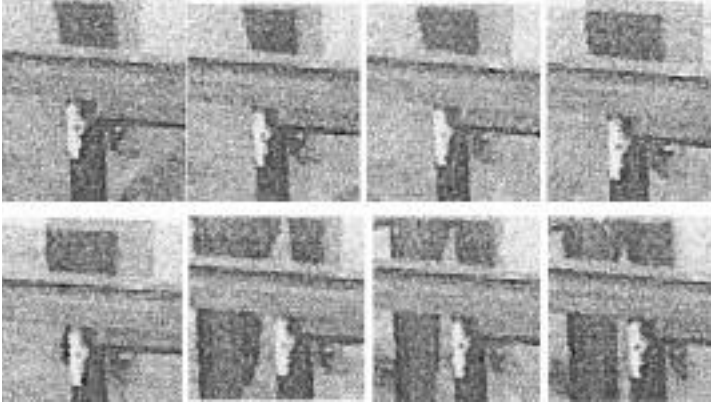
As is apparent from figures 1 and 2, the footage relied upon by Kelly was low resolution and the images were impeded by physical features such as a doorframe. The images in figure 2 have been cropped to focus on the image of the shooter but we have attempted to retain the resolution. The shooter's foot position varied, as might be expected, and the lower limbs

⁶⁸ The image used in figure 1 has been chosen because it offers one of the clearest pictures of the shooter's lower body and was relied upon by Kelly at trial, *Aitken* transcript, *supra* note 19 (18 November 2008) Haydn Kelly, cross-examination at 3570-72. Note, however, that this image does not show eversion of the left foot, which Kelly diagnosed from the images depicted in figure 2.

⁶⁹ *Aitken* transcript, *supra* note 19 (18 November 2008) Haydn Kelly, cross-examination at 3568 *et seq.* While Kelly emphasized throughout that the best way to view the CCTV footage was on an LCD screen, these still images were relied upon at trial. They are the highest quality available to us through the court record, although we have cropped them to remove the images of the victim that appeared in the original stills. All parties agreed that the CCTV footage was of low resolution.

did not always show the unusual features relied upon by the expert. Kelly did not offer clear testimony about the standards that he used to diagnose an abnormal gait feature, the relevance of the fact that the unusual gait features were not always present, or how well the CCTV footage corresponded with his “normal” practice that a complete normal gait cycle is the minimum required to diagnose podiatric abnormalities.⁷⁰

Figure 2: Sample images of shooter’s lower limbs relied upon by Kelly



These images are taken from the court record in *R v Aitken* (BCCA Registry number CA36854, Victoria). They correspond to Exhibit 17, Tab 2 at 89-96. CCTV footage date and time stamp 28.12.2004/23:11:23:234 – 23:11:24:175.

Regarding clinical experience as a sufficient warrant of reliability when the task of identifying abnormalities in gait is translated into the forensic context wholly overlooks the challenges that are presented by forensic casework. There is a very real risk of *slippage* as “specialised knowledge” and practices from clinical podiatry are transformed into *expertise* deemed capable of grounding the admission of incriminating opinions pertaining to the identity of offenders. The ability to identify and respond to problems with gait and posture in a clinical setting may not readily convert into an ability to accurately identify gait features in (low quality) CCTV recordings or to estimate the frequency of particular features in relevant (notably foreign) populations. Forensic comparison introduces a different range of issues and dangers.

Abbey expressly encourages trial judges to consider limiting the scope of admitted evidence.⁷¹ Satanove J’s decision to sever Kelly’s evidence regarding frequency of abnormal gait features from his evidence regarding the similarities between the known person and the shooter reflects a common judicial approach that we have described elsewhere as an

⁷⁰ *Ibid* at 3557-58, 3567-69, 3572-73.

⁷¹ *Abbey*, *supra* note 14 at paras 62-64.

admissibility compromise.⁷² It seems appropriate in principle that trial judges should apply the usual standards of threshold reliability to the task of delineating admissible evidence. In prior work, we have criticized admissibility compromises on two grounds. First, the evidence that is admitted pursuant to an admissibility compromise is rarely less susceptible to validation studies than the evidence that is excluded. In other words, reliability concerns persist in relation to the admitted testimony.⁷³ Secondly, restricting the expert to enumerating similarities perpetuates the many risks of contextual bias that have been identified in relation to forensic comparison evidence. These risks will be explained in greater detail in Part 3. Put simply, people are not very good at interpreting visual evidence of identity (even when such interpretation forms part of their daily work), and confidence is no guide to accuracy.⁷⁴ Furthermore, several studies suggest that jurors interpret verbal expressions (such as Kelly's testimony that there was a "very strong" similarity between the gait of the known person and that of the shooter) idiosyncratically, making it difficult to predict how a jury will interpret non-numeric expressions of uncertainty.⁷⁵ While numerical estimates of frequency should never be admitted unless they are demonstrably predicated on reliable databases, excluding an unreliable numerical estimate does not act as a panacea in relation to other well-founded reliability concerns. In the context of *Aitken*, for example, excluding Kelly's evidence regarding frequency did nothing to address legitimate concerns about whether Kelly was capable of identifying abnormal gait features from CCTV footage given the nature of that footage and the conditions in which it was captured.

Satanove J delivered her judgment before the Ontario Court of Appeal issued its judgment in *Abbey*. Accordingly, she did not have the benefit of the reformulated approach to admissibility set out by Doherty JA in 2009. Nor did she have access to the guidance offered by Doherty JA in relation to assessing the reliability of expert evidence. By the time the case came to

⁷² Edmond, Cole, Cunliffe and Roberts, *supra* note 6 at 105.

⁷³ *Ibid.*

⁷⁴ Indeed, some literature suggests that the risk of contextual bias is especially acute when dealing with photographic or visual evidence. See further Gary Edmond, Catherine Biber, Richard Kemp and Glenn Perter, "Law's Looking Glass: Expert Identification Evidence Derived from Photographic and Video Images" (2009) 20 *Curr Issues in Crim J* 337; Itiel E Dror and Simon A Cole, "The Vision in 'Blind' Justice: Expert Perception, Judgment and Visual Cognition in Forensic Pattern Recognition" (2010) 17 *Psychonomic Bull & Rev* 161; Richard Kemp, Nicola Towell and Graham Pike, "When Seeing Should Not Be Believing: Photographs, Credit Cards and Fraud" (1997) 11 *Applied Cognitive Psychology* 211.

⁷⁵ See Kristy Martire, Richard Kemp, Malindi Sayle and Ben Newell, "On the Interpretation of Likelihood Ratios in Forensic Science Evidence: Presentation Formats and the Weak Evidence Effect" (2014) 240 *Forensic Sci Int'l* 61.

the BC Court of Appeal, however, *Abbey* was at the heart of the arguments about the proper approach to the admission of expert testimony such as that offered by Kelly.

C) *The Court of Appeal Decision*

The BC Court of Appeal unanimously upheld Satanove J's decision to admit Kelly's evidence in Aitken's trial. In a judgment authored by Hall JA, the Court of Appeal relied on several lines of reasoning when confirming that Kelly's evidence had been properly admitted. First, the Court of Appeal agreed with Satanove J that Kelly's evidence was not properly characterized as scientific but that it was, instead, the product of professional experience. Secondly, the Court of Appeal purported to draw on *Abbey* to articulate the principles that should apply to assessing the reliability of evidence based on professional experience. Finally, the Court of Appeal relied heavily on the fact that Kelly's evidence had been admitted in English courtrooms in support of its admissibility in Canada. We discuss each of these lines of reasoning in turn, and add a note about the recurring dispute regarding the relevance of 'novel' science, which emerged once again in *Aitken*.

1) *The Appropriate Characterization of Kelly's Evidence: Science or Experience?*

Like the BC Supreme Court, the BC Court of Appeal characterized Kelly's opinion as non-scientific with the consequence that the reliability indicia associated with *Daubert* (and with scientific techniques in general) were said to be inapposite. Hall JA explained:

Abbey ... makes it clear that the *Daubert* factors are not essential to the reliability inquiry where the proffered evidence is based on specialized knowledge acquired through training or experience in a particular discipline:

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.

...

In my view, the forensic gait analysis provided by Mr. Kelly in the present case falls into the category of expert opinion evidence based on "specialized knowledge gained through experience and specialized training". In determining the admissibility of Mr. Kelly's evidence, the trial judge did not err in failing to consider indicia of scientific validity such as peer review, rate of error and adherence to a scientific method. These factors have limited relevance in a case like the one at hand where a witness's

expertise is gained over a period of years through observation and experience in the professional realm.⁷⁶

The Court of Appeal's approach transforms the need to attend to the reliability of the evidence into a question of nomenclature. Rather than ask whether the expert can actually do what he claims and how we know, the Court's focus was directed to the question of whether forensic gait analysis is scientific. By characterizing this technique or set of practices as non-science, the Court of Appeal suggested that scientific indicia of reliability are irrelevant to assessing the reliability of forensic gait analysis. Like Satanove J, the BC Court of Appeal offered no criteria by which the demarcation decision proceeded, or by which future categorizations should be made. What, we might wonder, determines whether some activity is scientific or non-scientific?⁷⁷ It seems from *Aitken* that the question of whether a subject, technique or set of practices constitutes "science" or is susceptible to formal evaluation is a matter for judicial declaration rather than analysis. Crucially, the classification of a technique as scientific or non-scientific does not address the more fundamental and illuminating questions of whether a given technique works and whether the analyst has *relevant* expertise.

As the foregoing discussion of *Daubert, J(J-L)* and *Abbey* suggests, a great deal depends on the characterization of a given field of expertise as scientific or non-scientific.⁷⁸ In *Aitken*, characterizing the evidence as predicated on experience rather than science transformed the admissibility determination from an inquiry using *Daubert*-type criteria and concern with validation to a question of whether the witness possesses the professional qualifications necessary to maintain a clinical practice. As

⁷⁶ *Aitken CA*, *supra* note 18 at paras 79-80, citing *Abbey*, *supra* note 14. In a section entitled "Confirming that the correct type of examinations and procedures have been selected," DiMaggio and Vernon stress that the "methods involved should have proven validity;" see *supra* note 23 at 174.

⁷⁷ This demarcation issue has been a vexed problem among historians, philosophers and sociologists of science. There is no simple formula to determine whether something is scientific and so, it is important in legal contexts (especially after *Mohan*, *supra* note 15, and *Trochym*, *supra* note 7) to focus on the reliability issues regardless of the classification. See generally, Rachel Laudan, ed, *The Demarcation between Science and Pseudo-science* (Blacksburg VA: Virginia Polytechnic and State University, 1983) and Thomas Gieryn, *Cultural Boundaries of Science* (Chicago: University of Chicago Press, 1999).

⁷⁸ The novelty of the so-called "field" should normally focus attention. Characterizing an emerging set of practices or new techniques as a "field" or part of a field may divert attention from the question of whether the technique works and how well. The NRC Report explains that every new technique should be validated, regardless of the field from which it originates; see NRC Report, *supra* note 2 at 113.

was true in the BC Supreme Court, the Court of Appeal offered no discussion of the applicability of the indicia of reliability such as peer review, rate of error and adherence to standardized scientific methods – such as independent validation studies. Focusing on the demarcation issue, in effect, enabled the judges at each level to avoid the need to grapple with the relationship between proficiency, methodology and reliability. Instead, the evaluation of Kelly’s evidence was based around more impressionistic and credibility-focused issues.

In *Aitken*, the effect of characterizing Kelly’s techniques as non-scientific was to relieve Kelly and the Crown of the need to satisfy a range of criteria that are relevant to formally evaluating the threshold reliability of forensic comparison and identification techniques.⁷⁹ Applying these criteria in this case would have allowed the Court to determine the value of and assess the risk of error in Kelly’s technique. Remarkably, the Court of Appeal declared forensic gait analysis non-scientific even though Kelly and others repeatedly claimed in front of the jury that his methods were scientific and that he was engaged in scientific research.

In *Abbey*, the decision that the expert’s methods were not properly characterized as scientific prompted the Ontario Court of Appeal to identify alternative, more appropriate, ways to assess the expert’s methodology.⁸⁰ Doherty JA held quite explicitly that characterizing the evidence as non-scientific did not dispense with the need to demonstrate reliability as a condition of admissibility: “I do not suggest that the Crown was not required to demonstrate threshold reliability. That reliability had to be determined, however, using tools appropriate to the nature of the opinion advanced by Dr Totten.”⁸¹ It is evident from *Abbey* that these tools extend well past an assessment of clinical experience.⁸² In *Aitken*, the BC Court of Appeal did not similarly require the Crown to use appropriate methods to demonstrate the reliability of Kelly’s opinion. Following

⁷⁹ NRC Report, *ibid* at 14ff. DiMaggio and Vernon commence the first substantial chapter “Forensic Podiatry Principles and Human Identification” in the following terms: “This chapter will introduce the reader to the scientific approach that is required to both comprehend and safely practice forensic podiatry;” see DiMaggio and Vernon, *supra* note 23, at 13, also 15, 16, 21.

⁸⁰ *Abbey*, *supra* note 14 at paras 116-19.

⁸¹ *Ibid* at para 118.

⁸² Factors listed as relevant to the reliability of qualitative evidence in *Abbey* include attention to the extent to which the expert is working in a recognized field or area of specialized training, the application of quality assurance measures and peer review, the retention and accessibility of data relied upon by the expert, the degree to which the expert’s methodologies are accepted in his or her field, and the degree to which the methodologies promote the reliability of the information gathered and relied upon by the expert; see *ibid* at para 119.

Aitken, experts who are not regarded by the Court as “scientific,” but whose experience is recognized as adequate, will seemingly be able to express mere *ipse dixit*. In our respectful opinion, this approach is at odds with *Mohan*, *J(J-L)*, *Trochym* and *Abbey*.

In the United States, in the aftermath of *Daubert*, several groups of forensic analysts confronted with *Daubert* challenges (especially around testing, error rates, publication and peer review), *re-characterised* themselves and their fields (or were re-branded by prosecutors) as “technicians” rather than “scientists” in order to avoid the application (and implications) of the *Daubert* criteria.⁸³ Initially it was argued, and accepted by several trial and appellate courts, that the *Daubert* criteria applied only to opinions derived from “scientific knowledge” and that opinions derived from “technical and other specialised knowledge” were effectively exempt.⁸⁴ That approach was decisively rejected through the US Supreme Court’s holding, in *Kumho Tire Co v Carmichael*, that all expert evidence should be reliable even if the *Daubert* criteria needed to be flexibly adapted to non-scientific forms of evidence.⁸⁵

It bears noting that, notwithstanding the legal characterization, the comparison technique relied upon by Kelly (like most other comparison and identification techniques) is readily susceptible to formal evaluation.⁸⁶ This merely reinforces the limits of classifications (science/non-science, for example) that do not attend to the validity and reliability of the underlying technique or the proficiency of analysts.

2) *Assessing the Reliability of Evidence that is Based on Professional Experience*

Having concluded that forensic gait analysis is non-scientific, the BC courts focused on other criteria to assess the evidence. Here, the training, qualifications and clinical experience of Kelly assume considerable prominence.

⁸³ This approach was subjected to excoriating critique by academic critics in the US. See e.g. Michael J Saks, “Banishing *Ipse Dixit*: The Impact of *Kumho Tire* on Forensic Identification Science” (2000) 57 Wash & Lee L Rev 879; Paul Giannelli and Ed Imwinkelried, *Scientific Evidence*, 4th ed (Albany, NY: Lexis Publishing Co 2007).

⁸⁴ Rule 702 of the US *Federal Rules of Evidence* (FRE) covers the admissibility of “scientific, technical and other specialized knowledge.”

⁸⁵ 526 US 137, 147 (1999).

⁸⁶ NRC Report, *supra* note 2 at 189. On the possibility of testing forensic science techniques, the NRC committee concluded: “Little rigorous systematic research has been done to validate the basic premises and techniques in a number of forensic science disciplines. The committee sees no evident reason why conducting such research is not feasible.”

In my view, the forensic gait analysis provided by Mr. Kelly ... falls into the category of expert opinion evidence based on “specialized knowledge gained through *experience and specialized training*.” In determining the admissibility of Mr. Kelly’s evidence, the trial judge did not err in failing to consider indicia of scientific validity such as peer review, rate of error and adherence to a scientific method. These factors have limited relevance in a case like the one at hand where a witness’s expertise is *gained over a period of years through observation and experience in the professional realm*.⁸⁷

Of the identified preconditions, it is primarily the requirement for a properly qualified expert that is in issue on this appeal. In her treatment of this criterion, the trial judge identified the proper test, which is set out at p. 25 of *Mohan*: “the evidence must be given by a witness who is shown to have acquired special or peculiar knowledge *through study or experience* in respect of the matters on which he or she undertakes to testify”. In my opinion, Mr. Kelly’s *experience and credentials satisfy this test*.⁸⁸

The Court further held that, based on precedent, the admissibility of the podiatrist’s evidence was not dependent upon the existence of a database capable of generating statistical probabilities, and that it was sufficient that his evidence was “*the consequence of personal experience*.”⁸⁹

As these quotes demonstrate, the BC Court of Appeal placed considerable emphasis on the relevance of Kelly’s qualifications and experience when determining the admissibility of his testimony. In the absence of evidence of actual reliability derived from formal evaluation of his techniques, Kelly and the prosecutor emphasized his formal qualifications, his extensive experience as a podiatrist – specializing in foot and ankle conditions, problems with knees and backs that are posture, stance and gait related, and biomechanics (“with a particular interest in sport”) – and his use of video as a diagnostic aid in the treatment of patients.⁹⁰ The BC Court of Appeal accepted that Kelly’s evidence that there was a very strong likeness between the shooter’s gait and that of Aitken was “merely a distillation of his factual observations” and that the introduction of the videos on which Kelly relied at trial permitted the jury to “utilize the knowledge acquired by Mr. Kelly through his training and experience in arriving at their own conclusions regarding the identity of the shooter.”⁹¹ This reasoning has at least two parts: first, Kelly’s clinical experience as a podiatrist was sufficient to equip him to point out unusual gait features; and second, that

⁸⁷ *Aitken CA*, *supra* note 18 at para 80 [emphasis added].

⁸⁸ *Ibid*, at para 73 [emphasis added].

⁸⁹ *Ibid* at para 95 [emphasis added].

⁹⁰ See Kelly’s professional website, online: <www.podiatry.co.uk> (accessed 10 April 2014).

⁹¹ *Aitken CA*, *supra* note 18 at paras 84, 86.

the jury was able to reach an independent judgment about the gait features identified by Kelly. In this section, we focus on the first of these propositions. We return to the second proposition below.

In accepting clinical qualifications and experience as a sufficient warrant of reliability, the BC Court of Appeal appears to have misunderstood Doherty JA's holding in *Abbey* that the reliability of an expert's opinion should be assessed according to the standards that are relevant to his or her field.⁹² Essentially, having found Kelly to be qualified to work as a clinical podiatrist (and to identify gait abnormalities in that context) and having identified his considerable experience in that field, the BC Court of Appeal treated those qualifications and that experience as conclusive of the question of whether Kelly could reliably perform the functions required of a forensic expert, and equally conclusive of the question of whether forensic gait analysis produces reliable forms of identification evidence. This approach is at odds with recommendations given by other forensic podiatrists in one of the few published works about the emerging field.⁹³ We believe that, in relying on qualifications and experience in a related but somewhat different clinical field, the Court adopted the wrong heuristic – or interpretive frame – for assessing the reliability of Kelly's forensic opinion. In fact, qualifications and experience are diversionary. Focusing on these factors diverts the courts' attention away from the need to focus on the reliability of the opinion offered and the underlying technique being used in the instant case.

Forensic gait analysis is a comparison exercise that is structurally similar to other forms of forensic comparison such as DNA analysis, latent fingerprints, ballistics, voice recordings, shoe, tire and foot impressions, bite marks, handwriting and so on.⁹⁴ In each of these fields, the task performed by the analyst is one of comparison between a sample that has some relevant connection with a crime or crime scene and a sample that is known to originate from a specific source. All of these comparison

⁹² *Abbey*, *supra* note 14 at para 119.

⁹³ DiMaggio and Vernon, *supra* note 23 at 16: "Fundamentally, although the scientific aspects of the podiatry knowledge base are used in clinical practice, in forensic podiatry work, the context of practice and the way that science is used in forensic work are fundamentally different." The authors, at 15, had previously warned that "caution is needed in the practice of forensic podiatry in order to ensure that the knowledge used is that which is scientific and robust and not those aspects of a podiatrist's knowledge which are tacit and also may be underdeveloped."

⁹⁴ The image examples might be even more problematic, as we explain below, because lawyers, jurors and judges might think they can do the comparison themselves. Studies suggest that lay people are highly error-prone when asked to compare fingerprints; see Tangen, Thompson and McCarthy, *supra* note 2.

techniques should, according to the NAS, be validated.⁹⁵ Experts should know and be willing to testify about the uncertainties in the technique and the error rate.⁹⁶ Whether one claims the imprimatur of science or not (and, as a reminder, Kelly did claim this), the principle at stake for any comparison-based evidence is how best to evaluate the technique and the expert's work.

During Aitken's trial, Kelly rejected the relevance of such factors as validation (for example, "blind testing" in conditions where the correct answer was known), the determination of indicative error rates, or the disclosure of limitations, while consistently maintaining that his methods were scientific.⁹⁷ What should one make of an individual who is recognized by the court as possessing qualifications and experience but who does not seem to have used appropriate methods in undertaking the analysis and developing his opinions? We consider that the appropriateness of an expert's methods and the correct characterization of the nature of the expert's claims are not simply a discretionary question that can be left to weight or decided by a jury.⁹⁸ Juries are not well-equipped to decide what constitutes science, when scientific methods should apply, and whether (untested) techniques are reliable.

One problem with relying on qualifications and experience as a substitute for a direct assessment of reliability is that qualifications and experience can be impressive even when their relationship with domain-relevant tasks is non-existent or unknown. A domain-relevant task is, for example, the task of identifying features of gait from a low resolution CCTV video of limited duration, given obscurities caused by clothing, structural elements, lighting and the circumstances of the walk. In *Aitken*, Kelly relied heavily on his qualifications and clinical experience, to the exclusion of any independent experimental evidence of his ability to perform the domain-relevant task. Rather than relying upon standardized techniques, validation studies and error rates, attention was directed to Kelly's involvement with the Forensic Podiatry Sub-Committee of the International Association for Identification (IAI).⁹⁹ This membership

⁹⁵ NRC Report, *supra* note 2 at 188-89.

⁹⁶ *Ibid* at 113-24

⁹⁷ Ironically, lack of familiarity with mainstream scientific methods may have made Kelly reluctant to concede limitations. Non-experts are characteristically ill-equipped to identify or respect limits on their knowledge; see Daniel Kahneman and Gary Klein, "Conditions for Intuitive Expertise: A Failure to Disagree" (2009) 64 *Am Psych* 515. Again, this lack of intuitive comprehension reinforces the need to validate techniques so that those using them understand why standards and practices are used.

⁹⁸ See e.g. *Trochym*, *supra* note 7.

⁹⁹ *Aitken* transcript, *supra* note 19 (17 November 2008) Haydn Kelly, cross-examination at 3405.

features prominently on Kelly's *curriculum vitae* and personal web page and was used when his qualifications were rehearsed during testimony. The Committee has produced a paper on the "Role and Scope of Practice" that includes some discussion of forensic gait analysis. The document explains what forensic podiatry is, or purports to be. Given the name of the Committee and its parent organization, one may well expect that the Committee would conduct research or establish scientific standards for the field of forensic podiatry.

The Committee does not, however, address fundamental scientific questions about whether specific techniques actually provide an effective means of identifying similarities or determining the significance of one or more similarities. The Committee has not published standards or made recommendations that might improve the performance of forensic gait analysis. Members do not appear, from documents published by the Committee and from our broader research of published literature, to have endeavoured to study their performance in routine comparison tasks. Indeed, it is far from obvious that all members of the Committee are aware of the kinds of advice and recommendations of authoritative scientific bodies such as the NAS or the US National Institute for Standards and Technology (NIST), and the applicability of reports such as Lord Campbell's inquiry into fingerprints in Scotland.¹⁰⁰ Rather, their preliminary efforts are descriptive, oriented toward carving out a niche for an emerging discipline ("This work is currently the exclusive domain of forensic podiatrists") and creating an idiosyncratic definition of "identification" that is apparently insensitive to decades of controversy in the mainstream (forensic) sciences.¹⁰¹ Similarly, while Kelly advertises himself as "Post Graduate Examiner in Forensic Human Identification and Researcher in Forensic Gait Analysis," he does not appear to have published any peer-reviewed articles about the "scientific" methods that he claims to have developed.

¹⁰⁰ Expert Working Group on Human Factors in Latent Print Analysis, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach* (US Department of Commerce, National Institute of Standards and Technology, National Institute of Justice, 2012); Anthony Campbell, *The Fingerprint Inquiry Report* (APS Group Scotland, 2011). DiMaggio and Vernon's recent text, *Forensic Podiatry*, *supra* note 23, does not feature a single reference to the NRC Report; but see note 107, *infra*.

¹⁰¹ Wesley Vernon is Chair of the IAI Forensic Podiatry Sub-Committee. Kelly described Vernon as "a colleague of mine" and was, presumably, familiar with Vernon's work; see *Aitken* transcript, *supra* note 19 (17 November 2008) Haydn Kelly, cross-examination at 3405. Kelly is thanked and discussed in DiMaggio and Vernon, *supra* note 23 at vii, 103-106.

One of the most prominent claims on Kelly's website is his entry in the *Guinness Book of Records* for being the first person to proffer forensic gait analysis in a criminal court.¹⁰² The BC Court of Appeal cited Kelly's experience in UK courtrooms as a further warrant of the reliability of his techniques.¹⁰³ Shortcomings in the Court's assumptions regarding the equivalence of English and Canadian rules regarding admissibility are discussed in the next section; for now, we note that – as the Goudge Report demonstrates¹⁰⁴ – repeated experience as an expert witness is no guarantee of the reliability of opinions offered by that expert.¹⁰⁵

The problem with all the information given at trial about Kelly's experience and professional networks is that it is basically “noise” in the attempt to assess the reliability of Kelly's opinion evidence. This information is, however, likely to bear on a layperson's assessment of opinion at trial.¹⁰⁶ While it seems likely that an experienced podiatrist might be able to observe similarities and differences between gaits of previously unknown persons, including those depicted in videos, there is no evidence that they perform well at this task, that they are sufficiently cautious when faced with uncertainty, or that their conclusions will be impervious to contextual bias.¹⁰⁷ We have already noted that their traditional methods and long experience are not directed toward identification. Equally, podiatrists engaged in forensic gait analysis do not receive appropriate feedback about the accuracy of their conclusions – the

¹⁰² Online: <www.podiatry.co.uk>. This claim is verified on the Guinness World Records, online <www.guinnessworldrecords.com>.

¹⁰³ *Aitken CA*, *supra* note 18 at para 73.

¹⁰⁴ *Supra* note 1, particularly at 11-14.

¹⁰⁵ See Emma Cunliffe, “Judging, Fast and Slow: Using Decision-Making Theory to Explore Judicial Fact Determination” (2014) 18 *Int'l J Evidence & Proof* 139 at 157-73 for a longer discussion of the psychological literature on expertise and the inadequacies of using experience as an expert witness as a heuristic for reliability.

¹⁰⁶ *Aitken CA*, *supra* note 18 at 81. These issues seem to have created difficulties for judges in the courts of British Columbia. Ironically, the Court of Appeal quoted *R v DD*, [2000] 2 SCR 275: “Faced with an expert's impressive credentials and mastery of scientific jargon, jurors are more likely to abdicate their role as fact-finders and simply attain to the opinion of the expert in their desire to reach a just result.”

¹⁰⁷ Studies suggest that training and experience may have limited value in improving abilities. For example, the ability of passport officers to determine whether two portrait photographs are of the same unfamiliar person is unrelated to the duration of employment; see David White *et al*, “Passport Officers' Errors in Face Matching” (2014) (under review). Another claim, this time from forensic podiatry, seems to have been mistaken. It was once believed that wear patterns on the soles of shoes “would be created by known foot pathologies.” Studies by Vernon and his colleagues found, however, that “this is not the case;” see DiMaggio and Vernon, *supra* note 23, 100-101. These later studies appear to demonstrate a level of awareness, among some forensic podiatrists, of orthodox scientific research methods.

very nature of their role in criminal trials is such that the ground truth of identification can rarely be known.¹⁰⁸ In contrast, podiatrists routinely receive feedback from patients in their clinical practice, and can tailor their diagnoses and interventions accordingly. Furthermore, it may be that even if they perform better than laypersons, clinical podiatrists are not sufficiently good (or error-free) at domain-relevant tasks to meet threshold expectations of reliability.¹⁰⁹

Highly qualified and experienced individuals involved in comparisons have previously been shown to have been mistaken about their ability to usefully discriminate, leading to wrongful convictions. Examples include bite mark and bullet lead comparisons and voice spectroscopy for voice comparisons.¹¹⁰ Significantly, those who proffered opinions derived from techniques that were eventually shown to have been unreliable include those with post-graduate and professional qualifications and experience in, respectively, dentistry, chemistry, linguistics and engineering. Accordingly, qualifications and experience are poor substitutes for validation. In the absence of a validated method and some indication of methodological limitations, error rates and individual proficiency, the courts are left to rely upon highly misleading heuristics such as qualifications, experience (without appropriate feedback), involvement in presumptively authoritative organizations and claims about unpublished research. These are not appropriate criteria to use to assess forensic science techniques and derivative opinions. Rather, courts should initially focus on the value of the underlying method.¹¹¹ Qualifications and experience are important for individuals attempting to validate methods and those using validated

¹⁰⁸ Convictions do not provide genuine feedback because in many cases we do not know if the person is factually guilty. See Cunliffe, *supra* note 105; Sharmila Betts, *A Critical Analysis of Medical Opinion Evidence in Child Homicide Cases* (PhD Thesis, University of New South Wales, 2013), online: <<http://handle.unsw.edu.au/1959.4/52522>>.

¹⁰⁹ The sort of error that a system with a reliability threshold should tolerate, is a separate question.

¹¹⁰ See Erica Beecher-Monas, “Reality Bites: The Illusion of Science in Bite-mark Evidence” (2008) 30 *Cardozo L Rev* 1369; National Academy of Sciences, *Forensic Analysis: Weighing Bullet Lead Evidence* (Washington, DC: The National Academies Press, 2005); National Research Council, *On the Theory and Practice of Voice Identification* (Washington, DC: National Academies Press, 1979).

¹¹¹ We note that judges in other jurisdictions, including those with reliability standards, tend to invest considerable, indeed inappropriate, confidence in experience. See Jennifer Groscup *et al.*, “The Effects of *Daubert* on the Admissibility of Expert Testimony in State and Federal Criminal Cases” (2002) 8 *Psychol, Pub Pol’y & L* 339 at 357, 363-5; Mark Page *et al.*, “Forensic Identification Science Evidence Since *Daubert*: Part I—A Quantitative Analysis of the Exclusion of Forensic Identification Science Evidence” (2011) 56 *J Forensic Sci* 1180.

methods. They do not substitute for formal validation of a comparison-based technique.

3) *The Acceptance of Forensic Gait Analysis by English Courts*

In support of its decision to admit Kelly's testimony, the Court drew upon English jurisprudence and the admission of forensic gait analysis evidence in several English trials. The BC Court of Appeal seemed reassured by these English decisions, and presented them as basically consistent with Canadian admissibility jurisprudence. For example:

... the Court in *Luttrell* rejected the use of scientific indicia of reliability as a general requirement for admissibility: "But a skill or expertise can be recognised and respected, and thus satisfy the conditions for admissible expert evidence, although the discipline is not susceptible to this sort of scientific discipline." This statement appears to be in accord with the analysis set forth in *Abbey*.¹¹²

Specifically, in response to forensic gait analysis, the Court found the approach adopted by the English Court of Appeal, in *R v Otway*, both similar and pertinent:

The criteria for admissibility of expert evidence in the U.K. as outlined and applied in *Otway* require logical relevance, a properly qualified witness and necessity and reliability. As well, there should be a consideration of the dangers associated with expert evidence in terms of its impact on the trier of fact. These criteria are very similar to the analysis applicable in Canada under *Mohan*.

As observed in *Otway*, each application to adduce expert evidence must be individually assessed in any given case. However, there is a high level of factual similarity between *Otway* and the present matter, that, in my opinion, makes the reasoning in *Otway* of assistance here.¹¹³

Recourse to English practice as support for the admissibility of Kelly's evidence offers a precedent-based approach to admissibility; relying on decisions rendered in a jurisdiction that uses quite different admissibility criteria to those that apply in Canada.¹¹⁴ English admissibility standards are lax and probably the most accommodating among advanced common

¹¹² *Aitken CA*, *supra* note 18 at para 91. *R v Luttrell*, [2004] EWCA Crim 1344 was a case where a lip reader, with a substantial error rate, was allowed to express her opinion about whether a suspect caught on CCTV (though without sound) had uttered the incriminating word "Raj." *Luttrell* is a deeply problematic case.

¹¹³ *Aitken CA*, *ibid* at paras 93-94. See *Otway v The Queen*, [2011] EWCA Crim 3.

¹¹⁴ For a comparative discussion of admissibility standards in Canada, England, the USA and Australia, see Edmond *et al*, *supra* note 6.

law jurisdictions. Despite the occasional use of the word, English jurisprudence and practice do not require reliability as a pre-condition to admissibility.¹¹⁵

The BC Court of Appeal's reliance on the decision in *Otway* raises two significant concerns when one considers the Supreme Court of Canada's case law regarding the proper approach to admissibility. First, in *Mohan*, Sopinka J stressed that the admissibility of expert testimony should be assessed on a case-by-case basis, with due regard for the costs and benefits of the evidence in the instant case. The burden of demonstrating admissibility rests on the party proffering the evidence.¹¹⁶ Secondly, in *Trochym*, Deschamps J further emphasized that even longstanding acceptance of the admissibility of a technique will not insulate expert evidence from a well-founded reliability challenge.¹¹⁷ The Supreme Court of Canada has stated clearly that a history of accepting expert testimony from a given field should not inoculate expert testimony from reliability challenges.

Given the confidence invested by the BC Court of Appeal in English case law and practice it is probably not surprising that there are no references within the judgment to a recent review and recommended changes to admissibility standards. In 2011, the Law Commission of England and Wales recommended changing admissibility standards because English judges had become too accommodating. According to its report:

... a number of recent criminal cases suggest that expert opinion evidence of doubtful reliability is being proffered for admission, and placed before the jury, too readily. This follows from the current *laissez-faire* approach to admissibility. It has even been suggested that there may be a "culture of acceptance" on the part of some trial judges, particularly in relation to evidence of a scientific nature.¹¹⁸

In response to the *laissez-faire* approach that prevails toward the admissibility of expert opinion evidence, the Law Commission prepared a draft bill stipulating that "expert opinion evidence is admissible in criminal proceedings only if it is sufficiently reliable to be admitted."¹¹⁹ Section 5 of the draft Criminal Evidence (Experts) Bill explains that this requires opinion evidence to "soundly based" and "the strength to be warranted."

¹¹⁵ Law Commission, *supra* note 6 at paras 2.12-2.16.

¹¹⁶ *Mohan*, *supra* note 15; *Trochym*, *supra* note 7.

¹¹⁷ *Trochym*, *ibid* at paras 31-32.

¹¹⁸ Law Commission, *supra* note 6 at paras 1.17, 1.27, 3.3.

¹¹⁹ *Ibid* at para 144.

To assist the trial judges, the bill lists the kinds of indicia that might inform the assessment, such as whether:

- (a) the opinion is based on a hypothesis which has not been subjected to sufficient scrutiny (including, where appropriate, experimental or other testing), or which has failed to stand up to scrutiny;
- (b) the opinion is based on an unjustifiable assumption;
- (c) the opinion is based on flawed data;
- (d) the opinion relies on an examination, technique, method or process which was not properly carried out or applied, or was not appropriate for use in the particular case;
- (e) the opinion relies on an inference or conclusion which has not been properly reached.¹²⁰

The influence of *Daubert* (and the amended FRE r702) on this list of factors should be obvious. After careful consideration and debate, including the publication of a discussion paper and two opportunities for stakeholders to offer input, the Law Commission settled on the need for all expert evidence to satisfy the reliability standard regardless of whether it was characterized as scientific, non-scientific or based on experience.¹²¹

For a variety of reasons the draft bill is unlikely to be enacted in England, and the *laissez-faire* approach that was criticized by the Law Commission will continue to dictate the admissibility of expert evidence in England and Wales.¹²² Accordingly, English case law and practice offers little assurance in relation to the reliability of expert evidence accepted in English courtrooms. Kelly's participation in previous prosecutions in England and Wales may add to his experience, confidence and performance in the witness box, but they reveal almost nothing about his actual abilities.

¹²⁰ *Ibid.*

¹²¹ *Ibid* at paras 5.71 ff.

¹²² The Ministry of Justice seems to have rejected the Law Commission's recommended reliability framework, without engaging with problems or recommendations, on the basis of impressions about costs; see Ministry of Justice, *The Government's response to the Law Commission report: "Expert evidence in criminal proceedings in England and Wales"* (Law Com No 325) (21 November 2013).

4) *A Note about Novelty*

Defence counsel in *Aitken* made the argument that Kelly's evidence should be subjected to a heightened assessment of reliability because it was novel science that had not previously been accepted in a North American courtroom. In *Mohan*, Sopinka J distinguished between the necessity standard, which applies to most expert evidence, and a requirement that novel scientific evidence be essential. Sopinka J held that novel scientific evidence should be subjected to a threshold reliability analysis and must be essential in the sense that the jury will be "unlikely to come to a satisfactory conclusion" without the evidence. We have already observed that in *J(J-L)* the Court held that the US Supreme Court decision in *Daubert* could assist a trial judge faced with assessing the reliability of novel scientific evidence.¹²³

Some uncertainty about the place of reliability in Canadian jurisprudence arose from the manner in which Sopinka J expressed the threshold reliability requirement in *Mohan*. Several commentators took reliability to be an element of the admissibility enquiry only when working with novel science. Further uncertainty arose from the fact that Sopinka J did not define novel scientific evidence in *Mohan*. *Trochym* largely puts these uncertainties to rest. In *Trochym*, as demonstrated in the passage quoted above, a majority of the Court held that a reliability inquiry was not confined to science that was new to the courtroom. Rather, a trial judge should consider reliability in every case in which the question is fairly posed. Whether one approaches this requirement on the basis that evidence which may be unreliable should be treated as novel science (or more robustly), reads a broad legal reliability standard into other aspects of Sopinka J's judgment in *Mohan*, after *Trochym* the centrality of reliability to every admissibility enquiry is firmly established. As we have already explained, the indispensability of reliability is similarly emphasized in *Abbey* and by Goudge JA in his *Report of the Inquiry into Pediatric Forensic Pathology Services in Ontario*.

In *Aitken*, the BC Court of Appeal rejected defence counsel's claim that forensic gait analysis constituted novel science. Satanove J had also rejected this argument at trial. According to Hall JA:

In evaluating Mr. Kelly's qualifications as an expert witness, the trial judge rejected the argument that his evidence was novel science. She observed, "Podiatry has been in existence for a thousand years and the expertise of a podiatrist to analyze an individual's gait has long been accepted and practiced in a clinical setting" (para. 34). Implicit in this conclusion is a determination that forensic applications of podiatry and

¹²³ *J(J-L)* *supra* note 40; *Daubert* *supra* note 38.

gait analysis do not render the practice “novel” for the purposes of the *Mohan* test. In my respectful opinion, the trial judge did not err in so holding.¹²⁴

Arguably, after *Trochym*, the only relevance of classifying a given field as “novel” is the stiffened requirement that the evidence must be essential in the sense articulated by Sopinka J in *Mohan*. The classification neither determines nor eliminates the need to engage in an analysis of the reliability of the evidence. Nonetheless, the approach adopted by the BC courts to the question of novelty is undesirable. This approach would enable almost any technique or proto-field to identify historical antecedents in order to avoid the more searching enquiry that should, as a matter of trial fairness, be conducted when a given technique is introduced into a criminal court for the first time.

Quite apart from the lack of support offered by Canadian case law for a precedent-based approach to admissibility, there are few principled reasons why the decision of the first trial or appellate courts to admit evidence or hear an admissibility challenge should dictate the admissibility of a technique on all future occasions. The first case may involve an unusual use, or the defence may want the evidence admitted for some tactical reason.¹²⁵ There is no reason why a decision to admit in one case should be determinative, especially if the defence is not resourced, prepared or capable of mounting a credible response.¹²⁶ Rather, courts should be vitally concerned with the reliability of the technique and derivative opinion and this concern should be ongoing.¹²⁷

Interestingly, podiatrists themselves seem to have few doubts about the novelty of gait analysis for the purposes of identification and the forensic application of many aspects of podiatric knowledge. In their publications, and in the publications of engineers developing biometric (that is, algorithmic) models of gait comparison, references to the novelty

¹²⁴ *Aitken CA*, *supra* note 18 at para 73. This use of novelty and the “field” of podiatry is revealing. By way of comparison, the admission of DNA techniques were not premised on the ability to trace biology back to Aristotle, Darwin or even Watson and Crick.

¹²⁵ For example, in *R v Phillion*, [1978] 1 SCR 18, the Supreme Court of Canada refused to permit an accused to rely upon the results of a polygraph test. Had the test results been admitted, future uses by the Crown may well have raised different considerations.

¹²⁶ See Goudge, *supra* note 1 at 457-62 (regarding the importance of supplying adequate legal aid funding for defence counsel and defence experts).

¹²⁷ We have identified in a previous article that courts tend to be reluctant to revisit the admissibility of an expert technique once that technique has been accepted and – worse – that initial caveats or limitations on admissibility tend to be overlooked in future uses. See further Edmond, Cole, Cunliffe and Roberts, *supra* note 6 at 102-103. *Trochym*, *supra* note 7 offers judicial support for the proposition that admissibility may always be challenged, even where a technique has previously been accepted.

of the approach are ubiquitous.¹²⁸ Indeed, the novelty of gait analysis is suggested by Kelly himself. We have already indicated that on his professional website Kelly claims to have “created a world first in forensic gait analysis in July 2000 ... recognized in Guinness World Records.”¹²⁹ Overall, however, we view debates around novelty as distracting. They divert attention from the issue of reliability – can the expert do what he claims and how do we know? Nothing in the purported millennia of podiatry substantially engages with or answers that fundamental question.¹³⁰

3. *Improving the Reliability Assessment*

In the previous section, we suggested that little light is shed on reliability by trying to determine whether a technique is novel (or established) or classifying it as a matter of professional experience (as opposed to “scientific”). Rather, and as emerges from *Aitken*, debates about such classifications are likely to encourage protracted and unhelpful demarcation disputes. Instead, there is a need to determine the kinds of criteria that will enable the technique and derivative opinion to be rationally evaluated. Reliance on forensic science techniques, particularly those involving comparisons linking a trace to an individual or object cannot be sustained by qualifications and experience. These techniques must be independently evaluated in conditions where ground truth (that is, the correct answer) is known. Notwithstanding its judicial classification as non-novel and non-scientific, forensic gait analysis is *susceptible* to formal evaluation according to the techniques described in *Daubert* and in the NRC Report. Susceptibility to evaluation should dictate whether the courts require such evaluation as a condition of admitting evidence. The question of whether a given expert witness is capable of designing, understanding or conceding the importance of the requisite studies is entirely different. Where a witness who uses a comparison-based technique rejects the need for research and formal evaluation, they are putting themselves at odds with the informed conclusions of the NAS, the NIST and numerous other attentive authoritative organizations.

¹²⁸ See e.g. DiMaggio and Vernon, *supra* note 23 at 21, 103, 114. Some examples of these references: “especially in relation to relatively new disciplines such as forensic podiatry;” “forensic gait analysis is the most recent subspecialty of forensic podiatry” and “forensic gait analysis is the most recent and fastest growing subspecialty of forensic podiatry.”

¹²⁹ *Supra* note 102.

¹³⁰ DiMaggio and Vernon, *supra* note 23, 114: “... gait analysis has an abundance of competing ‘theories’ and differences of opinion, few of which are truly research based.”

As a form of expert comparison evidence, forensic gait analysis should be evaluated before analysts are allowed to express incriminating opinions in future Canadian cases. Validation is essential because it enables the court to determine whether a technique works, how well and in what conditions. Validation provides information about performance and limitations and provides data to inform the expression of results. Evaluation is achievable and vastly preferable to the speculative impressions of an expert who, though trained and experienced in clinical work, may have little or no training in forensic techniques or the requirements of criminal investigations.

Recent reviews of the forensic sciences by peak scientific bodies (and others) confirm that pattern recognition or comparison techniques should be formally evaluated. The report prepared by the NRC insisted that:

Although some of the techniques used by the forensic science disciplines – such as DNA analysis, serology, forensic pathology, toxicology, chemical analysis, and digital and multimedia forensics – are built on solid bases of theory and research, many other techniques have been developed heuristically. That is, they are based on observation, experience, and reasoning without an underlying scientific theory, experiments designed to test the uncertainties and reliability of the method, or sufficient data that are collected and analyzed scientifically.¹³¹

And yet:

A body of research is required to establish the limits and measures of performance and to address the impact of sources of variability and potential bias. Such research is sorely needed, but it seems to be lacking in most of the forensic disciplines that rely on subjective assessments of matching characteristics. These disciplines need to develop rigorous protocols to guide these subjective interpretations and pursue equally rigorous research and evaluation programs.¹³²

Failure to attend to validity, reliability and the dangers posed by contextual bias in relation to comparison sciences has serious implications for the reliability of these techniques.

A court that is invited to rely upon a technique that has not been validated must speculate about whether the technique works. For example, Kelly's technique and abilities have not been subjected to evaluation along the lines described by the FBI (and the NAS) and so we have no idea about whether he can do what he claims nor how well. Consequently, the analyst

¹³¹ NRC Report, *supra* note 2 at 128.

¹³² *Ibid* at 8.

presents a subjective conclusion, but provides the jury with no means of rationally assessing his performance. Moreover, the lack of validation and the lack of systematic information about the prevalence of particular gait features (and their in(ter)dependence) means that any significance attached to apparent similarities is again speculative or impressionistic. Kelly wanted to say that the specific features such as abduction and eversion appear only in about one per cent of the population, but we have no idea of the accuracy of this figure nor whether the features he refers to are independent of one another or more likely to appear in combination. Moreover, we have no idea of whether approaching someone with the intention of killing them might affect gait. Setting aside the problems with low quality images, baggy clothing, and footwear, does it even make sense to compare gait in these circumstances?

When it came to expressing his opinion in court, Kelly relied upon a “six point qualitative scale” purportedly “developed by the Forensic Science Service in the United Kingdom, which identifies the degree of similarity as ‘none, weak, moderate, strong, very strong or extremely strong’.”¹³³ Because his opinion was not based on a validated technique we have no way of determining the value of this scale, which ordinarily reflects an expression of the rarity of the observed features within the general population. The lack of information about base rates raises the question of the basis upon which Kelly moved from the interpretation of a few features in the video of the shooter to suggesting that these (alleged) similarities suggested a “very strong likeness” with the implications that opinion carried for the likely identity of the shooter. Here, the fact that Kelly viewed comparison footage for a single suspect is far from insignificant.

Concerns about partisan bias were raised by the defence in *Aitken*. The BC Court of Appeal gave these concerns short shrift, reducing them to an argument about the possibility that Kelly was motivated primarily by profit:

Mr. Kelly’s fitness to give opinion evidence at trial also seems to be questioned on appeal on the grounds that he had a profit motive, charged high fees, engaged in improper billing practices, and was argumentative as a witness and not impartial. I agree with the respondent’s submission that these matters were properly considerations for the jury to entertain in assessing Mr. Kelly’s credibility. They are, in my view, matters of weight rather than of admissibility. It bears noting that the trial

¹³³ See Mike Redmayne *et al*, “Forensic Science Evidence in Question” (2011) Crim LRev 347.

judge instructed the jury to consider whether an expert was impartial in determining what weight to give to the opinion expressed by such person.¹³⁴

The defence questioned Kelly extensively on these issues on the *voir dire* and trial, particularly around his first trip to Canada – flying first class – and the issue of overcharging.¹³⁵

Less attention was paid by counsel and judges to the more insidious issues of contextual bias and cognitive contamination. With respect to the forensic sciences, these phenomena are caused by exposure to information, procedures and other influences that may influence analysis or interpretation in undesirable ways. Exposing the analyst to information about the case or the suspect, for example, has the potential to contaminate because it may suggest a particular outcome. While this is always a risk, that risk is needlessly increased when the analyst does not require the contaminating information in order to perform his or her role.¹³⁶ Complicating things, threats posed by contextual bias and cognitive contamination may operate subtly. Contaminating influences may operate unconsciously, and cannot necessarily be detected by a search for explicit bias or partisanship. The analyst might not know they have been exposed and might not be aware of influences. Even when they do know about exposure it is not always possible to put things out of one's mind, so to speak. Knowing about the dangers does not translate into an ability to resist the insidious effects of these forms of cognitive contamination. Moreover, in most situations an independent observer will not know, and will not be able to ascertain with certainty, if a specific result was the result of some bias or contamination.¹³⁷

Again it is useful to refer to what the NRC Report said about contextual bias:

Some initial and striking research has uncovered the effects of some [cognitive and contextual] biases in forensic science procedures, but much more must be done to understand the sources of bias and to develop countermeasures. ... The forensic science disciplines are just beginning to become aware of contextual bias and the

¹³⁴ *Aitken CA*, *supra* note 18 at para 75.

¹³⁵ *Aitken* transcript, *supra* note 19 (18 November 2008) Haydn Kelly, cross-examination at 3486ff, 3491-3501, (19 November 2008) Haydn Kelly, cross-examination at 3524-26.

¹³⁶ See Goudge, *supra* note 1 at 388-9.

¹³⁷ Hence the importance of shielding or blinding the analyst to information that is not strictly relevant to their analysis. Thorough documentation of all information shared with a given expert may assist with a subsequent inquiry into the possible operation of bias or contamination.

dangers it poses. The traps created by such biases can be very subtle, and typically one is not aware that his or her judgment is being affected.¹³⁸

Based on their concerns about the risks, the NRC Report recommended further research into contextual bias and the development of

standard operating procedures (that will lay the foundation for model protocols) to minimize, to the greatest extent reasonably possible, potential bias and sources of human error in forensic practice. These standard operating procedures should apply to all forensic analyses that may be used in litigation.¹³⁹

Kelly's testimony suggested that he is unfamiliar with methodological strategies that may reduce the risk of contextual bias. He made no attempt to shield himself from contaminating information about the case. He was provided with only one set of images – all of which the police believed to be of Aitken – to compare with the crime scene images. Kelly undertook his analysis with a Canadian police officer in his office, in circumstances where he was comparing the gait from the incident video with that of a single individual whom he knew the police believed to be the murderer.¹⁴⁰ When this issue was touched upon in cross-examination, in response to a defence question whether Kelly should have asked the police for video of a few foils, he responded “there was no need for that to be performed, in – in my view, with regards to my professional experience, the number of occasions I've examined people's gaits over time.”¹⁴¹ Kelly's responses sit awkwardly against the two previous extracts.

Revealingly, recent mistakes with fingerprint attributions seem to suggest that contextual bias is a real problem even where expertise has been demonstrated through validation studies.¹⁴² In recent cases in the United States (Brandon Mayfield) and Scotland (Shirley McKie), highly-experienced examiners who were exposed to domain-irrelevant information made mistakes. In subsequent studies cognitive scientists found that when exposed to domain-irrelevant information, that was suggestive of a particular

¹³⁸ NRC Report, *supra* note 2 at 184-85.

¹³⁹ *Ibid*, Recommendation 5 at 24.

¹⁴⁰ *Aitken* transcript, *supra* note 19 (18 November 2008) Haydn Kelly, cross-examination at 3474, 3488.

¹⁴¹ *Ibid* (18 November 2008) Haydn Kelly, cross-examination at 3475, (19 November 2008) Haydn Kelly, cross-examination at 3536. Defence counsel demonstrated the difficulty of addressing the issue of bias and contamination years after the initial analysis when he attempted to ask Kelly what he knew about the case at various points in time. Kelly had not taken notes of the relevant conversations with the RCMP and testified that he was unable to recall the contents of the discussions; see *ibid* (18 November 2008) Haydn Kelly, cross-examination at 3484.

¹⁴² See Tangen, Thompson and McCarthy, *supra* note 2.

conclusion, experienced examiners changed their opinions about whether two prints matched.¹⁴³

An additional problem, tightly coupled with the dangers from contextual bias, is the risk of cross-contamination and double counting of evidence. Where the forensic analyst is exposed to information about the case or the suspect that is not necessary for their analysis then any ensuing opinion is not independent, especially where that opinion is interpretive and derived from an impressionistic or experiential (that is, subjective) interpretation. The opinion does not provide independent corroboration of the information conveyed because the analyst was unnecessarily exposed to domain irrelevant information that might have (unconsciously) influenced the decision. An additional difficulty associated with both cross-contamination of evidence and contextual bias is the difficulty of exploring these issues at trial. How, for example, do you cross-examine a person about an unconscious threat that could have been avoided through a more appropriate process (such as the double blind processes used in medical research)? What should a trial judge say about the failure, or unwillingness, of an analyst to insulate themselves from influences that have a demonstrated tendency to produce mistakes?¹⁴⁴

Contextual bias is not a straightforward issue; it is complex and obscure. Most jurors, presumably like most judges, will dismiss as trivial (or matters for weight) the threats posed by contextual bias. And yet, medical researchers, air traffic controllers and those operating nuclear power stations, and increasingly forensic scientists, have endeavoured to design systems that avoid the threats of being misled by information that is not necessary for analysis.¹⁴⁵ In the context of *Aitken*, these dangers raise a non-trivial risk that Kelly was mistaken in his conclusions. Exposure to extraneous information might have led Kelly to be more confident than he ought to have been and to present opinions that were in actuality influenced by a range of incriminating evidence as independent opinions based exclusively on podiatric expertise. Moreover, the failure of the police to shield Kelly from highly suggestive incriminating information meant that investigators lost the chance to obtain independent corroboration or disconfirmation of their suspicions through Kelly's opinions. This issue was not developed or addressed at trial or on appeal. We have each written

¹⁴³ Itiel Dror, David Charlton and Ailsa Péron, "Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications" (2006) 156 *Forensic Sci Int* 174. See *R v Bornyk*, 2013 BCSC 1927, 7 CR (7th) 211.

¹⁴⁴ Dror, Charlton and Péron, *ibid*.

¹⁴⁵ See eg National Academy of Sciences, Institute of Medicine, Committee on Quality of Health Care in America, *To Err Is Human: Building A Safer Health System* (Washington, DC: National Academies Press, 1999).

elsewhere about the risks that arise when experts are exposed to domain irrelevant information in the course of performing their duties.¹⁴⁶

4. Conclusion

The simple principle that emerges from the NRC Report's review of expert comparison evidence is that such evidence should not be admitted without sufficient evidence of its reliability:

Two very important questions should underlie the law's admission of and reliance upon forensic evidence in criminal trials: (1) the extent to which a particular forensic discipline is founded on a reliable scientific methodology that gives it the capacity to accurately analyze evidence and report findings and (2) the extent to which practitioners in a particular forensic discipline rely on human interpretation that could be tainted by error, the threat of bias, or the absence of sound operational procedures and robust performance standards.¹⁴⁷

Regardless of the nomenclature adopted by an expert or any other participant in the criminal justice system, if the expert's testimony depends upon visual comparison, it is susceptible to validation studies, proficiency tests, and the adoption of procedures designed to reduce contextual bias. The judicial application of Canadian admissibility jurisprudence to forensic gait evidence adduced by the Crown in *Aitken* circumvented the "very important questions" identified in the NRC Report, thereby deflecting the BC courts from engaging with the reliability of this emerging form of comparison evidence. In this article, we have argued that the admission of forensic comparison and identification evidence should be contingent upon the expert's ability to answer the two questions set out by the NRC Report. This approach is appropriate as a matter of methodological principle and is supported by the approach taken to reliability in the Ontario Court of Appeal decision in *Abbey*.

In order to rationally evaluate expert comparison evidence, a court must have information about validation studies, error rates, blinding procedures adopted to avoid the dangers posed by contextual bias and cross-contamination, and proficiency. In the absence of this information, there are few grounds for an expert to adopt expressions of confidence such as "very strong likeness," and no means of determining the propriety

¹⁴⁶ Emma Cunliffe, *Murder, Medicine & Motherhood* (Hart Publishing: Oxford, 2011); Gary Edmond, Rachel Searston, Jason Tangen and Itiel Dror, "Contextual Bias and Cross-contamination in the Forensic Sciences: The Corrosive Implications for Investigations, Plea Bargains, Trials and Appeals" (forthcoming in *Law, Probability & Risk*).

¹⁴⁷ NRC Report, *supra* note 2 at 9.

of such conclusions. As the NRC Report recognized, traditional adversarial tools such as cross-examination have not safeguarded the reliability of expert comparison evidence, nor countered the influence of contextual bias within experts' opinions. Absent well-designed validation studies and proficiency tests, an expert's qualifications offer a misleading heuristic for reliability, particularly when those qualifications are not directly relevant to the task being performed in the case. Likewise, judicial limitation of an expert's choice of expressions does not address more fundamental concerns about reliability. Even for a relatively vague expression of similarities between trace and source, there is a very real risk that the fact-finder will attribute more weight to the evidence based on high qualifications, experience, and institutional affiliation. When expert comparison evidence is admitted in the absence of appropriate studies and standards, a trier of fact may be misguided about the apparently independent corroboration provided by other evidence or may simply defer to the expert. Admitting such evidence when there is a lack of information about whether the technique works, under what circumstances, and how well places unrealistic and undesirable pressures on trials and those evaluating evidence. Determining such limitations should fall to the analysts and institutions responsible for producing expert testimony, who are best placed to conduct such studies. Forensic analysts and their institutions should formally evaluate techniques and the results of formal evaluation and the application of appropriate procedures should guide admissibility and weight.

Finally, courts should be cautious about recognizing expert "fields" or conferring the imprimatur of admissibility upon claimed abilities unless the evidence provides good grounds to believe that there actually is a field in the sense that there is a network of well-qualified analysts who have adopted standards for the use of techniques and conducted systematic validation studies. Failure to require a reasonably established field undermines the burden of demonstrating threshold reliability, and makes it more difficult for the counter-party (usually the defence) to secure appropriate experts.¹⁴⁸ Admissibility should also depend upon evidence that the analyst actually possesses expertise in the specific task being performed for the court. Unfortunately, trials and appeals do not provide ideal conditions in which to make such assessments of expert performance. Validity and proficiency, like appropriate standards and procedures, need to be determined outside the context of particular cases, in conditions where the ground truth is known, and before evidence is admitted.

¹⁴⁸ Goudge JA supported the proposition that the existence of suitable experts who may confirm or cast doubt on the initial expert's conclusions is an appropriate criterion for admissibility; see *supra* note 1 at 489, citing *R v Melaragni* (1992), 73 CCC (3d) 348 (Ont Ct (Gen Div)) per Moldaver J.