Improving expert evidence: the role of open science and transparency

Jason M. Chin

TC Beirne School of Law, University of Queensland; Institute for Globally Distributed Open Research and Education (IGDORE)

Bethany Growns

School of Psychology, University of New South Wales

David T. Mellor

Center for Open Science

*We are indebted to Shari Diamond, Gary Edmond, Jeremy Freese, Glenn Harrell, Sarah Lawsky, and John Monahan for their insightful comments on previous drafts. We also thank the participants of the Second Annual Junior Faculty Forum for Law and STEM (Northwestern Pritzker School of Law) for helpful feedback. Many of the ideas herein were also discussed at the annual meeting of the Evidence-Based Forensics Initiative and with the Interdisciplinary MetaResearch Group (Fiona Fidler, Hannah Fraser, Felix Singleton Thorn, Elise Gould, and Bonnie Wintle). Sarah Hamid provided indefatigable editorial and research support.

Cite as: Jason M Chin, Bethany Growns & David T Mellor, "Improving expert evidence: the role of open science and transparency" (2019) 50:2 Ottawa Law Review 365.

Abstract

Both science and expert evidence law are undergoing significant changes. In this article, the authors compare these two movements – the open science movement and the evidence-based evidence movement. The open science movement encompasses the recent discovery of many irreproducible findings in science and the subsequent move towards more transparent methods. The evidence-based evidence movement is the discovery that many forms of expert evidence are unreliable, and that they have contributed to wrongful convictions. The authors identify similarities between these movements, which suggest how courts and legal actors may learn from the open science movement to produce more accurate results. Expert witnesses should comport themselves as rigorous open scientists to produce evidence that is more susceptible to evaluation. Parties should be subjected to more specific and rigorous disclosure requirements because research has shown that even leading scientists find it easy to discount and suppress findings that do not support their hypotheses. And trial judges, as gatekeepers, should not defer to the generally accepted practices that have proven insufficient in the mainstream sciences.

Table of Contents

Part I. Introduction

The way in which science is conducted and communicated is fundamentally changing. Scientists and journals are increasingly adopting practices aimed at making science more transparent, reproducible, and democratic.[1] In this article, we will demonstrate several parallels between this movement in science – the open science movement – and similar trends in expert evidence law. In particular, the genesis of many aspects of the open science movement was the realization that longstanding practices had failed, allowing spurious findings to reach general acceptance.[2] A similar pattern has been observed in several classic fields of expert evidence.[3] These parallels have significant consequences for law, a field where flaws in its truth-determining mechanisms have contributed to grave miscarriages of justice.[4] Throughout this article, our central thesis is that open science-inspired reforms align with and further the ideals of expert evidence: they help produce knowledge that is susceptible to critical evaluation.

The open science movement responded to the discovery of several results, many previously seen as robust and well-established, that could not be independently reproduced.[5]

---

[1] US, Committee on Toward and Open Science Enterprise, *Open Science by Design: Realizing a Vision for 21st Century Research* (Washington, DC: National Academies Press, 2018), online: <https://www.nap.edu/catalog/25116/open-science-by-design-realizing-a-vision-for-21st-century> [NASEM, Open Science Report]; Marcus R Munafò et al, "A Manifesto for Reproducible Science" (2017) 0021 Nature Human Behaviour 1 [Munafò].

[2] Leif D Nelson et al, "Psychology's Renaissance" (2018) 69 Annual Review of Psychology 511 at 512-514 [Nelson et al]. NASEM, Open Science Report, *ibid* at 31.

[3] Michael J Saks & David L Faigman, "Failed Forensics: How Forensic Science Lost Its Way and How It Might Yet Find It" (2008) 4 Annual Review of Law & Social Science 149 [Saks & Faigman]; Lisa Dufraimont, "New Challenges for the Gatekeeper: The Evolving Law on Expert Evidence in Criminal Cases" (2012) 58:3&4 Crim LQ 531 [Dufriamont]; Gary Edmond & Kent Roach, "A Contextual Approach to the Admissibility of the State's Forensic Science and Medical Evidence" (2011) 61:3 UTLJ 343 [Edmond & Roach]; Alan D Gold, *Expert Evidence in Criminal Law: The Scientific Approach*, 2nd ed. (Toronto: Irwin, 2009).

[4] See: Brandon L Garrett & Peter J Neufeld, "Invalid Forensic Science Testimony and Wrongful Convictions" (2009) 95:1 Va L Rev 1-97 [Garrett & Neufeld]; Ontario, *Report of the Kaufman Commission on Proceedings Involving Guy Paul Morin* (Toronto: Ministry of the Attorney General, 1998) (The Honourable Fred Kaufman) [Kaufman Report]; Ontario, Inquiry into Pediatric Forensic Pathology in Ontario: Report (Toronto: Ministry of the Attorney General, 2008) vols 1–4 (The Honourable Stephen T Goudge) [Goudge Report].

[5] See NASEM Open Science Report, *supra* note 1 at 31.

While these examples could at first be disregarded as outliers, the scientific community has come to acknowledge that they reflect an endemic problem. For instance, largescale attempts to reproduce established social scientific findings have only succeeded about 40-60% of the time (and have reported considerably weaker findings).[6] Reflecting these surprising (non)findings, a survey of 1,576 scientists in the journal *Nature* reported that 90% of those surveyed believed that science had a reproducibility problem.[7] Responses were consistent across fields, from chemistry to physics.[8]

These meta-scientific revelations have begun to inform and inspire reform. For simplicity, we use the term "open science movement" to refer to the totality of these developments. The movement, however, is broader than what we will focus on in this article, with an earlier (but ongoing) part of the campaign focused on access to scientific articles (i.e., removing paywalls to publicly-funded research). Rather, we will devote most of our analysis to transparency and openness as ways to improve the rigour of research methods by reducing undisclosed flexibility. This component of the movement has taken on various names in the literature, such as the "replicability crisis"[9] and "credibility revolution".[10]

In parallel with the open science movement, the forensic sciences (i.e., science designed to answer legal questions) and scientific evidence in court more broadly have also been subjected

[6] Open Science Collaboration, "Estimating the Reproducibility of Psychological Science" (2015) 349 Science 943 [OSC]; Colin F Camerer et al, "Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015" (2018) Nature Human Behaviour 637 [Camerer et al]; Andrew C Chang & Phillip Li, "Is economics research replicable? Sixty published papers from thirteen journals say "usually not."" (2015) Finance and Economics Discussion Series 2015-083. Washington, DC: Board of Governors of the Federal Reserve System, online:<http://dx.doi.org/10.17016/FEDS.2015.083> [Chang & Li].
[7] 52% classified it as a significant problem and 38% said it was a "slight crisis". Monya Baker, "1,500 scientists lift the lid on reproducibility" (2016) 533 Nature 452 [Nature Survey].
[8] *Ibid* at 453.
[9] Harold Pashler & Eric-Jan Wagenmakers, "Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence?" (2012) 7:6 Perspectives on Psychological Science 528.
[10] Simine Vazire, "Implications of the Credibility Revolution for Productivity, Creativity, and Progress" (forthcoming) Perspectives on Psychological Science [Vazire, Credibility Revolution].

to increased scrutiny.[11] Although open science and evidence law have almost never been explicitly linked,[12] many of the issues are remarkably similar. For instance, the challenges in science flow from cognitive biases that focus scientists on the data that confirm their hypotheses at the expense of those that do not (despite both sets of data having equal evidential).[13] These are the very biases present in expert evidence that have been uncovered by academics[14] and reports from peak bodies of scientists and jurists convened to address failures of the criminal justice system.[15] Moreover, both movements are associated with a lack of transparency[16] and a preoccupation with eminence over methodology.[17]

It is meaningful that mainstream scientists are being accused of many of the same practices that have resulted in wrongful convictions in law. Specifically, it means that it is insufficient that expert witnesses be directed to behave like scientists. Instead, they should

---

[11] See the sources at *supra* note 2.
[12] But see: Jason M Chin, "Psychological science's replicability crisis and what it means for science in the courtroom" (2014) 20:3 Psychology, Public Policy and Law 225 [Chin, Replicability Crisis]; Chuan-Peng Hu, "Open science as a better gatekeeper for science and society: a perspective from neurolaw" (2018) 63 Science Bulletin 1529.
[13] Munafò, *supra* note 1 at 1.
[14] D Michael Risinger et al, "The *Daubert*/*Kumho* implications of observer effects in forensic science: Hidden problems of expectation and suggestion" (2002) 90:1 California LR 1 [Risinger et al]; Gary Edmond et al, "Contextual bias and cross-contamination in the forensic sciences: the corrosive implications for investigations, plea bargains, trials and appeals" (2014) 14:1 Law, Probability and Risk 1 [Edmond et al]; Gary Edmond & Emma Cunliffe, "Cinderella Story? The Social Production of a Forensic 'Science'" (2017) 106:2 The Journal of Criminal Law & Criminology 219 [Edmond & Cunliffe]; Itiel E Dror & David Charlton, "Why Experts Make Errors" (2006) 56:4 Journal of Forensic Identification 600 [Dror, Context]; Itiel E Dror, David Charlton, & Alisa E Péron, "Contextual information renders experts vulnerable to making erroneous identifications" (2006) 156 Forensic Science International 74 [Dror, Mayfield].
[15] US, President's Council of Advisors on Science and Technology, *Forensic Science in Criminal Court: Ensuring Scientific Validity of Feature-Comparison Methods* (Washington, DC: Executive Office of the President, 2016) at 32-32 [PCAST Report]; US, National Research Council, *Strengthening Forensic Science in the United States: A Path Forward* (Washington, DC: National Academies Press, 2009) at 184-185 [NAS Report]; Goudge Report, *supra* note 4 at 43, 69, 79, 153-156, 374-377.
[16] In science, see Brian A Nosek et al, "Promoting an open research culture" (2015) 348:6242 Science 1422 [Nosek et al, TOP]. In law, see Gary Edmond et al, "Model forensic science" (2016) 48:4 Australian Journal of Forensic Sciences 496 at 497-499 [Edmond et al, Model Forensic Science].
[17] In science, see Robert K Merton, *The Sociology of Science* (Chicago: University of Chicago Press, 1973) [Merton]; Simine Vazire, "Our obsession with eminence warps research" (2017) 547:7 Nature [Vazire, Eminence]. In law, see: *Béland v R*, [1987] 2 SCR 398 at para 78, 42 DLR (4th) 641 [*Béland*]; *R v Mohan*, [1994] 2 SCR 9 at para 23, 114 DLR (4th) 419 [*Mohan*]; *Nguyen v R*, [2017] NSWCCA 4 at para 28.

behave like open scientists. In other words, expert witnesses should be expected to behave as *scientists should be expected to behave*: candidly sharing the results of research, avoiding appeals to status, and skeptically scrutinizing their own work and that of others.[18] By embracing these norms (rather than simply mainstream ones), expert witnesses can provide evidence that is both more trustworthy and more susceptible to rational evaluation.[19]

In Parts II and III, we will go on to describe the geneses of the open science movement and the evidence-based evidence movement, respectively. Then, in Part IV, we analyze these movements, identifying six points of comparison. These similarities suggest mutually applicable reforms – a commitment to transparency and openness can improve the accuracy of both science and expert evidence. Part V then delves into legal reforms. Part VI concludes with two ancillary benefits of open expert evidence: improved trust and efficiency.

Before delving into the substance of our article, we should provide a brief caution. While we will suggest that insights from the open science movement have much to offer fact-finding in court, we note that science and law do not share all of their values. Importantly, law must balance other interests, like procedural fairness, adversarial imbalance, and finality.[20] For instance, courts should be sensitive to the fact that a well-heeled corporate defendant facing a product liability claim would often be expected to have access to more sophisticated case-relevant scientific evidence than the plaintiff. On the other hand, the criminally accused often

[18] Robert K Merton, "The Normative Structure of Science", in Robert K Merton, ed, *The Sociology of Science: Theoretical and Empirical Investigations* (Chicago: University of Chicago, 1973).
[19] See generally, Gary Edmond, "Forensic Science Evidence and the Conditions for Rational (Jury) Evaluation" (2015) 39:1 Melbourne University LR 77 [Edmond, Rational Jury]; Alan W Mewett & Peter J Sankoff, *Witnesses* (Toronto: Carswell, 2018) at 16.3 [Mewett & Sankoff]
[20] See Gary Edmond & Mehera San Roque, "Just(,) Quick and Cheap? Contemporary Approaches to the Management of Expert Evidence" in Michael Legg (ed) *Resolving Civil Disputes* (LexisNexis, 2016) [Edmond & San Roque].

face resource constraints, making it difficult to hire a rebuttal expert. As a result, we do not suggest holding all parties to the highest standard of open scientific evidence.

Rather, as we will explain in Parts V and VI, openness of the foundations and application of expert knowledge simply results in evidence that is more susceptible to rational evaluation. Indeed, as noted above, the open science movement is sometimes referred to as the "credibility revolution"[21] because the reforms transcend value-laden categorizations of science and non-science, applying across fields of knowledge generation. Similarly, in law, academics have criticized rules that require slippery taxonomies of expert evidence (e.g., those that would give different scrutiny to science as opposed to what a court might characterize as non-science, or put less weight on evidence simply because it was generated for the purposes of litigation).[22] Openness, as we will demonstrate, cuts across quantitative and qualitative disciplines. Put simply – and as demonstrated by the meta-scientific findings we discuss below – openness makes the strengths and weaknesses of the expert's opinion more apparent and can therefore promote justice.

Part II. From crisis to renaissance: A brief history of the open science movement

> If a team of research psychologists were to emerge today from a 7-year hibernation, they would not recognize their field. Authors voluntarily posting their data. Top journals routinely publishing replication attempts, both failures and successes. Hundreds of researchers preregistering their studies. Crowded methods symposia at many conferences. Enormous increases in sample sizes. Some top journals requiring the full disclosure of measures,

[21] Vazire Credibility Revolution, *supra* note 10.
[22] See Jason M Chin, "*Abbey* Road: The (ongoing) journey to reliable expert evidence" (2018) 96:3 Canadian Bar Review 422 at fn 2 [Chin, *Abbey* Road]. As to litigation-driven science, see Susan Haack, "What's Wrong with Litigation-Driven Science? An Essay in Legal Epistemology" (2008) 38 Seton Hall LR 1053 at 1077 [Haack]: "The fact that research is litigation-driven in the stronger sense, I have argued, makes it likely to be biased. Biased research doesn't necessarily produce false results; nor does it necessarily produce false results more often than true.".

> conditions, exclusions, and the rules for determining sample sizes. Several multilab replication efforts accepted for publication before any data were collected. Overall, an unprecedented focus on replicability. What on earth just happened?[23]

The open science movement was spurred a by a surprising number of reports of published studies proving irreproducible (another thrust of open science is dedicated to making paywalled scientific journals available to the public, especially when the underlying research was publicly funded).[24] In other words, researchers attempted to recreate the findings of previous studies, but found inconsistent or considerably smaller effects. Such incidences were concentrated in pre-clinical and clinical medical research[25] and psychology.[26] Many fields, however, are struggling with the reproducibility of their findings.[27] These failures to reproduce inspired largescale systematic studies (mentioned in Part I), finding that studies published in eminent journals regularly proved irreproducible.[28]

These demonstrations of systemic problems within science raised difficult questions: most fundamentally, why were these studies, which carried the *indicia* of good science (e.g.,

[23] Nelson et al, *supra* note 2 at 17.2.
[24] NASEM, Open Science Report, *supra* note 1 at 23.
[25] Francis S Collins & Lawrence A. Tabak, "NIH plans to enhance reproducibility" (2014) 505 Nature 612 [Collins & Tabak]; Leonard P Freedman, Ian M Cockburn & Timothy S Simcoe, "The Economics of Reproducibility in Preclinical Research" (2015) 13:6 PLoS Biol [Freedman et al]; John PA Ioannidis, "Contradicted and Initially Stronger Effects in Highly Cited Clinical Research" (2005) 294:2 JAMA 218; C Glenn Begley & Lee M Ellis, "Raise standards for preclinical cancer research" (2012) 483 Nature 531.
[26] See Nelson et al, *supra* note 2 at 512-514.
[27] In cognitive neuroscience, see Denes Szucs & John PA Ioannidis, "Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature" (2017) PLoS Biol 1. In economics, see Chang & Li, *supra* note 6. In evolutionary biology, see Daiping Wang et al, "Irreproducible text-book 'knowledge': The effects of color bands on zebra finch fitness" (2018) 72:4 Evolution 961. In genetic association studies, see John PA Ioannidis et al, "Replication validity of genetic association studies" (2001) 29 Nature Genetics 306. Sociologists are also currently wresting with reforms in their field: Jeremy Freese & David Peterson, "Replication in Social Science" (2017) 43 Annu Rev Socio 147-165 [Freese].
[28] See the studies in *supra* note 6.

testing with a low reported error rate and publication in leading journals)[29] nevertheless false positive findings? Here, we note that many of these discoveries of false positives were not entirely surprising to some (in particular, sociologists of science), who long warned that science was more human and error prone than most realised.[30] Importantly, however, largescale replication attempts (in the past, replications of previous work were rare) gave teeth to the sociological concerns and encouraged rigorous meta-scientific research into the precise weaknesses in science.[31]

One of the most widely-studied of those weaknesses is what have come to been known as "researcher degrees of freedom" or "questionable research practices" (QRPs for short).[32] These are *undisclosed* choices that researchers can use to increase their chances of finding a result that meets conventional levels of statistical certainty:[33]

> In the course of collecting and analyzing data, researchers have many decisions to make: Should more data be collected? Should some observations be excluded? Which conditions should be combined and which ones compared? Which control variables should be considered?

To understand how such tactics might give a reported study a misleading sheen, imagine if a basketball team could, instead of respecting a predetermined time limit, strategically decide to stop a match when they were winning – it would bias the game in their favour. Indeed, in one

[29] And these are indeed three factors of scientific validity found in the U.S. Supreme Court's foundational expert evidence decisions, *Daubert v Merrell Dow Pharmaceuticals Inc*, 509 US 579 (1993) [*Daubert*]. We will discuss this decision in Part III.
[30] Simon A Cole & Alyse Bertenthal, "Science, Technology, Society, and Law" (2017) 13 Annual Review of Law and Social Science 351.
[31] Matthew C Makel, Johnathan A Plucker & Boyd Hegarty, "Replications in Psychology Research: How Often Do They Really Occur?" (2012) 7:6 Perspectives on Psychological Science 537; Brian A Nosek & Timothy M Errington, "Making sense of replications" (2017) 6 ELife e23383.
[32] 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 1359-1366 [Simmons et al].
[33] *Ibid* at 1359.

influential study, researchers found that use of just four[34] of these researcher degrees of freedom could inflate what would appear to be a 5% false positive rate to 60.7%.[35]

Studies examining the damage QRPs can do were lent additional force by survey studies in psychology, ecology, and evolutionary biology, finding that researchers in those fields (responding anonymously) used some questionable research practices at a high rate (50-60% in some cases).[36]

Reflecting the centrality of openness to scientific progress, the National Academies of Sciences, Engineering and Medicine (NASEM) issued a report in 2018 laying out a plan to improve science.[37] It acknowledged that "research conducted openly and transparently leads to *better science*" and that scientific findings "are more likely to be credible—or found wanting—when they can be reviewed, critiqued, extended, and reproduced by others".[38] We will now briefly detail these reforms as they have been expressed in peer review and publication guidelines, scientific methodology, as well as in key enhancements to the infrastructure of science (e.g., a central website to store data and promote collaboration (see below)).

Reforming practices at academic journals is crucial because they are one of the primary venues by which science is vetted and transmitted. Indeed, the journal *Nature* expressly admitted its own role in the replicability crisis: "The problems arise in laboratories, but journals such as this one compound them when they fail to exert sufficient scrutiny over the results that they publish, and when they do not publish enough information for other researchers to assess results

[34] They were: "flexibility in (a) choosing among dependent variables, (b) choosing sample size, (c) using covariates, and (d) reporting subsets of experimental conditions." See *ibid* at 1360.
[35] *Ibid* at 1361.
[36] Leslie John et al, "Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling" (2012) 23:5 Psychological Science 524 *supra* note 29 [John et al]. Hannah Fraser et al, "Questionable research practices in ecology and evolution" (2018) 13:7 PLoS ONE e0200303.
[37] NASEM, Open Science Report, *supra* note 1.
[38] *Ibid* at 90 [emphasis in original].

properly."[39] *Nature* went on to institute a host of improvements to its editorial policy, including enhanced reporting of methodology and abolishing space restrictions in those sections.[40] In medical research, these practices have been linked with modest improvements in reporting practices.[41]

More recently, a committee of researchers, journal editors, and funding organizations devised the Transparency and Openness Promotion ("TOP") Guidelines.[42] They include eight standards that can be implemented at three levels of rigour, from to encouragement to requiring the standard be met (with journal verification that it was met).[43] Several standards focus on transparency, in particular that of the author's data, analysis, materials, and research design. They also provide for replication (i.e., studies aimed at directly recreating previous studies) to address the barriers to funding and publishing important confirmatory research.

Further, the guidelines include a preregistration standard. Preregistration guards against QRPs by requiring that researchers must record their procedures *before* collecting data.[44] As we will discuss below, the preregistration record is uneditable and often housed in an online database. Such precautions prevent researchers from strategically supressing measures that did not support their hypothesis. This is not to say that deviations from the preregistered plan are not allowed and that the data collected from studies with deviations are worthless. Rather, it simply

---

[39] Editorial, "Announcement: Reducing Our Irreproducibility" (2013) 496 Nature 398 at 398.
[40] *Ibid*.
[41] Lucy Turner et al, "Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review" (2012) 1:60 Systematic Reviews 1; SeungHye Han et al, "A checklist is associated with increased quality of reporting preclinical biomedical research: A systematic review" (2017) 12:9 PLoS ONE e0183591.
[42] Nosek et al, TOP, *supra* note 16.
[43] *Ibid*.
[44] Nelson et al, *supra* note 2 at 519-520; David Mellor, Simine Vazire & D Stephen Lindsay, "Transparent science: A more credible, reproducible, and publishable way to do science" (forthcoming) in *Guide to Publishing in Psychology Journals*, Robert J Sternberg, ed; Brian A Nosek et al, "The Preregistration Revolution" (2018) Proceedings of the National Academy of Sciences 201708274 [Nosek et al, Preregistration].

ensures that the peer-reviewer and consumer of the science know that the plan changed, so that such deviations can be evaluated.

As of February 2009, nearly 5,000 journals have endorsed the TOP Guidelines and committed to reviewing their own standards within a year.[45] Many have implemented the guidelines at varying levels.[46] For instance, the journal *Science* recently revised its editorial policy to require that authors make their data available, subject to "truly exceptional circumstances".[47]

The changes in journals are scaffolded by new tools available to researchers.[48] Notably, the Open Science Framework (OSF) is a free web platform for open science.[49] It provides support and infrastructure for users at all stages of the research process. For instance, it allows researchers to conform with TOP by preregistering their study or by sharing data, code, and other digital materials. Such sharing is important because data analysis and the computation underlying research is increasingly complex and central to the scientific process. As a result, sharing can produce efficiencies and the chance that researchers will catch each others' mistakes.[50]

[45] The list of journals and publishers is one file with the Center for Open Science: <http://cos.io/TOP>. The obligations involved with endorsing the guidelines is also on file: < https://osf.io/pvf56/>.
[46] Center for Open Science, "Center for Open Science, online: <http://cos.io/TOP>.
[47] Science, "*Science*: Editorial policies", online: <http://www.sciencemag.org/authors/science-editorial-policies> [Science Editorial Policy].
[48] See Munafò, *supra* note 1 at 2-7 for a review.
[49] Center for Open Science, "Open Science Framework", online: <https://osf.io/>; Munafò, *ibid* at 4. See also "As Predicted", online: <https://aspredicted.org/>.
[50] Victoria Stodden, "Trust Your Science? Open Your Data and Code" (2011) Amstat News 21

Part III. The evidence-based evidence movement

As with the open science movement, the evidence-based evidence movement has been informed by past mistakes.[51] Most notably, legal scholars have documented the numerous miscarriages of justice attributable to expert witnesses giving invalid scientific evidence.[52] For instance, one U.S. study found invalid forensic science in 63% of cases in which forensic scientific testimony was tendered.[53] These revelations lent credence to longstanding worries that the practice of forensic science was subject to a host of errors and uncertainties. Throughout this period, the rules regulating the admission of scientific evidence were becoming formally more demanding (spurred by the admission of junk science in U.S. civil litigation).[54] In this part, we will briefly review these three ingredients: wrongful convictions, academic research focused on the forensic sciences, and the legal regulation of scientific evidence.

As with the U.S., several Canadian wrongful convictions have been caused, at least in part, by invalid or misleading forensic science. For instance, Guy Paul Morin's wrongful conviction, was in part based on scientifically invalid hair, fibre and blood testing.[55] About a decade later, Justice Goudge's pivotal *Inquiry into Pediatric Forensic Pathology in Ontario* (the "Goudge Report") found many failings in the work of pediatric forensic pathologist Charles Smith. His invalid testimony contributed to 14 wrongful convictions.[56] And in the child

[51] By using the term "evidence-based evidence movement", we are attempting to describe a general trend across several, sometimes related, areas of legal scholarship. See: David Paciocco, "Taking a 'Goudge' out of Bluster and Blarney: an 'Evidence-Based Approach' to Expert Testimony" (2009) 13:2 Canadian Criminal Law Review 135 [Paciocco]; D Michael Risinger, "Navigating expert reliability: Are criminal standards of certainty being left on the dock?" (2000) 64 Albany Law Review 99 [Risinger, Docks].
[52] See the sources at *supra* note 4.
[53] Garrett & Neufeld, *supra* note 4.
[54] *Daubert*, *supra* note29; *General Elec Co v Joiner*, 522 US 136 (1997); *Kumho Tire Co Ltd v Carmichael*, 526 U.S. 137 (1999) [*Kumho*].
[55] Kaufman Report, *supra* note 4 at 6-9.
[56] Goudge Report, *supra* note 4; Emma Cunliffe & Gary Edmond, "What Have We Learned? Lessons from Wrongful Convictions in Canada" in Benjamin Berger, Emma Cunliffe, and James Stribopoulos (eds) *To ensure that*

protection arena, many children were taken from their parents – the "capital punishment"[57] of child welfare law – based on invalid hair tests from the Motherisk program that purported to detect the use of drugs or alcohol in parents. These were detailed in the recent reports of Justice Lang and Justice Beaman.[58]

Wrongful convictions enlivened an existing scholarship evaluating the validity and reliability of the forensic sciences, and their suitability as inculpatory evidence.[59] Much of this research had warned that many forensic practices (e.g., fingerprint examiners comparing a found print to a suspect's) had never actually been tested and likely were more error prone and subjective than practitioners were acknowledging in court. of a suspect. Moreover, even if these foundations were established, forensic practitioners have long resisted practices designed to resist bias, such as blinding themselves (i.e., keeping themselves unaware) to the identity of accused and other visceral case details irrelevant to their task.[60] These warnings eventually reached critical mass in the form of a review by the National Research Council of the National Academies of Science (the "NAS Report").[61] The NAS Report was quite frank in its criticism of the forensic sciences and the failure of courts to regulate such evidence:[62]

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

Justice is Done: Essays in Memory of Marc Rosenberg (Toronto: Thomson Reuters, 2017) at 133 [Cunliffe & Edmond wrongful convictions].

[57] Ontario, *Harmful Impacts: The Reliance on Hair Testing in Child Protection Report of the Motherisk Commission* (Ontario: Ministry of the Attorney General, 2018) (The Honourable Judith C Beaman) xxii-xxiii, quoting Justice Susan E Lang [Beaman Report].

[58] Beaman Report, *ibid*; The Honourable Susan E Lang, Report of the Motherisk Hair Analysis Independent Review (Toronto: Ministry of the Attorney General, 2015) [Lang Report].

[59] See Jennifer L Mnookin, "The Courts, the NAS, and the Future of Forensic Science" (2010) 75:4 Brook LR 1209 at 1228-1234 [Mnookin].

[60] PCAST Report, *supra* note 15 at 31-32. And see Edmond et al, *supra* note 14.

[61] NAS Report, *supra* note 15.

[62] *Ibid* at 53.

The NAS Report was followed by 2016's (U.S.) President's Council of Advisors on Science and Technology report on forensic science, which found that little had changed since the earlier report.[63]

The third part of the evidence-based evidence story relates to the latter part of the NAS Report's conclusion – the (in)effectiveness of courts in regulating scientific evidence. During the early 1990s, courts in both the U.S. and Canada shifted from a deferential approach to expert evidence to one that is – at least formally – more hands-on.[64] In the U.S., the foundational case is *Daubert v Merrell Dow Pharmaceuticals*.[65] In *Daubert*, the Supreme Court overruled the previous doctrine that allowed expert evidence when it was generally accepted by the scientific community from which it came (this rule was in *Frye v United States*).[66]

The Court in *Daubert* held that trial judges must not defer to general acceptance as they had in *Frye*, but instead evaluate the science themselves to determine if it is sufficiently reliable to admit into court.[67] To assist trial judges in exercising their newly enhanced gatekeeping responsibility, the Supreme Court provided four factors: (1) whether and how the opinion had been tested; (2) the error rate associated with the opinion; (3) the peer review and publication status of the opinion; and (4) its acceptance within the relevant field of knowledge.[68]

Canadian courts responded to the proliferation of expert evidence in a similar way and along roughly the same timeline. In 1993, the Supreme Court of Canada issued its decision in *R v*

[63] PCAST Report, *supra* note 15 at 122.
[64] This was not an express response to forensic science, but a more general response to increasing expert evidence in court, see NAS Report, *supra* note 15 at 89.
[65] *Supra* note 329.
[66] 54 App DC 46, 293 F 1013 (1923).
[67] See *Daubert*, *supra* note 29 at 580, 589.
[68] *Daubert*, *supra* note 29 at 592-594.

Mohan.[69] *Mohan*, which still stands as Canada's leading expert evidence decision, clarified and strengthened the rules for admitting expert opinion. Justice Sopinka, writing for the court, held that expert evidence is only admissible if it meets four criteria: (1) relevance; (2) necessity in assisting the trier of fact; (3) no other exclusionary rule applies; and (4) the tendered expert is properly qualified.[70]

Relevance, under *Mohan,* includes both logical relevance and a balancing of the benefits and costs of admitting the evidence (sometimes referred to as legal relevance).[71] These factors include the reliability of the evidence and the ability of the jury to rationally evaluate the basis of the opinion.[72] Furthermore, Justice Sopinka remarked that "expert evidence which advances a novel scientific theory or technique" should receive special scrutiny, including meeting a threshold level of reliability and being essential to the trial.[73] Post-*Mohan* decisions elaborated on how reliability should be assessed,[74] with the Supreme Court expressly applying the *Daubert* factors in *R v J (J-L)*.[75]

Several post-*R v J (J-L)* developments should be noted. In 2015, the Supreme Court in *White Burgess Langille Inman v Abbott and Haliburton Co* reformulated the *Mohan* framework into a two-step test.[76] At the first step, the evidence's proponent must establish logical relevance, plus the three other threshold conditions from *Mohan*.[77] Also at the first stage, if the evidence is

---

[69] *Supra* note 17. See Sidney N Lederman, Alan W Bryant & Michelle K Fuerst, *The Law of Evidence in Canada*, 4th ed (Ontario: LexisNexis, 2014) at 783-790.
[70] *Mohan*, *supra* note 17 at paras 17-21.
[71] See *White Burgess Langille Inman v Abbott and Haliburton Co*, 2015 SCC 23 at paras 23-24, [2015] 2 SCR 182 [*White Burgess*].
[72] *Mohan*, *supra* note 17 at para 23; *White Burgess*, *ibid* at para 24.
[73] *Mohan*, *ibid* at para 32; *R v Dimitrov*, 68 OR (3d) 641, 181 CCC (3d) 554.
[74] Sankoff & Mewett, *supra* note 19 at 16.3(c)(i).
[75] 2000 SCC 51 at paras 9-15, [2000] 2 SCR 600 [*JLJ*].
[76] *White Burgess*, *supra* note 71 at para 23.
[77] They are: necessity, absence of an exclusionary rule, and a properly qualified expert. See *White Burgess*, *ibid* at para 23.

"based on novel or contested science or science used for a novel purpose",[78] then it must be scientifically valid and reliable pursuant to *Daubert*.[79] At the second stage, the trial judge must consider legal relevance, a calculus that, as noted above, includes the reliability of the evidence.[80] *White Burgess* is also well known for establishing that an expert's lack of independence and impartiality may be cause to exclude that evidence (rather than simply being a matter of weight).[81] Recent high-profile decisions have relied on these new rules to exclude expert evidence.[82] But while this new interest in excluding expert witnesses for partiality is promising, other courts have entertained creative ways to circumvent the expert evidence rules. For instance, science is regularly characterized as "specialized knowledge" and thus granted deference.[83]

In reviewing the Goudge Report, (now Justice) David Paciocco suggested it represented a paradigm shift. Judges should no longer allow experts to simply say "trust me" because that gave way to the many miscarriages of justice documented in the Report.[84] Instead, judges should insist that the expert "show me".[85] As we will detail in what follows, that is also the foundational prescription of open science reformists. Unfortunately, openness is still not regularly provided by experts, nor demanded by trial judges.

---

[78] *Ibid*.
[79] *Ibid*.
[80] *Ibid* at para 24.
[81] *Ibid* at paras 46-54. For a review of the pre-*White Burgess* law, see the lucid discussion in Sankoff & Mewett, *supra* note 19 at 16.8(c).
[82] See *Bruff-Murphy v Gunawardena*, 2017 ONCA 502 at paras 42-70, 414 DLR (4th) 65 [*Bruff-Murphy*]. See also *R v McManus* 2017 ONCA 188, 353 CCC (3d) 493 [*McManus*]; *JP v British Columbia (Children and Family Development)*, 2017 BCCA 308, BCLR (6th) 17 [*JP v BC*]. For a review, see *Forensic Science Evidence and Expert Witness Testimony: Reliability Through Reform*? Paul Roberts & Michael Stockdale, eds (Cheltenham, UK: Edward Elgar, forthcoming 2018)
[83] See Chin, *Abbey* Road, *supra* note 22 at fn 2.
[84] Paciocco, *supra* note 51 at 146, 155-156.
[85] *Ibid* at 156.

Part IV. Six parallel challenges

We will now discuss six commonalities shared by the open science movement and the evidence-based evidence movement: flawed incentives, excessive flexibility, motivated reasoning, immodesty, appeals to eminence, and some fraudulent practice. There are at least two reasons this comparison matters. First, it suggests shared remedies. The methods and principles that open science reformists have advocated for may inform how expert evidence ought to be produce and presented to courts. Second, these commonalities reinforce the notion that courts simply cannot defer to mainstream scientific norms and practices that have often proved inadequate. Rather, they should be aware of the reforms going on within science and, when appropriate, should hold experts to these new standards.

IV(a) A flawed incentive system

> Publishing…lies at the very heart of modern academic science—at levels ranging from the epistemic certification of scientific thought to the more personal labyrinths of job security, quality of life, and self-esteem.[86]

> Their livelihoods; their everyday relations with colleagues, superiors, and sometimes police; their personal identities – all are tied up in thinking and behaving as their group expects. And what is expected of them is to help complete the case against people suspected of having committed crimes.[87]

These two quotes, the first about academic scientists and the second about expert witnesses, are eerily similar. This is because both movements find their source in a misaligned incentive system.

[86] Michael J Mahoney, "Open Exchange and Epistemic Progress" (1985) 40 American Psychologist 29-39 at 30 [Mahoney].
[87] Michael J Saks & Barbara A Spellman, *The Psychological Foundations of Evidence Law* (New York: NYU Press, 2016) at 209 [Saks & Spellman].

In the open science movement, the culpable incentive is publication. Publication is the currency of the academic sciences, influencing promotion, grant success and personal well-being.[88] Unfortunately, a finding's publishability is often a poor gauge of its truth value.[89] Journals value novelty over rigour.[90] They have also not historically published replication research (i.e., findings confirming or disconfirming previous studies).[91] Researchers – seeking publication – are aware of these rules and play the game.[92] They conduct research with small sample sizes, do not attempt (or report) replications, and employ QRPs. Results that did not uncover an interesting finding (e.g., the result was not statistically significant) sit in file drawers.[93] As a result, the strength of published findings is likely overstated.

Flawed incentives also contribute to the problems identified in the evidence-based evidence movement. They flow from an adversarial culture that incentivizes favourable legal results over truth.[94] For example, the prosecutors who decide how to deploy forensic scientific expertise have drawn criticism for seeking convictions instead of serving the administration of justice.[95] Such behaviour has sometimes taken the form of failing to disclose exculpatory

[88] Mahoney, *supra* note 86.
[89] 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 [Nosek et al, Utopia].
[90] Roger Giner-Sorolla, "Science or Art? How Aesthetic Standards Grease the Way Through the Publication Bottleneck but Undermine Science" (2012) 7:6 Perspectives on Psychological Science 562 [Giner-Sorolla].
[91] Makel et al, *supra* note 31.
[92] Marjan Bakker, Annette van Dijk, & Jelte M Wicherts, "The Rules of the Game Called Psychological Science" (2012) 7:6 Perspectives on Psychological Science 543-554.
[93] Kay Dickersin, "The Existence of Publication Bias and Risk Factors for Its Occurrence" (1990) 263:10 JAMA 1385-1389 [Dickersin].
[94] See Saks & Spellman, *supra* note 87 at 208-209; David Paciocco would describe the misaligned incentive problem as an actual or perceived lack of independence between the expert and trial process: see David M Paciocco, "Unplugging Jukebox Testimony in an Adversarial System: Strategies for Changing the Tune on Partial Experts" (2009) 34 Queen's LJ 565 [Paciocco, Jukebox] at 573-574.
[95] Adam Benforado, Unfair: *The New Science of Criminal Injustice* (New York: Broadway Books, 2015) at 26-40.

evidence, as in the miscarriages of justice in the convictions of Roméo Phillion[96] and Donald Marshall Jr.[97]

Expert witnesses are not immune from biases flowing from the adversarial system.[98] Courts have frequently worried that expert testimony may be tinctured by the employment relationship with the party tendering them.[99] This "association bias" is heightened by "selection bias", the fact that experts are often chosen because they have a view favourable to that of the proffering party (David Paciocco refers to these two biases together as "adversarial bias").[100] The NAS Report, for instance, chronicled "significant concerns" with the independence of forensic scientists because they are often employed by the police.[101]

IV(b). Excessive flexibility

Misaligned incentives become a problem when there is flexibility to act on them. Open science researchers have documented "undisclosed flexibility" in the research process (see Part II) that gave way to the QRPs that many scientists used to artificially inflate the publishability of their findings.[102]

[96] *R v Phillion*, 2009 ONCA, 65 CR (6th) 255 [*Phillion*]. See Innocence Canada, online: <https://www.aidwyc.org/cases/historical/romeo-phillion/>.
[97] Nova Scotia, Royal Commission on the Donald Marshall, Jr., Prosecution, *Digest of Findings and Recommendations* (Halifax: Queen's Printer, 1989) [Donald Marshall Inquiry]; See also *R v Taillefer*, 2003 SCC 70, [2003] 3 SCR 307 [*Taillefer*]. Garrett and Neufeld, in their empirical study of wrongful convictions in the U.S., found many cases in which the prosecution failed to disclose exculpatory forensics: Garrett & Neufeld, *supra* note 4.
[98] Daniel C Murrie et al, "Are Forensic Experts Biased by the Side That Retained Them?" (2013) 24:10 Psychological Science 1889-1897.
[99] *Abinger v Ashton*, 17 LR Eq 358 at 374 (Ch 1873), quoted in *White Burgess*, *supra* note 71 at para 11.
[100] Paciocco, Jukebox, *supra* note 94 at 575-581.
[101] NAS Report, *supra* note 3 at 184; In the U.S., Garrett and Neufeld performed an empirical examination of wrongful convictions that included invalid forensic scientific evidence. The majority of experts in that study were employed by police crime labs: see Garrett & Neufeld, *supra* note 4 at 13.
[102] Simmons et al, *supra* note 32.

The same flexibility has contributed to issues with expert evidence. For example, the PCAST Report found great flexibility in how forensic examiners report their findings.[103] It also found a general failure to establish or hold to guidelines requiring that examiners be blind to potentially biasing information, such as the nature of the crime and the identity of the suspect.[104] In Canada, Justice Goudge singled out excessive flexibility as a contributor to the miscarriages of justice he reviewed: "Our systematic review of autopsy practices in Dr. Smith's years revealed the absence of any articulated principles…on which a set of best practices could be built."[105] Even validation studies (that scaffolds the work of examiners) in forensic science have drawn criticism along similar lines to that in mainstream science. For instance, the PCAST Report noted that unclear exclusion rules in such studies (e.g., should an examiner be excluded for unusually low performance, or could that be chalked up to a clerical error?) could bias their results.[106]

IV(c). Motivated reasoning

> …a major challenge for scientists is to be open to new and important insights while simultaneously avoiding being misled by our tendency to see structure in randomness. The combination of apophenia (the tendency to see patterns in random data), confirmation bias (the tendency to focus on evidence that is in line with our expectations or favoured explanation) and hindsight bias (the tendency to see an event as having been predictable only after it has occurred) can easily lead us to false conclusions.[107]

Flawed incentives and excessive flexibility provide fertile ground for motivated reasoning.[108] This term reflects the fact that our thought processes are not always (or often)

[103] PCAST Report, *supra* note 15 at 83. See also NAS Report, *supra* note 15 at 174.
[104] PCAST Report, *ibid* at 33.
[105] Goudge Report, *supra* note 4 at 44.
[106] PCAST Report, supra note 15 at 95.
[107] Munafò, *supra* note 1 at 1.
[108] Ziva Kunda, "The Case for Motivated Reasoning" (1990) 108:3 Psychological Bulletin 480-498.

rational, calculated and transparent.[109] Rather, cognition is inherently and implicitly motivated; our conclusions are influenced by processes like contextual bias (i.e., the effect of cues in the environment on our reasoning)[110] and confirmation bias (the effect of a pre-existing conclusion on our reasoning).[111] As these processes occur unconsciously, their operation is unimpeachable through cross- and self-examination.[112]

As reflected in the quote that began this subsection, scientists long believed that the strictures of science protected them against these biasing forces.[113] They were deceiving themselves. The incentive to publish is strong and most editors did not always hold authors to very rigorous and transparent standards.[114] This gratuitous flexibility in the scientific method contributed to an often-irreproducible body of research.[115]

In law, motivated reasoning is widely considered to have contributed to numerous wrongful convictions and accusations.[116] Subsequent research systematically exposing forensic professionals to biasing information finds that their decisions are indeed influenced by these often-irrelevant details.[117] In response, leading bodies like the NAS and PCAST have encouraged forensic scientists to adopt blinding procedures.[118]

---

[109] *Ibid*.
[110] Edmond et al, *supra* note 14; PCAST Report, *supra* note 15 at 31-32. Michael Risinger refers to these as "observer effects": see Risinger et al, *supra* note 14.
[111] Goudge Report, *supra* note 3 at 424-425; Raymond S Nickerson, "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises" (1998) 2:2 Rev Gen Psychol 175.
[112] Kathleen A Kennedy & Emily Pronin, "Bias Perception and the Spiral of Conflict" in Jon Hanson and John Jost, eds, *Ideology, Psychology, and Law* (Oxford: Oxford University Press, 2012).
[113] Munafò, *supra* note 1 at 1-2.
[114] Giner-Sorolla, *supra* note 90.
[115] Simmons et al, *supra* note 32.
[116] See PCAST Report, supra note 15 at 28; Goudge Report, *supra* note 4 at 153-156.
[117] Dror, Context *supra* note 14. PCAST Report, *ibid* at 31.
[118] PCAST Report, *ibid* at 96; NAS Report, *supra* note 15 at 188.

IV(d). Epistemological immodesty

The ubiquity of motivated reasoning should foster what some researchers have termed epistemological humility[119] and modesty.[120] In other words, scientists and expert witnesses should be careful to not overstate their conclusions and to couch them in the appropriate levels of uncertainty. Unfortunately, both scientists and expert witnesses have not always upheld this ideal.

Within mainstream sciences, researchers long employed QRPs but refused to acknowledge the uncertainty they produced.[121] Somewhat ironically, this was even found in the work of a psychologist who won a Nobel Prize for studying motivated reasoning. Another scientist re-analyzed[122] the studies in a chapter of Daniel Kahneman's book, *Thinking, Fast and Slow*.[123] He found that the reproducibility of many of the studies Kahneman relied on was limited because they used small sample sizes. Acknowledging his previous immodesty, Kahneman replied: "I placed too much faith in underpowered studies…there is a special irony in my mistake because the first paper that Amos Tversky and I published was about the belief in the 'law of small numbers,' which allows researchers to trust the results of underpowered studies with

[119] NAS Report, *supra* note 3 at 106, 142; Jennifer L Mnookin, "The validity of latent fingerprint identification: Confessions of a finger printing moderate" (2008) 7:2 Law, Probability and Risk 127.
[120] Model Forensic Science, *supra* note 18 at 497-499.
[121] See Nelson et al, *supra* note 2.
[122] Ulrich Shimmack, Mortiz Heene, & Kamini Kesavan, "Reconstruction of a Train Wreck: How Priming Research Went off the Rails" (2 February 2017), *Replicability Index* (blog), online: <https://replicationindex.wordpress.com/2017/02/02/reconstruction-of-a-train-wreck-how-priming-research-went-of-the-rails/>.
[123] Daniel Kahneman, *Thinking Fast and Slow* (London: MacMillan, 2001).

unreasonably small samples."[124] Kahneman's self-scrutiny is laudable, but has typically been uncommon.

Similarly, in expert evidence, academics have frequently accused expert witnesses of immodesty. For instance, forensic bitemark[125] and fingerprint analysts[126] regularly stated – *in court* – that they could match a found pattern to the accused to the exclusion of all the world. There is no scientific or even logical basis for this claim.[127] These are unsupportable and decidedly immodest claims. Justice Goudge recognized similar brazenness in Charles Smith's testimony. He testified in a "dogmatic"[128] manner and regularly strayed outside the bounds of his expertise.[129]

IV(e). A preoccupation with eminence

> The drive for eminence is inherently at odds with scientific values, and insufficient attention to this problem is partly responsible for the recent crisis of confidence in psychology and other sciences. Humans will always care about eminence. Scientific institutions and gatekeepers should be a bulwark against the corrupting influence of the drive for eminence.[130]

In the above quote, Simine Vazire implicates eminence in irreproducibility. This is because eminence bears a loose relationship with the actual truth value of scientific findings.

---

[124] "'I placed too much faith in underpowered studies:' Nobel Prize winner admits mistakes", *Retraction Watch* (blog) (20 February 2017), online: < http://retractionwatch.com/2017/02/20/placed-much-faith-underpowered-studies-nobel-prize-winner-admits-mistakes/>.
[125] Michael J Saks et al, "Forensic bitemark identification: weak foundations, exaggerated claims" (2016) 3:3 Journal of Law and the Biosciences 538 at 558-561.
[126] Mnookin, *supra* note 59 at 1225-1227.
[127] As Jennifer Mnookin has noted, this also implies both that the source pattern is unique and that the analysts never make errors. Mnookin, *ibid*.
[128] Goudge Report, *supra* note 4 at 16.
[129] *Ibid* at 14.
[130] Vazire, Eminence *supra* note 17; Merton, *supra* note 17 at 270.

But, it is human nature to overweight prestige.[131] Scientific safeguards have not always effectively controlled for eminence. For example, a 2017 study of research in computer science found that whether or not peer review was blind affected publication decisions (i.e., reviewers base their decisions, in part, on the work's author).[132]

Expert evidence shares this struggle with eminence. Before experts begin to provide their testimony to the factfinder, they are often led by the tendering lawyer through a lengthy review of their CV: impressive-sounding academic credentials, publications, journal editorships, and so on. This occurs unchallenged despite years of judicial admonishments against the danger that a lay jury will be unduly swayed by experts with impressive credentials.[133] Ironically, the factfinder is often not provided the most relevant *indicia* of "eminence" – the results of proficiency tests (i.e., is the expert actually good at the task at hand?).[134]

IV(f). Intentional and negligent misbehaviour

To this point, we have primarily described the phenomenon of experts and scientists operating with too much flexibility and succumbing to motivated reasoning. But these are not the only types of errors out there. Rather, some are intentional.[135] Transparency and openness reforms should be designed to identify and deter not just exploitation of researcher degrees of freedom, but intentional acts as well.

[131] Jon D Hanson & David G Yosifon, "The Situational Character: A Critical Realist Perspective on the Human Animal" (2004) 93:1 Georgetown LJ 1-179 at 6-13.
[132] Andrew Tompkins, Min Zhang & William D Heavlin, "Reviewer bias in single- versus double-blind peer review" (2017) 114:48 PNAS 12708.
[133] See: *Mohan*, *supra* note 17 at para 23.
[134] Brandon L Garrett & Gregory Mitchell, "The Proficiency of Experts" (forthcoming) 166 University of Pennsylvania Law Review [Garrett & Mitchell]; PCAST Report, *supra* note 15 at 57-59; Mnookin, *supra* note 59 at 1224-1225, 1235-1236, 1268-1275.
[135] Saks & Faigman, *supra* note 3 at 893.

Estimating the prevalence is of fraudulent research practices is difficult, but many researchers have attempted it.[136] Daniele Fanelli performed a meta-analysis (i.e., empirical review) of several such studies and found that 2% of scientists admitted to having "fabricated, falsified or modified" their data.[137] We are unaware of similar surveys of expert witnesses and it is precarious to infer intentionality simply because an expert gave scientifically invalid testimony.[138] Still, cases of intentional falsification have occurred.[139]

Part V. Open science lessons for legal actors

The primary takeaway from the above comparison is that, in the case of both mainstream science and expert evidence, individuals – influenced by misaligned incentives – took advantage of unreported flexibilities in their disciplines. Existing safeguards in both fields were poorly equipped from preventing misleading and invalid results from reaching orthodoxy. This similarity suggests that the problems found in expert evidence cannot simply be remedied by adherence to mainstream scientific methods and norms. Rather, legal reform should be attuned to the main response of open science reformists – transparency.[140] In science, new initiatives focus on transparently reporting flexibilities in the research process such that peer-reviewers and the public can rationally evaluate the science. As it happens, rational evaluation of evidence is also a

[136] Daniele Fanelli, "How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data" (2009) 4:5 PLoS ONE e5738.
[137] *Ibid* at 8.
[138] Garrett & Neufeld, *supra* note 4 at 76: "Even with the benefit of bench notes or laboratory reports, one may not be able to ascertain whether experts falsified or concealed test results."
[139] See Paciocco, Jukebox, *supra* note 94 at 582-584.
[140] This view also coincides with the focus on transparency of scientific processes proposed by Edward K Cheng & G Alexander Nunn, "Beyond the Witness: Bringing A Process Perspective to Modern Evidence Law" (forthcoming) Texas LR.

foundational principle of expert evidence law.[141] We will now suggest how open science may be embraced by legal actors: the experts, the advocates, and the gatekeepers.

V(a). Adopting open scientific practices: the role of expert witnesses

> Where courts do not regulate the content of expert testimony, and defendants typically do not have experts with which to effectively counter State-proffered forensic testimony in criminal trials, the scientific standards within the forensic sciences are the most important source for regulating the content of forensic science testimony.[142]

The experts themselves are best placed to ensure their evidence is susceptible to rational evaluation. Our advice to them is straightforward: adopt open scientific practices (e.g., open data, preregistration, thorough reporting of methodology, disclosure of flexibility in analysis). To make these recommendations more concrete, we will discuss them in the context of two recent controversial criminal law proceedings, one featuring expert social scientific evidence (*R v Abbey*) and one featuring expert forensic evidence (*R v Bornyk*).

In *Bornyk*, the forensic examiner's opinion exemplified many of the common criticisms of forensic science: he overclaimed and did not provide important knowledge to the court about the controversies in his field. But, as we will see, the academic sociologist in *Abbey*, who followed mainstream norms for the most part, did not do much better. His methodology was laden with QRPs that were only revealed a decade after he first gave testimony. In both cases, open science may have assisted in more efficiently bringing these limitations to light.

In *Abbey*, the Crown proffered an expert sociologist named Mark Totten. He gave evidence indicating that the accused's teardrop tattoo meant he had killed a rival gang

[141] *Mohan*, *supra* note 17 at para 23.
[142] Garrett & Neufeld, *supra* note 4 at 34.

member.[143] This evidence was central to the Crown's theory that the accused shot the deceased Simeon Peter because he mistook him for a member of a rival gang who had robbed him earlier.[144] Totten's evidence was excluded at the first trial in 2007 and Abbey was acquitted.[145] In 2009, the Court of Appeal for Ontario held that Totten's evidence should have been admitted and ordered a retrial.[146] Abbey was convicted at that retrial.[147] In 2017, the Court of Appeal ordered a third trial after fresh evidence cast serious doubt on Totten's research program.[148]

The fresh evidence came from an unlikely source – the Crown's own cross-examination of Totten in a subsequent case.[149] That case was *R v Gager* and Totten was proffered by the defence to opine that it was unlikely that the two co-accused were gang members.[150] The Crown in *Gager* found several methodological issues that had not been explored in any of the preceding *Abbey* trials but that were deeply relevant to its fair adjudication.[151] *Gager* ultimately provided Abbey with the fodder to bring an appeal based on fresh evidence in 2017.[152] In that decision, Justice Laskin (with Justices Doherty and Roberts concurring) described several weaknesses in Totten's evidence:[153]

[143] *R v Abbey* [2007], 73 WCB (2d) 411 at para 21, 2007 CarswellOnt 376 (Ont SC) [*Abbey* ONSC 2007]. A concise summary of the evolution of Totten's evidence can be found in *R v Abbey* 2017 ONCA 640 at para 41, 2017 CarswellOnt 12134 [*Abbey* ONCA 2017]. For a more thorough accounting of *Abbey*, see Chin, *Abbey* Road, *supra* note 22.
[144] *Abbey* ONSC 2007, *ibid* at paras 1-2; *R v Abbey*, 2009 ONCA 624 at paras 1-2, 10-11, 97 OR (3d) 330 [*Abbey* ONCA 2009].
[145] See *Abbey* ONCA 2017, *supra* note 143 at paras 13-15.
[146] *Abbey* ONCA 2009, *supra* note 144 at paras 25-149; see *Abbey* ONCA 2017, *ibid* at paras 13-41 for a summary of the proceedings.
[147] *Abbey* ONCA 2017, *ibid* at paras 16-36.
[148] *Ibid*.
[149] *Ibid* at para 5.
[150] 2012 ONSC 1472 at para 30-85, 100 WCB (2d) 285 [*Gager*].
[151] *Abbey* ONCA 2017, *supra* note 143 at para 38. See *Gager*, *ibid* at paras 29-96.
[152] *Abbey* ONCA 2017, *ibid* at para 45.
[153] See generally *ibid* at paras 69-106.

- Totten had double-counted some of his interviews across published studies, and thus he had interviewed fewer individuals than he had originally suggested.[154]
- He used a shifting definition of "gang member" to make it seem as if he had interviewed more such individuals than he had.[155]
- Some of the studies he relied on did not record whether the individual had been convicted of a homicide, and interview protocols did not include questions about tattoos, raising questions about the validity of conclusions based on those features.[156]
- None of the above was verifiable because Totten had destroyed his data.[157]

The Court of Appeal admitted the fresh evidence, allowed the appeal, and ordered a third trial.[158] Justice Laskin held that the fresh evidence casting doubt on Totten's opinion was cogent enough that it would have warranted Totten's exclusion if available at trial.[159] More specifically, the evidence undermined the opinion's reliability and thus its probative value.[160] In May of 2018, Abbey pled guilty to manslaughter and was sentenced to time served, plus one day.[161]

---

[154] *Ibid* at paras 101-104.
[155] *Ibid* at paras 72-82.
[156] *Ibid* at paras 66, 86, 92, 98-99. It may be that these questions were asked or it may be that Totten was, after the fact, remembering them in the way most beneficial to his client's case, see Nosek et al, Utopia, *supra* note 89 at 617: "Instead, 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. The reader may, as we do, recall personal examples of such motivated decisions—they are entirely ordinary products of human cognition."
[157] *Abbey* ONCA 2017, *supra* note 143 at para 95. This occurred in the *Gager* trial, but the Court of Appeal held that it was relevant to Abbey's fresh evidence application, see *ibid* at para 96. Note also that Totten's methodology was also unclear in that he claimed to have attended or conducted all interviews. There was not enough time for this to occur. *Ibid* at paras 105-106.
[158] *Ibid* at para 155.
[159] *Ibid* at para 109.
[160] *Ibid* at paras 117-125.
[161] Betsy Powell, "Accused in teardrop-tattoo case pleads guilty to manslaughter, released after almost 11 years" *The Star* (28 May 2018) online: <https://www.thestar.com/news/gta/2018/05/28/accused-in-teardrop-tattoo-case-pleads-guilty-to-manslaughter-released-after-almost-11-years.html>.

The experience with Mark Totten's evidence in *Abbey* and *Gager* demonstrates how easy it is to mislead the factfinder when using pre-open scientific (yet, in some cases, generally accepted) practices.[162] First, open science would have made it easier for the defence to identify the weaknesses in Totten's analysis. And second, it may have encouraged him to conduct his research more transparently to begin with. Indeed, as three researchers on the vanguard of open science have noted, the mere thought of showing how the sausage is made can inspire vigilance: "Public data posting not only allows others to verify the accuracy of the analyses, but also incentivizes authors to more carefully avoid errors."[163] Adherence to open scientific methods may have also made Totten's evidence more credible, providing valid explanations for what the Court interpreted as self-serving changes to his methodology.

Open data and preregistration of methodology may have made it easier to spot differences between Totten's original research and what he presented in court. For instance, it would have made it more apparent that Totten was double-counting participants (i.e., treating the same "gang member" as two data points). Specifically, if each study was recorded on the OSF with participants receiving anonymous identifiers, it would be easier to see the overlap in those identifiers. Note that with sensitive research like Totten's, investigators will have to carefully navigate ethical and confidentiality concerns when considering the openness of their data. For instance, simply using anonymous identifiers may not fully protect those in vulnerable populations. But investigators may consider sharing narratives stripped of identifying information or sharing some data to third-party protected repositories.[164] Indeed, the NASEM

[162] One might also argue that *Abbey* demonstrates the problems that result from conflating exploratory fieldwork with confirmatory (and inculpatory) evidence. At the very least, there should have been some mechanism in place to explain this difference to the factfinder.
[163] Nelson et al, *supra* note 1 at 525.
[164] See David Mellor, "Approved Protected Access Repositories" online: <https://osf.io/tvyxz/wiki/8.%20Approved%20Protected%20Access%20Repositories/>.

recently stated that confidentiality concerns will be one of open science's most significant hurdles – but work is already underway to address them.[165] Importantly, these questions should be asked and investigators should not simply assume that any level of open data is untenable.

Similarly, pre-registration of Totten's definition of gang-member (or a list of possible definitions) would have helped. Recall that Totten drew rebuke for changing his definition of gang-member over time to suit his conclusions. The trial judge in *Gager* also found that an additional category of gang-affiliation proposed by Totten in that case suggested bias.[166] On for the defence, Totten had suggested some individuals (implying this applied to the accused) were not gang-members, but "long term friends" of gang-members.[167] The trial judge was not impressed with this seemingly *ad hoc* formulation: "The reason Dr. Totten's postulation of the 'long-term friend' suggests bias is that the witness indicated in his cross-examination that he began contemplating the 'long term-friend' category of relationship at the beginning of his doctoral research in the 1990s, but had never committed this idea to paper prior to preparing his report in this case."[168] Preregistering his conceptualization (prior to initial data collection) of what it means to be in a gang (or a long-term friend) would have clarified Totten's evidence a great deal and could have bolstered its credibility.[169]

Finally, the defence and jury would want to know what was inside Totten's file-drawer. As we discussed above, the sciences have long suffered from a bias in publication known as the "file drawer effect".[170] This term describes the fact that interesting findings are published, while

[165] NASEM Open Science Report, *supra* note 1 at 41-44.
[166] *Gager*, *supra* note 150 at para 74.
[167] *Ibid*.
[168] *Ibid*.
[169] And preregistration can also assist when the data has already been collected, see Nosek et al, Revolution, *supra note* 44 at 4.
[170] Dickersin, *supra* note 93.

others languish in the researcher's archives (now likely a hard drive). Without open scientific reforms, there is simply no way to know how many interviews Totten conducted that did not make their way into his expert report. And accordingly, there is no way to know if some of those interviews were with individuals who got a tattoo for a reason other than killing a rival gang-member.

In parallel with *Abbey*, expert forensic evidence in the recent *Bornyk* proceedings in British Columbia revealed similar concerns in an area far afield from sociology.[171] The expert in *Bornyk*, a fingerprint examiner, identified the accused as the source of a print found at the crime scene.[172] In doing so, he cast fingerprint analysis as a rigorous and objective procedure that admitted of no error: "There's no errors allowed in fingerprint identification. That continues today. There is no errors permitted in fingerprint identification."[173] He did not mention the NAS Report, then four years old, nor the many studies finding that well-trained fingerprint examiners are susceptible to motivated reasoning.[174] He also failed to disclose weaknesses in his own comparison.[175]

Like the Crown's surprising cross-examination of Totten in *Gager*, the *Bornyk* trial also deviated from the usual path. The trial judge apprised himself of several leading reports about fingerprint identification, including the NAS Report.[176] He subsequently concluded that a "number of troubling aspects" arose from the expert's initial testimony.[177] These included the expert failing to disclose the subjectivity of fingerprint analysis, the role of unconscious bias,

[171] *R v Bornyk*, 2013 BCSC 1927, 7 CR (7th) 211 [*Bornyk* BCSC 2013]. See 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 Law Review 383-413 at 387-393 for a full review of the case [Edmond, Exogenous].
[172] *Bornyk* BCSC 2013, *ibid* at para 18.
[173] *Ibid* at para 23 [sic].
[174] *Ibid* at paras 17-31.
[175] *Ibid* at paras 55-58; Edmond Exogenous, *supra* note 171 at 393-397.
[176] *Bornyk* BCSC 2013, *ibid* at paras 32-33.
[177] *Ibid* at para 39.

"unexplained discrepancies" between the latent (i.e., found) print and Bornyk's print, and discrepancies between prints taken from Bornyk at various times.[178] The trial judge concluded that he was not convinced beyond a reasonable doubt and acquitted Bornyk. The Court of Appeal then reversed that judgment because the trial judge should not have taken judicial notice of the authoritative fingerprint reports.[179] Bornyk was convicted at the retrial.[180]

Abbey and *Bornyk*, cases drawing expertise from seemingly very different fields, both demonstrate the importance of openness through "before and after" comparisons. In *Abbey*'s "before", traditional (closed) scientific practices portrayed Totten's research as highly probative of the meaning of teardrop tattoos. Under the surface, however, lurked uncertainties that the factfinder could not know. But once Totten's practices were exposed (the "after"), it became clear he had taken advantage of the undisclosed flexibility to frame his data in misleading way. This resulted in the outright exclusion of his evidence and a third trial.

Bornyk is quite similar. There, the before constitutes the evidence prior to the trial judge becoming aware of the authoritative reviews of fingerprint identification. At this point, the evidence seemed extremely convincing: the analysis was objective, examiners were error-free, and motivated reasoning did not happen. The after, as described above, was quite different and seemed to substantially decrease the weight the trial judge accorded the evidence (although the second trial judge, also aware of these limitations, found it convincing enough to convict).[181]

It is important to note that academics have often urged the forensic sciences to behave more like the other sciences. For instance, a group of scientists, academics and forensic

[178] See *ibid* at paras 39-58.
[179] *R v Bornyk*, 2015 BCCA 28, 320 CCC (3d) 393.
[180] *R v Bornyk*, 2017 BCSC 849, 139 WCB (2d) 384.
[181] *Ibid*.

practitioners provided the following prescription: "The simplest advice we can offer to forensic practitioners is to use mainstream scientific methods and norms."[182] But here we saw a forensic scientist comporting himself in a very similar manner to Totten, and to many other mainstream scientists. We think it is time to amend the conventional wisdom: expert witnesses should aim not just to comport themselves as scientists, but rather *the most open, transparent, and methodologically rigorous scientists*.

V(b). Open science and the duty to fairly present the case: The role of the prosecution

The responsibility for ensuring that the expert's evidence is presented fairly does not end with the expert. This is especially true in criminal matters in which the party tendering the expert is the Crown.[183] In this subsection, we will suggest that the Crown's disclosure obligation should be construed very broadly and that it should take special care when tendering scientific evidence. This is because in science, it has proven much easier than most thought to present strong evidence of weak or non-existent effects by failing to disclose important limitations of the research. Similarly, in law, many wrongful convictions are attributable to the prosecution's failure to disclose exculpatory evidence that weakens its case.[184] In other words, these were failures of transparency, and thus open science reforms aimed at improving transparency may be helpful. In this vein, we will conclude this subsection with concrete reforms modeled off those that scientific journals are using to nudge researchers towards more full disclosure.

---

[182] Model Forensic Science, *supra* note 16 at 497. This entreaty was also implicit in Saks and Faigman's label "the non-science forensic sciences". See Saks & Faigman, *supra* note 3 at 150.
[183] *R v Boucher* [1955] SCR 16 at para 26, 110 CCC 263 [*Boucher*]; *R v Stinchcombe*, [1991] 3 SCR 326 at para 11, 83 Alta LR (2d) 193 [*Stinchcombe*]; Keith A Findley, "Innocents at Risk: Adversary Imbalance, Forensic Science, and the Search for Truth" (2008) 38 Seton Hall LR 893.
[184] In Canada, see the wrongful convictions of: Donald Marshall Jr., Donald Marshall Inquiry, *supra* note 134; Justice Edward MacCallum, *Report of the Commission of Inquiry into the Wrongful Conviction of David Milgaard*, online: <http://www.publications.gov.sk.ca/details.cfm?p=26267>; Thomas Sophonow, The Inquiry Regarding Thomas Sophonow: *The Investigation, Prosecution and Consideration of Entitlement to Compensation* (Winnipeg: Manitoba Justice, 2001); Roméo Phillion, *Phillion*, *supra* note 96.

Most fundamentally, lessons from the open science movement indicate that the scope of the Crown's disclosure obligation should be wider than it has been construed in recent jurisprudence. The duty to disclose all relevant evidence was recognized, in broad language, in the Supreme Court's landmark decision in *R v Stinchcombe*.[185] That duty, however, has been unevenly applied to the Crown's scientific evidence (especially when it is held by the police). For example, in *R v Taillefer*, the Supreme Court corrected a post-*Stinchcombe* appellate decision that had found that the Crown's failure to disclose a forensic dentist's earlier opinion linking the crime to an individual that was not the accused had not impacted trial fairness.[186] The Supreme Court quashed the conviction and the accused eventually recovered in a civil action against the attorney general.[187] Recently, however, the Supreme Court held in *R v Gubbins* that the maintenance records of breathalyzers do not fall within the Crown's disclosure obligation despite some evidence that they are relevant to their current operation.[188]

Gubbins is problematic because it runs counter to the meta-scientific discoveries we reviewed above. The sought-records bear on the reliability of breathalyzer evidence.[189] And motivated reasoning makes it easy for examiners to dismiss previous errors and mistakes as happenstance and thus fail to proactively disclose them. In other words, courts are providing too much discretion to the experts. In such cases, open science would limit that discretion and prescribe transparency: broad disclosure of the entire foundation (or lack thereof) of the scientific case against the accused. There is some precedent for this in the case law. One trial court noted: "The expert's report and any materials which contributed to the foundation of the

[185] *Stinchcombe*, *supra* note 183 at para 17.
[186] *Supra* note 97.
[187] *Duguay v Québec (Procurer Général)*, 2013 QCCS 4120, 246 ACWS (3d) 342.
[188] 2018 SCC 44, 76 Alta LR (6th) 213 [*Gubbins*].
[189] *Ibid* at para 79. Similarly, early challenges to breathalysers were thwarted by trade secret privilege. Courts that denied the privilege ultimately found that the tests were bugged: Rebecca Wexler, "Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System" (2018, forthcoming) Stanford Law Review at 49-50 [Wexler]

report or which are clearly relevant to the witness's credibility must be disclosed."[190] Furthermore, while the *Gubbins* majority worried about the inefficiency of ordering disclosure in all cases, at least one crime laboratory in the U.S. has found that "radical transparency" actually improved its efficiency.[191] That lab implemented an automatic system for making lab results available to public defenders.[192] *Gubbins* removes the motivation for the police to make similar advances that may have long term benefits.

The open science movement also suggests narrowing any litigation privilege[193] that may exist between the Crown and expert as to the foundations of the scientific opinion. No privilege is absolute and litigation privilege may be overridden by competing interests.[194] The interest of the accused in making full answer and defence in light of known limitations in scientific reporting may qualify as such a competing interest. Somewhat analogously, Jacob Sherkow, in the U.S. patent context, has cogently argued that the lessons from the open science movement suggest that courts should broaden their view of what is admissible evidence in patent disputes.[195] In short, he argues that evidence accumulated after the patent was awarded tending to show the initial findings were not reproducible is relevant and should be admitted. Under the current rules, such evidence is typically not admitted.

---

[190] *R v Friskie* [2001], 205 Sask R 208, 49 WCB (2d) 375. And, on the issue of privilege, Justice Binnie (dissenting, but agreeing with the majority on this point) said "once a witness takes the stand, he/she can no longer be characterized as offering private advice to a party. They are offering an opinion for the assistance of the court. As such, the opposing party must be given access to the foundation of such opinions to test them adequately." *R v Stone* (1999), 134 CCC (3d) at paras 99, 353 (SCC).
[191] *Gubbins*, *supra* note 188 at para 53; Nicole B Cásarez & Sandra G Thompson, "Three Transformative Ideals to Build a Better Crime Lab" (2018) 34:1 Georgia State LR 1007 at 1046 [Cásarez & Thompson].
[192] *Ibid* at 1045-1046. The ultimate goal of this program is a password-protected online portal for public defenders.
[193] Litigation privilege is the immunity from disclosing documents and communications when their dominant purpose is the litigation at hand, see *Lizotte v Aviva Cie d'assurance du Canada*, 2016 SCC 52, 404 DLR (4th) 389.
[194] See *General Accident Assurance Co. v Chrusz* [1999], 5 OR (3d) 321 at paras 145-146, 180 DLR (4th) 241 (Ont CA) and the authorities relied on therein.
[195] Jacob S Sherkow, "Patent Law's Reproducibility Paradox" (2017) 66 Duke Law Journal 845.

A goal of any reform should be to remove as much subjective judgment from the disclosure decision as possible. For example, Justice Goudge, observing many cases in which the Crown counsel was silent despite several warning signs in the conduct of Charles Smith, highlighted the need for "detailed guidelines and protocols" to help determine when "issues with expert witnesses" should be disclosed.[196]

Some reforms being implemented in science provide ideas for the content of the guidelines Justice Goudge proposed. These scientific developments require or incentivize scientists to positively confirm that they have disclosed their data and not changed their procedure in any way.[197] The reason such acknowledgments work is because it is psychologically more difficult to actively mislead than it is to simply remain quiet about the weaknesses of one's research.[198] The Crown may wish to use similar questions with its experts to ensure that it is receiving an accurate representation of the evidence. Or, in the spirit of preregistration, the Crown may wish to require that the expert produce a pre-determined analysis plan and record any deviations from that plan. Deviations from that pre-specified criteria would be disclosed as a matter of course, and the implications of that change could be openly evaluated in light of possible bias. Similarly, any decision to exclude observations or other data should be made explicit in the expert report.

These enhanced protocols should also be applied to the experts themselves. In this regard, the rules of court across Canada diverge significantly. For example, the *Federal Courts Rules* demand that experts, in their reports, provide "any caveats or qualifications necessary to render

[196] Goudge Report, *supra* note 3 at 454.
[197] Mallory C Kidwell et al, "Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency" (2016) 14:5 PLoS Biol e1002456 [Kidwell].
[198] See Nina Mazar & Dan Ariely, "Dishonesty in Everyday Life and Its Policy Implications" (2006) 25:1 Journal of Public Policy & Marketing 117.

the report complete and accurate, including those relating to any insufficiency of data or research and an indication of any matters that fall outside the expert's field of expertise".[199] The rules in Ontario, however, are less insistent and specific.[200] In the Motherisk Report, Justice Beaman noted the difference between the more searching federal rules and laxer *Family Law Rules*[201] and suggested the divergence contributed to the miscarriages of justice she studied:[202]

> The Rules do not require experts to include any information about the scientific limits of the method they are using, the possibility of contamination, or other issues that could affect the reliability of the opinions or test results. Had these requirements been in place, lawyers and judges may have been alerted to the need to probe the reliability of the Motherisk testing.

Like Justice Beaman, we suggest that rules modeled after the *Federal Court Rules* should be adopted across Canada. In the meantime, similar disclosure and acknowledgment that no conflicting results exist could also be turned into a carrot by awarding expert reports "badges". In science, some journal editors use badges to encourage authors to engage in open scientific methods. For instance, authors may earn badges if they make their data or methods open, and if they preregister their research plan and expectations.[203] This badge method is finding

---

[199] (SOR/98-106) at s 52.2 Schedule [*Federal Court Rules*]. See similar rules in Victoria, Australia, Forensic Evidence Working Group, Practice Note: Expert Evidence in Criminal Trials (County Court of Victoria, updated June 24, 2014) online: <https://www.countycourt.vic.gov.au/sites/default/files/forms/Practice%20Note%20-%20Expert%20Evidence%20in%20Criminal%20Trials_FINAL%20%28June%202014%29_0.pdf>; *Uniform Civil Procedure Rules* (2005), NSW, Reg 418 at Schedule 7, Expert witness code of conduct. The New South Wales rules have their root in the principles from *Ikarian Reefer*, see notes 293-4 and the text accompanying them.

[200] *Rules of Civil Procedure*, RRP 1990, Reg 194, r 4.1.01(1): "It is the duty of every expert engaged by or on behalf of a party to provide evidence in relation to a proceeding under these rules, (a) to provide opinion evidence that is fair, objective and non-partisan; (b) to provide opinion evidence that is related only to matters that are within the expert's area of expertise; and (c) to provide such additional assistance as the court may reasonably require to determine a matter in issue."

[201] O Reg 114/99, under the *Courts of Justice Act*, RSO 1990, c C43.

[202] Beaman Report, *supra* note 57 at 109.

[203] Center for Open Science, "Badges to Acknowledge Open Practices" online: <https://osf.io/tvyxz/wiki/home/>.

considerable success within science, with one study showing a ten-fold increase in open data after a journal instituted an open data badge.[204]

So, who might badges work in law? Mimicking preregistration, a court might wish to give a badge to an expert report when the expert engages in a pretrial meeting in which the judge, expert, and parties agree on the expert's analysis plan. In the context of real estate valuation, such a plan might include stating the scope of comparison properties that will be used to establish value. This way, the expert could not change the scope when he or she realizes the comparison properties are higher or lower in value than expected (and thus, ostensibly, conflicting with the client's position). As in science, experts wishing to ensure that their reports be viewed as fully credible (in an adversarial system) would be highly motivated to earn these badges.

Finally, if experts do not disclose the data and resources that support their opinions, this may call into question the admissibility of their evidence. For instance, police officers who do not disclose the source material for their opinion because of confidential informant privilege or, more generally, because it is hearsay, may see the probative value of their opinion reduced to the degree that it is inadmissible.[205] We will now turn to this question of admissibility and how it may be informed by open science principles.

V(c). Open science and expert evidence law: the role of the gatekeeper

In the face of evidence that defies rational evaluation, the next line of defence is exclusion or limitation of that evidence. Here, we will focus on two open scientific lessons for

[204] Kidwell, *supra* note 197.
[205] *R v Giles*, 2016 BCSC 294 at paras 165-173, 130 WCB (2d) 614; *R v O(F)* 2016 ONSC 724 at paras 14, 17, 129 WCB (2d) 142. In the U.S., a court came to a similar decision when trade secret privilege prevented the disclosure of the algorithm behind a DNA test, See Wexler, *supra* note 189 at fn 281.

judicial gatekeepers faced with that decision. First, it is perilous to defer too heavily to practices that are generally accepted by the scientific community. Within science, these practices introduced substantial error into the literature that are only now being corrected – many scientific disciplines are still not open.[206] Rather, gatekeepers should look to the best practices (which often coincide with open practices). Second, courts should attempt to limit the biasing effect of eminence information and instead focus on proficiency when possible.

Despite Canadian courts never endorsing a general acceptance standard (i.e., presuming evidence is reliable if it is accepted in the expert community from which it came),[207] flexibility in the current rules allows considerable deference to generally accepted science. This deference has been expressed in two ways.

First, recall that under *White Burgess*, trial judges are directed to give additional reliability scrutiny to expert evidence when it is "based on novel or contested science or science used for a novel purpose".[208] Therefore, one method of avoiding this scrutiny (and applying a superficial sheen of reliability) is to portray the expertise as old hat.[209] However, as Peter Sankoff has noted, the novelty criterion itself is often unhelpful because new methods and findings (e.g., the open science movement) can cast doubt on established findings.[210] As a result, "contested" science seems a better trigger for scrutiny. Troublingly, Chief Justice McLachlin

[206] NASEM Open Science Report, *supra* note 1 at 2.
[207] Even pre-*Mohan* decisions refused to adopt the *Frye* test. See: *R v Singh*, 23 WCB (2d) 558 at para 31, 1993 CarswellBC 3097: "The American *Frye* test [citation omitted] for the admissibility of new scientific evidence, general acceptance in the scientific community, is now seldom used in the United States and it is not the law in Canada…" [*Singh*].
[208] *White Burgess*, *supra* note 71 at para 23.
[209] See Jason M Chin & Helena Likwornik, "*R v Bingley* and the Importance of Scientifically Guided Legal Analysis" (2017) 43:1 Queen's LJ 33 at 49-50 [Chin & Likwornik], arguing that there has never been a consistent definition of novel science and so it is easy to fit a wide spectrum of evidence into that definition.
[210] Sankoff & Mewett, *supra* note 19 at 16.5(c)(i).

omitted the contested language in her recent enunciation of the expert evidence rules in *R v Bingley*.[211]

Second, courts sometimes construe social scientific evidence and some forensic evidence as "specialized knowledge" and then generally defer to the expert.[212] This move tracks back to the first *Abbey* appeal, which was a remarkably influential decision. The Court found that Totten's method was not scientific, but rather "specialized knowledge gained through extensive research".[213] As a result, the *Daubert* test was inapplicable: "It was not scientific. It was not novel. And it was not a theory."[214] This method of sidestepping the expert evidence rules has been used to admit a variety of dubious evidence.[215] As we will discuss below, open research, whether or not the subject matter of that research is something courts will characterize as science, is often more credible than that which is conducted opaquely. And, any lack of credibility in that evidence will be easier to spot.

The 2009 *Abbey* decision did not actually hold that "specialized knowledge" should avoid all scrutiny. Rather, Justice Doherty provided nine questions relevant to the reliability of such evidence.[216] In light of the meta-scientific research girding the open science movement, some of these questions are more helpful than others. The questions ranged from field-related factors (e.g., whether it is an accepted field and whether the examiner, using accepted methodologies, is accepted in that field) to those focused on the transparency of the methods and data themselves (e.g., whether the data is available for scrutiny and whether the methods are susceptible to

[211] *R v Bingley*, 2017 SCC 12 at para 17, 135 WCB (2d) 356 [*Bingley*]. See Chin & Likwornik, *supra* note 209.
[212] See Chin, *Abbey* Road, *supra* note 22.
[213] *Abbey* ONCA 2009, *supra* note 144 at para 108.
[214] *Ibid* at para 116.
[215] For a review, see Chin, *Abbey* Road, *supra* note 22 at 441-448.
[216] *Abbey* ONCA 2009, *supra* note 144 at para 119.

critical evaluation). After listing these questions, Justice Doherty justified them on the basis of a quote from *Kumho*:[217]

> The objective of that requirement [the gatekeeper function] is to ensure the reliability and relevancy of expert testimony. **It is to make certain that an expert, whether basing testimony upon professional studies or personal experience, employs in the courtroom the same level of intellectual rigour that characterizes the practice of an expert in the relevant field.**

A focus on that quotation from *Kumho* is problematic. It is the part of the judgment that has drawn the most criticism. Influential academics have noted that American decisions are often unduly focused on this quote, characterizing such judgments as "a dangerous trend and one certainly not endorsed by the Kumho Tire Court".[218]

Indeed, a review of those nine questions finds hints of general acceptance in several of them. There are, for instance, references to accepted methodologies and honouring the boundaries of the discipline. The answers to such questions may be misleading. Recall, for instance, that the journal *Nature* expressly admitted that it had not historically published enough information for peer reviewers to do their jobs. Similarly, we have reviewed many flexibilities in the research process that allow researchers to frame their findings to seem more credible than they are. These were then, and in many cases still are, generally accepted practices.[219] Instead,

[217] *Ibid* at para 120 [the emphasis and British-English spelling of rigour was added by the *Abbey* Court]; *Kumho*, *supra* note 54 at 152.
[218] David L Faigman et al, Modern Scientific Evidence (Thomson Reuters, 2016-2017) at §1:28 [Modern Scientific Evidence].
[219] NASEM Open Science Report *supra* note 1 at 12: "Sharing data, code, and other research products is becoming more common, but is still not routinely done across all disciplines".

experts in court should be held to the best research practices in their field, which are often open scientific standards.[220]

Similarly, application of the question "[t]o what extent is the proffered opinion based on data and other information gathered independently of the specific case or, more broadly, the litigation process?"[221] should be informed by open science research. As we noted in Part I, courts and academics regularly draw this distinction between research that is conducted for the purposes of litigation and research that is not. They suggest that litigation-driven research is more likely to be biased.[222] In this vein, Justice Doherty was to some extent correct in implying that evidence collected in the course of a criminal investigation is often subject to cognitive biases.[223] The same, however, is true of data collected long before litigation is contemplated. For instance, meta-scientists have documented pervasive use of the researcher degrees of freedom on topics far afield from litigation.[224] This was demonstrated by Totten's research in *Abbey* itself – his data were collected long before the *Abbey* case.

Indeed, Susan Haack has suggested that the litigation-driven characterization, while potentially useful, is a messy one. In short, she argues that there is greater potential bias in

[220] Chin, Replicability Crisis, *supra* note 12 at 225. Note that not all of Totten's practices should be classified as generally accepted. The 2017 decision generally characterized him as dishonest, but expressly refrained from making any finding of research fraud. See *Abbey* ONCA 2017, *supra* note 143 at para 124.
[221] *Abbey* ONCA 2009, *supra* note 178 at para 144.
[222] See Haack, *supra* note 22; *Daubert v Merrell Dow Pharms., Inc.*, 43 F 3d 1311 at 1317 (9th Cir 1995).
[223] Gary Edmond & Mehera San Roque, "Quasi-justice: Ad hoc expertise and identification evidence" (2009) 33 Crim LJ 8 at 32-33. See, e.g., the miscarriage of justice in the case of Jeffrey Gilham in which the trial court admitted unrepresentative "experiments" performed by the Crown's expert to demonstrate the rate at which fire spreads: *Gillham v R*, [2012] NSWCCA 131 at paras 158-198, (2012) 224 A Crim R 2.
[224] John et al, *supra* note 36. Research reports can also "spin" results to fit with the desires of the author: Kellia Chiu, Quinn Grundy, & Lisa Bero, "'Spin' in published biomedical literature: A methodological systematic review" (2017) 15:9 PLoS Biol e2002173.

litigation-driven research, but that does not mean that the evidence is unhelpful or invalid in many cases.[225]

Open scientific research lends additional force to Haack's comment and suggests that transparency presents a way to distinguish the credible litigation-driven research from the more dubious sort. For instance, there is a concern that drug company-funded research examining the safety of a drug will present a misleading picture of the science because the company indirectly controls the data collection, analysis, and peer-review process.[226] Open scientific standards, while certainly not a panacea for this bias, would assist in assessing the credibility of research; findings and conditions that did not support the company's preferred outcome would be more difficult to supress.[227] Indeed, vocal critics of the pharmaceutical industry have noted that large analyses of the medical literature (like that performed by the Cochrane Collaboration) are difficult to perform when research is systematically supressed (i.e., preclinical research that is not preregistered).[228]

Consider another distinction made in the *Abbey* judgments, that between qualitative and quantitative research. For instance, in the 2009 appeal, Justice Doherty appeared to accept Totten's insistence that concepts from *Daubert* like error rates and testing were inapplicable to qualitative research methods like his interviews.[229] However, at the 2017 appeal, Justice Laskin revisited this distinction, noting that the qualitative and quantitative parts of Totten's research

[225] Haack, *supra* note 22 at 1077.
[226] *Ibid* at 1067-69.
[227] And pre-registration would help distinguish between pre- and postdiction. For instance, consider a study that found ta drug had no overall effect on birth defects. However, if the cohort was divided up based on when the mother took the drug, there was a finding that at certain times, the drug did seem to increase the chances of birth defects. If the researchers had preregistered the prediction that there were certain sensitive times in which the drug was dangerous, that would significantly increase the credibility of the claim that the drug was dangerous. Otherwise, critics could (persuasively) argue that there is always a way to carve up the data such that there is a period in which it is not safe to take the drug.
[228] See Ben Goldacre, *Bad Science* (London: Fourth Estate, 1999) at 158-159.
[229] See *Abbey* ONCA 2009, *supra* note 144 at paras 47, 107-112.

were "intertwined".[230] Totten's conclusions drew force from the number of interviews he performed.

Justice Laskin's insight deserves further discussion. Much of the research we have relied on in our review of the open science movement has been quantitative – the researcher degrees of freedom that cause a field's actual numerical error rate to diverge from its reported error rate. But the same logic holds for non-quantitative research because statistics often just a method of standardized inference.[231] Open and transparent research is both numerically and logically more credible and trustworthy. In other words – and like the above discussion about litigation versus non-litigation driven science – open scientific norms and practices can be applied to both quantitative and qualitative research.

With Totten's research, for example, imagine if he had pre-recorded his definition of "gang member", provided evidence that there were not additional interviews sitting in his file drawer (through preregistration), and pre-recorded his reasons for performing his interviews. To any observer (familiar with inferential statistics or not), his findings would be more credible and better lend themselves to rational evaluation.

Some of the questions provided in *Abbey* more directly address open scientific principles.[232] These are whether the expert's reasoning and methods were "clearly explained by the witness and susceptible to critical examination" and whether the data was "accurately recorded, stored and available."[233] Similarly, meta-scientific research suggests that expert conclusions are more reliable if the underlying data were not just available by contacting the

[230] *Abbey* ONCA 2017, *supra* note 143 at para 122.
[231] Robert P Abelson, *Statistics Principled Argument* (New Jersey: Lawrence Erlbaum, 1995).
[232] *Abbey* ONCA 2017, *supra* note 143 at paras 93, 116.
[233] *Ibid*.

author (research shows many of these e-mails go unanswered),[234] but were truly available for scrutiny by peers.[235] And methods are indeed better "susceptible to critical examination" when they are fixed before the data are collected (i.e., preregistered).

We would, however, rethink the "same intellectual rigour" touchstone and provide a new one to use as an interpretative aid for all nine questions. All expert evidence, whether it originate from a hard science, a social science, or a forensic science, *should be transparently produced and reported*. This touchstone is not a far leap from the two questions we have endorsed and it should be easy to read the remaining questions with the theme of openness (e.g., "accepted and open methodologies"). We do not think that this interpretation should be controversial: openness flows directly from established principles of expert evidence law[236] and the academic work in this field.[237]

If experts are more transparent and gatekeepers are more attuned to open science principles, it will also make the job of cross-examining experts easier. Although we focused in Part V(b) on the party proffering the expert, the cross-examiner also has an important role to play. Excellent work providing tangible recommendations for cross-examining potentially unreliable and biased expertise already exists.[238] Therefore, we simply add that cross-examiners should take guidance from the open science practices described above: Did they change their methodology after seeing partial results? Did the peer-reviewers (of the scientific foundations of

[234] Jelte M Wicherts, Marjan Bakker, & Dylan Mosenaar, "Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results" (2011) 6:11 PLoS ONE e26828.
[235] See *JLJ*, *supra* note 75 at para 57.
[236] *Mohan*, *supra* note 17 at para 23; *Béland*, *supra* note 17 at para 116.
[237] See Edmond, Rational Jury, *supra* note 19; Paciocco, *supra* note 51 at 156.
[238] Gary Edmond et al, "How to cross-examine forensic scientists: A guide for lawyers" (2014) 39 Australian Bar Review 174; Paciocco Jukebox, *supra* note 94 at 599-608.

the opinion) have access to their data and detailed research materials? Is their methodology and data stored in a manner that demonstrates it was not changed after beginning the research?

Finally, lessons from open science also suggest that courts should regulate eminence rather than enable it. Eminence is, at best, a noisy measure of any fact relevant to adjudication.[239] Its biasing effect contributed to the replicability crisis and courts should seek to limit its effect.[240] *Abbey* provides an example of a (failed) appeal to eminence. In *Gager*, Totten declared himself "an expert witness and Canadian expert on gangs."[241] The trial judge chided him for this, saying that "the statement suggests to me that Dr. Totten was characterizing himself as the pre-eminent expert in the field."[242]

A curative may also flow from *Abbey*. In the original *Abbey* appeal, Justice Doherty suggested that the gatekeeper play an active role in establishing the scope of expert evidence: "The trial judge may admit part of the proffered testimony, modify the nature or scope of the proposed opinion, or edit the language used to frame that opinion…"[243] This recommendation has been endorsed by the Supreme Court at least twice and it offers an economical solution to managing eminence.[244] This is because, while it is impossible to ascertain all of the scientific issues that may arise as the expert gives evidence and is cross-examined,[245] eminence is low-hanging fruit for early scoping. In forensic science, for instance, an examiner's proficiency (i.e., his or her error rate on similar tasks) is much more important than his or her eminence.[246] For instance, awards and commendations may not necessarily capture the most important variable –

[239] Garrett & Mitchell, *supra* note 134.
[240] *Ibid*.
[241] *Gager*, *supra* note 150 at para 33.
[242] *Ibid* at para 38.
[243] *Abbey* ONCA 2009, *supra* note 144 at para 63. See also *R v Ranger*, 67 OR (3d) 1 at para 63, 178 CCC (3d) 375.
[244] *R v Sekhon*, 2014 SCC 15 at paras 46-48, 367 DLR (4th) 601; *Bingley*, *supra* note 211.
[245] See, for instance, *R v Kelly*, 2017 ONCA 920 at paras 30-35, 138 OR (3d) 241.
[246] Garrett & Mitchell, *supra* note 134.

how often the examiner tends to be wrong. In such cases, trial judges may find it useful to establish the boundaries of reportable eminence (i.e., sticking to proficiency) very early on, such as after a *voir dire* or even during case management meetings.

Part VI. Conclusion: Improving trust and efficiency in expert evidence

In this article, we began by describing the state of openness and transparency in science, our culture's prevailing means of producing knowledge. Metascientific research suggests that reforms aimed at openness can help generate a more reliable body of knowledge by curbing researchers' flexibility to come to a more publishable conclusion. We then turned to the legal system, where expert witnesses are generally in charge of generating and conveying knowledge. These experts, some who come from an academic scientific background and some who do not, labour under many of the same demands as mainstream scientists and make similar errors in similar ways. As a result, we suggested that openness and transparency should be treated as a guiding principle for expert evidence, the parties proffering it, and the courts evaluating it.

While we have mainly focused on openness and transparency in expert evidence as way to improve the reliability of that evidence (or to make it easier to see that it is not reliable), we will conclude with two additional benefits to open and transparent expert evidence: improved trust and efficiency. We hope these two benefits may encourage even the more reluctant adopters to adopt open and transparent practices.

Beginning with trust, recent (and ongoing) controversies may have damaged the public perception of several fields of expert evidence. As to experts generally, they are often accused of being partisan.[247] And in forensic science, concerns about its reliability appear to be reaching a

[247] Paciocco, Jukebox, *supra* note 94.

fever pitch. As has been extensively documented, past controversies called into question forensic science's credibility.[248] Public sentiment finally seems to be catching on to this information with the media documenting visceral mistakes made by current practices. This includes evocative and searching documentaries and extensive media coverage.[249] Recent empirical research reinforces the view that the public is beginning to see the forensic sciences as less believable.[250]

The ways in which forensic science's two most closely allied fields (i.e., mainstream science and law) have dealt with issues of trust may be instructive. In science, the open science movement is surely very much about the reliability and democratization of knowledge. But it also aims to improve the public's trust in science by laying bare its inner workings.[251] Otherwise, a lemons market could result, with consumers of science so uncertain about the quality of findings that they dismiss most results. Similarly, in many jurisdictions, law seeks to maintain accountability and thus trust in its processes through open courtrooms and published judgments.[252]

Experts (or those who generate research regularly used by experts) may wish to follow suit. For instance, although many journals (including forensic scientific journals) do not require publishing data and preregistering analysis plans, there is nothing stopping researchers from doing so. Such practices may assuage concerns that their studies used questionable research

[248] See the sources at *supra* note 3; Beaman Report, *supra* note 57; Lang Report, *supra* note 58
[249] See Center for Integrity in Forensic Sciences, Centre for Integrity in Forensic Sciences: Reform in Forensic Sciences, Crime Laboratories, and the Courtroom, online: <https://cifsjustice.org/#/main>; Last Week Tonight with John Oliver, 'Forensic Science' (HBO television broadcast 1 Oct., 2017); Making a Murderer (Netflix television broadcast, 2015); The Staircase (Netflix television broadcast, 2018). Jennifer Mnookin suggests forensic science is at a turning point with reform or staying the course being equally likely: Jennifer Mnookin, "The Uncertain Future of Forensic Science" (2018) 147:4 Daedalus 99.
[250] Gianni Ribeiro, Jason M Tangen & Blake M McKimie, "Beliefs about error rates and human judgment in forensic science (forthcoming) Forensic Science International.
[251] Simine Vazire, "Quality Uncertainty Erodes Trust in Science" (2017) 3:1 Collabra: Psychology 1.
[252] See Emma Cunliffe, "Open Justice: Concepts and Judicial Approaches" (2012) 40 Federal Law Review 385-411 at 388.

practices. As to the application of foundational research to specific cases, forensic labs may also wish to improve trust in their procedures by opening them. In this vein, forensics laboratories may note the successes of the Houston Forensic Science Center (HFSC), which was created after scandals plagued its predecessor organization.[253] The first "transformative ideal" adopted by the HFSC was transparency.[254] This ideal is expressed through measures like public board meetings, direct access for the public defenders' office to lab results, and an online portal open to the public with information about standard operating procedures, incidents, and the responses to those incidents.[255]

Beyond trust, expert witnesses – forensic scientists in particular – could take note of the efficiency gains open scientists have found in some of their reforms.[256] One of the most daunting tasks facing forensic science is the validation of many of its practices through largescale studies.[257] The OSF, which includes tools for combining research efforts and sharing data[258] could assist forensic researchers in conducting these studies across multiple labs.[259] As to the day-to-day work of forensic practice, transparency may also beget efficiency in the long run. Here again, the HSFC reports that their transparency reforms had the ancillary benefit of making their work more efficient.[260] In their case, providing a web portal for public defenders and the public significantly freed up administrative resources and cut down on freedom of information requests: "While the HFSC has pursued radical transparency as a way to strengthen public trust

[253] Cásarez & Thompson, *supra* note 191.
[254] *Ibid* at 1042-1043.
[255] *Ibid* at 1045.
[256] Munafò, *supra* note 1 at 2.
[257] PCAST Report, *supra* note 15 at 52.
[258] See, for instance, Center for Open Science, "Study Swap", online: <https://osf.io/view/StudySwap/>.
[259] Similarly, open scientific methods may assist in developing systems that combine the judgments of multiple examiners and machines (but such work in is in its nascent stages), see Kristy A Martire, Bethany Growns & Danielle J Navarro, "What do the experts know? Calibration, precision, and the wisdom of crowds among forensic handwriting experts" (forthcoming) Psychonomic Bulletin and Review.
[260] Cásarez & Thompson, *supra* note 191 at 1046.

in its operations, its commitment to transparency has resulted in an added benefit: the creation of a more efficient criminal justice system that saves time and money for all participants."[261]

Overall, we have endeavoured to outline the many benefits that may accrue to the legal system if it takes seriously the transparency and openness reforms going on in science. These include efficiency and a system that is more trustworthy and accountable. But perhaps most fundamentally, transparently produced and presented expert evidence helps fulfil the ideals of expert evidence: providing knowledge that is helpful to the court, but that is also assailable by the adverse party and understandable to the trier of fact.

[261] *Ibid.*