

The New Psychology of Expert Witness Procedure

Jason M Chin,* Mehera San Roque† and Rory McFadden‡

Abstract

Can procedural reforms effectively regulate expert witnesses? Expert procedures, like codes of conduct and court-appointed experts, remain controversial among academics and courts. Much of this discussion, however, has been divorced from the science of the reforms. In this article, the authors draw from emerging work in behavioural ethics and metascience that studies procedures analogous to those that are being used in courts. This work suggests that procedures can be effective, as they have been in science, if directed at key vulnerabilities in the research and reporting process. The authors' analysis of the metascience and behavioural ethics literature also suggests several nuances in how expert evidence procedure ought to be designed and employed. For instance, codes of conduct require specific and direct wording that experts cannot interpret as ethically permissive. Further, drawing on a recent case study, courts have an important role to play in establishing a culture that takes codes as serious ethical responsibilities, and not simply as pro forma requirements.

I Introduction

In response to the threat of partisan expert witnesses, legal systems have developed a variety of procedural mechanisms (for example, expert codes of conduct, concurrent evidence, and court-appointed experts) to help manage experts and maintain public trust in the courts.¹ These procedures have inspired considerable academic and professional debate, and uneven adoption by courts.² However, this

* Lecturer, Sydney Law School, University of Sydney, New South Wales, Australia; Institute for Globally Distributed Open Research and Education ('IGDORE'). Email: jason.chin@sydney.edu.au; ORCID iD: 0000-0002-6573-2670.

† Associate Professor, UNSW Law, University of New South Wales, Australia. Email: m.sanroque@unsw.edu.au. ORCID iD: 0000-0003-1352-9862.

‡ LLB candidate, TC Beirne School of Law, University of Queensland, Brisbane, Queensland, Australia. Email: rmcfa19@gmail.com. ORCID iD: 0000-0002-7350-2226.

¹ See Gary Edmond, 'After Objectivity: Expert Evidence and Procedural Reform' (2003) 25(2) *Sydney Law Review* 131 ('After Objectivity'); David Paciocco, 'Unplugging Jukebox Testimony in an Adversarial System: Strategies for Changing the Tune on Partial Experts' (2009) 34(2) *Queen's Law Journal* 565; Ian Freckelton et al, *Expert Evidence and Criminal Jury Trials* (Oxford University Press, 2016) 52–64. These reforms have also been inspired by the desire to make proceedings more efficient and cost effective, see New South Wales Law Reform Commission ('NSWLRC'), *Expert Witnesses* (Report No 109, June 2005) 40–41.

² See *ibid.* Compare the reluctance to enforce a code of conduct in *Chen v The Queen* (2018) 97 NSWLR 915 ('Chen'), with the earlier, stricter approach in *Commonwealth Development Bank of Australia Pty Ltd v Cassegrain* [2002] NSWSC 980 ('Cassegrain').

discussion has been almost entirely uninformed by empirical research.³ In contrast with this experience in law, several sciences are enthusiastically enacting procedural reforms, which are being robustly tested and which rely on a large body of psychological research.⁴ This new area of metascientific and psychological research provides a novel perspective on procedural reform, suggesting such reform can meaningfully contribute to the regulation of expert witnesses. It also suggests how procedures ought to be designed and implemented. In this article, we explore that connection and, in doing so, the possibilities and limits of expert witness procedure.

In law, procedural reform aimed at expert partisanship has been controversial, garnering professional and academic support,⁵ but also sceptical and critical commentary.⁶ In particular, the critics have pointed out that the focus on individual expert partisanship promotes a narrow understanding of current problems with expert evidence,⁷ and also that expert procedures were designed without the benefit of empirical testing and may have perverse effects.⁸ Moreover, in forensic science specifically, partisanship may be a less pressing concern than the fact that many practices have not been demonstrated to actually work.⁹

We seek to develop this discussion by highlighting an emerging corner of metascientific research (that is, the scientific study of science itself) that examines analogous procedural reform in science. These new procedures — grounded in the psychological study of ethical behavioural — have responded to a growing concern from many fields that many published studies cannot be reproduced by independent researchers.¹⁰ Such reforms include procedural modifications to the way scientists

³ See Paul Michell and Renu Mandhane, 'The Uncertain Duty of the Expert Witness' (2005) 42(3) *Alberta Law Review* 635, 663: 'One common element of the external controls discussed below is that they remain largely untested'.

⁴ Marcus R Munafò et al, 'A Manifesto for Reproducible Science' (2017) 1(1) *Nature Human Behaviour* 1; Tom E Hardwicke et al, 'Calibrating the Scientific Ecosystem through Meta-Research' (2020) 7 *Annual Review of Statistics and its Application* (advance); Jennifer K Robbennolt and Jean R Sternlight, 'Behavioral Legal Ethics' (2013) 45(3) *Arizona State Law Journal* 1107, 1116–24.

⁵ Justice Peter McClellan, 'New Method with Experts — Concurrent Evidence' (2010) 3(1) *Journal of Court Innovation* 259; Paciocco (n 1); *Caruana v Darouti* [2014] NSWCA 85, [123].

⁶ Gary Edmond and Mehera San Roque, 'Just(.) Quick and Cheap? Contemporary Approaches to the Management of Expert Evidence' in Michael Legg (ed) *Resolving Civil Disputes* (LexisNexis, 2016) ch 7; Edmond, 'After Objectivity' (n 1); GL Davies, 'Court Appointed Experts' (2005) 5(1) *Queensland University of Technology Law and Justice Journal* 89; Michell and Mandhane (n 3); Justice Garry Downes, 'Problems with Expert Evidence: Are Single or Court-appointed Experts the Answer?' (2006) 15(4) *Journal of Judicial Administration* 185.

⁷ Edmond and San Roque (n 6).

⁸ Edmond, 'After Objectivity' (n 1); Michell and Mandhane (n 3) 663–71.

⁹ Gary Edmond, 'Forensic Science Evidence and the Conditions for Rational (Jury) Evaluation' (2015) 39(1) *Melbourne University Law Review* 77. See also Emma Cunliffe, 'A New Canadian Paradigm? Judicial Gatekeeping and the Reliability of Expert Evidence' in Paul Roberts and Michael Stockdale (eds), *Forensic Science Evidence and Expert Witness Testimony: Reliability through Reform?* (Edward Elgar 2018) 310.

¹⁰ See Francis S Collins and Lawrence A Tabak, 'Comment: NIH Plans to Enhance Reproducibility' (2014) 505(7485) *Nature* 612. Moreover, the research underpinning these reforms understands what it means to be 'biased' or 'self-serving' in more nuanced ways than has traditionally been the case in the legal context. For a detailed account of the psychology of bias in the context of judges and judging more generally, see Gary Edmond and Kristy A Martire, 'Just Cognition: Scientific Research on Bias and Some Implications for Legal Procedure and Decision-Making' (2019) 82(4) *Modern Law Review* 633 ('Just Cognition').

typically see their findings reviewed by others and published. Importantly, these reforms have received empirical testing demonstrating they often work and have been endorsed by respected scientific bodies, which may increase their ethical and psychological force.¹¹ As we will discuss below, these insights from metascience help provide a roadmap for procedural reform in courts. Expert codes of conduct may especially benefit from recent research in metascience.

Our emphasis on codes of conduct — a procedural reform that spans civil and criminal trials in New South Wales (‘NSW’) — makes our analysis necessarily broad. That said, we recognise that criminal and civil litigation engage different policy considerations and practicalities (for example, the recent emphasis on efficiency in civil litigation).¹² As to the latter, criminally accused parties frequently cannot afford their own expert witnesses and must rely on the expert proffered by the Crown. So, in the criminal context, robust expert procedure may be especially important. Indeed, we will focus on criminal cases in our legal analysis.¹³ Any application of our suggestions should mind the significant policy gap between civil and criminal cases.

In Part II, we briefly set the scene with some of the most significant procedural reforms that have been introduced to manage the presentation, form and content of expert evidence and expert reports. Part III introduces new research in metascience and behavioural ethics (that is, the psychological study of the situational factors that influence ethical behaviour) that founds a procedural reform movement in science. As we discuss, these reforms are being eagerly adopted in many scientific fields. Part IV then begins the discussion about how revelations from metascience and behavioural ethics could be leveraged to improve expert evidence procedure, putting them on firmer (meta)scientific footing. In Part V, we conclude with some limitations that can be expected of even the most scientifically grounded expert procedural reforms.

II Legal Responses to the Problem of Expert Bias

The design and implementation of expert witness procedural rules and mechanisms need to be situated within the broader context of the admissibility rules and more conventional trial safeguards that also seek to regulate expert evidence. Admissibility regimes, whether restrictive or permissive, have not generally arisen or been designed to explicitly address problems of expert bias. Further, Australian courts have typically refrained from demanding that expert evidence be demonstrably reliable and adversarial safeguards such as cross-examination are not able adequately to fill this gap. This suggests there is a clear role for expert procedure — if carefully instituted and enforced — in managing expert partisanship and also in helping to address wider concerns relating to reliability and factual rectitude.

¹¹ 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 *Systematic Reviews* 60; BA Nosek et al, ‘Promoting an Open Research Culture’ (2015) 348(6242) *Science* 1422.

¹² Edmond and San Roque (n 6).

¹³ *Chen* (n 2); *R v Warwick* (No 33) [2018] NSWSC 1219 (‘*Warwick*’).

The rules of admissibility currently play a minimal role in limiting the admission of opinion evidence from witnesses designated as ‘experts’.¹⁴ The exception to the opinion rule in the *Uniform Evidence Law* (‘UEL’) requires that the expert possess, ‘specialised knowledge’ based on ‘training, study or experience’, and that the opinion is based on that knowledge.¹⁵ However, while some decisions have rejected an expert’s evidence on the grounds that the expert has not demonstrated a connection between their opinion and their ‘specialised knowledge’, these decisions have not addressed, at a more fundamental level, questions of reliability of expert opinion.¹⁶ Rather, Australian courts have resisted reading into s 79 a requirement that expert opinion be shown to be reliable, explicitly rejecting this argument in several cases.¹⁷

The light touch apparent in the application of s 79 is matched by the weakening of the protection offered by ss 135–7 of the *UEL*. These sections purport to allow the trial judge to exclude evidence where the probative value of the evidence is outweighed by the danger of unfair prejudice to the accused. Recently, however, the High Court of Australia in *IMM v The Queen* seemed to make it difficult for a trial judge to reject expert opinion evidence for lack of information about its reliability.¹⁸ In short, the High Court resolved conflicting appellate case law in NSW and Victoria by holding that the probative value of evidence (including its reliability) should be taken ‘at its highest’ for the ss 135–7 calculus.¹⁹ While the applicability of *IMM* to scientific evidence has not been conclusively established, at least one appellate decision hesitantly applied it to evidence it would have admitted anyway.²⁰

¹⁴ The deficiencies in the application of the current rules have been canvassed elsewhere. See, eg, Gary Edmond and Mehera San Roque, ‘Before the High Court: *Honeysett v The Queen*: Forensic Science, “Specialised Knowledge” and the Uniform Evidence Law’ (2014) 36(2) *Sydney Law Review* 323; Rachel A Searston and Jason M Chin, ‘The Legal and Scientific Challenge of Black Box Expertise’ (2019) 38(2) *University of Queensland Law Journal* 237.

¹⁵ *Uniform Evidence Law* (‘UEL’) s 79. In the majority of Australian jurisdictions, the admissibility of expert evidence is now regulated via ss 76–80 of the *UEL*. The *UEL* is incorporated into the laws of the Commonwealth (*Evidence Act 1995* (Cth)), the Australian Capital Territory (‘ACT’) (*Evidence Act 2011* (ACT)), NSW (*Evidence Act 1995* (NSW)), the Northern Territory (*Evidence (National Uniform Legislation) Act 2011* (NT)), Tasmania (*Evidence Act 2001* (Tas)) and Victoria (*Evidence Act 2008* (Vic)), as well as Norfolk Island (*Evidence Act 2004* (NI)). Queensland, Western Australia and South Australia have maintained a common law of evidence, supplemented by local legislation: see Searston and Chin (n 14). On the *UEL* reforms generally, see Jeremy Gans, Andrew Palmer and Andrew Roberts, *Uniform Evidence* (Oxford University Press, 3rd ed, 2019).

¹⁶ *Honeysett v The Queen* (2014) 253 CLR 122; *R v Tang* (2006) 65 NSWLR 681, 712 (‘*Tang*’).

¹⁷ *Tang* (n 16) 712 [137]; *Tuite v The Queen* (2015) 49 VR 196, 217 [70] (‘*Tuite*’) and, more recently, *Chen* (n 2) 926 [62]. A similar trend has occurred at common law, with courts unwilling to require that expert evidence be demonstrably reliable, with some exceptions: see Searston and Chin (n 14).

¹⁸ *IMM v The Queen* (2016) 257 CLR 300 (‘*IMM*’); Kristy A Martire and Gary Edmond, ‘Rethinking Expert Evidence’ (2017) 40(3) *Melbourne University Law Review* 967. See generally, David Hamer, ‘The Unstable Province of Jury Fact-Finding: Evidence Exclusion, Probative Value and Judicial Restraint after *IMM v the Queen*’ (2017) 41(2) *Melbourne University Law Review* 689. For commentary suggesting that reliability still ought to play a significant role in the s 137 calculus, see Stephen Odgers, *Uniform Evidence Law* (Thomson Reuters, 13th ed, 2018) 1249–50; Gary Edmond, ‘Icarus and the Evidence Act: Section 137, Probative Value and Taking Forensic Science Evidence ‘At Its Highest’’ (2017) 41(1) *Melbourne University Law Review* 106.

¹⁹ *IMM* (n 18) 313 [44], 314 [47].

²⁰ *Langford v Tasmania* (2018) 29 Tas R 68, 84–6 [51]–[57] (‘*Langford*’). See also discussion below of *Chen* (n 2): below nn 145–60 and accompanying text.

Decisions that employ a narrow view of ss 135–7 often fall back on reliance on (or the mere existence of) traditional trial safeguards to expose deficiencies in the evidence.²¹ In other words, courts regularly advert to the possibility of thorough cross-examination, judicial warnings, and rebuttal experts as ways to mitigate the risk of unfair prejudice that may arise in relation to expert evidence. However, this reliance may be misguided. Recent reviews of trial transcripts, for instance, find that cross-examination does not always assist in exposing controversies and uncertainties in expert evidence and, in some cases, may actually offer prosecution witnesses an opportunity to correct deficiencies in their evidence without consequences.²² Moreover, judicial warnings cannot themselves provide the knowledge needed to resolve a dispute, but rather provide general admonitions. And while rebuttal experts may be helpful, there are systemic limits in the defence’s ability to retain such experts in the criminal context.²³

A Expert Partisanship and Procedural Reform

Against the above backdrop, procedural mechanisms, which developed with the aim of controlling expert partisanship, are a relatively recent phenomenon. Just prior to the implementation of the first tranche of procedural reform, courts began to enunciate common law duties requiring the expert to remain independent from the litigation and to act in an impartial manner.²⁴ These duties responded to expressed (but not necessarily empirically-grounded) concerns about the effects of partisanship, costs, and delays in *civil* litigation.²⁵ Many of the concerns about partisanship merged with fears that ‘junk science’ was finding its way into courtrooms, presented by unscrupulous experts, willing to tailor their evidence to the needs of their instructing client.²⁶

In Australia, the procedural reforms were incorporated into ancillary legislation, court rules, and jurisdiction specific Practice Notes.²⁷ These

²¹ *R v Shamouil* (2006) 66 NSWLR 228, 239 [77]; *Tuite* (n 17) 236 [125]–[126]; *Collins Thomson Pty Ltd (in liq) v Clayton* [2002] NSWSC 366, [24]–[26]; *Fagenblat v Feingold Partners Pty Ltd* [2001] VSC 454, [8] (*‘Fagenblat’*); *Chen* (n 2) 928 [75]; *R v Cook* [2004] NSWCCA 52, [37]–[52].

²² Gary Edmond et al, ‘Forensic Science Evidence and the Limits of Cross-Examination’ (2019) 42(3) *Melbourne University Law Review* 858; Gary Edmond, Kristy Martire and Mehera San Roque, ‘Expert Reports and the Forensic Sciences’ (2017) 40(2) *University of New South Wales Law Journal* 590 (*‘Expert Reports’*); Searston and Chin (n 14).

²³ See Keith A Findley ‘Innocents at Risk: Adversary Imbalance, Forensic Science, and the Search for Truth’ (2008) 38(3) *Seton Hall Law Review* 893.

²⁴ See *National Justice Compania Naviera SA v Prudential Assurance Company Ltd (Ikarian Reefer)* [1993] 2 Lloyd’s Rep 68 (*‘Ikarian Reefer’*). For a prominent endorsement of *Ikarian Reefer* in Australia, see *Makita (Australia) Pty Ltd v Sprowles* (2001) 52 NSWLR 705, 739 [79]. See also NSWLRC (n 1) 27; Michell and Mandhane (n 3) 642–3.

²⁵ NSWLRC (n 1) 40–41.

²⁶ See Edmond, ‘After Objectivity’ (n 1) 141.

²⁷ In the ACT, see *Court Procedure Rules* 2006 (ACT) sch 1; in Queensland, see *Uniform Civil Procedure Rules 1999* (Qld) regs 426, 428, 429A, 429B; in Western Australia, see *Consolidated Practice Direction – Civil Jurisdiction* (WA) PD annexure 3; in South Australia, see *Supreme Court Civil Supplementary Rules 2014* (SA) pt 9 div 2; in Tasmania, see *Supreme Court Rules 2000* (Tas) ss 514–17; in the Northern Territory, see *Practice Direction No 4 of 2009: Expert Reports*, in Victoria, see *Supreme Court (General Civil Procedure) Rules 2015* (Vic) o 44. Expert procedures in

developments occurred alongside other reforms to procedure supported by the Australian Law Reform Commission's inquiry into civil justice.²⁸ In this regard, it is worth emphasising that the anxieties that gave rise to procedural reform, in Australia and elsewhere, arose in relation to civil litigation, and largely in response to expert evidence being called by plaintiffs.²⁹ As a notable example, in most jurisdictions codes of conduct, as well as rules relating to court-appointed experts and concurrent evidence, remain part of the rules relating to civil procedure. These rules are unevenly extended to the criminal courts without any adaption or modification referable to the different conditions of a criminal prosecution.

Few mechanisms have been developed that are specifically adapted to the particular context of an accusatorial prosecution. Only in Victoria, and only very recently, has a Practice Note been developed specifically for criminal trials.³⁰ Given this history, it is perhaps not surprising that of the three procedural reforms outlined below (codes, court-appointed experts and concurrent evidence), only the first has been readily or regularly incorporated into criminal procedural practice. Consequently, this article focuses on the development and enforcement of codes of conduct, but we also briefly consider court-appointed experts, as well as joint testimony and pre-trial meetings between experts.

B Codes of Conduct

The expert's overriding duty to the court and corresponding code of conduct fleshing out that duty have been described as the 'centrepiece' to the procedural reform effort.³¹ For the sake of brevity, we will generally focus our review and analysis on the procedures in NSW (but highlight some relevant differences). The NSW Expert Witness Code of Conduct ('*NSW Code*') was initially adopted as a schedule to the *Supreme Court Rules 1970* (NSW)³² and later incorporated into the *Uniform Civil Procedure Rules* ('*UCPR*').³³ It has been extended to apply in all criminal courts in NSW.³⁴

The codes often echo the expert's duties of independence and impartiality, with some elaboration. The *NSW Code*, for instance, begins with a general enunciation of the expert's 'paramount duty' to the court.³⁵ It goes on to require that experts affirm that they have read and agree to be bound by the Code, and then lists a number of expectations and requirements that must be complied with when

federal courts can be found in *Federal Court Rules 2011* (Cth) pt 23. We discuss procedural reform in NSW (and Victoria, to some extent) in more depth below, with a focus on criminal proceedings.

²⁸ These included enhanced case management and disclosure, see Australian Law Reform Commission ('ALRC'), *Managing Justice: A Review of the Federal Civil Justice System* (Report No 89, 2000).

²⁹ NSWLRC (n 1) 40–41.

³⁰ Supreme Court of Victoria, *Practice Note SC CR 3: Expert Evidence in Criminal Trials*, 30 January 2017 ('*Victorian Code*'), a reissue of the same Practice Note issued in 2014.

³¹ Edmond, 'After Objectivity' (n 1) 193.

³² Initially introduced as *Supreme Court Rules 1970* (NSW) sch 6.

³³ *Uniform Civil Procedure Rules 2005* (NSW) ('*UCPR*') sch 7, which was amended in 2016 to bring the wording in line with the *Harmonised Expert Witness Code* approved by the Council of Chief Justices.

³⁴ *Supreme Court Rules 1970* (NSW) pt 75 div 1, 3J; *District Court Act 1973* (NSW) s 171D.

³⁵ *UCPR* (n 33) sch 7 cl 2.

producing their report (as well as formal requirements for the report, like providing a summary if it is long).³⁶ These more focused attestations concern the foundations and limitations of the opinion. They ask the expert to connect the opinion to any research and testing performed,³⁷ as well as to any assumptions that have been made in formulating the opinion.³⁸ The expert must also confirm that all appropriate inquiries and qualifications have been made,³⁹ and whether or not any part should be considered preliminary due to insufficient information.⁴⁰

Proponents of codes have suggested that they will help with experts who would have been otherwise unaware of their duty to the court, and that they will encourage impartiality the same way normal oaths to tell the truth might be effective.⁴¹ Conversely, it has been pointed out that the effect of requiring explicit acknowledgement of a code may be limited to minor increases in frankness and changes in the form of expert reports.⁴² In particular, while they may raise the spectre of some remedial measure, there is a great deal of professional judgment that goes into forming opinions that is difficult to police.⁴³

C *Court-appointed Experts and Concurrent Evidence*

In addition to the codes, other reforms have been introduced that moderate the traditional adversarial presentation of expert evidence: court-appointed experts and concurrent evidence. All of these reforms have been justified on the basis that not only will they reduce delay and cost, but also they have the capacity to combat bias.⁴⁴ Court-appointed experts and concurrent evidence procedures are rarely deployed in criminal proceedings, but are now a regular feature of civil (including family) proceedings.⁴⁵ The *UCPR*, for instance, give the court power to appoint an expert, authorise that expert to make inquiries into specific issues, and limit the number of party experts who may be called upon to provide opinions on the same matter.⁴⁶

Support for court-appointed experts draws on the assumption that one of the key threats to expert independence and impartiality is the fact that parties typically

³⁶ Some have argued that including these purely procedural requirements ‘dilute’ codes of conduct, see NSWLRC (n 1) 141 [9.15].

³⁷ *NSW Code* (n 33) cl 3(g).

³⁸ *Ibid* cl 3(d).

³⁹ *Ibid* cls 3(i)–(j).

⁴⁰ *Ibid* cl 3(k).

⁴¹ See, eg, Paciocco (n 1) 589 (emphasis added): ‘Second, requiring experts to advert to that role when preparing and offering their opinions is apt to work in much the same way as oaths and affirmations are **believed** to work.’ Note that Paciocco does not refer to psychological research examining whether oaths and affirmations have an effect: see Part III of this article. The NSWLRC also suggests that codes might reduce unconscious bias, without explaining how that might work: (n 1) 74–5 [5.20].

⁴² ‘The duties sought to be imposed by these Rules are, in practice, unenforceable. Their expression is no more than a pious hope.’: Davies (n 6) 89. See also Edmond, ‘After Objectivity’ (n 1) 148–9; Edmond and San Roque (n 6) [7.13].

⁴³ *Ibid*.

⁴⁴ Edmond and San Roque (n 6) [7.6].

⁴⁵ Though it is worth noting that court-ordered reports in the context of fitness inquiries and of sentencing hearings may be considered analogous to court-appointed experts (and are an underexplored area).

⁴⁶ *UCPR* (n 33) r 31.46.

appoint those experts.⁴⁷ As a result, they may be selected for a particular view (that is, selection bias) and see their opinions tintured through their association with one side of the dispute. However, though court-appointed experts may assuage some concerns, they do raise other challenges. For instance, court-appointed experts may carry professional and ideological biases, but these may go unexplored because they seem more neutral.⁴⁸ More generally, a great deal of weight and importance will naturally attach to the judge's choice of expert, and in many cases that choice may determine the outcome of the dispute. Possibly as a result of these challenges, courts seem reluctant to exercise their power to appoint experts.⁴⁹

Turning to procedures involving two (or more) experts, the typical justification is the expectation that they will incline experts to moderate their views, discover areas of agreement, and abandon more tenuous claims.⁵⁰ In other words, proponents of these rules suggest that another expert's scrutiny will expose or ward off some biases and reduce reliance on 'junk science'.⁵¹ Similarly, pre-trial meetings between two experts may also help narrow issues and thus promote more efficient dispute resolution.⁵² However, sceptics point out that experts meeting before trial or providing concurrent evidence will still be subjected to pressures from the parties tendering them.⁵³ More fundamentally, just because two experts from a field agree, that is not necessarily a good reason to think their opinion is factually accurate. There are many matters within fields on which there is no expert consensus,⁵⁴ and there is limited empirical evidence that reforms concerning the use of multiple experts are working in the ways that were intended, especially in terms of whether such processes are more efficient overall.⁵⁵

III The Metascience and Psychology of Expert Procedure

While expert procedural reform has attracted a great deal of scrutiny, much of the existing discussion has not had the benefit of direct empirical research (which would be difficult to conduct in the context of courts).⁵⁶ Now, with several scientific fields enthusiastically implementing and testing their own procedural reforms, we have a new lens through which to evaluate expert procedure. In particular, we can say whether — in light of the experience in science — broadly analogous expert procedures can be expected to work, and under what conditions. We can also provide suggestions for empirically-guided improvements to our current web of procedural expert safeguards. We will begin by reviewing new insights from metascience with

⁴⁷ See Davies (n 6) 90–2.

⁴⁸ Michell and Mandhane (n 3) 666. For an encouraging, albeit limited, example of court-appointed experts being used to combat adversarial bias, see NSWLRC (n 1) 34–6 [3.37]–[3.42].

⁴⁹ See NSWLRC (n 1) 33–4 [3.33]–[3.36].

⁵⁰ See Edmond, 'After Objectivity' (n 1) 141.

⁵¹ *Ibid.*

⁵² ALRC, *Review of the Federal Civil Justice System* (Discussion Paper 62, 1999) 493–4 [13.44]–[13.46].

⁵³ Davies (n 6) 94–5; Michell and Mandhane (n 3) 670–71.

⁵⁴ Edmond, 'After Objectivity' (n 1) 150–51.

⁵⁵ Edmond and San Roque (n 6) [7.10]. See also Simon McKenzie, 'Concurrent Evidence in the Kilmore East Bushfire Proceeding' [2016] VicSCLRS 2, 13.

⁵⁶ Michell and Mandhane (n 3) 663.

a view to identifying how this emerging area of research could inform the regulation of expert witnesses in the legal context.

A *Metascience and Scientific Procedural Reform*

The field of metascience — the scientific study of science itself — is flourishing and has generated substantial empirical evidence for the existence and prevalence of threats to efficiency in knowledge accumulation.

Data from many fields suggests reproducibility is lower than is desirable; one analysis estimates that 85% of biomedical research efforts are wasted, while 90% of respondents to a recent survey in *Nature* agreed that there is a ‘reproducibility crisis’.⁵⁷

Over the past decade, scientists have begun to confront — with modern empirical tools — the biases in their research. As described in the above quote, this metascientific study has given credence to longstanding worries about the degree to which human nature can tincture the research and reporting process. In particular, it has provided evidence for a ‘reproducibility crisis’ whereby researchers have sought to reproduce an initial study’s findings by following its protocol as closely as possible, but failed to find evidence for the initial result.⁵⁸

The primary evidence for the reproducibility crisis comes from large-scale, multi-laboratory efforts seeking to reproduce peer-reviewed and published studies in eminent journals. For example, researchers recently re-performed, with large sample sizes, 21 social scientific studies originally published in *Nature* and *Science*. They found the same results as the original study in 13 of those replications (62%) and the size of the effects found were 50% smaller.⁵⁹ Similarly, in pre-clinical medical research (that is, studies performed before the drug is tested in humans), the findings of 53 landmark cancer studies could only be confirmed in six cases.⁶⁰ This finding contributed to the conclusion, expressed in the above quote, that 85% of pre-clinical medical research is wasted.⁶¹ Such findings recently prompted the United

⁵⁷ Munafò (n 4) 1 (references omitted).

⁵⁸ Ibid.

⁵⁹ Colin F Camerer et al, ‘Evaluating the Replicability of Social Science Experiments in *Nature* and *Science* between 2010 and 2015’ (2018) 2(9) *Nature Human Behaviour* 637. Similar results have been found in large-scale reproduction efforts focused on economics research: Andrew C Chang and Phillip Li, ‘Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say “Usually Not”’ (Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series Paper 2015–083). In the largest effort in psychology, the researchers attempted to reproduce the findings of 100 studies, finding the same result in just 39% attempted replications: Open Science Collaboration, ‘Estimating the Reproducibility of Psychological Science’ (2015) 349(6251) *Science* 943 aac4716.

⁶⁰ C Glenn Begley and Lee M Ellis, ‘Comment: Raise Standards for Preclinical Cancer Research’ (2012) 483(7391) *Nature* 531.

⁶¹ Above n 57 and accompanying text. See also Leonard P Freedman, Iain M Cockburn and Timothy S Simcoe, ‘The Economics of Reproducibility in Preclinical Research’ (2015) 13(6) *PLoS Biology* e1002165, 1. For similar problems in behavioural ecology, see Daiping Wang et al, ‘Irreproducible Text-Book “Knowledge”: The Effects of Color Bands on Zebra Finch Fitness’ (2018) 72(4) *Evolution* 961. For issues in neuroscience, see Anders Eklund, Thomas E Nichols and Hans Knutsson, ‘Cluster Failure: Why fMRI Inferences for Spatial Extent Have Inflated False-Positive Rates’ (2016) 113(28) *PNAS (Proceedings of the National Academy of Sciences of the United States of America)* 7900.

States ('US) Defense Advanced Research Projects Agency ('DARPA') to fund a project (in collaboration with an Australian meta-research group) to develop algorithms aimed at determining the *indicia* of studies with spurious findings.⁶²

More importantly — for our purposes — metascientists are developing an understanding of why some research findings prove more robust than others. Far from drawing awkward distinctions between science and 'junk-science' (a dichotomy that has been criticised in the legal sphere),⁶³ this research is studying the often-hidden practices by which researchers can make their (often spurious) research seem more superficially convincing. This research is, in turn, informing procedural reform in science.

These hidden practices are often referred to as questionable research and reporting practices ('QRPs').⁶⁴ They are termed 'questionable' because they fall below the level of research fraud, which appears generally uncommon (fraud is difficult to study, but most estimates put its prevalence at about 2% of researchers).⁶⁵ Rather, questionable practices rest in a grey area, with, in some cases, 60% of researchers anonymously admitting to using them.⁶⁶

QRPs allow researchers to portray their findings as speciously probative of their preferred conclusion and thus inflate their field's false discovery rate.⁶⁷ For example, one QRP is excluding outliers in an ad hoc way, giving rise to the possibility that these exclusions are driven by an unconscious desire to see one's hypothesis borne out.⁶⁸

A large simulation found that use of just four QRPs can allow researchers to take a random set of data and demonstrate any effect they wish in a way that meets traditional statistical standards of proof.⁶⁹ This study confirmed anecdotal findings showing that, with enough data and flexibility, any pattern can be made to appear superficially compelling.⁷⁰

⁶² Adam Rogers, 'Darpa Wants to Solve Science's Reproducibility Crisis with AI', *Wired* (online, 15 Feb 2019) <<https://www.wired.com/story/darpa-wants-to-solve-sciences-replication-crisis-with-robots/>>; *The RepliCATS Project: Collaborative Assessment for Trustworthy Science* (Website) <<https://replicats.research.unimelb.edu.au/>>.

⁶³ See Edmond, 'After Objectivity' (n 1) 147.

⁶⁴ Leslie K John, George Loewenstein and Drazen Prelec, 'Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling' (2012) 23(5) *Psychological Science* 524; Hannah Fraser et al, 'Questionable Research Practices in Ecology and Evolution' (2018) 13(7) *PLoS ONE* e0200303; Joseph P Simmons, Leif D Nelson and Uri Simonsohn, 'False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant' (2011) 22(11) *Psychological Science* 1359.

⁶⁵ Daniele Fanelli, 'How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data' (2009) 4(5) *PLoS ONE* e5738.

⁶⁶ John, Loewenstein and Prelec (n 64).

⁶⁷ Simmons, Nelson and Simonsohn (n 64).

⁶⁸ *Ibid* 1360.

⁶⁹ *Ibid* 1360–62.

⁷⁰ Eric-Jan Wagenmakers et al, 'An Agenda for Purely Confirmatory Research' (2012) 7(6) *Perspectives on Psychological Science* 632, 633–4.

A recent study by de Vries and colleagues provides a vivid demonstration of the combined effects of QRPs and other biases within science.⁷¹ They researched the published and unpublished literature studying one depression medication, finding that only about 50% of studies reported it effective. But this was not what the main message from the *published* studies showed. This was because de Vries and colleagues found rampant publication bias (for example, the phenomenon whereby studies showing some effect are more likely to be published than those that failed to find anything),⁷² QRP usage,⁷³ citation bias (that is, studies finding an effect are more likely to be cited than inconclusive or null findings),⁷⁴ and spin (that is, within a study, positive effects are emphasised and complicating factors are hidden in the body or footnotes).⁷⁵ Through the combined force of these factors, the published literature made it seem as if the treatment — which was only successful in 50% of studies — was effective in the vast majority of studies.

Drawing upon the above research, a host of new procedures are being employed to expose undisclosed flexibility in the research process. Many of these methods are fairly simple and not mandatory, but have found recent empirical support and endorsement by a 2018 report of the National Academies of Sciences, Engineering and Mathematics.⁷⁶ We will review these reforms now: preregistration and two modifications to the peer review process, checklists and pre-submission review.

Preregistration limits QRPs by asking researchers to pre-commit to the specifications of their studies before performing them and seeing the data. For example, researchers may pre-establish how they will exclude outliers beforehand to ensure they do not drop observations ad hoc out of an unconscious desire to confirm their hypothesis.⁷⁷ As to publication bias, preregistration can assist by creating a public record that a study has been performed (and what its specifications were) so that, if a null effect is found, there will still be a record even if the study is not published in a journal. These preregistrations are typically made on non-profit open science websites that include several focused questions about the planned research.⁷⁸

Early results for preregistration are promising. As we have noted, it has historically been very rare for journals to publish findings that did not support the

⁷¹ Y A de Vries et al, 'The Cumulative Effect of Reporting and Citation Biases on the Apparent Efficacy of Treatments: The Case of Depression' (2018) 48(15) *Psychological Medicine* 2453.

⁷² Ibid 2453. See also Kay Dickersin, 'The Existence of Publication Bias and Risk Factors for Its Occurrence' (1990) 263(10) *JAMA* 1385.

⁷³ In the de Vries study, this was only reporting positive findings in study, sometimes called 'reporting bias': de Vries et al (n 71) 2453.

⁷⁴ Ibid 2453.

⁷⁵ Ibid. See also Kellia Chiu, Quinn Grundy and Lisa Bero, "'Spin'" in Published Biomedical Literature: A Methodological Systematic Review' (2017) 15(9) *PLoS Biol* e2002173.

⁷⁶ National Academies of Sciences, Engineering, and Medicine, Committee on Toward an Open Science Enterprise, *Open Science by Design: Realizing a Vision for 21st Century Research* (National Academies Press, 2018)107–20.

⁷⁷ Brian A Nosek et al, 'The Preregistration Revolution' (2018) 115(11) *PNAS* 2600 ('Preregistration Revolution').

⁷⁸ Center for Open Science, *Open Science Framework (OSF)* <<https://osf.io>>; *AsPredicted* <<https://aspredicted.org>>.

researcher's hypothesis (only about 5–20% of such findings are published).⁷⁹ One recent study found that preregistered studies buck this trend, reporting such null result findings about 55% of the time.⁸⁰ These meta-scientific findings converge with the experience in medical research. After the US required preregistration for clinical trials, the percentage of large National Heart, Lung, and Blood Institute funded studies showing cardiovascular drugs had no effect changed drastically, from 43% to 92%.⁸¹ In tandem with these encouraging results about the efficacy of preregistration, some fields are reporting increases in their use.⁸²

Another way of encouraging researchers to more faithfully disclose the limitations of their methods and findings is by requiring or recommending they complete a checklist when submitting an article for peer-review.⁸³ While checklists may seem simplistic, they have proven surprisingly effective in applied contexts, such as in improving surgery outcomes.⁸⁴ In the research context, these checklists encourage authors to disclose important details that explain the limits of the study, like changes they made to the protocol after the study started and why they excluded any observations.⁸⁵

The most well-studied and widely-adopted checklist is produced by the Consolidated Standards of Reporting Trials ('CONSORT') initiative.⁸⁶ It is used in the reporting of clinical medical trials. Systematic analyses of CONSORT find that studies published in endorsing journals (which range from mere references to CONSORT to requiring authors submit the checklist along with their manuscripts) show improved reporting over a variety of measures.⁸⁷ These effects occur despite the checklist not being mandated or policed in many cases.⁸⁸ In fact, checklists can even encourage researchers to report weaknesses in their studies, like the failure to randomly assign animals to experimental conditions and to blind experimenters to

⁷⁹ Chris Allen and David MA Mehler, 'Open Science Challenges, Benefits and Tips in Early Career and Beyond' (2019) 17(5) *PLoS Biol* e3000246, 9.

⁸⁰ *Ibid.*

⁸¹ Robert M Kaplan and Veronica L Irvin, 'Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time' (2015) 10(8) *PLoS ONE* e0132382, 1.

⁸² For psychology, sociology, economics and political science, see David Birke et al, 'Open Science Practices Are on the Rise across Four Social Science Disciplines' (Presentation, Berkeley Initiative for Transparency in the Social Sciences, Annual Meeting, 30 November 2018) <<https://osf.io/kvbnh/>>.

⁸³ David Moher, Alison Jones and Leah Lepage, CONSORT Group (Consolidated Standards for Reporting of Trials), 'Use of the CONSORT Statement and Quality of Reports of Randomized Trials: A Comparative Before-and-After Evaluation' (2001) 285(15) *JAMA* 1992; Turner et al (n 11).

⁸⁴ J Bergs et al, 'Systematic Review and Meta-Analysis of the Effect of the World Health Organization Surgical Safety Checklist on Postoperative Complications' (2014) 101(3) *British Journal of Surgery* 150.

⁸⁵ See CONSORT, 'CONSORT 2010 Checklist of Information to Include When Reporting A Randomised Trial' (Checklist, 2010) 3b ('Important changes to methods after trial commencement (such as eligibility criteria), with reasons'); 13b ('For each group, losses and exclusions after randomisation, together with reasons') <<http://www.consort-statement.org/download/Media/Default/Downloads/CONSORT%202010%20Checklist.doc>>.

⁸⁶ *Ibid.*

⁸⁷ Turner et al (n 11); Moher, Jones and Lepage (n 83).

⁸⁸ Turner et al (n 11) 61; 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.

conditions.⁸⁹ We caution, however, that studies examining less widely adopted checklists (those without a widely respected backer like CONSORT) have found mixed or no support for their efficacy.⁹⁰

Pre-submission review (that is, registered reports) tweaks the typical peer-review timing by front-ending the review process.⁹¹ In other words, studies are reviewed before data are collected based on their rationale and methodology. If reviewers find the methods and planned analysis are sufficiently sound and accept the report, then publication is nearly guaranteed as long as the authors follow through with the approved plan. Pre-submission review therefore includes many of the benefits of preregistration by creating a record of the planned study and its specifications, thus making it possible to examine any discrepancies. It also has salutary motivational consequences. Researchers should be less likely to overstate their findings because the publication decision is independent of the results and the record of pre-accepted method would make any changes to that method more obvious anyway. Indeed, a recent study found that pre-publication review was associated with fewer retracted papers.⁹²

So, why are these new scientific procedures working? And what principles are they based on? We will now discuss two mechanisms, both of which help explain how scientific reforms might be extended and applied to expert evidence procedure. First, these procedures nudge researchers towards tempered claims and fuller disclosure by tying those acts to research ethics; ‘questionable’ research practices are becoming no longer questionable, but expressly unethical. Second, they help control and record unconscious bias in ways that pre-existing scientific safeguards failed to do.

B Behavioural Ethics

Research on behavioral ethics has flourished in recent years, providing much new insight into how people make ethical decisions; the dynamic and malleable nature of ethical preferences and behavior; and the variety of cognitive, situational, and social factors that influence ethical decisions.⁹³

As Robbennolt lays out in the above quote, a growing body of research is developing to explain the psychology of ethical decision-making. Robbennolt goes on to explain how behavioural ethics can be leveraged to help improve the ethics of lawyering.⁹⁴ Its implications for expert witnesses, to our knowledge, has gone unexplored. This gap is somewhat surprising given the impact that expert evidence has on the trial

⁸⁹ Han et al (n 88) 9.

⁹⁰ For a review, see Adrienne Stevens et al, ‘Relation of Completeness of Reporting of Health Research to Journals’ Endorsement of Reporting Guidelines: Systematic Review’ (2014) 348 *British Medical Journal* g3804.

⁹¹ SPJM Horbach and W Halfman, ‘The Ability of Different Peer Review Procedures to Flag Problematic Publications’ (2019) 118(1) *Scientometrics* 339, 350–1 (‘Ability of Peer Review to Flag Problems’). See generally SPJM Horbach and W Halfman, ‘The Changing Forms and Expectations of Peer Review’ (2018) 3 *Research Integrity and Peer Review* 8.

⁹² Horbach and Halfman, ‘Ability of Peer Review to Flag Problems’ (n 91) 350.

⁹³ Jennifer K Robbennolt, ‘Behavioral Ethics Meets Legal Ethics’ (2015) 11 *Annual Review of Law and Social Science* 75, 76.

⁹⁴ *Ibid*; Robbennolt and Sternlight (n 4).

process. It may be explained, however, by courts' typically narrow view of bias as purely based on adversarial (rather than cognitive psychological) processes.⁹⁵

In this subsection, we will review the foundations of behavioural ethics. This research illuminates the processes that encourage scientists and expert witnesses to overstate their findings and downplay the limits of their expertise. It explains how they can do these things and still see themselves as upstanding actors in their field. Importantly, it also suggests ways to improve this situation (and provides reasons why the new procedures discussed above in Part IIIA seem to be working).

Behavioural ethics generally seeks to move away from a purely dispositional approach to ethics — one focused on bad apples — to one that takes into account the situation and the way in which choices are framed.⁹⁶ Much of this research finds that individuals generally seek to behave ethically because they wish to maintain a positive self-concept and will even do so when it is costly (for example, whistleblowing at great personal risk).⁹⁷ However, various situational and cognitive factors can nudge us towards less careful ethical thinking and behaviour.⁹⁸ This perspective aligns with results from metascience, wherein outright fraud is seemingly rare, but questionable research practices pervade in an environment with 'considerable latitude for rationalization and self-deception'.⁹⁹ It also aligns with the circumstances in courtrooms whereby strong situational pressures may encourage experts to stretch the boundaries of their personal ethics.

Before we continue, however, we add a note of caution. Recall that Part IIA (above) reviewed the metascientific research finding that much of the published literature contains false positive findings and exaggerated results. As a result, the studies supporting the propositions below should not be taken without question. Still, in the below review of behavioural ethics, we have reviewed the studies to ensure they have not been contradicted by subsequent research. Moreover, research drawn from the same theoretical perspective has proven robust in the face of large-scale replication attempts.¹⁰⁰

⁹⁵ See Edmond and Martire, 'Just Cognition' (n 10).

⁹⁶ See Gerd Gigerenzer, 'Moral Satisficing: Rethinking Moral Behavior as Bounded Rationality' (2010) 2(3) *Topics in Cognitive Science* 528.

⁹⁷ See Nina Mazar and Dan Ariely, 'Dishonesty in Everyday Life and Its Policy Implications' (2006) 25(1) *Journal of Public Policy & Marketing* 117.

⁹⁸ Ibid 118–19. Our use of the word 'nudge' is intentional. The field of behavioural ethics is just one part of a broader social scientific movement finding that classic economics (and the assumption of a 'rational actor') fails to explain a great deal of human decision-making and behaviour. See, eg, the field of behavioural economics: Richard H Thaler and Cass R Sunstein, *Nudge* (Penguin Books, 2009).

⁹⁹ John, Loewenstein and Prelec (n 64) 524. Likewise, one account of the problems with expert evidence suggests they are doing just what these scientists are doing — 'Moreover, it is unlikely that most expert witnesses purposefully lie under oath; rather, they likely believe that they are just putting the best face on the truth.': Michell and Mandhane (n 3) 661.

¹⁰⁰ For example, Amos Tversky and Daniel Kahneman's choice framing research in behavioural economics (suggesting, as with behavioural ethics, that the situation impacts behaviour in a way that goes beyond the typical costs and benefits) was recently replicated with a sample size of 7,228: see Richard A Klein et al, 'Many Labs 2: Investigating Variation in Replicability Across Samples and Settings' (2018) 1(4) *Advances in Methods and Practices in Psychological Science* 443. It was also replicated with a sample size of 6,271: Richard A Klein et al, 'Investigating Variation in Replicability: A "Many Labs" Replication Project' (2014) 45(3) *Social Psychology* 142.

So, what are the processes by which average people, who are motivated to see themselves as good, might do morally questionable acts? Research has uncovered many, but we will focus on those especially relevant to expert witness practice.

One process documented by research is ‘ethical fading’, by which people operate on auto-pilot, losing sight of the ethical component of what they are doing.¹⁰¹ One way this may happen is through following scripts, preset ways of approaching tasks.¹⁰² This might include, in science, researchers following their field’s protocol of performing a study, discussing the results with their lab-mates, developing a narrative to frame the results, and then cutting away the findings that do not fit with this narrative.¹⁰³ Similarly, expert witnesses might follow the script of an adversarial trial without considering the ethics of their actions.

Researchers have uncovered several ways to combat fading by making salient the ethical part of the judgment.¹⁰⁴ For example, signing an honour code or reciting the Ten Commandments substantially reduces cheating, an effect that outweighs quadrupling the monetary incentive to cheat.¹⁰⁵ These manipulations appear to work by making participants more mindful of their internal standards and thus the possibility of betraying those standards. In the same vein, children are less likely to take more from a common pool when there is a mirror present.¹⁰⁶

Humans also appear adept at mentally recasting their behaviour to help fade the moral implications of the act. For example, one group of researchers found that people were more willing to cheat to obtain tokens than cash.¹⁰⁷ This was despite the fact that the tokens were directly redeemable for cash. The researchers posited that cheating to obtain tokens provides ‘room for interpretation’ of the participants’ actions, thus ‘making the moral implications of dishonesty less accessible’.¹⁰⁸ Supporting these findings, an analysis of 137 studies in behavioural ethics found the strongest predictor of ethical behaviour was a lack of opportunity to self-justify.¹⁰⁹

Additional research — more closely related to giving opinion evidence — concerns conveying information that is inherently imprecise and subject to multiple interpretations.¹¹⁰ In this field, psychologists find that people will make more

¹⁰¹ Robbenolt and Sternlight (n 4) 1120–24.

¹⁰² Robert P Abelson, ‘Psychological Status of the Script Concept’ (1981) 36(7) *American Psychologist* 715.

¹⁰³ This approach contains several QRPs, see Wagenmakers et al (n 70) 633–4.

¹⁰⁴ See Nina Mazar, On Amir and Dan Ariely, ‘The Dishonesty of Honest People: A Theory of Self-Concept Maintenance’ (2008) 45(6) *Journal of Marketing Research* 633.

¹⁰⁵ Ibid 635–7.

¹⁰⁶ Arthur L Beaman et al, ‘Self-Awareness and Transgression in Children: Two Field Studies’ (1979) 37(10) *Journal of Personality and Social Psychology* 1835.

¹⁰⁷ Mazar, Amir and Ariely (n 104) 637–8.

¹⁰⁸ Ibid 638.

¹⁰⁹ Nicola Belle and Paola Cantarelli, ‘What Causes Unethical Behavior? A Meta-Analysis to Set an Agenda for Public Administration Research’ (2017) 77(3) *Public Administration Review* 327.

¹¹⁰ Maurice E Schweitzer and Christopher K Hsee, ‘Stretching the Truth: Elastic Justification and Motivated Communication of Uncertain Information’ (2002) 25(2) *Journal of Risk and Uncertainty* 185; Russell Golman and Sudeep Bhatia, ‘Performance Evaluation Inflation and Compression’ (2012) 37(8) *Accounting, Organizations and Society* 534; Laetitia B Mulder, Jennifer Jordan and Floor Rink, ‘The Effect of Specific and General Rules on Ethical Decisions’ (2015) 126 *Organizational Behavior and Human Decision Processes* 115. See generally Jason Dana, Roberto A

moderate offers in sales and negotiations when they have less discretion in how they determine the value of what they are offering.¹¹¹ In other words, when a decision is based on increasingly elastic and unclear criteria, individuals can stray from strict ethical guidelines without jeopardising their moral self-concept (and will do so when motivated by external rewards like prestige and money). Similarly, strict rules appear to reduce the opportunity for rationalisation and, in turn, promote ethical behaviour.¹¹²

Ethical fading is helped along by various other situational factors. For example, time pressure and lack of sleep contribute to unethical decision-making.¹¹³ Social pressures, such as that from authority figures, have a similar effect.¹¹⁴ Even just observing one group member behaving unethically can negatively affect the observer's ethical decision-making.¹¹⁵ This changes the social norm and also allows the observer to rationalise away small divergences because there are people out there doing it much worse. More generally, the ethical culture of an organisation exerts a significant impact on individual decision-making.¹¹⁶ The culture's effect on the individual includes both the social impact of seeing peers and trusted advisors behave in a certain way, and seeing the way in which the system rewards or penalises those who adhere to a strict ethical code (or not).¹¹⁷ Further, when multiple parties carry some responsibility, a 'diffusion of responsibility' may cause each party to assume the other is policing them.¹¹⁸ This may occur in court where an expert may be more inclined to overstep because they assume a judge, opposing lawyer, or opposing expert will surely step in if they go too far.

Other factors like ethical blind spots and slippery slopes may be especially relevant to expert witnesses. Most of us have 'bias blind spots', tending to see ourselves as objective and others as biased.¹¹⁹ This tendency is related to a more general 'above-average effect' whereby we tend to be overconfident in our abilities.¹²⁰ In forensic science, one survey found that 71% of practitioners agreed that cognitive bias is a problem in forensics, but only 26% would concede that it impacted their own conclusions.¹²¹ Ethical slippery slopes may also contribute to expert witnesses pushing the bounds of ethical behaviour. For example, an expert

Weber and Jason Xi Kuang, 'Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness' (2007) 33(1) *Economic Theory* 67.

¹¹¹ Schweitzer and Hsee (n 110).

¹¹² And these results hold even when the situation is engineered such that the person cannot expect others to know about their behaviour — suggesting they are not driven by the possibility of reciprocation: Mulder, Jordan and Rink (n 110).

¹¹³ Robbenolt and Sternlight (n 4) 1140–42.

¹¹⁴ *Ibid* 1146–9.

¹¹⁵ Francesca Gino, Shahar Ayal and Dan Ariely, 'Contagion and Differentiation in Unethical Behavior: The Effect of One Bad Apple on the Barrel' (2009) 20(3) *Psychological Science* 393.

¹¹⁶ Robbenolt and Sternlight (n 4) 1165–7.

¹¹⁷ *Ibid*.

¹¹⁸ Michael J Saks and Barbara A Spellman, *The Psychological Foundations of Evidence Law* (New York University Press, 2016) 216.

¹¹⁹ Emily Pronin, Daniel Y Lin and Lee Ross, 'The Bias Blind Spot: Perceptions of Bias in Self Versus Others' (2002) 28(3) *Personality and Social Psychology Bulletin* 369.

¹²⁰ David Dunning, Chip Heath and Jerry M Suls, 'Flawed Self-Assessment: Implications for Health, Education, and the Workplace' (2004) 5(3) *Psychological Science in the Public Interest* 69, 72.

¹²¹ Jeff Kukucka et al, 'Cognitive Bias and Blindness: A Global Survey of Forensic Science Examiners' (2017) 6(4) *Journal of Applied Research in Memory and Cognition* 452, 454.

may fail to provide important cautions in one case, justify that act, and then continue along that path in future cases because it is easier to do that than admit that the first act was wrong.¹²²

C *Protection against Cognitive Biases*

Finally, the procedural reforms underway in science expressly acknowledge the aim of protecting the research process from cognitive biases.¹²³ Safeguards like randomly assigning participants to control and experimental conditions have long been orthodox in many fields. Still, less-obvious flexibilities in the research process (for example, QRPs) allowed researcher biases to creep in. Scientific reforms help protect against cognitive biases by: removing flexibility to slant results one way or another (for example, preregistration); recording the opportunities for bias to creep in (for example, checklists); and, removing the motivation to frame results in a more publishable way (for example, pre-submission review).

As we discussed in Part II, while legal procedural reform was initially designed to combat bias, it had a relatively narrow view of bias in mind — for example, witnesses selected for a certain view, expert witnesses taking on the role of advocates. Perhaps not surprisingly then, subsequent jurisprudence has generally ignored or downplayed more subtle forms of bias, such as unconscious contextual bias (that is, when irrelevant details of the case affect the expert’s judgment).¹²⁴ Or, it regards such bias as a matter that can simply be exposed to the fact-finder and corrected for by that fact-finder as they weigh the evidence.¹²⁵ This runs counter to the advice of peak scientific bodies and inquiries into the causes of wrongful convictions, which have been clear about the corrosive effects of unconscious bias in expert evidence.¹²⁶

Thus, one criticism of expert witness procedures is that — as currently designed and enforced — they do little to protect against unconscious bias.¹²⁷ In many cases, simply understanding that one’s duty is to the court will not prevent

¹²² Robbenolt and Sternlight (n 4) 1118–19.

¹²³ Munafò (n 4) 3 (Table 1). In forensic science, bias-limiting procedures exist, but have not yet reached mainstream: Itiel Dror et al, ‘Letter to the Editor — Context Management Toolbox: A Linear Sequential Unmasking (LSU) Approach for Minimizing Cognitive Bias in Forensic Decision Making’ (2015) 60(4) *Journal of Forensic Sciences* 1111.

¹²⁴ Edmond and Martire, ‘Just Cognition’ (n 10); Jason M Chin, Michael Lutsky and Itiel E Dror, ‘The Biases of Experts: An Empirical Analysis of Expert Witness Challenges’ (2019) 42(4) *Manitoba Law Journal* 21.

¹²⁵ Edmond, Martire and San Roque, ‘Expert Reports’ (n 22) 619–22.

¹²⁶ Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council, ‘Strengthening Forensic Science in the United States: A Path Forward’ (National Academies Press, 2009) 122–4 (‘NAS Report’); The President’s Council of Advisors on Science and Technology, ‘Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods’ (Executive Office of the President, 2016) 31–2 (‘PCAST Report’); Ontario Ministry of the Attorney General, *Inquiry into Pediatric Forensic Pathology in Ontario* (Report, 2008) 69 (vol 1), 79 (vol 1), 447–8 (vol 3) (‘Goudge Report’).

¹²⁷ For instance, expert witnesses sometimes acknowledge the possibility of unconscious bias, but deny that they are susceptible to it: see Edmond, Martire and San Roque, ‘Expert Reports’ (n 22) 620, quoting *JP v DPP (NSW)* [2015] NSWSC 1669, [23].

experts from being exposed to biasing details (or even seeking them out).¹²⁸ In Part IV, we will suggest that expert procedures could be designed and employed in a way that helps address unconscious bias.

IV Improving Expert Evidence Procedure

To summarise the above, the peer-review and reporting process within science has not historically been designed in a way that effectively regulated the findings being published. Questionable, but not necessarily fraudulent, practices have long been used, allowing the biases of the researcher to impact published findings. New reforms, based in part on behavioural ethics, are being tested and employed to help control these biases. In this way, Part III suggests that, by analogy, procedure can be an effective way to control experts as witnesses by encouraging them to be transparent about the limits of their opinions. We now turn to the potential application of this research to expert evidence procedure. These applications range from establishing a general culture that takes expert procedure seriously, to specific tweaks to existing procedural mechanisms.

A Culture Changes and the Courts

Even if codes are reformulated to more effectively engage the ethics of experts (as we discuss in Part IVB below), their effectiveness will be hindered by a wider ethical culture that tolerates ‘experts’ overstating their claims and does not effectively enforce the requirements of the codes of conduct. Here, as we noted above, legal-behavioural ethicists find that it is not just the explicit rules that matter, but the broader system and what it appears to value:

Importantly, the ethical culture of an organization depends not only on its expressed ethical codes and policies but also far more broadly on its systems and practices. Just as group norms may have a negative impact, so too may group norms set the stage for attorneys to do the right thing.¹²⁹

Robbennolt and Sternlight go on to explore the ethical cultures that arise at law firms that may hinder ethical lawyering; a similar analysis could be undertaken with respect to expert witnesses. Consider, for instance, forensic pathology’s overarching value as prescribed by Cordner, a pathologist who led the Victorian Institute of Forensic Medicine (Australia) for years and provided measured opinions in many criminal trials: ‘forensic pathology expertise should be focused on minimising or avoiding any adverse outcomes associated with its contribution.’¹³⁰ Then compare that statement with the conclusion of an inquiry in Ontario (Canada)

¹²⁸ However, as we will discuss below, early civil decisions at least adverted to ‘human psychology’ and construed the policy behind expert codes of conduct as the avoidance of preformed decisions. In *United Rural Enterprises Pty Ltd v Lopmand Pty Ltd*, Campbell J generally agreed that codes of conduct had an important role to play in heading off bias whereby the expert may, without knowledge of the Code, form a view that is ‘difficult to retreat from ... This can happen as a matter of ordinary human psychology’: [2003] NSWSC 870, [15] (*Lopmand*). See also *Kirch Communications Pty Ltd v Gene Engineering Pty Ltd* [2002] NSWSC 485, [14].

¹²⁹ Robbennolt and Sternlight (n 4) 1165–6.

¹³⁰ Stephen Cordner, ‘Expert Opinions and Evidence: A Perspective from Forensic Pathology’ (2015) 17(2) *Flinders Law Journal* 263, 268.

into the wrongful convictions that resulted from the work of forensic pathologist Dr Charles Smith:

He acknowledged that, when he first began his career in the 1980s, he believed that his role was to act as an advocate for the Crown and to “make a case look good.” He explained that the perception originated, in some measure, from the *culture of advocacy* that he said prevailed at SickKids at the time.¹³¹

In terms of culture, courts, as the final arbiter for what is admitted into evidence, may play some role in establishing the ethical force ascribed to expert procedures. In other words, if judges take procedures like codes of conduct seriously and demand they be carefully followed, then lawyers will be more exacting in ensuring that their experts understand and follow the codes. The experts, in turn, may be more likely to attend to the specifics of the codes and take seriously their ethical qualities. Such a stance towards codes of conduct may also make experts see them as more legitimate forms of authority.¹³² By analogy, the CONSORT checklist may be effective because it is endorsed by a widely-respected organisation. This position contrasts to one in which codes of conduct are seen as pro forma by all actors.

Here, however, diverging from what we have prescribed, the trend has been for courts to take codes of conduct increasingly *less* seriously, perhaps robbing them of their ethical force. In the early days of the *NSW Code*, courts gave some consideration as whether to forgive non-compliance with the Code. Two early cases, for instance, suggested that failure to read and sign the Code could be cause to exclude experts, unless they knew of the Code’s provisions when forming their opinions but simply failed to follow it formally.¹³³ The onus appeared to be on the breaching party to justify non-compliance with the Code, and it appeared to be a heavy onus.¹³⁴

This position gave way to increasingly flexible views on the *NSW Code*’s role. Non-compliance could be forgiven, for instance, if there were objective differences between the views of the experts that the court could parse (for example, the experts were clearly relying on different data).¹³⁵ Another decision held that an expert who was not initially aware of the *NSW Code* could give testimony because, among other reasons, he ‘would not have changed his approach or opinion’ had he known of the Code.¹³⁶ He was also aware of a similar code in South Australia.¹³⁷ The same reasoning appeared in a contemporaneous decision, with the Court expressly stating it was moving from the earlier ‘strict compliance’¹³⁸ view of the *NSW Code* to one with more ‘leeway’.¹³⁹ Here, we note that numerous psychology studies have demonstrated that individuals (including scientists themselves) cannot examine their own unconscious biases, making it unlikely that the expert would

¹³¹ ‘Goudge Report’ (n 126) 16–17 (vol 1) (emphasis added).

¹³² Robbennolt and Sternlight (n 4) 1127–8.

¹³³ *Cassegrain* (n 2); *Barak Pty Ltd v WTH Pty Ltd* [2002] NSWSC 649, [4]–[5].

¹³⁴ *Cassegrain* (n 2) [4], [14].

¹³⁵ *Lopmand* (n 128) [16], [19].

¹³⁶ *Langbourne v State Rail Authority* [2003] NSWSC 537, [13].

¹³⁷ *Ibid.*

¹³⁸ *Jermen v Shell Co of Australia Ltd* [2003] NSWSC 1106, [32].

¹³⁹ *Ibid* [27].

know whether his or her opinion would be different if he or she had fully accept the applicable code.¹⁴⁰

Courts have also considered whether non-compliance with the *NSW Code* could result in exclusion under ss 135 and 137 of the *Evidence Act 1995* (NSW). One prominent appellate decision suggested exclusion could result from a ‘sufficiently grave breach’.¹⁴¹ This was, however, before the High Court of Australia’s decision in *IMM*, which directed judges to take evidence’s credibility and reliability at its highest when assessing its probative value (thus reducing the likelihood it *could* be excluded under the trial judge’s discretion).¹⁴² As noted above, while there is, as yet, no clear answer in the jurisprudence, it would appear that violation of the requirements in a code of conduct would be seen to reflect on either the reliability evidence, or the credibility of the witness (or both) and thus unavailable as a factor to be considered in the ss 135 and 137 balancing task.¹⁴³ We are only aware of one Australian decision (which was pre-*IMM*) to exclude an expert under the trial judge’s discretion (*UEL* or *Christie*) for failing to follow a code of conduct.¹⁴⁴

The most recent extensive analysis of the *NSW Code* is the NSW Court of Criminal Appeal’s recent decision in *Chen*.¹⁴⁵ The outcome in *Chen*, a drug trafficking case, hinged on the translations of phone transcripts between the accused and others in Fuqing, an under-described Chinese dialect.¹⁴⁶ These translations were provided by a Crown’s witness Ms Yang, who was an accredited interpreter in Mandarin.¹⁴⁷

Despite initially claiming that she had complied with the Code, Yang admitted on cross-examination that she had not in fact encountered the *NSW Code* prior to that cross-examination.¹⁴⁸ Critically, the most incriminating of her translations was the designation of a word spoken by the accused in an intercepted phone call, of the Fuqing word ‘la’ to ‘granule’, which was the particular form in

¹⁴⁰ For a review of research detailing one’s inability to explore one’s own biases, see Timothy D Wilson, *Strangers to Ourselves: Discovering the Adaptive Unconscious* (Belknap Press, 2002). See also Pronin, Yin and Ross (n 119).

¹⁴¹ *Wood v The Queen* (2012) 84 NSWLR 581, 620 [729] (‘Wood’). See also *Chen* (n 2) 923 [34]–[45].

¹⁴² *IMM* (n 18) 313 [44], 314 [47].

¹⁴³ Indeed, a Tasmanian court recently struggled to determine whether an expert’s reliance on an unknown source (which may have breached the *NSW Code*, if heard in that jurisdiction) could reduce the opinion’s probative value post-*IMM*. The Court noted that *IMM* seemed to indicate that the expert’s reliance on the unknown source could not impact its probative value, but that this seemed inconsistent with other parts of the majority’s opinion in *IMM: Langford* (n 20) 85–6 [56]. Ultimately, the Tasmanian court did not need to decide on this issue: at 86 [57].

¹⁴⁴ *Kyluk Pty Ltd v Chief Executive, Office of Environment and Heritage (NSW)* (2013) 298 ALR 532, 546–7 [62]–[68]. Canadian courts regularly exclude experts for failing to follow codes of conduct, see Jason M Chin & Rory McFadden, ‘Expert Witness Codes of Conduct for Forensic Practitioners: A Review and Proposal for Reform’ *Canadian Journal of Law and Justice* (forthcoming) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3505769>.

¹⁴⁵ *Chen* (n 2).

¹⁴⁶ *Ibid* 929 [80]. Fuqing is a tonal dialect related to the better described Fuzhou dialect, see ‘Dialect: Fuzhou’, *Glottolog 4.0* (Web Page catalogue of languages and families) <<https://glottolog.org/resource/languoid/id/fuzh1239>>.

¹⁴⁷ *Chen* (n 2) 926 [65].

¹⁴⁸ *Ibid* 919 [18].

which the pseudoephedrine (the supply of which was the subject of the charge) was produced.¹⁴⁹

The facts in *Chen* are such that it would have been expected that the witness called by the Crown as an expert would and should have been both aware of the Code and have formulated an opinion with a view to comply with it. It also appears to be the type of case in which adherence to a code of conduct would be of benefit if the aim is to prevent both partisanship and more subtle forms of cognitive bias. Indeed, there were many opportunities for bias to creep into the interpreter's judgment because she was involved in the case at an early stage, was present when a search warrant was executed, and was aware of the physical appearance of the drug in question.¹⁵⁰ Furthermore, translation is a subjective task in which early-stage exposure to biasing information may be particularly dangerous.¹⁵¹ Indeed, translation affords the translator wide discretion, especially in instances in which a word has no obvious equivalent.¹⁵² In *Chen*, it was common ground that 'la' has no direct English translation, with its appropriate translations not even limited to small-sized items.¹⁵³ Yang gave no explanation as to why 'granule' was 'the most verbatim translation of how the word "la" was used' in the context, rather than some other equally viable, connotatively neutral English equivalent.¹⁵⁴

Attending closely to a strongly worded code of conduct may have made Yang's ethical duties more salient to her and guided her interactions with the police and her approach to the translation. It may have also caused her to offer alternative translations of 'la' and disclose more uncertainty in her conclusion. As it was, an unusually robust challenge aimed at Yang's opinion and impartiality drew out many of these details.¹⁵⁵ Accused parties do not always (or typically) have access to such assistance.¹⁵⁶

Unfortunately, the appellate court did not engage with the *NSW Code's* role in regulating the knowledge proffered into court. The Court did not consider its

¹⁴⁹ Ibid 918–9 [14]. In its closing address, the Crown said: 'So what does it really come down to? It comes down to this word "la":' at 929 [80]. However, 'la' appears to be a classifier, that is to say a word with highly abstract meaning used in combination with nouns. As noted in *Chen*, the way that it would need to be translated or interpreted into English *would* vary according to context (at 924 [47]). But this does not justify, without more, the selection of the highly specific English word of granule.

¹⁵⁰ Ibid 925 [54].

¹⁵¹ This will especially be the case when the word being interpreted or translated is polysemous: see discussion in David Wayne Gilbert, 'Electronic Surveillance and Systemic Deficiencies in Language Capability: Implications for Australia's National Security' (PhD Thesis, RMIT University, 2014) ch 5 <<http://researchbank.rmit.edu.au/eserv/rmit:161277/Gilbert.pdf>>. See also, generally, Samuel Nunn, "'Wanna Still Nine-Hard?': Exploring Mechanisms of Bias in the Translation and Interpretation of Wiretap Conversations' (2010) 8(1) *Surveillance & Society* 28. On the effect of biasing information in listening more generally, see Helen Fraser and Bruce Stevenson, 'The Power and Persistence of Contextual Priming: More Risks in Using Police Transcripts to Aid Jurors' Perception of Poor Quality Covert Recordings' (2014) 18(3) *The International Journal of Evidence & Proof* 205.

¹⁵² Furthermore, translations are not susceptible to a single theory of qualitative assessment. See generally: Juliane House, *Translation Quality Assessment: A Model Revisited* (Gunter Narr Verlag, 1997).

¹⁵³ *Chen* (n 2) 929 [78]–[79].

¹⁵⁴ Ibid 929 [79].

¹⁵⁵ Ibid 925 [52]–[57].

¹⁵⁶ See Edmond and San Roque (n 6).

potential to control and reveal bias, or the importance of disclosing flexibility and uncertainty in the expert's process (for example, QRPs). Rather, the Court embarked in a lengthy statutory analysis that served to undermine, rather than strengthen, the role of the Code as an effective mechanism to regulate or restrain the type of evidence presented to the fact finder under the guise of expert opinion evidence.¹⁵⁷ The Court ultimately concluded that the *Supreme Court Act 1970* (NSW), unlike the *Evidence Act 1995* (NSW), was primarily concerned with procedure and practice and thus failure to follow the Code was not itself a matter of admissibility.¹⁵⁸ As to s 137 of the *Evidence Act 1995* (NSW), the Court did not seem to accept submissions that cognitive bias may have rendered Yang's translation of little probative value.¹⁵⁹ Instead, the Court noted that trial safeguards could ably handle any danger it posed.¹⁶⁰ This reasoning is simply not in step with the current scientific position that would suggest courts should establish a culture in which codes of conduct are serious ethical matters (by excluding the evidence of experts who breach them) and that broadly analogous 'codes' in science produce salutary effects.

B Reforming Expert Procedure

While judicial enforcement of codes of conduct is a logical first step in bringing them in line with empirical research, there is still work to do in reforming the specific provisions of those codes. From behavioural ethics, we saw that codes may be effective by making the moral and ethical component of the expert's job salient and thus avoiding ethical fading. Here, it is worth noting that the notion of self-concept maintenance has long (implicitly) been applied to the task of controlling fact witnesses through the use of oaths. This oath serves at least two purposes: to remind the witness of his or her internal standard for honesty, and to make salient the possibility of an external punishment for lying (that is, the rule against perjury).¹⁶¹

Matters of expert opinion are more complicated because they are naturally subject to more judgment and interpretation than factual recollections. As a result, sanctions against experts who provide problematic opinions have met resistance.¹⁶² Moreover, time pressures and the social-adversarial structure of the legal system may make expert witnesses especially susceptible to ethical fading — for example, rationalising behaviour as in the client's best interest or that it should be caught by the supervising lawyer.¹⁶³ That said, there may be considerable room to improve

¹⁵⁷ *Chen* (n 2) 919–23 [15]–[34].

¹⁵⁸ *Ibid* 923 [33].

¹⁵⁹ See *ibid* 924–28 [43]–[75]. Similarly, at the leave to appeal application before the High Court of Australia, the panel did not engage with submissions about cognitive bias: Transcript of Proceedings, *Chen v The Queen* [2018] HCATrans 240. *Chen* was subsequently followed in *Warwick* (n 13) to forgive an expert's failure to be aware of the *NSW Code*. In *Warwick*, the Court merely referred to *Chen*'s discussion of the jurisdiction of the *Supreme Court Act 1970* (NSW) and did not consider at all the effect of cognitive bias (let alone behavioural ethics and metascience): *Warwick* (n 13) [41]–[42].

¹⁶⁰ *Chen* (n 2) 928 [75].

¹⁶¹ Edmond, 'After Objectivity' (n 1) 140.

¹⁶² *Ibid* 148–9; NSWLRC (n 1) 158–9; Michell and Mandhane (n 3) 661. See *Wood v New South Wales* describing the difficulty in sufficiently associating the expert witness with the prosecution so as to support a claim of malicious prosecution, and in the potential application of expert witness immunity: [2018] NSWSC 1247, [206], [564], [568].

¹⁶³ Robbennolt and Sternlight (n 4) 1140–42.

expert testimony by engaging the expert's self-concept, and to do so through specific, carefully drafted codes of conduct. These may operate in much the same way that specific and demanding checklists have been shown to be effective in the mainstream sciences.¹⁶⁴

First, as with oaths, an expert code of conduct can serve to remind the expert of his or her duty to the court. It should interrupt the typical script and any favourable social comparisons that the expert may make (for example, by comparing him or herself to an unscrupulous expert). Similarly, the codes should make it clear that while there are safeguards in place, the expert is solely responsible for the content of his or her opinion (to avoid diffusion of responsibility).

Further, codes should be drafted in a way that makes self-justification as difficult as possible. Simply reminding an expert of a flexible and amorphous duty may not be very effective because those acts easily give way to rationalisation (a strong predictor of straying from strict ethical duties).¹⁶⁵ In the legal sphere, expert witnesses have a great deal of flexibility in reporting their results, allowing them to portray them in a misleading way.¹⁶⁶ Reforms in Australia have done little to curb this flexibility.¹⁶⁷

So, how do the current codes of conduct stack up in light of this research? In short, they show some promise, but there is still much to improve on. For instance, consider cl 3(i) of the *NSW Code*:

a declaration that the expert has made all the inquiries which the expert *believes are desirable and appropriate* (save for any matters identified explicitly in the report), and that *no matters of significance* which the *expert regards as relevant* have, *to the knowledge of the expert*, been withheld from the court¹⁶⁸

This item is on the right track in reminding the expert of his or her duty to make inquiries and report relevant findings. But, it still provides several avenues for rationalisation. It is not hard for an expert in an area to construct reasons why a line of research that may have called into question the client's preferred outcome was not 'desirable' or 'appropriate'. For example, forensic practitioners sometimes fail to discuss parts of reports that cast doubt on their methods.¹⁶⁹ Moreover, those accustomed to the adversarial trial may be particularly adept at developing such rationalisations.¹⁷⁰

On the other hand, consider a construction we devised that more directly engages the expert's personal ethics and removes some chances for rationalisation: 'I [the expert hand-writes his or her name here], have conducted and reported all

¹⁶⁴ Turner et al (n 11); Han et al (n 88).

¹⁶⁵ Belle and Cantarelli (n 109).

¹⁶⁶ NAS Report (n 126) 185–6.

¹⁶⁷ Edmond, Martire and San Roque, 'Expert Reports' (n 22) 604–19.

¹⁶⁸ (emphasis added). The *Victorian Code* (n 30) 6.1(h) provision is nearly identical and reads: 'a declaration that the expert has made all the inquiries and considered all the issues which the expert believes are desirable and appropriate, and that no matters of significance which the expert regards as relevant have, to the knowledge of the expert, been withheld'.

¹⁶⁹ Searston and Chin (n 14).

¹⁷⁰ Adam Benforado, *Unfair: The New Science of Criminal Injustice* (Broadway Books, 2016) 85.

inquiries that the strongest critic of my opinion would make. ___ [initial here].’ This revamped provision removes some ambiguity in the original’s language. It also removes some subjectivity as to how to deem what is relevant, moving it from solely the expert’s judgment to that of the strongest critic.¹⁷¹ The latter helps avoid the ethical fading technique of comparing oneself to a less diligent person. Alternatively, coder reformers may prefer simply a ‘strong’ critic — the key is to move towards a more objective observer.

Another problem with the *NSW Code* is that items like cl 3(i) are also part of a long list of ‘requirements’ that can easily be glossed over and thus lose their ethical salience (especially for repeat players). For that reason, the expert should not just provide an omnibus signature confirming adherence with the Code (thus permitting various rationalisations, such as it *generally* being complied with), but should actively confront each aspect of the Code. For that reason, experts should write their initials at each step.¹⁷²

Metascientific reforms like checklists and preregistration are also confronting one of the prime QRPs that produce false positive findings — conducting tests and making observations, but only reporting them if they are consistent with the researchers’ interests.¹⁷³ On this point, the CONSORT checklist has been effective in encouraging researchers to distinguish between planned and unplanned analyses so that it will be clearer when they have tried various analyses to find the one that gets a publishable result.¹⁷⁴

Current codes of conduct are not well-designed to avoid post-hoc framing of the facts on which experts base their opinions. For instance, the *NSW Code* cl 3(g) requires that an expert report contain:

any examinations, tests or other investigations on which the **expert has relied**, identifying the person who carried them out and that person’s qualifications¹⁷⁵

While this item seems admirably aimed at encouraging the expert to report the source of his or her opinion, it is deficient. It allows the expert to rationalise away non-reporting of examinations, tests, or investigations that they conducted or were aware of, but on which they did not directly rely.

In light of the social science reviewed in Part IIIA above, cl 3(g) of the *NSW Code* might be amended to:

I [the expert hand-writes his or her name here] have reported all examinations, tests, or other investigations conducted by myself or others since I was contacted to provide an opinion, despite the outcomes of those inquiries.
___ [initial here]

¹⁷¹ The critical reviewer standard may also be used for other items that advert to the expert’s judgement about what ought to be reported, such as cls 3(j) and (k) of the *NSW Code* (n 33).

¹⁷² We acknowledge that simply initialling each item provides a minimal safeguard. However, we think that, in conjunction with other reforms aimed at engaging the expert’s ethics, initialling can help.

¹⁷³ Munafò (n 4) 4–5.

¹⁷⁴ Turner et al (n 11) 65. The CONSORT Checklist asks scientists to report ‘[r]esults of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory’: CONSORT (n 85).

¹⁷⁵ *NSW Code* (n 33) cl 3(g) (emphasis added). See cl 6.1(g) of the *Victorian Code* (n 30) for its analogous provision.

This construction not only makes salient the requirement of reporting tests, but it removes the flexibility of reporting only those tests that the expert ‘relied’ on. It demands, very clearly, that the expert report not just those tests that support the opinion, but those that might not.

While asking expert witnesses to decisively attest that they have reported all tests (and controversies) is an improvement, scientists are now acknowledging that such attestations may not always go far enough in curbing their biases. Rather, as we discussed above, it is very easy for researchers to deceive themselves into thinking that the conclusion they came to was unavoidable (that is, hindsight bias), thus convincing themselves that it is perfectly ethical to not report the weaknesses in their design, analysis, and data.¹⁷⁶ Moreover, they may simply ‘forget the details’ of the study that supported an alternative hypothesis.¹⁷⁷ In other words, experts, like scientists, generally understand the importance of impartiality but may require constraints on their reasoning to fulfil that aim: ‘The values of impartiality and objectivity are pervasive, particularly for scientists, but human reasoning is not reliably impartial or objective.’¹⁷⁸

Following from the trend in science, courts may wish to supplement codes of conduct with elements of preregistration to help reveal and control post-hoc rationalisations of analytic choices. In other words, experts may be asked — before they are exposed to the facts of a case — to specify how they will go developing their opinion. In the civil sphere, this might involve experts explaining how they go about valuing real estate before they are told the precise property they are valuing.¹⁷⁹ This would prevent them from choosing, after the fact, the methodology that leads to their preferred outcome and then rationalising away (or forgetting) reasons for applying a different methodology. We note that it may be perfectly appropriate for experts to shift methodologies in some circumstances. Preregistration (likely at an early case conference) would at least create a record of the initial choice and compel the expert to explain the decision to change.

In the criminal sphere, many choices that forensic scientists make have been criticised for being overly fact-driven and post-hoc.¹⁸⁰ For example, one forensic scientist who gave evidence in the trial of Jeffrey Gilham, an acknowledged wrongful conviction, applied a controversial methodology to find a pattern of stab

¹⁷⁶ Nosek et al, ‘Preregistration Revolution’ (n 77) 2600–1.

¹⁷⁷ Brian A Nosek, Jeffrey R Spies and Matt Motyl, ‘Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth over Publishability’ (2012) *Perspectives on Psychological Science* 7(6) 615, 617.

¹⁷⁸ Nosek et al, ‘Preregistration Revolution’ (n 77) 2601.

¹⁷⁹ For example, the discounted cash flow (‘DCF’) approach to valuation is often viewed suspiciously by courts for being outside the mainstream, but there may be reasons it is preferable in some cases: see *Downie v Sorell Council* (2005) 141 LGERA 304. If an expert were to state a preference for the DCF before knowing the specific property in question, then this may provide reason to think it was not chosen simply because it was beneficial to the client’s case. Similarly, battles are often fought in valuation cases over what properties are comparable to the instant property: see *Warren v Lawton (No 3)* [2016] WASC 285, [50]–[51]. If experts were to say before they see the property how they choose comparator properties, then this may also shed light on the reliability of their analysis.

¹⁸⁰ See ‘PCAST Report’ (n 126) 10; Jason M Chin and D’Arcy White, ‘Forensic Bitmark Identification Evidence in Canada’ (2019) 52(1) *University of British Columbia Law Review* 57.

wounds across the bodies of two deceased individuals.¹⁸¹ This suggested a common assailant. In short, the forensic scientist found that if she focused on one subset of the wounds (those made after the deceased parties were disabled) and disregarded the others, there seemed to be a pattern.¹⁸² The prosecution relied considerably on this conclusion.¹⁸³

The forensic practitioner's process and conclusions were sharply criticised by another forensic scientist¹⁸⁴ and largely dismissed by the NSW Court of Criminal Appeal.¹⁸⁵ She was interpreting data (stab wounds) that were highly uncertain and varied, and was attempting to identify a pattern within that noise. Such situations are ripe for apophenia (that is, seeing structure in randomness), which she may have succumbed to by focusing her opinion on the subset of the wounds that showed some similarity. However, if she had pre-specified (before seeing the wounds) what constitutes a pattern and what does not, it would have been easier for the jury to assess the probative force of her opinion at the first instance.¹⁸⁶

Gilham is not uncommon; experts often must conduct analyses in an ad hoc manner. For instance, in *Wood v The Queen*, an expert had a woman thrown repeatedly into a swimming pool, in an effort to support the prosecution's contention that a deceased could have been thrown to her death by the accused (alone or in the company of another).¹⁸⁷ These types of experiments may benefit from measures such as preregistration and code of conduct compliance to encourage not only those results that favour the expert's theory are reported.

In contrast to the *NSW Code*, the *Victorian Code* is more implicitly attuned to the behavioural ethics of providing expert opinions, especially in requiring disclosure of controversies in the field:

Where an expert is aware of any *significant and recognised disagreement or controversy* within the relevant field of specialised knowledge, which is *directly relevant* to the expert's ability, technique or opinion, the expert *must disclose* the existence of that disagreement or controversy.¹⁸⁸

This item has the right idea in demanding that experts disclose disagreements or controversies. Its language does, however, provide some ethical leeway in providing the disagreement must be 'directly' relevant to the opinion and 'significant and recognised'.

¹⁸¹ *Gilham v The Queen* [2012] NSWCCA 131, [280] ('*Gilham*'). See generally Gary Edmond, David Hamer and Andrew Ligertwood, 'Expert Evidence after *Morgan*, *Wood* and *Gilham*' (September/October 2012) 112 *Precedent* 28.

¹⁸² *Gilham* (n 181) [281].

¹⁸³ *Ibid* [248]–[250].

¹⁸⁴ *Ibid* [351]–[370].

¹⁸⁵ *Ibid* [346].

¹⁸⁶ See Mehera San Roque, "'A Woman Like You': Gender, Uncertainty and Expert Opinion Evidence in the Contemporary Criminal' (2013) 3(2) *feminists@law* 1, 6. It is important to note here that the evidence presented by the different experts in this case was framed by questionable Crown tactics including a failure to call a material expert witness who had provided an opinion to the Crown that contradicted the evidence led at trial as to the significance of the 'pattern'.

¹⁸⁷ *Wood* (n 141) 590 [45].

¹⁸⁸ *Victorian Code* (n 30) 6.2 (emphasis added).

Finally, revelations from metascience and behavioural ethics reinforce existing worries that court-appointed experts and concurrent procedures may not be as useful as once thought. Chiefly, they bring all of their field's norms and practices — many of which can be questionable — into the courtroom. As we saw above, peer-review, which is something of an analogue for experts testifying concurrently (with, of course, many key differences), has not worked as well as it could have in many fields. In particular, there is vast publication bias where null findings are much less likely to be published. Moreover, several fields are finding that their false discovery rates are higher than expected, and perhaps close to 50%.¹⁸⁹ With experts concurrently reviewing each other's work based on deficient field standards, there is little reason to think this procedure will help very much. Similarly, court-appointed experts may be unbiased in that they are not paid or selected by a party, but they may still use the questionable practices of their field.¹⁹⁰ None of this is to say that these practices are useless. For instance, in many cases it will be likely that a court-appointed expert will be subjected to fewer adversarial pressures than those that are party-appointed.

That said, the metascientific reforms going on in the mainstream sciences should lend both some optimism and some tangible guidance. For example, checklists that have found some demonstrable success may be used alongside court-appointed experts to encourage them to go beyond their field's standards and present their findings with the appropriate cautions and levels of uncertainty. Here we note that Edmond and colleagues have discussed the NSW Police's revised template for preparing their expert reports, which is an effort to better comply with the *NSW Code*.¹⁹¹ While this new template still understates the error in the field, it appears to be an improvement over practitioners' previous way of providing evidence.¹⁹² Similar industry-specific templates that correspond more closely to checklists like CONSORT may represent important and empirically-justified reforms to expert procedure.

V Conclusions and the Limits of Procedure

In this article, we have made a case for procedure. Specifically, we think that procedure that is designed with reference to effective scientific procedural safeguards and that is enthusiastically enforced by courts may provide serious benefits to the trial process. Now, we should mitigate our own conclusions.

First, we do not mean to suggest that expert procedure is a silver bullet in ensuring factual rectitude. Yes, procedure may encourage experts to be transparent about the weaknesses of their methods (for example, codes of conduct) and control some biases that were not previously controlled (for example, preregistration). But

¹⁸⁹ See the sources at n 59 above.

¹⁹⁰ Edmond, 'After Objectivity' (n 1) 153–4.

¹⁹¹ Edmond, Martire and San Roque, 'Expert Reports' (n 22) 604.

¹⁹² Ibid: 'The first thing to say about the Revised Certificate template is positive'.

even an expert who discloses some of his or her opinion's weaknesses may still strongly hold to that opinion in a way that is persuasive to the factfinder.¹⁹³

Second, the effectiveness of codes can be undermined by failures of counsel to take advantage of the tools offered, as well as countervailing procedural reforms that may discourage comprehensive reporting.¹⁹⁴ These countervailing trends create barriers and disincentives for parties, and anecdotally raise the possibility of higher costs being imposed on defendants and their representatives who request more comprehensive reporting.¹⁹⁵

Third, there is the applied shortcoming that we saw in *Chen* and its progeny whereby, despite the critical outward posture adopted towards the partisan expert, courts are reluctant to enforce codes of conduct (and this often occurs in the criminal context).¹⁹⁶ Indeed, beyond the codes of conduct cases we discussed above, Australian courts have been careful to note that even serious questions of bias go to the weight, rather than the admissibility, of the evidence.¹⁹⁷ Similarly, expert procedures can be circumvented when witnesses are characterised as providing lay opinion or 'ad hoc' expert opinion based on repeated exposure to case-specific facts.¹⁹⁸

Against this sobering backdrop, we want to argue nonetheless that we should not be vacating the procedural field. Rather — especially given weakened admissibility rules and trial safeguards that do not demonstrably work — there is a place for procedure to fill the gaps. Science, our culture's chief means of answering factual questions, is undergoing serious changes, with many of these changes being procedural in nature. Law should be aware of this revolution within science and take advantage of the research it has produced — research that reinforces the importance of taking procedure seriously.

¹⁹³ Consider, for instance, the High Court's recent decision in *Lee v Lee* (2019) 372 ALR 383 ('*Lee*'). In *Lee*, one expert admitted that bloodstain pattern analysis is a 'notoriously inexact science': at 391 [34]. Still, that same expert refused to resile from her position when made aware of evidence casting serious doubt on her opinion. The High Court ultimately described her opinion as 'to a high degree, improbable': at 398 [63].

¹⁹⁴ See, eg, discussion in Carol McCartney, 'Streamlined Forensic Reporting: Rhetoric and Reality' (2019) 1 *Forensic Science International: Synergy* 83; Gary Edmond, Sophie Carr and Emma Piasecki, 'Science Friction: Streamlined Forensic Reporting, Reliability and Justice' (2018) 38(4) *Oxford Journal of Legal Studies* 764.

¹⁹⁵ In NSW, there is a new 'Early Appropriate Guilty Pleas' regime that includes a 'short form' of forensic reporting, which, like the Streamlined Forensic Reporting regime in the UK, significantly reduces the level of detail initially provided in relation to forensic evidence: see NSWLRC, *Encouraging Appropriate Early Guilty Pleas* (Report 141, December 2014) and Legal Aid of New South Wales, 'Early Appropriate Guilty Pleas' (website) <<https://www.legalaid.nsw.gov.au/for-lawyers/resources-and-tools/early-appropriate-guilty-pleas>>. The impact of the NSW regime has yet to be evaluated.

¹⁹⁶ *Chen* (n 2); *Warwick* (n 13); Edmond and Martire, 'Just Cognition' (n 10).

¹⁹⁷ *Fagenblat* (n 21) [7]; *Flavel v South Australia* (2007) 96 SASR 505, 523 [102]; *Haoui v The Queen* (2008) 188 A Crim R 331, 354 [127].

¹⁹⁸ See Gary Edmond, Kristy Martire and Mehera San Roque 'Unsound Law: Issues with ('Expert') Voice Comparison Evidence' (2011) 35(1) *Melbourne University Law Review* 52; *Nguyen v The Queen* (2017) 264 A Crim R 405, 415–17 [41]–[51].