THE CANADIAN BAR REVIEW

LA REVUE DU BARREAU CANADIEN

Vol. 96 2018 No. 3

ABBEY ROAD: THE (ONGOING) JOURNEY TO RELIABLE EXPERT EVIDENCE

Jason M Chin*

Canadian courts draw a tenuous distinction between expert scientific evidence and what they characterize as specialized knowledge gained through the expert witness' experience, training, and research. This characterization is based on unclear criteria and has significant consequences. Notably, specialized knowledge regularly receives considerably less scrutiny than that which is characterized as science, while still often serving as powerful inculpatory evidence in criminal trials. Moreover, specialized knowledge is often provided by figures that carry an air of authority, like police officers and scientists. This article focuses on the leading opinion on specialized knowledge, the Court of Appeal for Ontario's decision in R v Abbey. An analysis of Abbey's application to three fields of contested specialized knowledge (including the evidence the Abbey Court admitted, but which fresh evidence revealed as fundamentally unreliable) provides two general insights. First, while Abbey could be interpreted as providing for a flexible and probing analysis of all expert evidence, courts have often relied on it to justify giving almost no scrutiny to specialized knowledge. Second, this review of the post-Abbey jurisprudence suggests that scrutiny focused on the transparency of the expert's data and analysis, and whether that analysis can reliably be applied to the relevant factual question, may provide a valuable way to evaluate expertise.

* TC Beirne School of Law, University of Queensland; School of Psychology, University of Queensland. Many of the ideas herein were developed collaboratively at the annual meeting of the Evidence-Based Forensics Initiative, Wollongong, NSW. I am especially grateful for discussions with Emma Cunliffe and Gary Edmond. Thank you also to two anonymous reviewers for remarkably perceptive feedback. Rory McFadden and Michael Lutsky provided unflinching editorial and research support.

Les tribunaux canadiens opèrent une distinction ténue entre la preuve d'expert scientifique et ce qu'ils qualifient de connaissances spécialisées acquises par le témoin expert au moyen d'expérience, de formation et de recherche. Cette caractérisation, fondée sur des critères imprécis, est lourde de conséquences. En particulier, les connaissances spécialisées font régulièrement l'objet d'un examen moins minutieux par rapport à ce qu'on qualifie de science, alors qu'un poids considérable leur est souvent accordé en tant qu'élément de preuve incriminante dans le cadre de procès criminels. Qui plus est, ce sont souvent des personnes affichant une certaine autorité qui fournissent ces connaissances spécialisées, tels les policiers et les scientifiques. Le présent article porte sur le jugement de principe en matière de connaissances spécialisées, soit la décision de la Cour d'appel de l'Ontario dans l'arrêt R v Abbey. Deux conclusions générales ressortent d'une analyse de l'application de cet arrêt dans trois domaines de connaissances spécialisées contestées (dont la preuve que la Cour a admise dans l'affaire Abbey, mais qui a été jugée comme étant fondamentalement non fiable une fois de nouveaux éléments de preuve produits). Premièrement, même si l'arrêt Abbey peut être interprété comme proposant une analyse souple et approfondie de toute preuve d'expert, les tribunaux se sont fondés sur cet arrêt à plusieurs reprises pour justifier un examen très peu rigoureux des connaissances spécialisées. Deuxièmement, la présente étude de la jurisprudence qui a suivi l'arrêt Abbey semble indiquer qu'un tel examen axé sur la transparence des données fournies par l'expert et l'analyse qu'il en fait, et la question de savoir si cette analyse est applicable de façon fiable aux questions de fait pertinentes, peut s'avérer un moyen très utile d'évaluer l'expertise.

Contents

1. Introduction

Expert evidence is both immensely important to legal adjudication and equally dangerous. Experts, veiled in their authority and in their (often) superior knowledge, can exert great influence over the trier of fact. One method of safeguarding the trial against prejudicial expert evidence is to exclude it when it is not demonstrably reliable. When determining how to scrutinize the reliability of expert evidence, Canadian courts frequently engage in the (misleading and unhelpful) task of characterization. They characterize expertise as either science or "specialized knowledge" flowing from the witness' training, experience, and research.[1] Both this categorization system and the fact that specialized knowledge regularly receives insufficient scrutiny have received considerable criticism.[2] The leading guidance on the task of assessing specialized knowledge is found

---

[1] See e.g. *R v Abbey*, 2009 ONCA 624 at paras 108–09, 97 OR (3d) 330 [*Abbey* 2009], which distinguished between the "product of scientific inquiry" and "specialized knowledge gained through experience and specialized training in the relevant field." See also *R v Aitken*, 2012 BCCA 134 at paras 79–80, 92 CR (6th) 384 [*Aitken*] comparing an opinion that is "scientific in nature" to specialized knowledge; *R v Awer*, 2016 ABCA 128 at paras 32–49, 37 Alta LR (6th) 62 [*Awer* ABCA], rev'd 2017 SCC 2, [2017] 1 SCR 83 [*Awer* SCC]; Sidney N Lederman, Alan W Bryant & Michelle K Fuerst, *The Law of Evidence in Canada*, 4th ed (Ontario: LexisNexis, 2014) at 808–09 [Lederman]. Formally, expert evidence receives additional scrutiny if it is "novel or contested science or science used for a novel purpose": *White Burgess Langille Inman v Abbott and Haliburton Co*, 2015 SCC 23 at para 23, [2015] 2 SCR 182 [*White Burgess*]. Even this categorization, however, has been called into doubt by recent Supreme Court jurisprudence and the definition of novel science has never been clear: see Jason M Chin & Helena Likwornik, "*R v Bingley* and the Importance of Scientifically Guided Legal Analysis" (2017) 43:1 Queens LJ 33 at 38–39, 47–50 [Chin & Likwornik]. Although, within this article, I am critical of a categorical approach to expert evidence and ultimately advocate for more flexibility, I will use the term "specialized knowledge". This is because the term was featured heavily in *Abbey* (which is this article's focus) and has been widely used in the subsequent case law.

[2] Emma Cunliffe & Gary Edmond, "Gaitkeeping in Canada: Mis-Steps in Assessing the Reliability of Expert Testimony" (2013) 92:2 Can Bar Rev 327 at 361 [Cunliffe & Edmond, "Gaitkeeping"]: "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')." See also Lisa Dufraimont, "New Challenges for the Gatekeeper: The Evolving Law on Expert Evidence in Criminal Cases" (2012) 58:3&4 Crim LQ 531 at 546 [Dufraimont, "Gatekeeper"] saying *R v Abbey* provided for a move away from a "simplistic and value-laden dichotomy between 'hard' and 'soft' science"—I will suggest that while *Abbey* should (and could) have had this effect, much of the time this has not been the case; Gary Edmond & Kent Roach, "A Contextual Approach to the Admissibility of the State's Forensic Science and Medical Evidence" (2011) 61:3 UTLJ 343 at 399 [Edmond & Roach]: "whether we can develop useful means of demarcating science from other types of knowledge and experience are distractions"; Jason M Chin & Scott Dallen, "*R. v. Awer* and the Dangers of Science in Sheep's Clothing" (2016) 63 Crim LQ 527 at 548 [Chin & Dallen].

in a provincial appellate decision, *R v Abbey* (*Abbey* 2009).[3] *Abbey* recently gained attention for another reason—fresh evidence revealed that the expert evidence admitted in that decision was deeply flawed. In this article, I will evaluate how *Abbey* has been interpreted and applied, and then suggest how it ought to be interpreted and applied.

There is much at stake in the decision to admit or exclude the evidence that courts have frequently characterized as specialized knowledge. Notably, and spanning fields as diverse as social science and police experience, specialized knowledge regularly appears as inculpatory evidence in criminal trials.[4] The trial judge's gatekeeper role is especially important in such cases because the defence is often underfunded and thus at a disadvantage in cross-examining the expert evidence and calling its own witness.[5] Still, for the reasons I will discuss throughout this article, courts have not always rigorously scrutinized specialized knowledge.[6]

---

3 *Abbey* 2009, *supra* note 1.

4 In Canada, see *Abbey* 2009, *supra* note 1; *Awer* SCC, *supra* note 1; *Aitken*, supra note 1; Edmond & Roach, *supra* note 2 at 391–95; Cunliffe & Edmond, "Gaitkeeping", *supra* note 2. Evidence that would likely be characterized as specialized knowledge in current Canadian courts was present in the wrongful conviction of Guy Paul Morin: see Ontario, *The Commission on Proceedings Involving Guy Paul Morin: Report* (Toronto: Ministry of the Attorney General, 1998) vol 1 (Commissioner: The Honourable Fred Kaufman) at 56–57. This was evidence from a dog behaviour expert. See also Jason M Chin, Jan Tomiska & Chen Li, "Drawing the Line Between Lay and Expert Opinion Evidence" (2017) 63:1 McGill LJ 1 [Chin, Tomiska & Li] for a trend whereby evidence similar to that of the dog expert's is being tendered and admitted under the more permissive rules for lay opinion. For a review of specialized knowledge evidence in the United States, see David H Kaye, David E Bernstein & Jennifer L Mnookin, "Beyond Kumho Tire: Validating Nonscientific Knowledge", in *The New Wigmore: A Treatise on Evidence* (Alphen aan den Rijn: Aspen Publishers, 2018) at §10.3 [Kaye, Bernstein & Mnookin]; D Michael Risinger, "Defining the 'Task at Hand': Non-Science Forensic Science After *Kumho Tire Co. v. Carmichael*" (2000) 57:3 Wash & Lee L Rev 767 [Risinger]; David E Bernstein, "Expert Witnesses, Adversarial Bias, and the (Partial) Failure of the *Daubert* Revolution" (2008) 93:2 Iowa L Rev 451 at 480–89 [Bernstein], referring to specialized knowledge evidence as "connoisseur" evidence. Note that the effect of culture and context makes generalizations across the US and Canadian evidence jurisprudence precarious: see David M Paciocco, "Context, Culture and the Law of Expert Evidence" (2001) 24:1 Adv Q 42. In Australia, see Kristy A Martire & Gary Edmond, "Rethinking Expert Opinion Evidence" (2017) 40:3 Melbourne UL Rev 967 [Martire & Edmond].

5 See Edmond & Roach, *supra* note 2; Risinger, *supra* note 4 at 773; Ontario, *Harmful Impacts: The Reliance on Hair Testing in Child Protection Report of the Motherisk Commission* (Ontario: MAG, 2018) (Commissioner: The Honourable Judith C Beaman) at 65: "In many cases, the Motherisk hair testing and interpretation chart were admitted on consent at trial with no cross-examination".

6 Other ways of controlling expert evidence are excluding it when it is not beyond the knowledge of the factfinder (i.e., when it is not necessary) and excluding it for bias. For necessity, see David M Paciocco, "Coping with Expert Evidence About Human Behaviour" (1999) 25:1 Queen's LJ 305; Lisa Dufraimont, "Regulating Unreliable Evidence: Can Evidence

This article examines the impact of *Abbey* 2009 on three forms of contentious expert evidence: social science, forensic science, and police experience. I seek to achieve two general aims. First, I will suggest that *Abbey* 2009 has not had the impact that it could have had and attempt to explain why that is the case. In particular, there was a hope that *Abbey* 2009 would have been used by trial courts as guidance in flexibly assessing the reliability of expert evidence in a way that did not require a nebulous exercise in categorizing the evidence as scientific or not.[7] Unfortunately, many post-*Abbey* courts have not taken up that challenge. Rather, the decisions frequently do not engage with *Abbey* at all or simply refer to it as a means to avoid exerting closer scrutiny by simply characterizing the evidence as non-scientific. These courts appear to either see the reliability factors established by *Abbey*, discussed below, as inapplicable, or they overemphasize the portions of the judgment that were deferential to the expert's field and experience within that field.[8] In other words, *Abbey* has simply served as an off-switch for any substantive reliability scrutiny.

Second, this exploration into the post-*Abbey* expert evidence jurisprudence provides some insight into how courts ought to evaluate expert evidence. In particular, *Abbey*'s guidance regarding the openness and transparency of expert knowledge production has been especially useful in providing the legal scaffolding for recent judgments that examined expert evidence with a critical eye.

Because my two aims—an evaluation of *Abbey*'s effect and the lessons that flow from its failures and successes—permeate the remainder of this article, it may be useful to provide a roadmap for the reader. Following

---

Rules Guide Juries and Prevent Wrongful Convictions?" (2008) 33:2 Queen's LJ 261 [Dufraimont, "Regulating Unreliable Evidence"]; Emma Cunliffe, "Without Fear or Favour? Trends and Possibilities in the Canadian Approach to Expert Human Behaviour Evidence" (2006) 10:4 Int'l J Evidence & Proof 280; Jason M Chin & William E Crozier, "Rethinking the Ken Through the Lens of Psychological Science" Osgoode Hall LJ [forthcoming in 2018] [Chin & Crozier]. For bias, see Bernstein, *supra* note 4; David M Paciocco, "Unplugging Jukebox Testimony in an Adversarial System: Strategies for Changing the Tune on Partial Experts" (2009) 34:2 Queen's LJ 565; Paul Michell & Renu Mandhane, "The Uncertain Duty of the Expert Witness" (2005) 42:3 Alta L Rev 635; Dufraimont, "Gatekeeper", *supra* note 2 at 551–56; Emma Cunliffe, "A New Canadian Paradigm? Judicial Gatekeeping and the Reliability of Expert Evidence" in Paul Roberts & Michael Stockdale, eds, *Forensic Science Evidence and Expert Witness Testimony: Reliability Through Reform?* (Cheltenham: Edward Elgar, forthcoming in 2018) [Cunliffe, "Paradigm"].

7 See above, n 2.

8 See Cunliffe & Edmond, "Gaitkeeping", *supra* note 2 at 351, making this observation in their analysis of the *Aitken* case: "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".

this introduction, I will detail *Abbey* itself and contextualize it in the rest of Canada's expert evidence landscape in Part 2.

Part 3 begins my exploration of *Abbey*'s impact on three fields of expert evidence. Part 3 discusses *Abbey*'s impact on social scientific evidence, primarily on the sociological expertise at issue in the *Abbey* proceedings and in a case in which the same expert was proffered by the defence. In these cases, the courts engaged with the reliability of the evidence but may have given too little emphasis to transparency and too much to the appearance that the expert followed the standards of the field. In Part 4, I examine forensic scientific evidence. Here, *Abbey* has frequently served as an exemption to any substantive reliability scrutiny by allowing parties to avoid the scrutiny that novel or contested science typically receives. For evidence that is not susceptible to formal testing, I suggest reliable expertise may be inferred in some circumstances. Part 5 then examines *Abbey* in the context of police knowledge. I find that courts often do not apply *Abbey* at all, and simply advert to the expert's general experience and status as a police officer. However, when they do engage with *Abbey*, demanding a transparent knowledge base provides some safeguards as to reliability. Part 6 concludes with a summary of my main findings through what I describe as a "transparent proficiency" approach to expert evidence.

2. Abbey and the Reliability of Specialized Knowledge

Abbey 2009 was decided during a sensitive time in the development of Canada's law of expert opinion evidence. It was heard just months after Justice Goudge's *Inquiry into Pediatric Forensic Pathology in Ontario*, which detailed, among other things, scientifically invalid practices that contributed to at least 14 wrongful convictions.[9] At the same time, Canadian courts were subjecting evidence from the experimental sciences to increasing scrutiny.[10]

---

[9] Ontario, *Inquiry into Pediatric Forensic Pathology in Ontario: Report* (Toronto: Ministry of the Attorney General, 2008) vols 1–4 (Commissioner: The Honourable Stephen T Goudge) [Goudge Report]; Emma Cunliffe & Gary Edmond, "What Have We Learned? Lessons from Wrongful Convictions in Canada" in Benjamin Berger, Emma Cunliffe & James Stribopoulos, eds, *To Ensure that Justice is Done: Essays in Memory of Marc Rosenberg* (Toronto: Thomson Reuters, 2017) at 133. For the role of expert evidence in wrongful convictions generally, see Bruce A MacFarlane, "Wrongful Convictions: Determining Culpability When The Sand Keeps Shifting" (2014) 47:2 UBC L Rev 597 [MacFarlane, "Wrongful Convictions"].

[10] *R v J(LJ)*, 2000 SCC 51, [2000] 2 SCR 600 [*J(LJ)*]; *R v Trochym*, 2007 SCC 6, [2007] 1 SCR 239; Goudge Report, *supra* note 9 at 475–87; Lederman, *supra* note 6 at 803–18. See also Dufraimont, "Gatekeeper", *supra* note 2 at 533–47 for a lucid review of this law.

A) Situating *Abbey*: The Beginning of the Road

The foundations of modern expert opinion law in Canada are found in *R v Mohan*.[11] In that case, Justice Sopinka provided four requirements: (1) relevance; (2) necessity in assisting the factfinder; (3) no other applicable exclusionary rules; and (4) a properly qualified expert.[12] An opinion passing those hurdles must also withstand the trial judge's residual discretion to exclude evidence when the opinion's costs (e.g., potential to confuse and overwhelm the factfinder) outweigh its benefits to the trial process (e.g., assistance to the factfinder).[13] Further, Justice Sopinka said that expert opinions advancing a "novel scientific theory or technique" should be able to withstand "special scrutiny" to determine if they are reliable.[14]

Following *Mohan*, two Supreme Court of Canada cases elaborated on what special reliability scrutiny entailed. These two decisions followed the US Supreme Court's landmark decision, *Daubert v Merrell Dow Pharmaceuticals Inc*.[15] *Daubert* added to the trial judge's gatekeeping responsibility, holding that the judge should take neither the expert's nor the field of expertise's word for it that the science is reliable.[16] The US Supreme Court provided four factors to assist in this gatekeeping endeavor: (1) whether and how the science had been tested; (2) the error rate of the science; (3) the peer-review and publication status of the science; and (4) whether the science is generally accepted in the scientific community from which it came.[17]

Daubert was applied by the Supreme Court of Canada in *R v J (LJ)*.[18] The Supreme Court characterized an expert's psycho-physiological tests (designed to ascertain whether a person is a pedophile) as novel science.[19] Among other shortcomings, the opinion was insufficiently tested and error-prone, and thus inadmissible.[20] In *R v Trochym*, decided two years before *Abbey* 2009, the Supreme Court again relied on the *Daubert* factors. This time, the Court excluded testimony from a witness because her memories

---

11 *R v Mohan*, [1994] 2 SCR 9, 89 CCC (3d) 402 [*Mohan*].

12 *Ibid* at paras 16–49.

13 *Ibid* at paras 22–24.

14 *Ibid* at para 32. Further, note that opinion evidence is presumptively inadmissible and so the party tendering bears the burden of proving its admissibility.

15 509 US 579 (1993), 113 S Ct 2786 [*Daubert*].

16 Prior to *Daubert*, the leading federal case in the US was *Frye v United States*, 54 App DC 46, 293 F 1013 (1923), which provided a more deferential standard whereby judges determined if the science was generally accepted in the scientific community. *Frye* was never the law in Canada: *R v T(JE)*, 25 WCB (2d) 490 at para 74, 1994 CarswellOnt 3370 [*T(JE)*].

17 *Daubert*, *supra* note 15 at 592–94.

18 *J(LJ)*, *supra* note 9.

19 *Ibid* at paras 33–36.

20 *Ibid* at paras 50–55.

had been retrieved and clarified through the non-novel (but contested) practice of hypnosis.[21]

Before addressing the expert opinion in *Abbey*, it is worth elaborating on the case's place in the pantheon of opinion evidence decisions. First, *Abbey*'s influence has gone well beyond its treatment of specialized knowledge. The decision, written by Justice Doherty, has been widely cited for the direction that trial courts should establish the scope of expert evidence and actively ensure experts do not go beyond that scope.[22] Moreover, *Abbey* employed a two-step formulation of the *Mohan* test. The two-step process drew favour from courts across Canada and was eventually adopted, with minor modifications, by the Supreme Court in *White Burgess Langille Inman v Abbott and Haliburton Co*.[23]

White Burgess is the most current thorough enunciation of the expert evidence rules. Under the first step of the test, the opinion must meet four preconditions: *logical* relevance, absence of an exclusionary rule, a properly qualified expert, and necessity (note *Abbey* had relegated necessity to the second stage).[24] Further, novel or contested science must receive special reliability scrutiny (likely under *Daubert*; *White Burgess* left the treatment of anything that could not be characterized as novel or contested science, ostensibly, to *Abbey*).[25] If the evidence passes the first step, only then does it receive the discretionary costs-benefits weighing, which also includes reliability and any bias or partiality the expert may possess.[26]

Second, after *Daubert*, US courts have also struggled with expert evidence that does not lend itself easily to formal testing.[27] The leading case on this issue is *Kumho Tire Co Ltd v Carmichael*, a 1999 decision of the

---

21 *Supra* note 10.

22 *Abbey* 2009, *supra* note 1 at paras 62–70. Although this was not a new idea, see *R v Ranger*, 67 OR (3d) 1, 178 CCC (3d) 375; Helena Likwornik, "Overstepping and Sidestepping: The Expert Evidence Dance" (2017) 35:4 Adv J 24 [Likwornik].

23 *White Burgess*, *supra* note 1. Another important post-*Abbey* development is the Supreme Court's decision in *R v Sekhon*, 2014 SCC 15, [2014] 1 SCR 272 [*Sekhon*]. That case will be explored in Part 5. For a helpful description of *Abbey*'s reformulation of the Mohan test, see Lisa Dufraimont's annotation accompanying *R v Boswell*, 2011 CarswellOnt 2428.

24 *White Burgess*, supra note 1 at para 23.

25 *White Burgess* cites the paragraphs of *J(LJ)* detailing the *Daubert* analysis.

26 According to *White Burgess*, a threshold level of partiality is required at the first step (a properly qualified expert). A review of the post-*White Burgess* case law suggests that partiality may be a fertile ground for challenging experts going forward: see Cunliffe, "Paradigm", *supra* note 6.

27 For commentary on the *Daubert-Kumho* interstitial period, see Edward J Imwinkelried, "The Next Step After *Daubert*: Developing a Similarly Epistemological Approach to Ensuring the Reliability of Non-Scientific Testimony" (1994) 15:6 Cardozo L Rev 2271.

US Supreme Court.[28] In that case, the Court held that *all expert evidence* should have a reliable basis and thus the *Daubert* factors should be flexibly applied to these opinions.[29] The Court also remarked that it was important that the witness use "the same level of intellectual rigor" as experts in his or her field.[30] Influential American commentators and academics have criticized a post-*Kumho* trend whereby some courts have seemed overly reliant on these standards within the relevant field, thus allowing dubious specialized knowledge to pass the gatekeeper.[31] D. Michael Risinger has labeled standards which defer to the internal quality controls of a field as "guild tests".[32] *Kumho*, as we will see below, had a clear impact on the Court of Appeal's decision in *Abbey* 2009.

B) *R v Abbey*: The Tattoo and the Sociologist

Abbey is widely regarded as Canada's leading decision on specialized knowledge. This is not to say that it was the first Canadian decision to suggest a broad role for reliability in the assessment of all expert evidence.[33] These cases, however, did not elaborate on how reliability should be assessed and none has been as influential as *Abbey*.[34] Moreover, it should be noted

---

28 526 US 137 (1999), 119 S Ct 1167 [*Kumho*].

29 *Ibid* at 141, 147: "In *Daubert*, this court held that Federal Rule of Evidence 702 imposes a special obligation upon a trial judge to 'ensure that any and all scientific testimony … is not only relevant, but reliable.' The initial question before us is whether this basic gatekeeping obligation applies only to 'scientific' testimony or to all expert testimony. We, like the parties, believe that it applies to all expert testimony".

30 *Ibid* at 152.

31 For a review, see David L Faigman et al, *Modern Scientific Evidence: The Law and Science of Expert Testimony*, revised ed (Eagan: Thomson Reuters, 2016) at §1:28; Risinger, *supra* note 4 at 775–78. Similarly, in Australia, academics have eschewed a general acceptance standard for one focused on demonstrable reliability: see Gary Edmond, "Forensic Science Evidence and the Conditions for Rational (Jury) Evaluation" (2015) 39:1 Melbourne UL Rev 77.

32 Risinger, *supra* note 4 at 770.

33 *T(JE)*, *supra* note 16 at paras 73–77; *R v McIntosh*, 35 OR (3d) 97 at paras 14–19, 117 CCC (3d) 385 [*McIntosh*]; *R v F(DS)*, 43 OR (3d) 609 at para 45, 169 DLR (4th) 639; *R v K(A)*, 45 OR (3d) 641 at paras 84–87, 176 DLR (4th) 665 [*K(A)*]. For a review of reliability's general role in the admissibility decision, see David M Tanovich, "*R v. Hart*: A Welcome new Emphasis on Reliability and Admissibility" (2014) 12 CR (7th) 298 at 301–03.

34 Note *McIntosh* serves as a *de facto* bar to eyewitness identification expertise, holding that such expertise failed *Mohan*'s necessity criterion. It was also critical of the field's (applied cognitive psychology) scientific status. See Dufraimont, "Regulating Unreliable Evidence", *supra* note 6 at 269–70; Chin & Crozier, *supra* note 6.

that some pre-*Abbey* decisions denied there was any reliability requirement (and some continue to do so, as I will detail in Parts 3–5).[35]

R v Abbey encompasses a series of murder trials and appeals set in the context of violent gang wars in the Toronto area during the early 2000s.[36] Warren Abbey, an alleged gang member, was accused of murdering Simeon Peter. Under the Crown's theory, Abbey shot Peter, confusing him with a rival gang member who had robbed Abbey earlier. Identity was the central issue at trial.[37] Peter's killer had followed him down a Toronto area street and shot him at least three times.[38]

Abbey was acquitted at the first trial after Justice Archibald excluded the Crown's key expert witness, a sociologist named Dr. Mark Totten.[39] Totten would have opined that a teardrop tattoo Abbey received after the killing meant he had killed a member of an opposing gang: "it is clear to me that Warren Abbey's teardrop tattoo on his right cheek below the eye represents the fact that he killed a rival gang member, most likely in 2004."[40] Alternatively, Totten would provide more general evidence about the three possible meanings of teardrop tattoos within gang culture: the bearer had suffered the loss of a family or fellow gang member, the bearer had served time in prison, or the bearer had killed a rival gang member.[41]

Justice Archibald, at the first trial, excluded the totality of Totten's opinion pursuant to his gatekeeping discretion.[42] In particular, he found

---

35 See *R v O(N)*, 2009 ABCA 75 at para 21, 2 Alta LR (5th) 72 [*O(N)*], citing *R v Chan* (1993), 145 AR 304 at para 9, 87 CCC (3d) 25: "The only requirement for the admission of expert opinion is that the 'expert witness possesses special knowledge and experience going beyond that of the trier of fact' … The admissibility of [expert] evidence does not depend on the means by which the skill was acquired." Troublingly, some still deny a reliability requirement: see *R v Woodcock*, 2010 ONSC 671 at para 12, 87 WCB (2d) 630 [*Woodcock*].

36 See *R v Abbey*, 2017 ONCA 640 at paras 13–41, 140 OR (3d) 40 [*Abbey* 2017] for a summary of the proceedings.

37 *Ibid* at para 13.

38 *Abbey* 2009, *supra* note 1 at para 7.

39 *R v Abbey* (2007), 73 WCB (2d) 411 at paras 13–16, 2007 CarswellOnt 376 (WL Can) (Sup Ct J) [*Abbey* 2007]. In a 2012 case, he described himself as "an expert witness and Canadian expert on gangs": *R v Gager*, 2012 ONSC 1472 at para 33, 100 WCB (2d) 285 [*Gager* Totten].

40 *Abbey* 2017, *supra* note 36 at para 61. Alternatively, the Crown proffered Totten to opine that the tattoo carried three possible meanings: (1) killing a rival gang member; (2) losing a loved one or gang member or; (3) spending time in prison. Due to the other facts of the case, (1) was the only plausible option of the three. See *Abbey* 2007, *supra* note 39 at para 18.

41 *Abbey* 2009, *supra* note 1 at paras 33–54, 98–99.

42 *Abbey* 2007, *supra* note 39 at para 93.

that Totten's evidence was not reliable[43] and that it cut right to the ultimate issue (i.e., identity).[44] As to reliability, Justice Archibald characterized Totten's evidence as "novel science" and thus applied the *Daubert* factors to "measure the reliability of his methodology."[45] First, while Totten had interviewed approximately 290 gang members, it was unclear if those experiences could be applied to Abbey. More specifically, Totten's sample was not random[46] and other than screening out some gang members who seemed like they might lie, his inclusion criteria were unclear.[47] As a result, no reasonable error rate could be ascribed to Totten's conclusion. In other words, there was no way of systematically comparing Totten's gang research to the context of the killing.[48]

Justice Doherty, writing for the Court in *Abbey* 2009, held that the trial judge erred in excluding Totten's evidence and he did so for two general reasons: the opinion overreached but could have been circumscribed as to be admissible, and it should not have been judged against a rigid scientific standard (i.e., *Daubert*). As to the first reason, Justice Doherty seemed to agree with the lower court decision that Totten could not opine on the reason that Abbey got his particular tattoo (the Crown's primary position).[49] He held, however, that Totten *could* provide framework evidence about the three possible meanings of teardrop tattoos in gang culture.[50]

The distinction drawn by the Court of Appeal is well known in evidence law scholarship. David Faigman, John Monahan, and Christopher Slobogin refer to the problem as "G2i", or the difficulty of reasoning from group (G) data to the individual (i) level.[51] To drastically simplify this line of research, evidence researchers tend to agree that there is insufficient data in many fields to reliably individuate group findings to specific cases.[52]

---

43 For a succinct summary of Justice Archibald's reliability analysis, see *Abbey* 2007, *ibid* at para 4.

44 *Ibid* at para 46.

45 *Ibid* at para 49. For the totality of the *Daubert* analysis, see paras 47–88.

46 *Ibid* at paras 50–58.

47 *Ibid* at para 78.

48 *Ibid* at paras 59–62.

49 *Ibid* at para 61.

50 *Ibid* at paras 69–70.

51 John Monahan, Laurens Walker & Gregory Mitchell, "Contextual Evidence of Gender Discrimination: The Ascendance of 'Social Frameworks'" (2008) 94:7 Va L Rev 1715; David L Faigman, John Monahan & Christopher Slobogin, "Group to Individual (G2i) Inference in Scientific Expert Testimony" (2014) 81:2 U Chicago L Rev 417 [Faigman G2i].

52 For example, applied cognitive psychologists can provide reliable evidence about the factors that influence eyewitness memory, but they cannot say, with any degree of certainty, how likely a particular eyewitness' memory is to be accurate. See Faigman G2i, *supra* note 51 at 432–34.

The second major mistake the trial court made, according to the Court of Appeal, was excluding the group-level research. This was Totten's conclusion about the three potential meanings of teardrop tattoos, drawn from the years of field research that Totten had conducted. The Court held that Justice Archibald should not have applied the *Daubert* factors because Totten's research was not scientific in nature: "It was not scientific. It was not novel. And it was not a theory."[53] Rather, Justice Doherty characterized the evidence as "specialized knowledge gained through extensive research."[54]

Importantly, the specialized knowledge characterization was not the end of the analysis. Justice Doherty held that the opinion still needed to be reliable, but that the reliability inquiry should be conducted according to the "nature of the opinion".[55] He then provided nine questions to assist with that inquiry.[56] Approximately half of these factors deferred to the standards of the field of expertise. The others were directed at accurate recording of any data, clarity of reasoning, use of methodology, and collecting data independent of the case. I will revisit these factors throughout this article, so they are worth reproducing in full now:

- To what extent is the field in which the opinion is offered a recognized discipline, profession or area of specialized training?

- To what extent is the work within that field subject to quality assurance measures and appropriate independent review by others in the field?

- What are the particular expert's qualifications within that discipline, profession or area of specialized training?

- To the extent that the opinion rests on data accumulated through various means such as interviews, is the data accurately recorded, stored and available?

- To what extent are the reasoning processes underlying the opinion and the methods used to gather the relevant information clearly explained by the witness and susceptible to critical examination by a jury?

---

53 *Abbey* 2009, *supra* note 1 at para 116.
54 *Ibid* at para 108.
55 *Ibid* at para 118.
56 *Ibid* at para 119.

- To what extent has the expert arrived at his or her opinion using methodologies accepted by those working in the particular field in which the opinion is advanced?

- To what extent do the accepted methodologies promote and enhance the reliability of the information gathered and relied on by the expert?

- To what extent has the witness, in advancing the opinion, honoured the boundaries and limits of the discipline from which his or her expertise arises?

- To what extent is the proffered opinion based on data and other information gathered independently of the specific case or, more broadly, the litigation process?[57]

Critically, Justice Doherty went on to quote *Kumho* and emphasize the much maligned "same intellectual rigour" language (i.e., the portion of the judgment that hinted at a guild test).[58]

Here, it may be useful to summarize the framework that *Abbey* now fits into.[59] Under *White Burgess*, expert evidence must be logically relevant, necessary, absent an exclusionary rule, and from a properly qualified expert who meets a threshold level of unbiasedness, impartiality and independence.[60] Then, at the second stage, the benefits of that evidence must outweigh the costs, a calculus that includes, necessity, reliability and bias. Novel or contested scientific evidence, which has never been clearly defined,[61] receives special scrutiny under *Daubert*. Other expert evidence is (notionally) then subject to the guidelines in *Abbey* 2009.

Applying the nine reliability factors, the appellate court found that Justice Archibald had given insufficient weight to the fact that Totten's opinion seemed to have used the well-established social-scientific method of drawing inferences from interviews.[62] Further, the defence had not challenged the recording of the data (although formally, the onus is on the party tendering opinion evidence to establish its admissibility) and Totten's

---

57 *Ibid.*

58 *Ibid* at para 120.

59 For a more thorough summary, see Lisa Dufraimont's annotation accompanying the Carswell report of *White Burgess*, *supra* note 1 (2015 CarswellNS 313) (WL Can). See also Chin & Likwornik, *supra* note 1 at 47–48.

60 See *White Burgess*, *supra* note 1 at paras 23–24.

61 Chin & Likwornik, *supra* note 1 at 49–50.

62 *Abbey* 2009, *supra* note 1 at para 123.

opinion was not too complex for the jury to scrutinize.[63] Justice Doherty also noted that the data had been collected independently years prior to the case.[64] As a result, the opinion was admissible. Totten gave his evidence in the second trial and Abbey was convicted.[65]

In many cases, this would be the end of the story: an expert had provided (seemingly) powerful inculpatory evidence; the defence had not produced its own expert but had cross-examined the expert about the reliability of his method; much judicial ink had been spilled about the admissibility of the opinion. Indeed, this is often much more than an accused can expect.[66] In *Abbey*, however, the case went on to follow an even more aberrant path—in the years after Abbey's conviction, Totten provided expert evidence for a co-accused in *R v Gager*.[67] In that case, the Crown's probing cross-examination of Totten revealed serious flaws in his methodology that the earlier *Abbey* proceedings had failed to unearth.

3. Social Science and the Importance of Transparency

The revelation from *Gager* that much of Totten's evidence was fundamentally flawed suggests that Totten's use of accepted standards within sociology may have been misleading. In this section, I will review the flaws in Totten's opinion and demonstrate that (besides some seemingly overt dishonesty) they are not very different from widespread scientific practices that have been criticized as part of a movement in science towards more transparency and openness.[68] More generally, this section finds that the social sciences typically do draw some substantive reliability analysis. That analysis, however, should include more attention to the transparency of the expert's methods, following from emerging guidelines from scientists, publishers, and funding bodies.[69]

In *Gager*, Totten's evidence would have supported the defence's contention that the co-accused (Gager) was not a gang member, but merely a friend or affiliate of a gang. In assessing Totten's evidence, Justice Clark

---

63 *Ibid* at paras 122, 125.

64 *Ibid* at para 124. Recent meta-scientific research finds that even previously collected research can easily be framed in a misleading way: see n 94 below.

65 *Abbey* 2017, *supra* note 36 at paras 63–68.

66 See above, n 5.

67 *Supra* note 39.

68 Marcus R Munafò et al, "A Manifesto for Reproducible Science" (2017) 1:1 0021 Nature Human Behaviour 1–9 (article number 0021) [Munafò]. For a cogent analysis of the impact of irreproducible science on patent law, see Jacob S Sherkow, "Patent Law's Reproducibility Paradox" (2017) 66:4 Duke LJ 845 [Sherkow].

69 Marcia McNutt, "Taking up TOP" (2016) 352:6290 Science 1147 [TOP Guidelines].

applied the nine factors in *Abbey* 2009.[70] His analysis revealed an important tension in those factors: Totten appeared, for the most part, to be using accepted methods, but he seemed to have twisted those methods to arrive at the conclusion he was seeking.

As to the standards in the field, Justice Clark noted that sociology was a recognized discipline with quality assurance measures.[71] Moreover, he found that, by and large, Totten "arrived at his opinions using methods accepted by those working in the sociological field"[72] and "honoured the boundaries and limits of his discipline."[73] Despite these concessions to accepted practices, Justice Clark was deeply critical of Totten's data collection and retention practices (the fourth *Abbey* 2009 factor).[74] First, he found that Totten's raw data had been destroyed, thus making it difficult to test his claims. This would have been useful because Totten's summary and analysis evinced many inconsistencies. For instance, Totten (as he did in *Abbey*) made conclusions about the meaning of tattoos, but many of the studies he relied on did not seem to be about tattoos, nor did their protocols seem to include questions about tattoos.[75]

Justice Clark further noted that the number of actual gang members in Totten's sample was dubious because he had shifted his definition of what it means to be a gang member over time. As a result, by his most recent definition, only "10 to 20" of the 90 putative gang members presented in his report would now qualify as such and "[s]urprisingly, he could not be more

---

70 *Gager* Totten, *supra* note 39 at paras 29–95.

71 *Ibid* at para 31.

72 *Ibid* at para 56.

73 *Ibid* at para 63; The decision in *Gager* is replete with descriptions of Totten's methods as well in keeping with generally accepted sociological practices. See *ibid* at para 32: "he has both academic credentials as a sociologist along with considerable knowledge and experience concerning street gangs acquired within pursuit of his discipline." *Ibid* at para 44: "Dr. Totten made much in his evidence on the *voir dire* of the fact that his writings are all peer reviewed. While that may be true, the fact remains that, except for two ongoing studies, for the most part, the data upon which his opinions rest has been destroyed." *Ibid* at para 55: "Given the witnesses' academic training, I am confident that he can clearly explain the methodology he employed to gather the data he did and the reasoning he used to interpret it in such a way that it can be critically examined by the jury." *Ibid* at para 62: "With the exception of the use of 519 as the overall sample size and the gathering of the data concerning the prevalence of tattoos other than the tear drop, the methodologies employed would appear to be in keeping with those recognized in the field of sociology".

74 *Abbey* 2009, *supra* note 1 ("To the extent that the opinion rests on data accumulated through various means such as interviews, is the data accurately recorded, stored and available?" at para 119).

75 *Gager* Totten, *supra* note 39 at para 45. Justice Clark remarked: "absent any record of what was asked, it strikes me that the validity of any data gathered in this fashion is questionable at best".

precise than that."[76] Further, some aspects of the data collection seemed plainly implausible. Totten said that he had interviewed 309 participants in a month, a number that, according to the Court's math, would have made it difficult or impossible to have also slept and eaten.[77]

Finally, Justice Clark questioned a category of gang affiliation (i.e., "long term friend" of the gang) that Totten used to suggest that Gager was not in the relevant gang, but rather tightly allied with it.[78] In particular, Totten modified a diagram he had used in previous publications to add that long term friend category.[79] He also admitted that he had never written about this category prior to his expert report in *Gager*.[80] According to Justice Clark, this apparently *ad hoc* characterization suggested partiality.[81]

While Justice Clark accepted at several points that Totten had used the methodologies of his field, he did not do so universally.[82] In particular, Justice Clark found it improbable that academic sociologists would inflate their sample sizes by exploiting flexible categories like gang membership and draw such strong conclusions about a tattoo's meaning when such data did not seem to be planned or recorded.[83] While Justice Clark was correct to independently question Totten's methodological rigour, he would likely be surprised to find out that recent studies find that academic psychologists, ecologists, and biologists have admitted to using similar methods at rates approaching 60%.[84] I will discuss these findings further below.

Totten's cross-examination provided the fresh evidence Abbey needed to support an appeal in 2017.[85] The Court ordered a third trial, finding that in light of the weaknesses exposed in *Gager*, Totten's evidence was too unreliable to pass a gatekeeper's weighing of the evidence's risk of prejudice against its probative value.[86] More specifically, as in *Gager*, Totten's sample

---

76 *Ibid* at para 49.

77 *Ibid* at para 54. See Risinger, *supra* note 4 at 788, describing an expert who testified as to having performed over a million document examinations, a similarly impossible figure.

78 *Gager* Totten, *supra* note 39 at paras 64–75.

79 *Ibid* at para 64.

80 *Ibid* at para 74.

81 *Ibid* at para 75.

82 *Ibid* at paras 56–61.

83 Justice Clark questioned whether it was acceptable to rely on data that was two decades old: *ibid* at para 61.

84 See n 93, below. Further, current norms in many sciences do not require the scientist keep a public record of data (although it is a current reform proposed as part of the open science movement). Justice Clark expressed surprise at this practice of Totten's but seemed to accept the practice.

85 *Abbey* 2017, *supra* note 36.

86 *Ibid* at para 121.

size (i.e., the number of interviews he conducted) was unverifiable and likely lower than he represented at trial.[87] This was because he used a flexible definition of gang member and counted the same interviews as two separate data points, thus inflating his reported sample size.[88] And, also as in *Gager*, there was no record—and it did not seem likely—that the interviewees were explicitly asked about their tattoos.[89]

In holding that Totten's evidence should be excluded, Justice Laskin, writing for the Court, adverted to the list of nine reliability factors in *Abbey* 2009.[90] He also referenced the Goudge Report's recitation of 14 factors, which Justice Goudge extracted from *R v Johnston*.[91] Justice Laskin noted two of those factors bore on appeal before the Court: the data being recorded accurately (from *Abbey* 2009), and the opinion accurately representing the data and studies on which it is based (this factor does not actually appear in either list).[92]

As I mentioned above, some of the tactics Totten used to make his data seem more credible are unfortunately not entirely uncommon in science. They are known as "researcher degrees of freedom" or, in other words, flexibilities in the scientific process that do not amount to outright fraud but bias the scientific literature.[93] They include excluding data in a manner that supports the researcher's hypothesis, deciding to end data collection after a statistically significant result has been found, and using flexible definitions of key measures.[94]

Recent scientific reforms being implemented as part of the "open" or "reproducible science" movement can help control for (or at least illuminate) these researcher degrees of freedom and allow for rational evaluation of the evidence.[95] For instance, several leading journals now require authors to

---

87 *Ibid* at para 117.

88 *Ibid* at paras 72–82, 101–04.

89 *Ibid* at paras 98–100.

90 *Ibid* at para 116.

91 (1992), 69 CCC (3d) 395 at para 74, 12 CR (4th) 99; Goudge Report, *supra* note 9 at 488.

92 *Gager* Totten, *supra* note 39 at para 116.

93 Joseph P Simmons, Leif D Nelson & Uri Simonsohn, "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant" (2011) 22:11 Psychological Science 1359 at 1359–60 [Simmons]; Leslie K John, "Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling" (2012) 23:5 Psychological Science 524 [John et al]; Hannah Fraser et al, "Questionable Research Practices in Ecology and Evolution" (2018) 13:7 PLoS One e0200303, online: <osf.io/ajyqg/> [Fraser].

94 Simmons, *supra* note 93 at 1359, 1361.

95 Munafò, *supra* note 68.

publish their raw data and materials to improve peer review and general scrutiny in the scientific community.[96] Further, leading groups of scientists prescribe a practice known as "preregistration" whereby researchers publicly commit to a set of procedures that they cannot edit after data collection has started.[97] The idea is not to constrain science, but to make any changes (like those Totten apparently made) transparent. I do not mean to say that open science would have served as a panacea in *Abbey*, but rather that—even today—orthodox standards in many scientific guilds do not include these important reforms.[98]

As to research fraud and data fabrication, the *Abbey* 2017 Court expressly refrained from weighing in on Totten's behaviour: "it would be unfair to make the positive finding that Abbey urges us to make: Totten fabricated or concocted part of his research, or gave deliberately misleading testimony."[99] Still, the Court pulled few other punches, recounting Totten's "false" reply in *Gager* to the question of whether he ever used the same interviews in different published studies.[100]

Indeed, inferring intentionality in bad science is challenging and should not be taken lightly. Perhaps the most important thing to note is that research finds that, depending on the specific activity, up to 60% of academic scientists anonymously admit to using researcher degrees of freedom (i.e., mainstream but error-prone scientific practices).[101] Similar studies find that only about 1% of researchers intentionally falsify or fabricate their data.[102] As a result—and as I propose in Part 6—legal approaches aimed at reducing researcher degrees of freedom may help in promoting reliable expert evidence.

Moving beyond Totten's evidence in *Abbey* and *Gager*, framework evidence from social scientists tendered by the defence have not seen

---

96 TOP Guidelines, *supra* note 69; Science, "*Science:* Editorial Policies", online: <www.sciencemag.org/authors/science-editorial-policies> [Science Editorial]; Royal Society, "Data Sharing and Mining", online: <royalsociety.org/journals/ethics-policies/data-sharing-mining/>.

97 Brian A Nosek et al, "The Preregistration Revolution" (2018) 115:11 Proceedings National Academy Sciences US 2600 [Preregistration Revolution].

98 Jason M Chin, "Psychological Science's Replicability Crisis and What it Means for Science in the Courtroom" (2014) 20:3 Psychol Pub Pol'y & L 225; Jason M Chin, "What Irreproducible Results Mean for the Law of Scientific Evidence" (2016) 35:1 Adv J 17.

99 *Abbey* 2017, *supra* note 36 at para 125.

100 *Ibid* at para 101.

101 John et al, *supra* note 93 at 525; Fraser, *supra* note 93.

102 Daniele Fanelli, "How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data" (2009) 4:5 PLoS ONE e5738. See also Sherkow, *supra* note 68 at 855 reporting that fraud was "extremely rare".

their evidence treated any more favourably post-*Abbey*.[103] Some in the criminal law bar thought that *Abbey* would assist in admitting legally relevant academic psychology. Prior to *Abbey*, the most influential case to discuss the admissibility of the social scientific evidence was *R v McIntosh*. And *McIntosh* still stands as a controversial bar to the admission of expert psychological evidence on the fragility of eyewitness memory.[104] *McIntosh* took a dim (and mistaken) view of psychology, questioning whether it was a recognized body of knowledge.[105] *Abbey*, on the other hand, appeared (at least facially) to be more accepting of social science.[106]

Abbey was not the game-changer that some wished for. For example, the Court of Appeal for Ontario had the chance to reconsider the exclusion of eyewitness memory experts in *R v Frimpong*.[107] The defence submitted that *Abbey* overruled the *McIntosh* view of social science.[108] The appellate court disagreed.[109] Similarly, in *R v Jeanvenne*, an Ontario trial court held that *Abbey* had not changed the law as to experts who would opine on the factors

---

[103] Courts, by and large, at least seem to see the applicability of *Abbey* 2009 to social scientific expert evidence (unlike with police expertise, see Part 5, below). In sociology, see *R v Shafia*, 2016 ONCA 812, 341 CCC (3d) 354. In history, see *Ross River Dena Council v Canada (AG)*, 2014 YKSC 53, 249 ACWS (3d) 119. In criminology, see *R v Orr*, 2015 BCCA 88, 18 CR (7th) 158. In *Orr*, a criminologist would have explained why an alleged victim of human trafficking did not report her plight to the police immediately. The Court of Appeal for British Columbia held that while the evidence did not need to meet scientific standards, it was from the "behavioural or 'soft' sciences" and thus *Abbey* applied: *Orr* at paras 2, 16–32. Courts may, however, take a permissive approach to admitting social science evidence in bench trials. Recently, in *R v Comeau*, 2018 SCC 15 at para 40, 420 DLR (4th) 199, the Supreme Court noted that part of a historian's expert evidence did not meet *Mohan*'s necessity criterion for, in essence, being a legal opinion. The transparency approach that I discuss in Part 6 would also help determine if the historian had surveyed a sufficient amount of the historical literature in forming his opinion.

[104] Dufraimont, "Regulating Unreliable Evidence", *supra* note 6 at 269–70; Jill Copeland, "Helping Jurors Recognize the Frailties of Eyewitness Identification Evidence" (2002) 46:2 Crim LQ 188; Chin & Crozier, *supra* note 6.

[105] "I would have to be persuaded that the subject is a recognized branch of psychology": *McIntosh*, *supra* note 33 at para 19. To the contrary, it is very much a recognized branch of psychology. For instance, the Society for Applied Research in Memory and Cognition (SARMAC) has existed since 1997: SARMAC, online: <www.sarmac.org/about/aims-and-origin/>.

[106] See *R v Jeanvenne*, 2011 ONSC 7244 at para 50, 286 CCC (3d) 65 [*Jeanvenne*]. This was the defence's argument in *R v Frimpong*, 2013 ONCA 243, CR (7th) 242 [*Frimpong*], *R v Frimpong*, Appellant's Factum, Court file No C53147 at para 61, online: <osf.io/ukdqw/> [*Frimpong* Factum].

[107] *Frimpong*, *supra* note 106.

[108] *Frimpong* Factum, *supra* note 106 at para 61.

[109] *Ibid* at paras 23–28 .

that produce false confessions, a form of evidence that is also difficult (if not impossible) to admit into Canadian courts.[110]

While there does seem to be some imbalance in where *Abbey*'s permissive effect has been felt, cases like *Frimpong* and *Jeanvenne* do not hinge on the reliability of experts, but whether their evidence is beyond the factfinder's ken. In other words, memory and confessions experts are excluded because courts believe their evidence to be common sense (i.e., about everyday memory processes and social pressure) and thus not "necessary" pursuant to *Mohan*.[111] There is, in fact, little doubt that the field of eyewitness memory research has produced a tested and reliable basis of knowledge.[112]

Courts might be more likely to find that a psychologist's opinion was beyond the factfinder's ken if he or she could opine on the psychology of a particular witness or confessor rather than on the general psychological processes. However, these researchers have largely stayed within the boundaries of their discipline.[113] And when they have not, courts have excluded them.[114] For example, the defence in *R v Pearce* tendered an expert who psychologically assessed the accused and opined that he was prone to false confessions (i.e., he did not simply describe the factors that produce false confessions).[115] The Court applied *Daubert* to this minimally tested evidence and found it was not demonstrably scientifically valid, and thus should be excluded.[116]

4. Forensic Science, the Off-Switch, and the Science of Expertise

Abbey's impact on the admission of forensic scientific evidence is possibly the most baffling. Courts have not taken up Justice Doherty's suggestion of

---

110 *Jeanvenne*, *supra* note 106 at para 50.

111 *McIntosh*, *supra* note 33 at para 20.

112 *State v Henderson*, 208 NJ 208 (2011) at 283, 27 A 3d 872: "The research presented on remand is not only extensive, but as Dr. Monahan testified, it represents the 'gold standard in terms of the applicability of social science research to the law.' Experimental methods and findings have been tested and retested, subjected to scientific scrutiny through peer-reviewed journals, evaluated through the lens of meta-analyses, and replicated at times in real-world settings".

113 Faigman G2i, *supra* note 51 at 432–34.

114 I am not aware of a Canadian case in which an eyewitness expert has attempted to provide an individuating opinion. Clinical psychologists or psychiatrists periodically attempt to provide unreliable individuating opinions about the effects of abuse on memory and reporting of abuse: [*K(A)*], *supra* note 33; *JP v British Columbia (Children and Family Development)*, 2017 BCCA 308, [2017] 12 WWR 639.

115 *R v Pearce*, 2014 MBCA 70, 310 Man R (2d) 14.

116 *Ibid* at para 88.

tailoring the reliability analysis to the nature of the opinion.[117] Rather, they have engaged in various exercises in characterization to find that a testable forensic scientific practice is not novel or is more appropriately labeled "experiential". Then, they have avoided any substantive reliability inquiry by apparently construing *Abbey* 2009 as a guild test and adverting to the expert's membership in a profession or field of study. Or, they simply admit the expert based on his or her qualifications and experiences (a "sufficient experience" test).[118] In other words, courts simply hit the "off-switch" on any substantive reliability scrutiny.[119] As academics and peak scientific bodies have noted, this is a dangerous practice.[120] I will only briefly review this phenomenon—it has been detailed elsewhere[121]—and then suggest that when expertise is truly not susceptible to formal testing, courts may look to whether the expert obtained that expertise in a reliable manner.

In a case lucidly detailed by Emma Cunliffe and Gary Edmond,[122] the Court of Appeal for British Columbia in *R v Aitken* held that a "forensic gait" expert had been properly admitted at trial despite the fact that the trial judge did not consider the scientific validity of this highly contested practice.[123] In short, forensic gait analysts compare footage of an unknown person walking to a known person and determine, based on the way they walk, if they are the same individual (or share gait characteristics, as was the case in *Aitken*).[124] The Court of Appeal found that the trial court properly admitted the putative expert because, among other reasons, the expertise was specialized knowledge and so *indicia* of scientific validity were inapplicable.[125] It did not apply the nine reliability factors to forensic gait analysis nor give any express consideration of reliability.

In similar manner to that in *Aitken*, the Crown has also relied on *Abbey* to tender a forensic practice developed by police officer Steve Horwood and his

---

117 See e.g. Cunliffe & Edmond, "Gaitkeeping", *supra* note 2; Chin & Dallen, *supra* note 2.

118 Risinger, *supra* note 4 at 769–71.

119 A similar trend has been occurring in the US: Kaye, Bernstein & Mnookin, *supra* note 4 at §10.3.2, concluding: "A number of opinions then simply assert that the *Daubert* deficiencies are not fatal because the traditional forensic identification testimony satisfies *Kumho Tire*".

120 US, National Research Council, *Strengthening Forensic Science in the United States: A Path Forward* (Washington: National Academies Press, 2009) at 53, online: <www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf> [NAS Report]; Cunliffe & Edmond, "Gaitkeeping", *supra* note 2.

121 Cunliffe & Edmond, "Gaitkeeping", *supra* note 2; Chin & Dallen, *supra* note 2.

122 Cunliffe & Edmond, "Gaitkeeping", *supra* note 2.

123 *Aitken*, *supra* note 1 at para 74.

124 Cunliffe & Edmond, "Gaitkeeping", *supra* note 2 at 333.

125 *Aitken*, *supra* note 1 at paras 79–80

protégé Scott Ferguson to identify the individuals in videos who are carrying a gun and to opine on the characteristics of gun-carrying individuals.[126] In two cases, courts have excluded the identification portion of these opinions (for going to the ultimate issue) but allowed them to describe gun carrying *indicia*.[127] These range from common sense (e.g., a gun-sized bulge in the clothing) to unsupported folklore (e.g., keeping one's gun-armed side away from nearby police).[128] In admitting this expertise in *R v Edwards*, an Ontario trial court did no more than recount Officer Ferguson's years in the police force and the fact he had learned from Horwood and gone on to teach the method himself.[129] There was no reported discussion of whether these predictors are reliable. Further, as I will now address in the context of *R v Awer*, there was no discussion of whether the officers developed their purported specialized knowledge in a reliable way. In other words, based on the environment in which they trained and studied, could they be expected to have developed valid cues to identify armed individuals?

The opinion at issue in *Awer* was a DNA expert's conclusion that a sample of DNA found on a swab of the accused's penis came from a "wet" source of the complainant's DNA.[130] This opinion called into question the defence's theory of a secondary transfer or, in other words, that Awer may have touched something with the complainant's DNA on it and then touched himself. The expert, Mr. Stephen Denison, mainly relied on an anecdote to support the wet transfer mechanism: "I have not seen a dry stain transfer that total amount of DNA."[131] The defence expert disagreed, saying that the science was too unsettled to determine if a transfer was secondary or direct based on the amount of DNA found. The trial judge, however, was not persuaded and appeared to place great weight on Denison's anecdote in convicting Awer.[132]

On appeal, Awer challenged the admissibility of the Crown expert's opinion as unreliable anecdotal evidence. A majority of the Court of Appeal for Alberta disagreed and simply hit the off-switch, refraining from applying any reliability scrutiny. *Daubert*, according to the majority, did not apply because the expert, like Dr. Totten, was drawing on specialized knowledge rather than novel science: "The evidence of Dr. Totten, like that of Denison here, was of specialized knowledge flowing from personal life

---

126 *Woodcock*, *supra* note 35; *R v Edwards*, 2016 ONSC 5434, 132 WCB (2d) 566 [*Edwards*].

127 *Woodcock*, *supra* note 35 at paras 4, 21; *Edwards*, supra note 126 at para 36.

128 *Edwards*, *supra* note 126 at para 8.

129 *Ibid* at paras 13–17.

130 *Awer* ABCA, *supra* note 1 at para 9. See a review of this case in Chin & Dallen, *supra* note 2; Likwornik, *supra* note 22.

131 *Awer* ABCA, *supra* note 1 at para 9.

132 *Awer* SCC, *supra* note 1 at para 4.

experience in relation to activities of gangs. For Denison it arose in the practical application of DNA investigations and related science."[133] Despite this analogy the Court did not apply the *Abbey* 2009 factors to Denison's evidence or consider whether his previous practical applications provided him with reliable expertise.

The defence also relied on the Supreme Court's 2014 *R v Sekhon* decision.[134] In *Sekhon*, the Court held as inadmissible a police officer's opinion that in his years of work, he had never encountered a drug courier who did not know he or she was carrying drugs (i.e., a blind courier).[135] The Supreme Court, quoting the dissent in the lower court, said that such evidence was unduly prejudicial because it was endorsed by the "superficial attractiveness" of an expert.[136] Yet, at the same time, the merely anecdotal personal observation seemed dispositive of the accused's *mens rea*.[137] The defence in *Awer* argued that, like the police officer's drug courier opinion, Denison could not found his opinion on the bald anecdote that in all his years analysing DNA, he had not seen a dry source transfer as much DNA as he observed in the instant case.[138] The majority of the Alberta court again disagreed. First, the trial judge's weighting of the evidence deserved discretion[139] and second, *Sekhon* was distinguishable because the anecdote in that case went to the ultimate issue—Sekhon's *mens rea*.[140]

At the Supreme Court, Awer again argued that Denison's evidence should not be admitted. The Supreme Court allowed the appeal but importantly did not hold that Denison's anecdote should have been excluded. Rather, the Court held that the trial judge should have been more measured in the relative scrutiny applied to the Crown and defence expert.[141] This imbalance unduly shifted the burden of proof to the accused. While this decision itself is not very informative about the meaning of *Abbey* and *Sekhon* going forward, a review of oral arguments may provide insight into the Supreme Court's current thinking on expert evidence. Note, however, that these comments are not in any way binding and may have been posed rhetorically as part of the discussion with counsel.

---

133 *Awer* ABCA, *supra* note 1 at para 38. Justice Berger dissented, saying that Denison's evidence did not meet scientific standards: see paras 89–132.

134 *Supra* note 23.

135 *Ibid* at paras 49–51.

136 *Ibid* at para 50.

137 *Ibid*.

138 *Awer* ABCA, *supra* note 1 at para 50.

139 *Ibid* at paras 56–57.

140 *Ibid* at paras 59–60.

141 *Awer* SCC, *supra* note 1 at para 6.

First, Justice Rowe suggested that there are circumstances in which experience-based expertise is valuable. In doing so, he drew a contrast between experimental science and the experience of a sea pilot.[142] His description of the sea pilot's experience is worth quoting in full:

> The testimony of a master mariner concerning the handling of a vessel in a particular sea state or in a particular situation in a port. You cannot have a controlled study there. You're not going to smash vessels into a dock and do a statistical analysis of it.
>
> But if you have a mariner who has basically spent their whole life on the water and says look, I docked vessels thousands of times and I know what happens and the limits in which you can control it. That to me, in the right circumstances, is proper expert evidence … the publication of studies, the reference to peer reviewed literature, the strict application of the scientific method it seems to me is not a prerequisite for expert opinion in all instances.[143]

Here, Justice Rowe appears to be relating the "harbour pilot" analogy from *United States v Starzecpyzel*.[144] An important point, unfortunately not explored in oral arguments, was that the harbour pilot example is distinguishable from much of the expert evidence that passes the *Abbey* test (including evidence in *Awer*, *Woodcock*, *Edwards*, and much of the police expertise discussed in Part 5). The harbour pilot receives immediate and accurate feedback about his performance each time he docks. The mere fact that he has not crashed provides objective evidence of his proficiency. Denison, however, is unlikely to receive such unambiguous feedback from his DNA tests. In many cases, he may never learn if the source of a sample is dry or wet and, in fact, may be less likely to learn the source of dry secondary transfers because those occur in cases where the complainant's account is difficult to corroborate.[145]

---

[142] Supreme Court of Canada, "Webcast of the Hearing on 2017-01-17: Nihal Awer v Her Majesty the Queen" (17 January 2017) at 20m:30s, online (video): <www.scc-csc.ca/case-dossier/info/webcastview-webdiffusionvue-eng.aspx?cas=37021&id=2017/2017-01-17--37021&date=2017-01-17&fp=n&audio=n> [*Awer* Webcast]: "Some things lend themselves to the application of the scientific method and others simply do not. Chemical analysis does".

[143] *Ibid* at 20m:40s.

[144] (1995) 880 F Supp 1027 at 1029, 1043 NY Southern Dist Ct (1995). For a review, see D Michael Risinger & Michael J Saks, "Science and Nonscience in the Courts: *Daubert* Meets Handwriting Identification Expertise" (1996) 82:1 Iowa L Rev 21 at 33–34.

[145] For example, wet source transfers may occur in more straightforward cases in which other evidence eventually corroborates the mode of transfer. Dry transfers may occur in cases in which it is difficult to directly implicate an assailant. Therefore, dry transfers may actually transfer more DNA than Denison thinks, but he never received the feedback to know this.

Justice Moldaver, who authored the *Sekhon* decision, picked up on the harbour pilot example and related it to *Sekhon*. First, Justice Moldaver expressed that lower courts had read the holding—as against anecdotal evidence—too broadly.[146] Instead, he indicated that *Sekhon* should be read narrowly, applying specifically to anecdotes going to the accused's mental state. In particular, he said that the harbour pilot is a "whole different matter" than "what's in somebody's mind based on every time you have arrested someone for murder."[147]

Going forward—and assuming these comments made during legal arguments are representative of the Supreme Court's thinking—there may be a sizable contested area between potentially admissible harbour pilot-type situations (i.e., specialized knowledge from circumstances when the tendered expert receives immediate and unambiguous feedback about his or her performance) and clearly inadmissible *Sekhon*-type situations, in which the expert never learns what was in the mind of the accused. Where courts draw the line between permissible and impermissible expertise will be important and, in my view, should be informed by the science of expertise.[148]

Several insights from cognitive scientific research on expertise merit a brief review. First, while the amount of experience is a factor,[149] it is often a misleading factor because of the importance of the learning environment.[150] As reflected in the harbour pilot example, timely and objective feedback is

---

146 *Awer* Webcast, *supra* note 142 at 27m:45s: "There is no legitimate inference that can be drawn from the fact that every murder that I have investigated—and if you look at the examples we used in *Sekhon they are all related to mental state, every one of them and to the extent that's not been considered, it really should be*—because you cannot possibly say that because in every murder I have investigated, the accused was found to have the requisite intent for murder, therefore this accused had the requisite intent for murder in the context of we'll say a domestic assault or something" [emphasis added]. In the *Awer* Supreme Court judgment, Justice Moldaver similarly noted "qualitative" distinction between *Awer* and *Sekhon*. See *Awer* SCC, *supra* note 1 at para 2.

147 *Awer* Webcast, *supra* note 142.

148 See William G Chase & Herbert A Simon, "The Mind's Eye in Chess" in William G Chase, ed, *Visual Information Processing* (London, UK: Academic Press, 1973) 215 [Chase & Simon]; Robin M Hogarth, Tomás Lejarraga & Emre Soyer, "The Two Settings of Kind and Wicked Learning Environments" (2015) 24:5 Current Directions in Psychological Science 379 [Hogarth]. For a review, see Daniel Kahneman & Gary Klein, "Conditions for Intuitive Expertise: A Failure to Disagree" (2009) 64:6 American Psychologist 515 [Kahneman & Klein]. For a comparison between the scientific view of expertise and Australian expert evidence law, see Martire & Edmond, *supra* note 4. For a comparison between the scientific view of expertise and Canadian lay opinion law, see Chin, Tomiska & Li, *supra* note 4.

149 Chase & Simon, *supra* note 148. Several factors contribute to this number, including the complexity of the task and the learning environment.

150 Kahneman & Klein, *supra* note 148 at 519–24. Hogarth, *supra* note 148.

a key aspect of a learning environment because it allows the individual to connect some relevant cue (e.g., the speed of the harbour pilot's vessel) to the ultimate conclusion (e.g., whether that was a safe speed at which to dock). In this respect, Daniel Kahneman and Gary Klein give the counterexample of "technical"[151] stock analysts who have no non-public information about a company, but instead deceive themselves into thinking they have actual expertise based on industry folklore about patterns of price changes (e.g., arbitrary "resistance levels" that drive stock prices—these are what psychologists call illusory correlations).[152]

Further, expertise is typically not fractionated—it does not carry over to tasks that may seem closely related. For example, weather forecasters are proficient at predicting routine weather, but are much poorer at predicting more aberrant events like hail.[153] Given these findings, job status and general professional progression (i.e., forms of social recognition) are often poor indicators of domain specific forms of expertise.[154] Moreover, a particular danger arises when professionals "who know how to use their knowledge for some purposes attempt to use the same knowledge for other purposes."[155] This can result in the highly prejudicial "illusion of validity" whereby the expert gives a confident opinion about something he or she is not actually skilled in.[156]

These findings from the study of expertise should inform the admission of experience-based expertise. For instance, courts should be less concerned with the number of years a purported expert has held some job adjacent to the expertise[157] and more concerned with the way in which he or she developed the expertise. In this way, Justices Rowe and Moldaver were correct in suggesting the "anecdote" characterization is itself not very useful. For instance, it may have been that the concealed gun experts had viewed hundreds of videos in which they ultimately learned if individuals were

---

151 Technical analysts contrast with fundamental analysts, who seek to outperform other investors by performing novel research on companies and industries.

152 Kahneman & Klein, *supra* note 148 at 520; Kahneman and Klein also provide the example of the multimillion dollar business of baseball decisions (at 521). Executives long relied on largely invalid *indicia*, such as the player's attractiveness, and only recently began relying on predictive statistical measures. See also Paul Slovic, "Psychological Study of Human Judgment: Implications for Investment Decision Making" (2001) 2:3 J Psychology & Financial Markets 160.

153 Kahneman & Klein, *supra* note 148 at 522.

154 Martire & Edmond, *supra* note 4 at 974–75.

155 Kahneman & Klein, *supra* note 148 at 523.

156 *Ibid*.

157 For examples of this problematic approach, see *R v M(C)*, 2010 ONSC 4819 at para 13, 90 WCB (2d) 192; *R v Pham*, 2013 ONSC 4903 at para 32, 300 CCC (3d) 111 [*Pham*]; *R v Reid*, 2017 ONSC 4082 at paras 29–30, 140 WCB (2d) 649 [*Reid*].

indeed carrying guns. Or, more dangerously, they may have simply been repeating unvalidated heuristics passed down through the years as a way to rationalize unconscious reliance on other cues, like the race of the suspect. Importantly, as I will discuss in Part 5, questions probing actual expertise are not asked enough of the time.

5. Police Expertise and a (Possible) New Role for Transparency

An examination of police expertise post-*Abbey* demonstrates many of the themes detailed so far. First, courts rarely engage in any substantive reliability analysis, instead simply reciting the officer's qualifications and experiences (i.e., sufficient experience) or standards of the police guild. And following from the above discussion of the science of expertise, courts rarely attempt to connect the officer's experiences to their ability to reliably produce the opinions they are tendered for. However, when an expert purports to possess some gang-related knowledge, courts do sometimes engage with the *Abbey* reliability factors. This may simply be due to the superficial similarity of the expertise (i.e., Totten was a gang expert). Those courts have found the transparency-related factors particularly useful, a trend I will pick up on in Part 6 in my description of a transparent proficiency-based approach.

With an increasing amount of police expertise proffered by the Crown, determining the reliability of that expertise is of special importance.[158] The Crown frequently calls on police officers to provide inculpatory opinions about drug and gang related charges. With drugs, the opinion often supports a charge of possession for the purpose of trafficking.[159] For example, the officer may describe the typical tactics of traffickers, the amount of the drug they carry and sell, and the equipment they use.[160] The officer may also be proffered to describe the consumption patterns of drug users to counter the defence's suggestion that the accused merely possessed the drug for personal use.[161] Sometimes, the officer would go so far as to definitively state that the accused possessed the drug to traffic it (or give that opinion on a

---

158 See Joëlle Anne Moreno, "What Happens When Dirty Harry Becomes an (Expert) Witness for the Prosecution?" (2004) 79:1 Tul L Rev 1; D Michael Risinger, "The Irrelevance, and Central Relevance, of the Boundary Between Science and Non-Science in the Evaluation of Expert Witness Reliability" (2007) 52:4 Vill L Rev 679. The admissibility of police as expert witnesses is a relatively new phenomenon and appears to be the result of an organized public image campaign on the part of police groups: Anna Lvovsky, "The Judicial Presumption of Police Expertise" (2017) 130:8 Harv L Rev 1995.

159 See *Controlled Drugs and Substances Act*, SC 1996, c 19, s 5(2). For a review of the role of experts in these cases, see Bruce A MacFarlane, Robert J Frater & Croft Michaelson, *Drug Offenses in Canada*, 4th ed (Toronto: Canada Law Book, 2015) at 17:60.60 [MacFarlane et al]; Chin & Dallen, *supra* note 2 at 537–38.

160 See e.g. *O(N)*, *supra* note 35 at para 12.

161 See e.g. *R v Jacobs*, 2014 ABCA 172 at paras 42–50, 577 AR 3 [*Jacobs*].

hypothetical set of facts identical to the Crown's case).[162] As to gangs, police also provide several inculpatory opinions: *indicia* of gang membership,[163] whether an individual is in a gang,[164] and the meaning of gang slang or coded language.[165]

Following from the discussion of experiential expertise above, it is possible to develop guidelines for how difficult-to-test police expertise should and should not be scrutinized for reliability. On the less reliable end of the spectrum is a general reference to time spent as a police officer and the experiences that follow from that (e.g., conversations with informants, internal coursework). Such analysis provides meagre assurance that the officer can reliably perform tasks like identifying drug traffickers (or the characteristics of them).

Rather, courts should inquire into how experience translates into reliable proficiency in the specific judgment the expert is making. For instance, undercover officers will learn quickly if slang means what he or she thinks it means and if a drug purchase of a certain size is customary. This is because if the officer was mistaken, he or she would unequivocally find out. Other roles in investigations in which the officer regularly receives feedback can also be expected to produce expertise. Officers may, for instance, consistently learn if other evidence corroborated the suspicion that a suspect was involved in trafficking. On the other hand, mere conversations with arrestees and informants about their drug use carry no guarantees of reliability. The officer may also be siloed into a particular role and thus rarely learn about the outcome of investigations. Finally, as demonstrated by the expert in *Abbey*, results of interviews and investigations may be misremembered or

---

[162] See e.g. *R v Piechotta*, 2016 BCPC 463 at para 21, 141 WCB (2d) 51 [*Piechotta*]: "Based on my review of the investigative materials, exhibits, speaking with the investigator, my experience and for reasons I have articulated throughout this report. [sic] It is my opinion that the drugs seized from this investigation were for the purposes of trafficking at the 'retail/distribution' level"; *R v Dominic*, 2016 ABCA 114 at para 10, 34 Alta LR (6th) 219 [*Dominic*]: "In the course of his examination-in-chief, Larson was presented with a hypothetical scenario matching the facts for the Crown's case. He opined that the facts indicated cocaine possession for the purpose of trafficking".

[163] See e.g. *R v Sheriffe*, 2015 ONCA 880 at para 108, 333 CCC (3d) 330.

[164] See e.g. *R v Gager*, 2012 ONSC 388 at paras 224–48, 99 WCB (2d) 325 [*Gager* Police] (the officer's opinion about membership in gangs was excluded in this case).

[165] See e.g. *R v A(T)*, 2015 ONCJ 624, 125 WCB (2d) 631 [*A(T)*]. Decoding drug-related slang is also a regular area of expertise: see *R v Lucas*, 86 WCB (2d) 42, 2009 CarswellOnt 7776 (WL Can) [*Lucas*]; *R v Fabos*, 2015 ONSC 8013, 128 WCB (2d) 358 [*Fabos*]. Other jurisdictions have also struggled with the admissibility of police experts about gangs. In Australia, see *R v Cluse*, [2014] SASCFC 97, 120 SASR 268. In the UK, see *Myers v R*, [2015] UKPC 40, [2015] 3 WLR 1145. See generally MacFarlane et al, *supra* note 159 at 17:60.60.

otherwise distorted. Therefore, transparency should be demanded, even of police experts relying on their experiences.

A) Police Drug Expertise

This subsection details *Abbey*'s lack of impact in cases considering the admissibility of police drug experts.[166] Rather, the Supreme Court's decision in *R v Sekhon*—a trafficking case—and its exclusion of a police officer who provided anecdotal evidence about drug traffickers seems to have had more of an effect. Courts subsequently scrutinized such anecdotal evidence more closely. This may have led to a practice whereby the police undertook more systematic surveys and interviews of drug users and traffickers. As *Abbey* and *Gager* demonstrated, however, this practice carries its own dangers, even when the expert is a trained social scientist.

After *Abbey*, courts continued to sidestep reliability analysis by characterizing the expertise as non-scientific and then broadly adverting to the officer's years of experience.[167] One Ontario trial court at least cited *Abbey*'s reliability analysis but, puzzlingly, did not apply it to the facts and instead focused on general police experience.[168] Another Ontario decision developed 11 factors for use in assessing police experts' qualifications.[169] These factors generally focus on experience and training in the form of coursework and membership in organizations, without any express consideration of reliability.[170] Such decisions sharply contrast with the

---

166 Prior to *Abbey*, there was no consistent approach to the reliability of police experts. For instance, the Court of Appeal for Alberta did not consider reliability when reversing a trial decision that had excluded a police officer because he was not demonstrably more experienced than an average police officer. Instead, the Court simply examined the officer's experience as a police officer relative to what a factfinder might know. The officer opined on, among other things, drug consumption patterns, packaging for distribution, and pricing: *O(N)*, *supra* note 35 at paras 12, 14, 24. Similarly, a British Columbia court characterized drug consumption knowledge as not novel science and resultingly evaded the question of reliability altogether. *R v Petavel*, 2006 BCSC 1931 at para 14, 72 WCB (2d) 35. These decisions are difficult to reconcile with contemporaneous decisions that demanded more structured and scientific experience. See *R v Klassen*, 2003 MBQB 253 at para 26, 179 Man R (2d) 115; *R v Jerrett*, 242 Nfld & PEIR 348 at para 28, 719 APR 348, demanding the evidence pass *Daubert*. Post-*White Burgess*, police officers who are involved in the investigation or have significant experience with the accused may be excluded for bias and partiality, see *R v McManus*, 2017 ONCA 188 at paras 56–75, 353 CCC (3d) 493.

167 *Fabos*, *supra* note 165 at para 30. See Also *Dominic*, *supra* note 162; *Piechotta*, *supra* note 162 at paras 81, 162.

168 *R v Murphy*, 2010 ONSC 109 at para 50, 86 WCB (2d) 866 [*Murphy*].

169 *Pham*, *supra* note 157 at para 13.

170 The Court eventually found the police officer had "sufficient expertise" based on general advertence to his interactions with "literally hundreds" of potential drug users in vehicle stops and "numerous" drug users without considering how these may have produced

one decision I am aware of to have expressly applied *Abbey*'s reliability analysis. In that case, *R v Sriskanda*, the court noted that the *Abbey* factors were an imperfect fit with the officer's purported specialized knowledge.[171] Still, there were no quality assurance measures (factor 2) and information passed to the officer through others was not reliably stored and available (factor 3).[172] As a result, this officer's opinions on the risks associated with consuming and producing a marijuana by-product were inadmissible.[173]

Perhaps because it was a drug trafficking case itself, the Supreme Court's decision in *R v Sekhon* seems to have been more influential than *Abbey* in cases of drug expertise.[174] Across several decisions it appears to have at least nudged courts to consider whether a police officer's years of anecdotal observations could scaffold his or her expertise.[175] This issue is especially likely to arise when the expert would opine on drug addiction and casual use (as opposed to trafficking) because most officers obtain their knowledge about these areas from unstructured interviews and conversations with confidential informants and individuals arrested for other reasons. Following *Sekhon*, courts began to exclude such evidence under the theory that anecdotally-derived specialized knowledge requires additional scrutiny.[176]

These worries about anecdotal drug use evidence may have inspired the Crown in *R v Reid* to tender a police officer who purported to have performed a more systematic study of drug users.[177] He opined on the behaviours of drug traffickers and users. This officer, Detective Constable Charlie Rau, questioned 40 marijuana users about their use.[178] His participants consisted of arrestees, those identified when executing search warrants, and confidential informants.[179]

---

reliable expertise: *ibid* at paras 32–81. Several other post-*Abbey* decisions gave little or no attention to reliability: see *R v Lee*, 2014 ONCJ 640, 117 WCB (2d) 598 [*Lee*]; *R v Tremblett*, 2012 NSPC 121, 325 NSR (2d) 6; *R v Edison*, 2015 NBQB 74, 433 NBR (2d) 267.

171 2016 ONCJ 667 at para 33, 134 WCB (2d) 578.

172 *Ibid*.

173 *Ibid* at paras 23–41.

174 *Sekhon*, *supra* note 23.

175 *Jacobs*, *supra* note 161; *R v Mulaj*, 2014 ONSC 4405, 114 WCB (2d) 635 [*Mulaj*]; *R v Pico*, 2016 ONSC 1470, 129 WCB (2d) 340 [*Pico*]. See also *R v Gausal*, 2017 BCSC 1192, 141 WCB (2d) 620.

176 *Jacobs*, *supra* note 161 at para 58; *Mulaj*, *supra* note 175 at para 44; *Pico*, *supra* note 175 at para 92.

177 *Supra* note 157.

178 *Ibid* at para 11.

179 *Ibid*.

In assessing the reliability of Rau's opinion, the Court expressly declined to follow guidance from *Daubert*[180] and the Goudge Report.[181] As to the Goudge Report, the trial judge easily dismissed it: "The Goudge Inquiry dealt with the evidence of a disgraced forensic pathologist. I am dealing with a police officer."[182] This view ignores the vast literature the Goudge Report inspired and reflected, finding systemic uncertainties in expert evidence.[183] Instead, the Court characterized Rau's evidence as specialized knowledge pursuant to *Abbey*.[184] Despite the clear similarity between Rau and Totten's interview methods, the Court did not apply the *Abbey* factors.

Instead of relying on the *Abbey* factors, the trial judge found sufficient expertise in Rau's years of experience as a police officer who attends courses and converses with "fellow officers, arrested persons, confidential informants, found-ins, and others."[185] The difficulty in the Court's approach towards Rau's interviews is viscerally illustrated by Mark Totten's evidence in *Abbey* and *Gager*. As discussed, Totten exploited shifting inclusion criteria and definitions of key terms (e.g., "gang member") to support the opinion he was hired to give. To qualify a police officer with ostensibly less training in social science who labours under the same (or worse) adversarial bias seems dangerously imprudent.[186]

B) Police Gang Expertise

While *Abbey*'s influence on the admission of drug expertise has been notably limited, the decisions in the gang context have at times taken on a different tenor.[187] This may be due to the simple fact that Totten's gang expertise in *Abbey* is factually similar to the expertise of police gang experts who also purport to understand gang culture. In any event, the gang cases are useful

---

180 *Ibid* at para 25.

181 *Ibid* at paras 23, 28.

182 *Ibid* at para 28.

183 See David Paciocco, "Taking a 'Goudge' out of Bluster and Blarney: An 'Evidence-Based Approach' to Expert Testimony" (2009) 13:2 Can Crim L Rev 135; NAS Report, *supra* note 120; US, President's Council of Advisors on Science and Technology, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods* (Washington: Executive Office of the President, 2016) [PCAST Report]; Alan D Gold, *Expert Evidence in Criminal Law: The Scientific Approach*, 2nd ed (Toronto: Irwin Law, 2009).

184 *Reid*, *supra* note 157 at paras 25–27. *Abbey* 2009, *supra* note 1 at paras 104–17.

185 *Reid*, *supra* note 157 at paras 29–30.

186 Both Totten and Rau's methodologies were not purely qualitative because they drew force from the number of interviews performed.

187 The following decisions applied the factors in *Abbey* 2009 or compared an expert witness to Totten: *R v Myles* [2011], 97 WCB (2d) 377 at paras 26, 52, 2011 CarswellOnt 10352 (WL Can) (Sup Ct J); *Gager* Police, *supra* note 164 at paras 165–214; *R v Giles*, 2016 BCSC 294 at paras 120–22, 130 WCB (2d) 614 [*Giles*]; *R v O(F)*, 2016 ONSC 724 at paras 17–18, 129 WCB (2d) 142 [*O(F)*].

in showing what occurs when courts begin demanding demonstrable reliability from police officers. In short, they find that these opinions are seriously flawed.

In 2016, the British Columbia Supreme Court in *R v Giles* was much more critical of a police expert who based his opinion on his "omnibus" experience and training,[188] without demonstrating the reliability of that knowledge.[189] For instance, after a lengthy comparison to Totten's opinion in *Abbey* 2009, the Court in *Giles* found that the expert's failure to retain his notes was problematic, leading to the possibility that his opinion was not representative of his actual expertise.[190] Somewhat ironically in light of *Abbey* 2017, the Court held up Totten's evidence as a model of transparency and diligence: "an important aspect … in relation to Dr. Totten's work in *Abbey* 2009 was that the information upon which the opinion was based had been 'accurately recorded and memorialized.'"[191] Moreover, the expert in *Giles* relied on literature that seemed unsystematically chosen, except for the fact that it supported his opinion.[192]

Giles bears many similarities to the Ontario Superior Court of Justice's recent decision in *R v O(F)*.[193] In *O(F)*, the Court excluded the Crown's police gang expert, who would have opined on the characteristics and customs of Toronto street gangs, for many reasons related to the reliability of his expertise. Notably, while the officer had debriefed many alleged gang members, these debriefings were not produced for the defence.[194] And again, in this pre-*Abbey* 2017 decision, the Court drew a somewhat portentous parallel to Totten's work:

> [T]he officer did not give any evidence that he had employed any techniques or methods to ensure, to the degree possible, the veracity of the information that he received through these debriefs. I contrast this absence of validation with the evidence that was given regarding the techniques and methods employed by Dr. Totten in *Abbey* to ensure that the methodology that he followed was valid.[195]

The officer had also not done any undercover work, attended conferences, or reviewed the academic literature on gangs.[196]

---

188 *Giles*, *supra* note 187 at paras 141, 147, 152, 153, 156.
189 *Ibid* at paras 120–46.
190 *Ibid* para 130.
191 *Ibid*.
192 *Ibid* at paras 123–26.
193 *Supra* note 187.
194 *Ibid* at para 17.
195 *Ibid*.
196 *Ibid* at para 16.

All of this is not to say that *Abbey* has been widely effective in encouraging trial judges to seek reliable gang expertise. For instance, and somewhat puzzlingly, the *Abbey* remand judge did not apply a reliability analysis to a police officer's opinion that Abbey was a gang member.[197] Further, the trial judge in *Gager* who was deeply critical of the reliability of Totten's research did not apply the same scrutiny to a police officer who gave substantively the same evidence.[198] It appears both judges expected more out of the sociology guild than they did the police guild. However, some objective level of reliability ought to be demanded and provided when an authority provides such powerful inculpatory opinion evidence in a criminal trial.[199]

6. Towards a Transparent Proficiency Approach

> The Royal Society's motto 'Nullius in verba' is taken to mean 'take nobody's word for it'.[200]

---

197 Rather, that decision was excluded because it went to the ultimate issue and would have forced a cross-examination on inadmissible hearsay. The expert's opinion on the meaning of Abbey's tattoo was excluded for its basis in anecdote. See *R v Abbey*, 2011 ONSC 1260 at paras 41–49, 82 CR (6th) 385.

198 For instance, Justice Clark noted that police expertise should not be held to a standard that is too high because it is predominantly experiential and typically not based on systematic research. On the factors that proved nearly fatal to Totten's admissibility—whether the underlying data was recorded and open to scrutiny—Justice Clark seemed less concerned with the fact that the police officer's (Detective Backus) interviews with confidential informants were privileged and thus not available to the defence. Similarly, when considering whether Backus honoured the boundaries of his discipline, Justice Clark was not particularly concerned that Backus refused to resile from opinions even when "objectively viewed, their foundations appear to be weak." See *Gager* Police, *supra* note 164 at paras 167–83, 193. The same stubbornness on the part of Totten drew much more criticism: see *Gager* Totten, *supra* note 39 at para 42. Justice Clark ultimately allowed Backus to opine on the identity of the gangs that appeared to be involved in the case, but not conclude as to whether the co-accused were members of that gang. See *Gager* Police, *supra* note 164 at paras 215–66.

199 Beyond *Abbey* and *Gager*, several other forms of police gang expertise continue to evade the gatekeeper with little or no reliability analysis. For instance, the Nova Scotia Supreme Court qualified a police officer as an expert on motorcycle gangs over the defence's protestations that he did not have "*reliable* 'specialized knowledge'": *R v Howe*, 2017 NSSC 213 at para 30, 140 WCB (2d) 648 [emphasis in original]. The Court justified its decision merely on the basis of the officer's long experience without delving into whether that experience had a reliable basis. See also *R v Sappleton*, 2010 ONSC 5704, 91 WCB (2d) 699; *R v Valentine* (2009), 87 WCB (2d) 518, 2009 CarswellOnt 8872 (WL Can) (Sup Ct J); *R v Ali*, 2011 BCSC 1850, 103 WCB (2d) 68; *R v Nurse*, 2014 ONSC 2409, 117 WCB (2d) 378; *R v Patterson*, 2015 ONSC 660, 120 WCB (2d) 42; *A(T)*, *supra* note 165. But note that some consideration of reliability may be tacit in a court's analysis: see *R v Farah*, 2016 ONSC 2874 at paras 33, 41, 129 WCB (2d) 577.

200 See The Royal Society, "History of the Royal Society", online: <royalsociety.org/about-us/history/>.

Despite its promise, *Abbey* 2009 seems to not have had the effect that commentators hoped for.[201] In other words, it has not served as an adaptive approach to a panoply of expertise (i.e., one tailored to the "nature of the opinion").[202] Rather, the specialized knowledge characterization is an off-switch for any substantive evaluation of the reliability of experts. In some cases, *Abbey* has allowed for the uncritical admission of untested (yet testable) evidence. In other cases, the say-so of the expert has been replaced by the say-so of the group. The most deferential parts of the judgment have—too often—ruled the day.

Ultimately, it may be that Supreme Court guidance is needed for courts to take specialized knowledge more seriously—in both providing assistance in how to scrutinize specialized knowledge and in rethinking the unhelpful dichotomy between novel or contested science, and specialized knowledge. But, in the interest of adding to the judicial toolkit and in service of a more flexible approach, I offer suggestions based on the transparent proficiency of experts. As I noted above, this suggestion is informed by both the failures and successes of *Abbey*. Moreover, transparency builds off of the text of *Abbey* 2009 itself, which includes an inquiry into the storage, recording, and availability of the underlying data (although, I will suggest broader transparency, including the parameters of the expert's analysis). Transparent proficiency can also be applied in a manner that is sensitive to the expert's opinion and thus is not reliant on blunt distinctions between novel science and specialized knowledge.

Transparency responds to the fact that mere reference to conversations, interviews, and various other experiences the witness has engaged in may be unreliable and misleading. Even scientists working under ostensibly stricter safeguards have frequently fooled themselves and their peers by exploiting undisclosed flexibility in their data and analysis. Brian Nosek, Jeffrey Spies, and Matt Motyl cogently describe how these biases operate in science: "we might remember the gist of what the study was and what we found. Forgetting the details provides an opportunity for reimagining the study purpose and results to recall and understand them in their best (i.e., most publishable) light."[203] There is no reason to think the same is not true among expert witnesses.

The primary meta-scientific reform flowing from Nosek and colleagues' work would enhance transparency. For example, the journal *Science* now

---

201 Dufraimont, "Gatekeeper", *supra* note 2; Edmond & Roach, *supra* note 2.

202 *Abbey* 2009, *supra* note 1 at para 118.

203 Brian A Nosek, Jeffrey R Spies & Matt Motyl, "Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability" (2012) 7:6 Perspectives on Psychological Science 615 at 617 [footnotes omitted].

requires that authors make their data and methods publicly available.[204] And, as mentioned above, pre-registration is becoming increasingly common.[205] While it is certainly important that scientists build a reliable body of knowledge, science can self-correct in most cases. Such opportunities (e.g., appeals or fresh evidence) are much more limited in law—it does not seem like so much of a stretch to suggest that experts providing powerful inculpatory opinions should be held to a standard of transparency. Indeed, as I described above, there is some precedent now that a failure to disclose the foundations of an opinion (for instance, based on confidential informant privilege) can impair the probative force of that opinion so greatly that it is inadmissible.[206]

Along with transparency, courts should determine if the witness has proficiency in the precise task at hand (i.e., to reliably provide the conclusion they are seeking to provide). This requires more than "omnibus"[207] experience as a member of a guild. For instance, it is not enough that a witness has decades of experience as a forensic scientist. If the question is determining whether a DNA sample came from a wet source based on the amount of DNA present, then the witness must be able to reliably do that—or, as we will see, provide convincing indirect evidence that he or she can do that.[208]

The concept of proficiency is an important and growing area of study among academics.[209] Formal proficiency testing (e.g., a fingerprint examiner's accuracy in identifying fingerprints in realistic circumstances according to a validated test) was also recently prescribed by a report of the US President's Council of Advisors on Science and Technology (the "PCAST") on forensic science.[210] Surprisingly, however, courts do not seem to require that experts establish their proficiency, even in fields where tests are routinely performed and when the expert's practices are described

---

204 Science Editorial, *supra* note 96.

205 Preregistration Revolution, *supra* note 97.

206 *Giles*, *supra* note 187; *O(F)*, *supra* note 187. See also *Gager* Police, *supra* note 164 at para 255. Rebecca Wexler, "Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System" (2018) 70:5 Stan L Rev 1343, n 281.

207 *Giles*, *supra* note 187.

208 Further, Likwornik, *supra* note 22, suggests that defining the task at hand can also provide courts with analytic clarity. In the context of the *Awer* trial, Likwornik suggests that the relevant issue was if the DNA transfer was direct or from a secondary source. If the Court had established this as the scope of the evidence, then it would have been even clearer that Denison's opinion lacked probative value.

209 Brandon Garrett & Gregory Mitchell, "The Proficiency of Experts" [forthcoming 2018] 166:4 U Pa L Rev 901 [Garrett & Mitchell]; Chin & Likwornik, *supra* note 1 at 50–52.

210 PCAST Report, *supra* note 183 at 57–59.

as scientific.[211] On the other hand, cogent academic work suggests that examiners in several forensic sciences should not be permitted—as they are—to describe their processes as objective and scientific, because that is misleading.[212] Rather, they should simply provide results of their proficiency tests. Analogously, years of largely unrelated police and forensic experience and accreditations may only mislead the trier of fact, who should be focused on the expert's proficiency on the specific task at hand.

Abbey 2009 may have helped the Canadian case law down a path towards proficiency at the task at hand by encouraging trial judges to define that task at an early stage. This exercise promotes analytic clarity and sets up an inquiry into what exactly the expert must demonstrate proficiency in.[213] Recall that Justice Doherty held that the (first) trial judge in *Abbey* should have established the permissible scope of Totten's evidence as the three general meanings of teardrop tattoos (and not a case-specific judgment of Abbey's tattoo).[214] Several decisions have followed *Abbey* to exclude conclusory opinions about the accused (but note several have not followed this rule).[215] Courts typically justify these decisions by referring to the old ultimate issue rule, saying that while that rule is no longer in general effect, opinions going to the ultimate issue are especially prejudicial.[216]

I suggest that a more useful and parsimonious justification for excluding these conclusory opinions is that they are judgments about which the expert

---

211 See Garrett & Mitchell, *supra* note 209; Gary Edmond et al, "Admissibility Compared: The Reception of Incriminating Expert Evidence (i.e., Forensic Science) in Four Adversarial Jurisdictions" (2013) 3 U Denver Criminal L Rev 31; Gary Edmond, David Hamer & Emma Cunliffe, "A Little Ignorance is a Dangerous Thing: Engaging with Exogenous Knowledge Not Adduced by the Parties" (2016) 25:3 Griffith L Rev 383.

212 Gary Edmond, Matthew B Thompson & Jason M Tangen, "A Guide to Interpreting Forensic Testimony: Scientific Approaches to Fingerprint Evidence" (2014) 13:1 Law, Probability and Risk 1; Jennifer L Mnookin, "The Courts, the NAS, and the Future of Forensic Science" (2010) 75:4 Brook L Rev 1209 at 1219; Bernstein, *supra* note 4 at 481: "Much 'forensic science' testimony is actually connoisseur testimony disguised as science. If one asks (as this author has) fingerprint experts, forensic anthropologists, polygraph examiners, and many other forensic 'scientists' what basis the jury ultimately has to trust their testimony, the answer is that the jury must rely on their training and years of experience" [footnotes omitted]; Matthew B Thompson & Jason Tangen, "The Nature of Expertise in Fingerprint Matching: Experts Can Do a Lot with a Little" (2014) 9:12 PLoS ONE e114759.

213 Likwornik, *supra* note 22.

214 *Abbey* 2009, *supra* note 1 at paras 62–70. That framework evidence was still assailable for a lack of transparency, as the courts in *Gager* and *Abbey* 2017 eventually decided.

215 See *Lucas*, *supra* note 165 at para 9. Note, however, that several courts have chosen to not exclude the individuating portion of the expert's opinion: *R v Tulloch*, 2016 ONSC 3667, 131 WCB (2d) 37; *R v Robertson*, 2017 BCSC 1451, 141 WCB (2d) 621 [*Robertson*]; *Lee*, *supra* note 170; *Murphy*, *supra* note 168; *Dominic*, *supra* note 162.

216 *Awer* ABCA, *supra* note 1 at para 59.

cannot be expected to be reliably proficient. For example, while the expert may be proficient in the pre-scoped expertise (e.g., globally, as a police officer or forensic scientist), the expert has not established that he or she can reliably identify drug traffickers or armed individuals in CCTV footage. Moreover, it seems illogical to admit unreliable evidence simply because of its distance from the ultimate issue.[217] This invites (counterproductive) dispute about how far evidence needs to be from the ultimate issue. In *Awer*, for instance, the Court of Appeal held that the opinion on the source of the DNA should have been admitted because it was not about the ultimate issue.[218] The Supreme Court, however, delved deeper into the case and noted the opinion was essential to the trial judge's decision.[219] The question of ultimate issue, in this case, was little more than a distraction—the evidence was baseless and should have been excluded.

The proper manner of demonstrating proficiency at the task at hand, the onus of which lies on the party tendering the evidence, will depend on the nature of the expertise. For instance, forensic identification can typically be subjected to formal proficiency tests that are as challenging as the instant case. If a forensic gait analyst purports to be able to reliably identify individuals (or reliably compare gait characteristics) based on their gait in videos of similar quality to the case, then that expert should be able to demonstrate that in a proficiency test. Pursuant to the transparency requirement, the conditions of these tests (e.g., the quality of the video) should be disclosed. For judgments that cannot be formally tested, valid *indicia* of proficiency may substitute in some cases. For instance, when a witness has amassed experience in circumstances in which the feedback is immediate and unambiguous (as with the harbour pilot), the trial judge may properly infer proficiency.

Perhaps most critically, transparent proficiency helps remedy an information asymmetry that has been at the root of many miscarriages of justice.[220] Experts, cloaked in the garb of authority, will almost always know more about the subject matter of their opinion than both the judge and jury. This presents the risk that both will uncritically accept the expert's opinion.

---

217 See for instance *Robertson*, *supra* note 215. In that case, the police expert opined that the location of the crime was a place of drug trafficking (an individuating opinion). However, there were no drug trafficking charges. Rather, the Crown relied on the drug trafficking evidence to suggest the accused must have known he was receiving stolen goods. Despite this evidence not going to the ultimate issue, the court excluded the individuating portion of the evidence. While the court did not justify the exclusion on these grounds, I suggest a good reason for exclusion in *Robertson* is that the expert was not demonstrably (or impliedly) proficient in determining if specific locations are places of drug trafficking.

218 *Awer* ABCA, *supra* note 1 at para 59.

219 *Awer* SCC, *supra* note 1 at para 4.

220 MacFarlane, "Wrongful Convictions", *supra* note 9.

Indeed, as Justice Doherty noted in *Abbey* 2009, "the most important risk is the danger that a jury will be unable to make an effective and critical assessment of the evidence."[221] By laying bare the foundations of the opinion and the expert's ability to provide any judgments related to that foundation, the factfinder stands a much better chance.

221 *Abbey* 2009, *supra* note 1 at para 90.