

# RETHINKING EXPERT OPINION EVIDENCE

KRISTY A MARTIRE\* AND  
GARY EDMOND†

*This article approaches expert opinion evidence from a scientific, specifically cognitive science, perspective. Decades of scientific research on expertise presents a picture of expertise that bears limited resemblance to the categories and practices used by legal institutions to regulate the admission, presentation and evaluation of expert evidence (ie, opinions based on specialised knowledge). This article seeks to explain why legal institutions should direct more attention to scientifically-based criteria and insights, rather than the somewhat crude set of legal proxies developed by common law judges, if they hope to credibly regulate forensic science and medicine evidence in ways that enhance factual rectitude and fairness.*

## CONTENTS

I	Introduction: Expert Performance .....	968
II	Regulating the Admission of Expert Opinion Evidence .....	969
II	Scientific Approaches to Expertise .....	971
	A Evaluating Expertise: Validity, Reliability, Limitations .....	982
IV	Legal Approaches to Evaluating Expertise .....	983
	A Using (Legal) Proxies Instead of an Expert Performance Criterion ....	984
	B Assessing Expert Performance Data.....	987
	C Efficacy of Trial Safeguards and Appeals .....	989
V	Remediating Our Jurisprudence.....	991
	A Concurrent Evidence and Recent Procedural Reform .....	995
VI	Conclusion .....	996

\* BA (Syd), MPsych, PhD (UNSW); Senior Lecturer and Australian Research Council DECRA Fellow, School of Psychology, UNSW Sydney. This research was supported by the following Australian Research Council grants: DE140100183 (Martire) and LP160100008 (Martire and Edmond).

† BA (Hons) (Wollongong), LLB (Hons) (Syd), PhD (Cantab); Professor, School of Law, UNSW Sydney; Research Professor (Fractional), Northumbria Law School, Northumbria University; Chair, Evidence-Based Forensics Initiative.

## I INTRODUCTION: EXPERT PERFORMANCE

Decades of study have revealed the hallmarks of expertise and the conditions under which it can be acquired. Within the relevant scientific literature there are two distinct intellectual strands: heuristics and biases ('HB') and natural decision-making ('NDM'). While the opinions of these 'schools' differ in some important respects, both consider objectively successful performance of domain specific tasks as a requirement for the assignment of 'expert' status.<sup>1</sup> In this article, we explore whether legal approaches to the assessment and evaluation of expertise pay sufficient attention to this key characteristic, particularly in response to opinions proffered by forensic scientists.

In Part II, we briefly summarise the rules regulating the admissibility of expert opinion evidence in Australia.<sup>2</sup> In Part III, we review scientific literatures pertaining to expertise — both the conditions leading to its acquisition as well as its characteristics.<sup>3</sup> In doing so, we identify indicia that lawyers and judges should consider when endeavouring to identify and evaluate expertise. In Part IV, we analyse legal engagement with experts and expertise, highlighting the fundamental disconnect between scientific and legal approaches and, in the process, explain some of the limitations inherent in traditional legal practice. Part V explains why legal rules, procedures and safeguards, with the potential to remediate the reception of expert opinion evidence, have not operated as expected. Finally, in the Conclusion, we argue that demonstrable ability, under the guise of *reliability*, should be given explicit attention, and that courts will benefit from an approach to expertise that engages with scientific research and evidence-based indicia.<sup>4</sup>

<sup>1</sup> See Daniel Kahneman and Gary Klein, 'Conditions for Intuitive Expertise: A Failure to Disagree' (2009) 64 *American Psychologist* 515.

<sup>2</sup> These are treated in more detail elsewhere: see Gary Edmond, 'The Admissibility of Forensic Science and Medicine Evidence under the Uniform Evidence Law' (2014) 38 *Criminal Law Journal* 136; Gary Edmond, 'Specialised Knowledge, the Exclusionary Discretions and Reliability: Reassessing Incriminating Expert Opinion Evidence' (2008) 31 *University of New South Wales Law Journal* 1.

<sup>3</sup> We leave the not entirely unrelated question of judicial expertise for another occasion.

<sup>4</sup> We appreciate that the High Court seems to have rejected recourse to 'reliability' in relation to s 137 (and implicitly s 79) of the uniform evidence legislation in *IMM v The Queen* (2016) 257 CLR 300, 306 [16]–[17], 314 [48] (French CJ, Kiefel, Bell and Keane JJ) ('*IMM*'). For reasons this article makes clear, inattention to reliability (and validity) and proficiency represents an undesirable response to both the identification and evaluation of scientific, medical and technical evidence, particularly forensic science evidence. For reasons made clear below, rational endeavours to gauge the probative value of opinions based on specialised knowledge (including 'at its highest') require courts to engage with evidence of validity, reliability and demonstrable performance. See Gary Edmond, 'Icarus and the *Evidence Act*: Section 137,

## II REGULATING THE ADMISSION OF EXPERT OPINION EVIDENCE

Historically, common law judges regulated expertise on the basis of relatively superficial, though fairly straightforward, considerations such as whether a potential witness possessed formal qualifications (eg, as a doctor or engineer) or appropriate training or experience in a legally-recognisable profession or ‘field’, whether the evidence appeared pertinent to facts in dispute, and was considered (by the trial judge to be) capable of assisting the tribunal of fact.<sup>5</sup> From the 19<sup>th</sup> century, in conjunction with greater specialisation and emerging legal apprehension about bias and partisanship (especially in civil proceedings), the ‘qualification’ and ‘field’ requirements were often supplemented.<sup>6</sup> When admitting and evaluating expert opinion evidence, courts often refer — in ways that are neither consistent nor principled — to a range of additional considerations in their attempts to identify and evaluate expert opinions. Supplementary factors include: how long a procedure has been in use;<sup>7</sup> whether the procedure or witness had previously been admitted (or even used in a criminal investigation);<sup>8</sup> whether the evidence was developed for legal proceedings;<sup>9</sup> whether the evidence is admissible in other jurisdictions;<sup>10</sup>

Probative Value and Taking Forensic Science Evidence “at Its Highest” (2017) 41 *Melbourne University Law Review* (forthcoming).

<sup>5</sup> See, eg, *Buckley v Rice Thomas* (1554) 1 Plowden 118, 124–5; 75 ER 182, 192–3 (Saunders J); *Folkes v Chadd* (1782) 3 Dougl 157; 99 ER 589. For modern common law manifestations of this practice, see *R v Turner* [1975] 1 QB 834; *Clark v Ryan* (1960) 103 CLR 486.

<sup>6</sup> See Tal Golan, *Laws of Men and Laws of Nature: The History of Scientific Expert Testimony in England and America* (Harvard University Press, 2004).

<sup>7</sup> *JP v DPP (NSW)* [2015] NSWSC 1669 (11 November 2015) [35] (Beech-Jones J), quoting *R v Tang* (2006) 65 NSWLR 681, 713 [144] (Spigelman CJ) (*‘Tang’*); *Dasreef Pty Ltd v Hawchar* (2011) 243 CLR 588, 604 [39] (French CJ, Gummow, Hayne, Crennan, Kiefel and Bell JJ) (*‘Dasreef’*).

<sup>8</sup> *Honeysett v The Queen* (2013) 233 A Crim R 152 (*‘Honeysett’*).

<sup>9</sup> A classic example from the United States is *Daubert v Merrell Dow Pharmaceuticals Inc*, 43 F 3d 1311 (9<sup>th</sup> Cir, 1995) (*‘Daubert’*). See also Peter W Huber, *Galileo’s Revenge: Junk Science in the Courtroom* (Basic Books, 1991). Interestingly, in recent months, the appellate judge who doubted the need to apply the *Daubert* criteria and consider whether the evidence was prepared for litigation to the forensic sciences seems to have experienced an epiphany: Alex Kozinski, ‘Rejecting Voodoo Science in the Courtroom’, *The Wall Street Journal* (online), 20 September 2016 <<http://www.wsj.com/articles/rejecting-voodoo-science-in-the-courtroom-1474328199>>.

<sup>10</sup> Consider the Canadian engagement with forensic gait analysis in England: *R v Aitken* [2012] BCCA 134 (2 April 2012), quoting *Otway v The Queen* [2011] EWCA Crim 3 (14 January 2011).

the plausibility of the claim;<sup>11</sup> whether opposing parties (especially in criminal proceedings) have access to expert advice;<sup>12</sup> the availability of alternatives;<sup>13</sup> and more recently, whether the witness complied with the terms of expert witness codes of conduct and practice directions.<sup>14</sup> Such supplementary considerations are more commonly used to inform the assessment of probative value (or weight) than admissibility, per se.<sup>15</sup>

After the introduction of the uniform *Evidence Acts* beginning in 1995,<sup>16</sup> the admissibility of expert opinion evidence in most Australian jurisdictions came to be regulated by s 79 rather than the common law. An exception to the exclusionary opinion rule,<sup>17</sup> s 79(1) of the uniform *Evidence Acts* provides for the admission of opinions based on ‘specialised knowledge’:

If a person has specialised knowledge based on the person’s training, study or experience, the opinion rule does not apply to evidence of an opinion of that person that is wholly or substantially based on that knowledge.

Section 79(1) has been authoritatively interpreted as a two-part test:

Section 79(1) states two conditions of admissibility: first, the witness must have ‘specialised knowledge based on the person’s training, study or experience’ and,

<sup>11</sup> Consider the stab wound evidence in *Gilham v The Queen* (2012) 224 A Crim R 22, 38 [152]–[153]. See also the voice identification and comparison evidence as ad hoc expertise in *R v Leung* (1999) 47 NSWLR 405; *Li v The Queen* (2003) 139 A Crim R 281.

<sup>12</sup> *R v Jung* [2006] NSWSC 658 (29 June 2006). See also *R v Madigan* [2005] NSWCCA 170 (9 June 2005) for judicial reluctance to recognise expert evidence adduced by the defendant.

<sup>13</sup> *Tang* (2006) 65 NSWLR 681, 709 [120] (Spigelman CJ), where the alternative would have been leaving the images for the jury.

<sup>14</sup> *Wood v The Queen* (2012) 84 NSWLR 581, 619–20 [728] (McClellan CJ at CL), discussing the *Uniform Civil Procedure Rules 2005* (NSW) r 31.23, sch 7 (*‘Expert Witness Code of Conduct’*). For the most elaborate practice direction in Australia, see the Supreme Court of Victoria, *Practice Note No 2 — Expert Evidence in Criminal Trials*, 25 June 2014.

<sup>15</sup> For a more detailed review, see Gary Edmond, ‘Legal versus Non-Legal Approaches to Forensic Science Evidence’ (2016) 20 *International Journal of Evidence and Proof* 3.

<sup>16</sup> The uniform *Evidence Acts* comprise seven Australian statutes: *Evidence Act 1995* (Cth); *Evidence Act 2011* (ACT); *Evidence Act 1995* (NSW); *Evidence Act 2004* (Norfolk Island); *Evidence (National Uniform Legislation) Act 2011* (NT); *Evidence Act 2001* (Tas); *Evidence Act 2008* (Vic). They are substantially similar to the *Evidence Act 1995* (Cth); however, not entirely identical. Queensland, South Australia and Western Australia have not adopted the uniform legislation. To prevent confusion, we refer to the Commonwealth legislation when citing the uniform *Evidence Acts*.

<sup>17</sup> Uniform *Evidence Acts* s 76(1) states: ‘[e]vidence of an opinion is not admissible to prove the existence of a fact about the existence of which the opinion was expressed.’

secondly, the opinion must be ‘wholly or substantially based on that knowledge’.<sup>18</sup>

Despite the differences between the wording of s 79(1) and common law concepts, the conspicuous omission of ‘expert’ and ‘field’ and the fresh emphasis on ‘knowledge’, most Australian courts have not dramatically altered their admissibility practice since the introduction of the uniform *Evidence Acts*.<sup>19</sup> Some even continue to endorse obscure common law concepts such as ad hoc expertise — notwithstanding the conspicuous absence of ‘knowledge’.<sup>20</sup> Our concerns in this article are primarily oriented to the admission and evaluation of scientific, medical and other types of technical expertise, with a particular emphasis on the assessment of the opinions of forensic scientists.

### III SCIENTIFIC APPROACHES TO EXPERTISE

Neither the common law nor the jurisprudence and practice that has emerged around s 79(1) require an expert witness (or the party calling them) to demonstrate that the ‘training, study or experience’, or any resultant ‘specialised knowledge’, manifest in the witness displaying measurably *superior performance* in the relevant domain.<sup>21</sup> Rather, courts tend to assume that training, study or experience begets specialised knowledge and that this

<sup>18</sup> *Honeysett* (2014) 253 CLR 122, 131 [23]. See also *Dasreef* (2011) 243 CLR 588, 602–3 [32] (French CJ, Gummow, Hayne, Crennan, Kiefel and Bell JJ); *HG v The Queen* (1999) 197 CLR 414, 427 [38]–[39] (Gleeson CJ).

<sup>19</sup> Cf *Dasreef* (2011) 243 CLR 588, 604 [37] (French CJ, Gummow, Hayne, Crennan, Kiefel and Bell JJ):

The admissibility of opinion evidence is to be determined by application of the requirements of the *Evidence Act* rather than by any attempt to parse and analyse particular statements in decided cases divorced from the context in which those statements were made.

See generally Australian Law Reform Commission, *Uniform Evidence Law*, Report No 102 (2005).

<sup>20</sup> See *Nguyen v The Queen* [2017] NSWCCA 4 (2 February 2017). Cf Gary Edmond, Kristy Martire and Mehera San Roque, ‘Unsound Law: Issues with (“Expert”) Voice Comparison Evidence’ (2011) 35 *Melbourne University Law Review* 52; Gary Edmond and Mehera San Roque, ‘Quasi-Justice: Ad Hoc Expertise and Identification Evidence’ (2009) 33 *Criminal Law Journal* 8.

<sup>21</sup> The need for ‘reliability’ was explicitly rejected in *Tang* (2006) 65 NSWLR 681, 712 [137] (Spigelman CJ); *Tuite v The Queen* [2015] VSCA 148 (12 June 2015) [10] (‘*Tuite*’); *IMM* (2016) 257 CLR 300, 314 [48] (French CJ, Kiefel, Bell and Keane JJ). But see *Honeysett* (2014) 253 CLR 122, 136–7 [38]–[42].

qualifies the proposed witness as an 'expert'.<sup>22</sup> By contrast, scientific approaches focus on an actual ability, along with independent evidence of that ability, as integral requirements for assigning expert status.<sup>23</sup> Where there is no demonstrably superior performance, there is no expert.

Cognitive scientists (and others) have been studying the acquisition of expertise, along with indicia of expert performance, across a diverse range of activities for more than 50 years. Beginning with studies of chess grand masters<sup>24</sup> and clinical psychologists<sup>25</sup> in the 1940s and 1950s, two schools of thought gradually coalesced.<sup>26</sup> The first, the HB approach, focuses on factors that limit the 'expertise' of practitioner judgements. The other, an approach based around NDM, highlights the role of intuition in facilitating 'expert' outcomes in complex decision environments.<sup>27</sup> Notwithstanding a variety of differences between these two intellectual communities, their respective definitions of 'expertise' share a common component. Both require demonstrably successful performance on tasks specific to the expert domain.<sup>28</sup>

In the NDM tradition, this history of successful outcomes is usually assessed by peer evaluation against a relative standard of success. Inherent in the

<sup>22</sup> *Dasreef* (2011) 243 CLR 588, 604 [37] (French CJ, Gummow, Hayne, Crennan, Kiefel and Bell JJ).

<sup>23</sup> In order to determine whether a person is a good archer, for example, we would want to see them shoot a number of arrows, on target, on a standard range. Ribbons, trophies and even Olympic medals might be used as proxies for performance, perhaps even very informative proxies of performance (at some particular stage or stages). 'Proxies' operate as particularly good evidence only when they provide a direct indication of performance relative to others (ie, through competitions) or some objective standard (eg, proximity of the arrow in relation to the bullseye). However, where the proxy is membership of an archery club, or perhaps even being an office bearer in an archery club, or even selector, we cannot assume that the individual is a better archer than non-members. Further, depending on the specific activity, performance might improve, deteriorate or remain reasonably stable over time. A gold medal for archery at the 1984 Los Angeles Olympics might not reveal very much about current ability as an archer. To assess post-Olympic ability would require more recent evidence of performance. Additionally, expertise as an archer reveals little about abilities in other domains. Generations of ancient Cretan archers confirm that expertise in archery reveals nothing about the ability to accurately fire a rifle at a target.

<sup>24</sup> Adriaan D de Groot, *Thought and Choice in Chess* (Mouton Publishers, 2<sup>nd</sup> ed, 1978).

<sup>25</sup> Paul E Meehl, *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence* (University of Minnesota Press, 1954).

<sup>26</sup> Other intellectual traditions in this area include the 'traditional' and 'expert-performance' approaches: see K Anders Ericsson and Tyler J Towne, 'Expertise' (2010) 1 *Wiley Interdisciplinary Reviews: Cognitive Science* 404. These schools roughly correspond to the NDM and HB traditions (respectively) with regard to relative versus objective definitions of expertise.

<sup>27</sup> For an authoritative review, see Kahneman and Klein, above n 1.

<sup>28</sup> *Ibid.*

NDM approach is an emphasis on defining ‘expertise’ in relation to ‘the field’.<sup>29</sup> Those who are accepted as the highest relative performers are considered expert even though their success rate may not be objectively high.<sup>30</sup> For example, experts may be ‘operationally defined as those who have been recognized within their profession as having the necessary skills and abilities to perform at the highest level.’<sup>31</sup>

In the HB tradition, successful outcomes are defined against an optimality criterion, such as objective (or ground) truth.<sup>32</sup> This is a significantly more demanding criterion whereby expert status is afforded only where the performance of the decision maker is optimal, or better than mathematical models, decision rules or algorithms. For example, Herling defines human expertise as:

displayed behaviour within a specialized domain and/or related domain in the form of consistently demonstrated actions of an individual that are both optimally efficient in their execution and effective in their results.<sup>33</sup>

Neither definition affords expert status to individuals or opinions based on the mere possession of knowledge derived from some form of training, study

<sup>29</sup> We tend to use the terms field, domain, discipline and profession interchangeably. Our concern is not with social or professional classification or recognition, but rather with the ability of individuals to do specific tasks.

<sup>30</sup> The question of whether performance is sufficient to warrant the admission of expert opinion into a legal proceeding is a policy question for judges. Self-evidently, where the level of performance is not much above the performance of ordinary persons, there are numerous dangers in admitting an individual into a legal proceeding, particularly a criminal proceeding, and conferring the (somewhat specious) attribution ‘expert’. There are risks and costs associated with expert opinion evidence, and these should be considered before evidence is admitted. Limited defence resourcing and the ineffectiveness of conventional legal safeguards should also inform admissibility decision-making. For example, these dangers are demonstrated by proponents of emerging areas of legally recognised ‘expertise’, such as forensic gait comparison; often suggesting that slightly enhanced performance over novices ought to provide them with access to courtrooms. See, eg, Ivan Birch et al, ‘The Identification of Individuals by Observational Gait Analysis Using Closed Circuit Television Footage’ (2013) 53 *Science and Justice* 339, 342. Cf the more critical approach in Gary Edmond and Emma Cunliffe, ‘Cinderella Story: The Social Production of a Forensic “Science”’ (2016) 106 *Journal of Criminal Law and Criminology* 219.

<sup>31</sup> James Shanteau, ‘Competence in Experts: The Role of Task Characteristics’ (1992) 53 *Organizational Behavior and Human Decision Processes* 252, 255.

<sup>32</sup> See Kahneman and Klein, above n 1, 519.

<sup>33</sup> Richard W Herling, ‘Operational Definitions of Expertise and Competence’ (2000) 2(1) *Advances in Developing Human Resources* 8, 20 (emphasis altered).

or experience (be it formal or informal).<sup>34</sup> Fundamentally, expertise is determined by the performance of a skill at either a relatively high (in NDM) or objectively high (in HB) level. This expectation applies to a wide range of practitioners from nursing and medical diagnosis to playing tennis and fire scene management. Furthermore, when this expectation is not fulfilled, it is common to observe a disconnect between social assignments of expertise and verified or validated expertise as defined in the HB or NDM traditions. For example, analyses of clinical psychologists and stockbrokers' intuitive judgments about risk of dangerousness or the future value of stocks (respectively) have not met objective standards for expertise in that the performance of individuals from relevant fields is routinely worse than basic algorithms or simple decision rules.<sup>35</sup>

It is important to appreciate that the disconnect between social recognition (and beliefs) and validated expertise can arise as a consequence of popular impressions or naïve realist views of fields and professions. People from outside a field may hold the simplistic belief that professionals occupy a particular position or role *because they are expert* in representative tasks in their domain.<sup>36</sup> They may believe that promotion is dependent on the achievement of objectively high performance of skills that seem archetypal within the field or profession. While this perception may be accurate in many domains (eg, test pilots fly exceptionally well and mathematicians are excellent at maths problems), upon closer inspection it is clear that it is not invariably the case. Or, more precisely, it is clear that the performance standards determining certification and career progression in some domains are not necessarily the same as those that may seem inherent to the role.

<sup>34</sup> It is, however, important to note that there is debate regarding whether the 'demonstrated' skill has been sufficiently defined and measured (measurable) to support an attribution of expertise. For example, proponents of the 'expert-performance' approach and those of the 'traditional' approach may disagree about the specific nature of the 'expertise' demonstrated by billionaires, senators or child prodigies. For detailed discussion, see K Anders Ericsson, 'Why Expert Performance is Special and Cannot be Extrapolated from Studies of Performance in the General Population: A Response to Criticisms' (2014) 45 *Intelligence* 81; Ericsson and Towne, above n 26.

<sup>35</sup> Ericsson and Towne, above n 26, 405.

<sup>36</sup> Collins describes how those with advanced research degrees in physics and mathematics, regulating access to research funding from agencies such as the United States' National Science Foundation, may not understand the dynamics of knowledge production and social ordering in specialist sub-groups: Harry Collins, 'Public Experiments and Displays of Virtuosity: The Core-Set Revisited' (1988) 18 *Social Studies of Science* 725. For a more detailed account, see Harry Collins, *Gravity's Shadow: The Search for Gravitational Waves* (University of Chicago Press, 2004).



For example, research reveals that in Australia, one can become and remain a passport examiner without ever having to demonstrate expertise comparing persons in photographs or a person to a photograph.<sup>37</sup> It might seem incredible that passport officers are not selected or promoted on the basis of their ability to correctly identify faces. Nevertheless, that was the situation until very recently.<sup>38</sup>

Similarly, highly experienced and respected forensic psychologists do not achieve high status by making many correct or more correct predictions about future dangerousness than their peers.<sup>39</sup> Even though these skills seem integral to professional practice, individuals in these domains progress by being adequate (or strong) performers on other tasks. The forensic psychologist may become respected in the field by competently using actuarial assessment tools, by thinking critically and engaging in evidence-based practice, by building strong rapport with their clients, by being a good colleague and co-worker, and by being a clear communicator; rather than being a relatively or highly accurate predictor of future dangerousness. To those outside the profession it is not always obvious that certification and progression is based on *other* important (and sometimes not so important) skills. External evaluators do not see the full picture and may not appreciate the range of practical, professional, and institutional factors at play. They may assume that entry and elevation through the ranks is based on skill in particular crucial tasks, rather than a broad range of professional competen-

<sup>37</sup> Recent research found that Australian passport officers were no more accurate at standardised face-matching tasks than first year university students. Significantly, experience as a passport officer made no difference to performance: see David White et al, 'Passport Officers' Errors in Face Matching' (2014) 9(8) *PLoS ONE* 1, 3–4.

<sup>38</sup> Until quite recently, Australian courts allowed anthropologists to testify about similarities between persons accused of crimes and persons of interest in images (such as CCTV recordings) for purposes of identification. It is not entirely clear, following *Honeysett*, whether anatomists (and others) who devote additional time to examining the images might yet testify (possibly as an ad hoc expert). Ironically, under *Honeysett*, it is quite likely that a passport examiner would be entitled to interpret CCTV images in order to identify a person of interest (or describe similarities between a person of interest and the accused). Most Australian courts (and *Honeysett* is exemplary) do not direct attention to what would seem to be the fundamental questions: (i) can this witness actually do what is claimed? (ii) How good are they? And (iii) how do we know? See *Honeysett* (2014) 253 CLR 122, 138–9 [47]–[48]; Gary Edmond, 'A Closer Look at *Honeysett*: Enhancing Our Forensic Science and Medicine Jurisprudence' (2015) 17 *Flinders Law Journal* 287.

<sup>39</sup> Assessment of future dangerousness is now commonly considered a core component of the forensic (rather than clinical) subspecialty of psychological practice: Kirk Heilbrun and Stephanie Brooks, 'Forensic Psychology and Forensic Science: A Proposed Agenda for the Next Decade' (2010) 16 *Psychology, Public Policy and Law* 219.

cies. This example helps to illustrate why ‘socially recognized’ expertise is insufficient on its own (regardless of whether expert status is assigned from inside or outside the field; and regardless of how the ‘field’ is defined).<sup>40</sup> It nicely juxtaposes scientific approaches to expertise. From both HB and NDM perspectives, the central determinant of expertise is demonstrable superiority in performance on some specific task.<sup>41</sup>

Interestingly, the observation that an individual may be expert in some tasks in their domain but not others is one of the hallmarks of expertise established through scientific research. While there are many different characterisations of the constituent components of expertise, there is general agreement about several features. Summarising the research, Bédard and Chi proposed five invariants of expertise:

First, experts, by definition, know more about their domain than do novices. Second, experts not only know more, but their knowledge is better organized. Third, on the basis of their greater knowledge and better organization, experts perform better than novices in domain-related tasks. Fourth, experts’ skill is domain specific: [t]here is very little transfer to unrelated domains. Finally, there are also many situations in which experts do not excel.<sup>42</sup>

The first and the second indicators relate to the acquisition of what might be described as background and foundational knowledge.<sup>43</sup> The third indicator is the performance dimension previously discussed.<sup>44</sup> The fourth and fifth indicators relate to the scope of expertise and its generalisability within and across domains. This raises the important issue of the ‘expert claim’.<sup>45</sup>

<sup>40</sup> See Ericsson, above n 34.

<sup>41</sup> Kahneman and Klein, above n 1.

<sup>42</sup> Jean Bédard and Michelene T H Chi, ‘Expertise’ (1992) 1 *Current Directions in Psychological Science* 135, 138–9.

<sup>43</sup> Although, it is worth noting that courts do not usually attempt to assess the amount or organisation of a witness’s knowledge. Background knowledge may be part of ‘specialised knowledge’ under s 79(1) of the uniform *Evidence Acts*, but many opinions may draw upon information, commitments, beliefs and knowledge that forms part of a domain or tradition and may not deal with a specific ability. When dealing with a specific ability, there is a need for independent evidence of validity and/or performance. This is consistent with the reference to ‘study or investigation’ linked to ‘specialised knowledge’: *Honeysett* (2014) 253 CLR 122, 131 [23].

<sup>44</sup> Note that the conceptualisation of successful performance described by Bédard and Chi is consistent with the NDM framework: Bédard and Chi, above n 42.

<sup>45</sup> See generally Kristy A Martire and Richard I Kemp, ‘Considerations When Designing Human Performance Tests in the Forensic Sciences’ (2017) 49 *Australian Journal of Forensic Sciences* (forthcoming).

As we explained in our example of a highly competent forensic psychologist who is unlikely to be 'expert' in the estimation of future risk (on the HB definition), determination of 'expertise' according to a performance standard must be specific. It cannot be broad, because superior performance in one task within an expert domain does not necessarily generalise to another domain or other tasks within the same domain. Our forensic psychologist may well be expert in the administration of actuarial risk assessment tests, but as a result of the uncertainty and complexity of the task at hand, may never be able to accurately predict the risk posed by *an individual* with a high degree of objective accuracy, nor perform better than an algorithm.<sup>46</sup>

The determination of whether a professional has expertise must therefore be assessed against the specific claim being made.<sup>47</sup> Here, it is important to focus on specific skills and competencies rather than more general training, qualifications or experience. If the claim of the psychologist in our example is that they are expert in the completion of actuarial risk assessment, or that they perform the assessments more competently than novices, then the skill that needs to be examined to establish expertise is the administration of the assessment tool either relative to optimal execution or comparative competence. If, however, the claim is that they are experts in the prediction of dangerousness, or that they make better predictions than novices, then the skill that needs to be examined to establish expertise is the accuracy of their predictions compared to optimal performance or the performance of the less experienced.<sup>48</sup>

The importance of assessing ability against the specific claims being made applies to all domains. The claims being made by forensic scientists in the pattern matching disciplines (eg, DNA profiling, latent fingerprints, ballistics, handwriting, voice and image comparison) are particularly revealing because they are relatively straightforward, potentially quite probative, and are among

<sup>46</sup> It is important to note that this is not a failing of the practitioner. Rather, it is an unavoidable consequence of the complexity of the task.

<sup>47</sup> Training and qualifications may affect the nature of the claims put forward by a practitioner, as well as their validity.

<sup>48</sup> We are agnostic on the level of performance required to warrant legal recognition and admission as an expert. However, we suggest that performance should be substantially better than novices because of the costs and dangers associated with introducing expert evidence in criminal proceedings, especially evidence adduced by the state and represented by the prosecutor and (perhaps) the courts, as expert. These are issues that warrant consideration in addition to s 79(1) under ss 135 and 137 of the uniform *Evidence Acts*. See also above n 30.

the easiest to empirically test.<sup>49</sup> Although not necessarily made explicit, forensic scientists in pattern matching disciplines make claims, as experts, about their ability to link a trace to its source. For example, associating deposited semen with a known DNA profile (in terms of a likelihood or probability), matching a crime scene latent fingerprint to the perpetrator's finger (to the exclusion of all others with no possibility of error), linking a spent casing to a specific gun (to a reasonable degree of ballistic certainty), assigning a writing sample to its author, assigning spoken words to an individual, or finding similarities between a person of interest and a suspect in images.<sup>50</sup> These claims to expertise ought to be evaluated against criteria of accurate associations (eg, matching a print to the correct finger) or relatively more accurate associations than novices.<sup>51</sup> Otherwise, we cannot be sure that they offer actual assistance to decision-makers.<sup>52</sup> In reality, the specificity of the forensic scientist's skill may require an even more precise definition of the expert claim to ensure that there is no unjustified slippage from one task to another.<sup>53</sup> For example, different procedures are required to develop and

<sup>49</sup> See, eg, The President's Council of Advisors on Science and Technology, 'Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods' (Report, Executive Office of the President (US), September 2016) ch 5 ('PCAST Report'); National Research Council (US), *Strengthening Forensic Science in the United States: A Path Forward* (National Academies Press, 2009) ch 5.

<sup>50</sup> For a surprisingly critical review of these and other domains in the United States ('US'), see National Research Council (US), above n 49; Expert Working Group on Human Factors in Latent Print Analysis, 'Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach' (Report, National Institute of Standards and Technology, US Department of Commerce, February 2012) ch 2; PCAST Report, above n 49. On the use of specific terminologies in the US, see Simon A Cole, 'More Than Zero: Accounting for Error in Latent Fingerprint Identification' (2005) 95 *Journal of Criminal Law and Criminology* 985; Simon A Cole and Gary Edmond, 'Science without Precedent: The Impact of the National Research Council Report on the Admissibility and Use of Forensic Science Evidence' (2015) 4 *British Journal of American Legal Studies* 585.

<sup>51</sup> The PCAST Report recommends 'black-box' empirical studies to establish the 'foundational validity' of a method: PCAST Report, above n 49, 47–8.

<sup>52</sup> See *Smith v The Queen* (2001) 206 CLR 650. The PCAST Report suggests that in the absence of appropriate empirical tests of foundational validity (ie, performance relative to the expert claim), assertions about the significance of apparent similarities may be meaningless. At ibid 46, the PCAST Report states:

Without appropriate estimates of accuracy, an examiner's statement that two samples are similar — or even indistinguishable — is scientifically meaningless: it has no probative value, and considerable potential for prejudicial impact. Nothing — not training, personal experience nor professional practices — can substitute for adequate empirical demonstration of accuracy.

<sup>53</sup> See Gary Edmond et al, 'How to Cross-Examine Forensic Scientists: A Guide for Lawyers' (2014) 39 *Australian Bar Review* 174.

analyse fingerprints that have been left in blood as compared to those made by sebaceous oils.<sup>54</sup> Thus, the forensic scientist may actually be claiming accuracy in development or analysis of oil-to-ink rather than blood-to-ink comparisons. Ultimately, adequate specification of the claim is central to the ability to demonstrate verifiable expertise relevant to a fact in issue.

Skilled intuition is another widely discussed feature of expertise.<sup>55</sup> This reflects the observation that the judgement and decision-making of experts usually becomes an automatic process of recognition (ie, identifying a familiar problem and recalling an appropriate response) that may be difficult to articulate.<sup>56</sup> Importantly, both experts and non-experts also engage in intuitive but *imperfect* automatic responses, susceptible to heuristics and biases that undermine outcome success.<sup>57</sup> This makes the presence of intuition and the inability to articulate the reasoning behind decision-making commonplace in the judgement of experts, but also in non-experts.<sup>58</sup> Whether the intuition was skilled or imperfect can ultimately only be established by recourse to the performance criterion discussed above.

<sup>54</sup> Bonnie Marchant and Christina Tague, 'Developing Fingerprints in Blood: A Comparison of Several Chemical Techniques' (2007) 57 *Journal of Forensic Identification* 76, 76–7.

<sup>55</sup> See, eg, Kahneman and Klein, above n 1.

<sup>56</sup> This is an issue for a system that depends primarily on transparency and questioning to evaluate the claimed expertise. Consider the reasoning in *Davie v Magistrates of Edinburgh* [1953] SC 34 (28 November 1952) 40 (Lord Cooper P).

<sup>57</sup> See generally Itiel Dror, 'Cognitive Bias in Forensic Science' in Mark Licker et al (eds), *McGraw-Hill Yearbook of Science and Technology 2012* (McGraw-Hill, 2012) 43; Itiel E Dror, 'The Paradox of Human Expertise: Why Experts Get it Wrong' in Narinder Kapur et al (eds), *The Paradoxical Brain* (Cambridge University Press, 2011) 177.

<sup>58</sup> There are species of expert who have abilities but may not possess relevant knowledge or insight. Some sport stars might be able to do extraordinary things without being able to explain them. Similarly, the ability to remember faces seems to be an ability that has a genetic base. Interestingly, legal admissibility rules and practices (eg, cross-examination) might have difficulty accommodating tacit and intuitive abilities. On 'super-recognisers', see generally Richard Russell, Brad Duchaine and Ken Nakayama, 'Super-Recognizers: People with Extraordinary Face Recognition Ability' (2009) 16 *Psychonomic Bulletin and Review* 252; Anna K Bobak, Peter J B Hancock and Sarah Bate, 'Super-Recognisers in Action: Evidence from Face-Matching and Face Memory Tasks' (2016) 30 *Applied Cognitive Psychology* 81; David J Robertson et al, 'Face Recognition by Metropolitan Police Super-Recognisers' (2016) 11(2) *PloS ONE* 1. See also the law-related discussion in Gary Edmond and Natalie Wortley, 'Interpreting Image Evidence: Facial Mapping, Police Familiars and Super-Recognisers in England and Australia' (2016) 3 *Journal of International and Comparative Law* (forthcoming). For a general discussion of tacit knowledge, see Michael Polanyi, *The Tacit Dimension* (Anchor Books, 1967); Harry Collins, *Tacit and Explicit Knowledge* (University of Chicago Press, 2010).

Notwithstanding the previous discussion, research suggests that the likelihood of expertise might be estimated (imperfectly) based on some of the characteristics of the profession or domain. Some areas are more likely to generate experts than others. Two conditions have been accepted as necessary for the development of expert level skill: the skill must be obtained in high-validity environments, and these environments must afford ample opportunities for learning.<sup>59</sup> ‘Opportunities for learning’ simply refers to the amount of exposure an individual has to the cues and outcomes in the relevant context (eg, hours spent playing chess, flying planes or comparing passport photos).<sup>60</sup> Understanding high and low validity environments is also necessary, though somewhat more challenging.

A high-validity environment is one where there is a strong causal relationship between the cues available to the learner (or expert), and the outcome they are trying to produce.<sup>61</sup> This means that the learner can acquire the information or behaviour that will achieve the desired outcome — for example, the moves to win a chess game or to identify the symptoms that indicate an infection. There may be some uncertainty surrounding the extent to which the desired outcome will be achieved in any given instance (the same move won’t always win a chess game, and a fever may predict both infections and stroke), although uncertainty is not incompatible with validity in this context.

Things are not as straightforward with low-validity environments. These tend to be complex and unpredictable — for example, many factors contribute to whether an individual released from custody will reoffend. No one cue strongly predicts a reoffending outcome particularly well, and the presence or absence of cues will vary over time after release, potentially changing the outcome after a prediction has already been made.<sup>62</sup> This means that it will always be difficult to develop expertise in this type of prediction, because the environment does not support learning.

Feedback is another mechanism that can facilitate the acquisition of expertise. Where feedback is provided, clearly linking outcomes with judgements or decisions, and where the feedback about performance (eg, around outcomes and associations) is accurate and abundant, then conditions are considered to

<sup>59</sup> Kahneman and Klein, above n 1, 519.

<sup>60</sup> On comparing passport images, see White et al, above n 37.

<sup>61</sup> Kahneman and Klein, above n 1, 520.

<sup>62</sup> *Ibid* 523.

be supportive or *kind* to skill acquisition.<sup>63</sup> Conversely, where feedback on performance is limited, inaccurate or absent, the environment can be classed as unsupportive or *wicked*.<sup>64</sup> In addition, where the training environment and the professional practice environment differ substantially in terms of task complexity or task representativeness, skills developed in the training setting may not generalise well to professional practice, limiting the utility of the learning.<sup>65</sup>

In light of these considerations, many forensic scientists are likely to encounter ‘wicked’ rather than ‘kind’ learning environments. Forensic scientists tend to develop their skills on simplified case examples but often provide testimony or reports in complex real-world scenarios.<sup>66</sup> They do not receive accurate feedback on the opinions they provide in genuine casework because the truth is not known and is usually unknowable. Moreover, forensic scientists may receive misleading feedback in the form of trial verdicts, which take into account many case factors and may or may not reflect the accuracy of their opinions. In some domains the strong relationship between the cues being attended to by practitioners and the outcomes have yet to be confirmed. For example, hair morphology has recently been discarded as a predictor of identity following a recent Federal Bureau of Investigation (‘FBI’) review.<sup>67</sup> Similarly, indentation patterns and bruising are no longer considered useful for determining the origin of a bite mark.<sup>68</sup> These were once socially recognised by investigators and courts, but errors and limitations identified

<sup>63</sup> Robin M Hogarth, Tomás Lejarraga and Emre Soyer, ‘The Two Settings of Kind and Wicked Learning Environments’ (2015) 24 *Current Directions in Psychological Science* 379.

<sup>64</sup> Ibid; Robin M Hogarth, *Educating Intuition* (University of Chicago Press, 2001) 90–1, 217–19.

<sup>65</sup> Hogarth, Lejarraga and Soyer, above n 63.

<sup>66</sup> See David L Faigman, John Monahan and Christopher Slobogin, ‘Group to Individual (G2i) Inference in Scientific Expert Testimony’ (2014) 81 *University of Chicago Law Review* 417.

<sup>67</sup> See Michael P Kortan et al, ‘FBI Testimony on Microscopic Hair Analysis Contained Errors in at Least 90 Percent of Cases in Ongoing Review’ (Media Release, Federal Bureau of Investigation, 20 April 2015) <<https://www.fbi.gov/news/pressrel/press-releases/fbi-testimony-on-microscopic-hair-analysis-contained-errors-in-at-least-90-percent-of-cases-in-ongoing-review>>.

<sup>68</sup> Erica Beecher-Monas, ‘Reality Bites: The Illusion of Science in Bite-Mark Evidence’ (2009) 30 *Cardozo Law Review* 1369, 1383–4; Mark Page, Jane Taylor and Matt Blenkin, ‘Expert Interpretation of Bitemark Injuries — A Contemporary Qualitative Study’ (2013) 58 *Journal of Forensic Sciences* 664; Mary A Bush, Howard I Cooper and Robert B J Dorion, ‘Inquiry into the Scientific Basis for Bitemark Profiling and Arbitrary Distortion Compensation’ (2010) 55 *Journal of Forensic Sciences* 976; Michael J Saks, ‘Forensic Bitemark Identification: Weak Foundations, Exaggerated Claims’ (2016) 3 *Journal of Law and the Biosciences* 538.

through innocence projects and scientific reviews have questioned the existence of relevant expertise among, respectively, microscopists and certified odontologists.<sup>69</sup> Given that environmental factors may make the development of expertise in the forensic sciences particularly challenging, it is all the more important to demonstrate expertise through a performance criterion.

#### A *Evaluating Expertise: Validity, Reliability, Limitations*

Validation studies or human performance tests are empirical investigations that aim to establish whether and to what extent an individual (or system) can do what is claimed or meet a defined performance standard (be it optimal or relative).<sup>70</sup> The idea of validity goes beyond the assessment of human performance, and can be used to describe the extent to which any technique, procedure or analysis does what it is intended to do in a scientifically robust manner.<sup>71</sup> From a scientific point of view, the reliability of a human, apparatus or procedure is captured by the extent to which the results, observations or conclusions are reproducible (eg, can be obtained across examinations or re-examinations).<sup>72</sup> The factors that undermine or augment the validity or reliability of a procedure define its limits and provide the basis for the quantification and estimation of the likelihood or rate of erroneous or misleading results (ie, error rate).<sup>73</sup> Information about all of these factors can be obtained from assessments against a performance criterion. For example, by determining whether a fingerprint examiner can correctly associate a crime scene impression with a known print, how often, and under what conditions performance is improved or impaired, we can begin to speak to the validity and reliability of the examiner and/or the procedure (depending on how the test is designed) as well as the limitations.

<sup>69</sup> Very recently, the PCAST Report concluded that foundational validity has not been established for complex-mixture DNA analysis, bite mark analysis, firearms analysis, footwear analysis or hair analysis: PCAST Report, above n 49, ch 5.

<sup>70</sup> Ibid 52.

<sup>71</sup> For a detailed discussion of validity, see Thomas D Cook and Donald T Campbell, *Quasi-Experimentation: Design and Analysis Issues for Field Settings* (Rand McNally College Publishing, 1979) ch 2. The PCAST Report defines the *foundational validity* of a technique in terms of its reliability, repeatability, reproducibility and accuracy: ibid 47–8.

<sup>72</sup> See Itiel E Dror, 'A Hierarchy of Expert Performance' (2016) 5 *Journal of Applied Research in Memory and Cognition* 121.

<sup>73</sup> PCAST Report, above n 49, 51–2.



Information of this kind is central to the rational evaluation of an opinion or procedure.<sup>74</sup> Its absence is at the forefront of concerns about many kinds of forensic science evidence expressed by the US National Academy of Sciences ('NAS')<sup>75</sup> and most recently the US President's Council of Advisors on Science and Technology ('PCAST').<sup>76</sup> It is not our aim, in this article, to provide detailed explanation of these constructs, or the way to obtain data about them.<sup>77</sup> Rather, we direct attention to these measures, as indicators of genuine expertise (and its boundaries), by way of counterpoint to some of the social criteria traditionally utilised by legal institutions.

#### IV LEGAL APPROACHES TO EVALUATING EXPERTISE

Lawyers and judges are vitally concerned with the actual skills, ability and knowledge possessed by individuals who might be recognised and relied upon as experts. Legal institutions care deeply about the rectitude of verdicts and the expert opinions upon which they are increasingly based. Such commitments are consistent with the HB definition of expertise; concerned as it is with the objective accuracy and optimal performance of judgements and decisions. Furthermore, those admitted as experts are expected to perform substantially better than ordinary persons (such as judges and jurors) in order for their opinions to be considered relevant, admissible and able to assist in fact-finding.<sup>78</sup> Under s 135 of the uniform *Evidence Acts* courts may be sensitive to the resource implications of adducing, admitting and contesting evidence by balancing these costs against the benefits of the opinion. Sections 135 and 137 should sensitise courts to the dangers and risks to the accused (and others) flowing from the admission of opinions that are not demonstrably expert (and therefore not *known* to be probative).<sup>79</sup>

<sup>74</sup> Obtaining the appropriate kinds of information is just the beginning. Gigerenzer suggests that many highly trained medical doctors, including specialists, struggle with the reported results of experiments and scientific research even in their domain: Gerd Gigerenzer, *Simply Rational: Decision Making in the Real World* (Oxford University Press, 2015) ch 5.

<sup>75</sup> National Research Council (US), above n 49, 7–8.

<sup>76</sup> PCAST Report, above n 49.

<sup>77</sup> For more information about designing human validation/performance trials, see Martire and Kemp, above n 45; PCAST Report, above n 49, 47–54.

<sup>78</sup> The relevance of expert testimony (see uniform *Evidence Acts* ss 55, 56) is discussed in Edmond et al, 'How to Cross-Examine Forensic Scientists', above n 53.

<sup>79</sup> Logically, with most scientific, medical and technical forms of evidence there is a need to know how probative an opinion is before you can begin to consider its 'highest' probative

It is our contention that where the court has not assessed the purported expertise of the witness against an appropriate performance criterion, its relevance and probative value, along with efficiency concerns and even the rectitude of the verdict, might not be susceptible to rational evaluation.

*A Using (Legal) Proxies Instead of an Expert Performance Criterion*

Despite the apparent correspondence between legal aspirations and scientific definitions of expertise, common law standards and the jurisprudence around statutory rules are not particularly well suited to the identification of those with expertise or the regulation of their opinions. Failings are acute in relation to opinions based on scientific, medical and technical forms of knowledge. Rather than seek information from validation studies which directly address performance and its limitations with respect to a particular procedure or claim, courts tend to rely on alternative (and indirect) factors with much lower utility.<sup>80</sup>

When deciding whether to admit opinion evidence in criminal proceedings, Australian courts routinely consider: formal qualifications and/or training; the existence and longevity of a 'field'; historical use of a procedure; the practitioner's years of experience and involvement in previous investigations or prosecutions; recognition or certification by non-legal institutions; prior legal admission; the plausibility of the claim; demeanour and the resilience of the practitioner in response to cross-examination (on the *voir dire*, or during trial); whether the defendant has access to expert assistance; and, even the strength of the overall case.<sup>81</sup> While some of these factors are undoubtedly of relevance to the assessment of expertise, they are not necessarily strong or direct indicators. Placing reliance on such factors may convince a court or decision-maker to afford expert status to a witness who has not actually demonstrated expert performance.<sup>82</sup>

More specifically, institutional recognition and certification are social indicators of expertise that may or may not provide insight into the posses-

value. Otherwise, the attribution of a highest probative value is speculative. It might appear reasonable, but it remains just a guess.

<sup>80</sup> Highly specific claims might be more common in the forensic sciences (eg, DNA profiling and fingerprint comparison) than forensic pathology or medicine more generally. Moreover, procedures are often developed to constrain both access to information and the level of discretion.

<sup>81</sup> See generally Edmond, 'Legal versus Non-Legal Approaches to Forensic Science Evidence', above n 15.

<sup>82</sup> *Ibid* 24–5.

sion of expertise or level of performance. If formal certification requires a demonstration of ability that can be compared to an objective outcome, threshold or peer group — providing the test items have sufficient similarity to the actual task — then the certification may support the attribution of expertise. Conversely, if the certification is like the proficiency tests utilised in many Australian (and other) forensic science communities, this inference might not be sound. In those cases, the tests are not comparable to the practitioner's casework and, beyond the mere fact of certification, do not provide information capable of establishing superiority of performance. Many proficiency tests are so simple that virtually everyone achieves the same correct result.<sup>83</sup>

Most of the cues relied on in Australian courts are social indicators of expertise, which may or may not correspond to actual performance superiority. What is more, lengthy use of social indicia has contributed to the legal recognition of practices that are not known to be valid or reliable and opinions not known to be expert (or based on knowledge). Many forensic science procedures, particularly those developed before the advent of DNA profiling, have not been formally evaluated to determine whether they produce objectively superior performance in relation to the expert claim. Instead of requiring evidence that directly supports performance, and derivative claims (which is *relatively* difficult to satisfy), courts substituted requirements that were less demanding.<sup>84</sup> Reified through long use, courts have grown accustomed to, even confident in the value of, their misguided legal proxies.<sup>85</sup> Moreover, the value of legal proxies appears to be reinforced through routine use, their conversion into precedent, and forensic science

<sup>83</sup> For a discussion of the limits of proficiency tests commonly used and relied upon in the forensic sciences, see PCAST Report, above n 49, 68; National Research Council (US), above n 49, 206–8. Seemingly oblivious to limitations and the written advice of commercial proficiency test providers, the Australian National Institute of Forensic Sciences — an organisation primarily funded by police — recently attempted to calculate error rates for a variety of forensic procedures using the results of simplistic proficiency tests: Australia New Zealand Policing Advisory Agency, 'NIFS Presentations', *ANZPAA NIF News* (Melbourne), October 2016, 6.

<sup>84</sup> Daniel Kahneman and Shane Frederick, 'Representativeness Revisited: Attribute Substitution in Intuitive Judgment' in Thomas Gilovich, Dale Griffin and Daniel Kahneman (eds), *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge University Press, 2002) 49. For an insightful discussion of the role of substitution in judicial decision-making, see Emma Cunliffe, 'Judging, Fast and Slow: Using Decision-Making Theory to Explore Judicial Fact Determination' (2014) 18 *International Journal of Evidence and Proof* 139.

<sup>85</sup> They are also confident about the effectiveness of legal safeguards, even though safeguards have not produced widespread legal awareness of systemic problems with many forensic sciences.

providers scrupulously attending to legal expectations. Unfortunately, legal reliance on proxies and institutional traditions of acceptance has resulted in the admission of opinions that are not demonstrably valid or reliable.<sup>86</sup> Very occasionally, this becomes clear after admission, reliance and conviction.<sup>87</sup> Conspicuous examples of questionable admissions include microscopic hair comparison, bullet lead comparison, facial mapping, bite-mark comparisons and voice spectroscopy.<sup>88</sup> This evidence was admitted and relied upon as *expert opinion* even though practitioners in these domains, some with formal qualifications and long experience doing the very task in question (eg, comparing hairs or dentition with bite marks), did not perform much better than chance or ordinary persons (ie, those without relevant training, study or experience and without *knowledge* in the domain).<sup>89</sup>

We accept that when assessing knowledge claims it is common for assessors to consider things such as the qualifications and experience of the individual, their position and institution, the esteem they are afforded, and their past performances at trial. These factors are not irrelevant to expertise, but with most types of scientific, medical and technical evidence, they provide indirect insights at best (as explained in relation to the forensic psychologists and passport officers above). Experience, institutional recognition and long use assist with the assessment of an opinion *once expert performance has been established*. Weak legal categories and associated admission practices may have deleterious consequences beyond the courtroom. Legal recognition and admission may legitimate investigative activities, police departmental structures and hierarchies, in-house training and traditional procedures.<sup>90</sup>

<sup>86</sup> It is questionable whether such opinions are based on 'specialised knowledge'.

<sup>87</sup> See, eg, Commonwealth/Northern Territory, Royal Commission of Inquiry into Chamberlain Convictions, *Report of the Commissioner; The Hon Mr Justice T R Morling* (1987) ch 16; *Gilham v The Queen* (2012) 224 A Crim R 22; *Wood v The Queen* (2012) 84 NSWLR 581; Acting Justice Brian Martin, 'Inquiry into the Conviction of David Harold Eastman for the Murder of Colin Stanley Winchester' (Report of the Board of Inquiry, 29 May 2014); *Eastman v DPP (ACT) [No 2]* (2014) 9 ACTLR 178; *R v Keogh [No 2]* (2014) 121 SASR 307.

<sup>88</sup> National Research Council (US), *On The Theory and Practice of Voice Identification* (National Academies Press, 1979); National Research Council (US), *Forensic Analysis Weighing Bullet Lead Evidence* (National Academies Press, 2004). See also above n 50.

<sup>89</sup> We have italicised knowledge because the dentists doing bite mark comparison had specialised knowledge. The problem is that their knowledge of anatomy and ability to manage the health of mouths did not transfer to discriminating between, and matching, bite marks — especially on bodies.

<sup>90</sup> See David A Harris, *Failed Evidence: Why Law Enforcement Resists Science* (New York University Press, 2012) 164–7; Jennifer L Mnookin et al, 'The Need for a Research Culture in the Forensic Sciences' (2011) 58 *UCLA Law Review* 725.

Directing emphasis toward social indicators — such as certification, accreditation, seniority, job title, case counts and so forth — might be palatable to courts but reveals little, if anything, about ability or performance.<sup>91</sup>

### B Assessing Expert Performance Data

Courts in the common law tradition have impaired their understanding and analysis of expertise through the focus on individual cases, rather than types of evidence or systemic problems with types of evidence.<sup>92</sup> This orientation has compromised the ability of judges and courts to recognise generic problems with expert evidence, enable systematic engagement with authoritative scientific and medical institutions and literatures, and improve mechanisms for admitting, representing, evaluating and reviewing.<sup>93</sup> For example, trial courts and courts of appeal have uneven records in recognising the value of general information that might challenge social definitions of expertise in favour of a performance standard consistent with HB and NDM traditions. Whereas validation and reliability studies are likely to produce indicative levels of performance, including estimates of the level of error in some specific conditions, courts have tended to: (a) not require this information; or (b) be unreasonably dismissive of such studies and their implications.<sup>94</sup> Courts have tended to advance concerns that studies will rarely if ever test expert performance approximating the conditions operating in the specific case.<sup>95</sup>

<sup>91</sup> Validation studies and performance testing are not always required for certification or accreditation. According to the PCAST Report, above n 49, 55:

Importantly, good professional practices — such as the existence of professional societies, certification programs, accreditation programs, peer-reviewed articles, standardized protocols, proficiency testing, and codes of ethics — cannot substitute for actual evidence of scientific validity and reliability.

<sup>92</sup> There are exceptions, such as the United Kingdom ('UK')'s response to injured children in *R v Harris* [2006] 1 Cr App R 5. Cf Emma Cunliffe, *Murder, Medicine and Motherhood* (Hart Publishing, 2011).

<sup>93</sup> See, eg, *Aytugrul v The Queen* (2012) 247 CLR 170; Justice Peter McClellan and Amber Doyle, 'Legislative Facts and Section 144 — A Contemporary Problem?' (2016) 12 *Judicial Review* 421. Cf Gary Edmond, David Hamer and Emma Cunliffe, 'A Little Ignorance is a Dangerous Thing: Engaging with Exogenous Knowledge Not Added by the Parties' (2016) 25 *Griffith Law Review* 1.

<sup>94</sup> See, eg, *Li v The Queen* (2003) 139 A Crim R 281, 294–5 [106]–[111] (Ipp JA). Worth noting is also the High Court's response to independent research in the New South Wales Court of Criminal Appeal by McClellan CJ at CL in *Aytugrul v The Queen* (2012) 247 CLR 170.

<sup>95</sup> See Faigman, Monahan and Slobogin, above n 66. We note that pharmaceuticals and therapeutics are routinely tested and used notwithstanding the inability to test them on every type of potential patient: Steven Epstein, *Inclusion: The Politics of Difference in Medical Re-*

While it is obvious that it would be ideal to have a validation study or performance test focused on the precise conditions of every case, it is highly unlikely that case conditions can ever be faithfully reproduced or resources made available for such testing.<sup>96</sup> However, the absence of such studies need not be debilitating for the analysis of forensic science expertise. Inferences about the efficacy of a medical treatment, for example, are routinely made on the basis of data derived from a large number of treatment instances, each of which differs to some extent from the presenting case.<sup>97</sup> A doctor never knows if a generally efficacious approach will work for a particular patient. Rather, they know about the likely outcomes associated with the treatment, and the conditions under which it tends to be more or less efficacious. Doctors and patients routinely make decisions on that basis.

Likewise, if courts have access to performance test data from forensic practitioners applying a particular methodology in a forensic science discipline (eg, a visual comparison of latent fingerprints), and the data shows conditions under which performance on the task is relatively better (with high quality impressions) or worse (with low quality impressions), as well as whether practitioners perform the task better than an untrained novice (eg, juror or judge), then informed inferences might be drawn by the court. These might apply both to the issue of admission as well as to what the latent fingerprint examiner can legitimately opine.<sup>98</sup> Specifically, the court can assess whether: (a) there is expertise in the procedure as claimed (ie, finger-marks can accurately be attributed to their source); and, (b) there is likely to be sufficiently reliable performance in the current case (based on the case features) to seek an 'expert' opinion.<sup>99</sup> To the extent that the claim to expertise in relation to a specific task (or domain) is supported, the court would likely also want to be assured that the practitioner providing an opinion was sufficiently proficient in the procedure to be classified as an expert.

For example, where a study of latent fingerprint examiners reported a small number of errors in good quality latent prints, the court might reasonably infer that performance would not be improved where a latent fingerprint

*search* (University of Chicago Press, 2007). We also note that, putting aside regulatory requirements, the failure to test would expose manufacturers to negligence and liability actions.

<sup>96</sup> Indeed, there are dangers in studying particular fact scenarios or cases, or trying to reproduce, in order to prove or disprove, the circumstances in a specific case.

<sup>97</sup> See generally Gigerenzer, above 74.

<sup>98</sup> As well as the determination of probative value 'at its highest': *IMM* (2016) 257 CLR 300, 314 [47] (French CJ, Kiefel, Bell and Keane JJ).

<sup>99</sup> The PCAST Report refers to (a) as *foundational validity* and (b) as *validity as applied*: PCAST Report, above n 49, 43.

is partial and distorted.<sup>100</sup> Similarly, where the study did not allow the use of enhancing tools or did not include verification, one could infer that performance would not be worse if these were available.<sup>101</sup> Thus, notwithstanding potential limitations, validation studies and other information about performance can and should mediate the admission and evaluation of expert opinion in individual cases.<sup>102</sup> Experimentally derived evidence, as the study of Australian Passport Officers revealed, will almost always be better than judicial (or juror) impressions of apparent plausibility and witness credibility.<sup>103</sup>

### C Efficacy of Trial Safeguards and Appeals

Trial safeguards can work, and they can work well; however, they have not *consistently* exposed substantial weaknesses with many types of expert opinion evidence; especially forensic science evidence.<sup>104</sup> They have not brought the kinds of issues explored in this article — many recommended by authoritative scientific organisations — or their implications, into the consciousness of lawyers, judges and policymakers.<sup>105</sup> Where lawyers, judges and fact-finders do not focus on empirical evidence of expert performance (eg, validity, reliability and proficiency), the potential efficacy of the safeguards can be undermined or misguided.<sup>106</sup> Most lawyers are not technically

<sup>100</sup> Jason M Tangen, Matthew B Thompson and Duncan J McCarthy, 'Identifying Fingerprint Expertise' (2011) 22 *Psychological Science* 995.

<sup>101</sup> Verification is a part of the Analysis, Comparison, Evaluation and Verification ('ACE-V') process employed by most fingerprint bureaus. The process as implemented has been subjected to criticism in both the National Research Council and National Institute of Standards and Technology reports: National Research Council (US), above n 49; National Institute of Standards and Technology, above n 50.

<sup>102</sup> See Gary Edmond, Matthew B Thompson and Jason M Tangen, 'A Guide to Interpreting Forensic Testimony: Scientific Approaches to Fingerprint Evidence' (2014) 13 *Law, Probability and Risk* 1.

<sup>103</sup> White et al, above n 37.

<sup>104</sup> Gary Edmond and Mehera San Roque, 'The Cool Crucible: Forensic Science and the Frailty of the Criminal Trial' (2012) 24 *Current Issues in Criminal Justice* 51.

<sup>105</sup> Authoritative scientific and technical organisations and entities include PCAST (US), the National Academy of Sciences (US), the National Institute of Standards and Technology (US), the Royal Society (UK), and the Forensic Science Regulator (England and Wales). Cf the Australian approach embodied in *Tang* (2006) 65 NSWLR 681; *IMM* (2016) 257 CLR 300.

<sup>106</sup> These approaches are accentuated by changes (mostly reductions) to the resourcing of the defence.

sophisticated and so they often avoid detailed methodological explorations of expert performance during the trial.<sup>107</sup> This, as we have explained, tends to have undesirable implications for fact-finding.

Furthermore, because courts have unwittingly relied on a range of simplistic and potentially misguided proxies for admitting and evaluating expertise, judges are not well positioned to guide fact-finders on the assessment of expert opinion. There are three serious issues here. First, notwithstanding their ongoing use, along with regular expressions of confidence by appellate courts, there is little evidence that judicial directions and warnings about expert evidence have the kind of effects suggested by judges.<sup>108</sup> Secondly, most of the comments offered by trial judges are banal. Trial judges tend to proffer comments that do not direct attention to issues that are central to assessing expertise and gauging probative value. In particular, model directions and warnings do not refer to validation, reliability and performance. They do not draw attention to error rates. Even if directions and warnings (and cross-examination) did direct attention to oversights and omissions, without the information produced through formal evaluation, decision-makers are not well positioned to evaluate the kinds of opinion proffered by forensic scientists.<sup>109</sup> Thirdly, comments (and directions) along with the fact of admission may give conflicting signals. Where a trial judge warns the jury to be careful, or draws attention to the fact that a procedure has not been validated, the reasons are not usually developed in ways that are appropriate for comprehension and evaluation. Moreover, any formal 'message' might be thwarted by a range of implicit contra-indicators. The fact that the state employs the forensic practitioner, routinely deploys them in investigations, and allows them to perform the untested procedure, in conjunction with the practitioner being called by the prosecutor and admitted by a trial judge (perhaps on many previous occasions), might be understood to suggest that claimed expertise is genuine notwithstanding any alternative impressions generated by cross-examination, rebuttal witnesses or vague judicial dampeners.

Warnings tend to be *thin*. They do not embody or convey the centrality of performance standards to mainstream scientific conceptualisations of expertise. Moreover, the trial judge, in an attempt to be fair (or to save the incriminating opinion), might draw attention to the lack of validation studies

<sup>107</sup> These are frequently characterised by judges on appeal as tactical decisions, even though lawyers (and judges) often do not know better and are not resourced.

<sup>108</sup> Warnings and directions are not known to cure problems with evidence.

<sup>109</sup> See especially Gary Edmond, 'Forensic Science Evidence and the Conditions for Rational (Jury) Evaluation' (2015) 39 *Melbourne University Law Review* 77.



in the same breath as reference to the forensic practitioner's extensive experience or the strength of the case. Judges in common law jurisdictions need to change their traditional and inadequate practices to directly embrace the issue of whether the individual possesses demonstrable expertise in the specific domain or domains. Several jurisdictions have attempted this (eg, the United States and Canada have adopted reliability tests) but they are yet to clearly explain their admissibility jurisprudence in these terms.<sup>110</sup>

## V REMEDIATING OUR JURISPRUDENCE

Misdirected reliance on legal proxies rather than scientific indicia of expertise could be corrected if lawyers, trial judges and courts of appeal augmented the jurisprudence around 'specialised knowledge' from *Honeysett v The Queen* ('*Honeysett*').<sup>111</sup> According to the High Court:

'Specialised knowledge' is to be distinguished from matters of 'common knowledge'. Specialised knowledge is knowledge which is outside that of persons who have not by training, study or experience acquired an understanding of the subject matter. It may be of matters that are not of a scientific or technical kind and a person without any formal qualifications may acquire specialised knowledge by experience. However, the person's training, study or experience must result in the acquisition of *knowledge*. The *Macquarie Dictionary* defines 'knowledge' as 'acquaintance with *facts, truths, or principles*, as from study or investigation' ... and it is in this sense that it is used in s 79(1). The concept is captured in Blackmun J's formulation in *Daubert v Merrell Dow Pharmaceuticals Inc*: 'the word "knowledge" connotes more than subjective belief or unsupported speculation ... [It] applies to any body of known facts or to any body of ideas inferred from such facts or accepted as truths on good grounds.'<sup>112</sup>

This explanation of 'specialised knowledge' provides the foundation for an admissibility jurisprudence that, unlike its predecessors, is suited to legal practice and criminal proceedings in the 21<sup>st</sup> century. The High Court

<sup>110</sup> For some insight into the impact of reliability standards in criminal proceedings in the US in the wake of the NAS report, see Cole and Edmond, above n 50. In Canada, appellate courts have read some of the *Daubert*-style requirements down: *R v Abbey* (2009) 97 OR (3d) 330; *R v Aitken* [2012] BCCA 134 (2 April 2012); Emma Cunliffe and Gary Edmond, 'Gaitkeeping in Canada: Mis-Steps in Assessing the Reliability of Expert Testimony' (2014) 92 *Canadian Bar Review* 327.

<sup>111</sup> (2014) 253 CLR 122.

<sup>112</sup> *Ibid* 131–2 [23] (emphasis in original) (citations omitted).

endorsed the definition from *Daubert v Merrell Dow Pharmaceuticals Inc*<sup>113</sup> (and *Kumho Tire Co Ltd v Carmichael*),<sup>114</sup> where the Supreme Court of the United States imposed a validity and reliability requirement on the admission of ‘scientific’ and ‘technical’ evidence, because of the presence of the word ‘knowledge’ in r 702 of the *Federal Rules of Evidence*, 28 USC (1975).<sup>115</sup>

The High Court in *Honeysett* drew attention to the need for knowledge: ‘training, study or experience must result in the acquisition of knowledge’.<sup>116</sup> Subsequently, knowledge is linked to ‘facts, truths, or principles’, and importantly, ‘as from study or investigation’.<sup>117</sup> Thus, when it comes to scientific and technical forms of evidence, the High Court has provided the foundations for a more functional and publicly credible admissibility jurisprudence. That is, for forensic science evidence (especially procedures in regular use) courts should expect references to formal studies (ie, ‘study or investigation’) that demonstrate ‘expertise’ via superior performance and provide information about the limits of the procedure.<sup>118</sup> While expert witnesses will usually possess knowledge in an area or on some issue, where they are interpreting data and proffering opinions, courts should require additional knowledge pertaining to validity and reliability. Where an expert is called upon to report or testify based on some ability, courts should insist on demonstrable evidence of that specific ability.

Unfortunately, the High Court did not provide criteria or a particularly serviceable exemplification in *Honeysett*. The Court accepted that the witness (a professor of anatomy) had a general knowledge of anatomy. However, the anatomist in question was deemed not to be an expert in image interpretation or comparison for the purpose of identification — that is, the specific task at hand.<sup>119</sup> Without referring to the absence of formal studies or evidence of

<sup>113</sup> 509 US 579 (1993).

<sup>114</sup> 526 US 137 (1999).

<sup>115</sup> The original term in the United States *Federal Rules of Evidence* was ‘scientific, technical or other specialized knowledge’: *Federal Rules of Evidence*, 28 USC r 702 (1975).

<sup>116</sup> (2014) 253 CLR 122, 131 [23] (emphasis altered),

<sup>117</sup> *Ibid* (emphasis altered).

<sup>118</sup> Forensic science institutions and courts should also direct their attention to the dangers created by human factors: see, eg, D Michael Risinger et al, ‘The *Daubert/Kumho* Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion’ (2002) 90 *California Law Review* 1; Bryan Found, ‘Deciphering the Human Condition: The Rise of Cognitive Forensics’ (2015) 47 *Australian Journal of Forensic Sciences* 386.

<sup>119</sup> *Honeysett* (2014) 253 CLR 122, 139 [48]. The question of whether he was an ad hoc expert, from having spent time looking at the images, rather than from any demonstrated perfor-

proficiency, the Court excluded his opinions about the images.<sup>120</sup> An explanation that was consistent with this approach, though of far greater service to lawyers, experts and judges, would have been to explain that the anatomist's procedure had not been validated and there was no evidence that his performance was superior to the tribunal of fact.<sup>121</sup> Significantly, in the absence of knowledge about the value of his procedure and conclusions, we cannot say that the opinion was based on 'specialised knowledge' because it was not linked to 'training, study or experience' by the required 'study or investigation'.<sup>122</sup>

Earlier, in civil proceedings, the High Court expressed a willingness to use familiar proxies to avoid the need to delve into contests around 'knowledge':

The way in which s 79(1) is drafted necessarily makes the description of the requirements very long. But that is not to say that the requirements cannot be met in many, perhaps most, cases very quickly and easily. That a specialist medical practitioner expressing a diagnostic opinion in his or her relevant field of specialisation is applying 'specialised knowledge' based on his or her 'training, study or experience', being an opinion 'wholly or substantially based' on that 'specialised knowledge', will require little *explicit articulation or amplification* once the witness has described his or her qualifications and experience, and has identified the subject matter about which the opinion is proffered.<sup>123</sup>

This is a fairly recent example of a senior court suggesting that when it comes to 'established' disciplines, there may be no need to direct attention to the question of whether the individual possesses the requisite expertise. The High Court seemed to indicate, in this civil appeal, that the traditional proxies suffice. Weakly diagnostic social indicia are said to be sufficient for the task.

mance, was strategically avoided by the court yet remains potentially open to future prosecutors. Use of the category *ad hoc expert* directs no attention to evidence of expertise.

<sup>120</sup> Edmond, 'A Closer Look at *Honeysett*', above n 38, 300. Revealingly, the decision does not address the question of the anatomist's actual abilities: at 297–8.

<sup>121</sup> See especially *Smith v The Queen* (2001) 206 CLR 650, 655 [11] (Gleeson CJ, Gaudron, Gummow and Hayne JJ):

The fact that someone else has reached a conclusion about the identity of the accused and the person in the picture does not provide any logical basis for affecting the jury's assessment of the probability of the existence of that fact when the conclusion is based only on material that is not different in any substantial way from what is available to the jury.

In the absence of formal evaluation, it is not possible to determine whether anatomists, physical anthropologists or military intelligence officers can outperform ordinary persons.

<sup>122</sup> *Honeysett* (2014) 253 CLR 122, 131 [23]. See also uniform *Evidence Acts* s 79(1).

<sup>123</sup> *Dasreef* (2011) 243 CLR 588, 604 [37] (French CJ, Gummow, Hayne, Crennan, Kiefel and Bell JJ) (emphasis added).

This *might* be comprehensible with respect to a highly regulated domain like modern biomedicine; where most areas are highly specialised and increasingly attentive to evidence-based practices available through public repositories.<sup>124</sup> Biomedicine is tightly regulated and monitored, medical practitioners are subject to professional review mechanisms and negligence suits, such that validity, reliability and proficiency *might* be subsumed within the certification to practice (or membership). Moreover, as opposed to the forensic sciences, largely monopolised by the state, in civil proceedings plaintiffs and defendants have access to medical and other forms of expert assistance. However, we should never forget that social indicators of expertise operate as a proxy or a short cut. That is, they stand in the place of demonstrable domain specific expertise without always guaranteeing its existence. While the use of legal proxies might be reasonable in some civil disputes, they are less well suited to criminal prosecutions. Relying on proxies is inappropriate where, as with much forensic science and medicine evidence adduced by the state, validity, reliability and proficiency have not been established.

With respect to those practices in forensic science and medicine, where study or investigation has not been undertaken to produce relevant 'knowledge', formal qualifications, job titles, longstanding practices, experience and prior admission should not be allowed to stand in place of (published) data demonstrating superior performance in respect of the expert claim. We should be careful when substituting social indicia such as whether a set of practices are 'established' (or longstanding) for demonstrable ability derived through validation and performance testing. We should not assume that juries will appreciate these differences — after all, trial and appellate judges have not. Rather, in many cases there will be a need for 'explicit articulation or amplification' of the 'specialised knowledge' on which the opinion is based, along with how it is connected to the specific 'subject matter' and the 'training, study or experience'.<sup>125</sup>

Notwithstanding our advocacy for demonstrable expertise, admissibility determinations do not have to be all or nothing. In many cases, particularly in the comparison or pattern matching forensic sciences, the issue is not whether a procedure works or does not work. Rather, we are concerned with how well it works, in what conditions, and how should opinions be expressed to accurately capture the value of opinion evidence in ways that facilitate lay

<sup>124</sup> See, eg, the Cochrane collaboration and its systematic review of studies to provide a foundation for evidence-based medicine: *Homepage* (2017) Cochrane <<http://www.cochrane.org>>.

<sup>125</sup> See *ibid*; *Honeysett* (2014) 253 CLR 122; *HG v The Queen* (1999) 197 CLR 414.

comprehension. It may be that, to the extent that they are willing to admit some expert opinions, trial judges may only need to moderate the strength of the expert claim.<sup>126</sup> This may require probabilistic forms of expression and opinions that incorporate or acknowledge limitations, uncertainty and the ubiquitous threat of error. This is the kind of information that studies focusing on a performance criterion will generate.

#### *A Concurrent Evidence and Recent Procedural Reform*

In cases where there is disagreement between experts, the experts may be required to meet (pre-trial), perhaps in the absence of the lawyers, to identify and ideally eliminate areas of disagreement. Where disagreement cannot be eliminated, the experts may be required to explain why they cannot agree, and should the matter proceed to trial, they may be required to testify concurrently. Concurrent evidence, known colloquially as ‘hot-tubbing’, enables the experts to testify in a joint session.<sup>127</sup> In addition to cross-examination and re-examination, experts giving concurrent evidence may be asked to comment on answers provided by other witnesses and may even ask each other questions.<sup>128</sup>

Pre-trial meetings and concurrent evidence are not objectionable on their face, but they do introduce a range of new sociological and psychological issues. The main issue is that the procedures, designed to enhance efficiency and reduce court time and the number of cases that go to court, do not necessarily direct attention to questions related to establishing or moderating expert performance. The procedures used predominantly in civil litigation and by tribunals do not require that participants engage with knowledge or its foundations. Where two (or more) witnesses are presenting opinions without demonstrating expertise, pre-trial meetings and concurrent evidence process-

<sup>126</sup> *Tang* (2006) 65 NSWLR 681 is not a good guide here, because we should know that a putative expert performs better than jurors before we consider moderating their opinion in order to allow them to testify. Perceived utility or need cannot overcome the requirement that opinions must be based on knowledge: see Simon A Cole, ‘Splitting Hairs? Evaluating “Split Testimony” as an Approach to the Problem of Forensic Expert Evidence’ (2011) 33 *Sydney Law Review* 459.

<sup>127</sup> See Steven Rares, ‘Using the “Hot Tub”: How Concurrent Expert Evidence Aids Understanding Issues’ (Paper presented at New South Wales Bar Association Continuing Professional Development Seminar, Bar Association Common Room, 23 August 2010); Gary Edmond, ‘Merton and the Hot Tub: Scientific Conventions and Expert Evidence in Australian Civil Procedure’ (2009) 72 *Law and Contemporary Problems* 159.

<sup>128</sup> Ian Freckelton et al, *Expert Evidence and Criminal Jury Trials* (Oxford University Press, 2016) 54–5 [3.34]–[3.35].

es may not bring that oversight to light. Thus, where procedures have not been appropriately evaluated, legal procedures that do not engage with the fundamentals of expertise would seem to perpetuate traditional means of admitting and evaluating expert opinion.<sup>129</sup> We should be careful not to mistake institutional efficiencies for enhanced responses to expertise. Concurrent evidence has considerable potential, but where procedures and expert claims have not been appropriately evaluated, it cannot overcome that lacuna. It may help to identify such oversights and limitations, but that is yet to be demonstrated.

Finally, when thinking about procedural reform and conventional trial safeguards, it is increasingly significant that most persons accused of a criminal offence will not have the benefit of expert assistance, even if they are tried and instruct their counsel to contest the expert opinion evidence assembled against them.<sup>130</sup> Not only should glib assertions and uncritical commitment to the value of conventional trial safeguards be avoided, in the context of an inadequately resourced criminal justice system and impecunious defendants, the need for prosecutors to proffer demonstrably reliable forensic science evidence is more important than ever before.

## VI CONCLUSION

Legal approaches to expertise in criminal proceedings are misguided. They rely heavily on models of expertise that arose in the enlightenment, where the dominant forms of relevant knowledge, namely natural philosophy, medicine and early manifestations of engineering, were predominantly gentlemanly pursuits.<sup>131</sup> Legal recourse to expert evidence and its gradual expansion through recognition of emerging fields, areas of specialisation and experience, enabled courts to admit an ever-expanding array of putatively expert opinions. Accommodating legal responses were undermined as permissive rules, initially extended to the social equals of judges (eg, university trained elites),

<sup>129</sup> Where all parties are represented, a trial judge might think it fair to admit the evidence even if there might be validity and reliability problems. On the practical limits of trial safeguards, see Edmond and San Roque, 'The Cool Crucible', above n 104.

<sup>130</sup> Recent studies suggest that only the state calls expert witnesses in most cases: see Freckelton et al, above n 128, 123 [6.23].

<sup>131</sup> These conceits are still visible, for example, in continued legal privileging of psychiatry over clinical or forensic psychology. Historically, the role of trust between university-educated elites seems to have been a feature of legal engagement: see generally Steven Shapin, *A Social History of Truth: Civility and Science in Seventeenth Century England* (University of Chicago Press, 1994) ch 1.

were used to obtain opinions, however marginal and inexpert, that supported partisan interests. It was not long after the introduction of our modern instantiation of the expert opinion rule in the late 18<sup>th</sup> century that judicial concerns about expert venality and partisanship emerged fully-fledged.<sup>132</sup>

In general, though especially where opinion evidence is challenged, courts should be expecting to see specialised knowledge demonstrated through study or investigation. That is, they should be expecting to see the party adducing an expert opinion to bring scientific literatures that support the procedure (eg, validation and performance studies) and the specific type of application. For many procedures, formal scientific validation is the type of study that produces relevant 'specialised knowledge'. In relation to most forensic science and medicine evidence, opinions should be based on such knowledge. This knowledge should exist separately from the expert, should be publicly available, and should be made available to the court, ideally included in expert reports.<sup>133</sup> Courts should construe the need for 'training, study or experience' to confirm the specific witness's ability to meet expert performance standards and competently use valid procedures relevant to their claims.

Inattention to scientifically accepted criteria for identifying expertise (requiring superior performance, either relative to an objective standard or compared with novices or lay persons) has meant that, to various degrees, the opinions received in legal proceedings may be speculative, perhaps mere *ipse dixit*, and unrepresentative of what is known beyond the courtroom. Standards for the admission of expert opinion evidence in criminal proceedings are excessively liberal. We accept that there may be institutional reasons for tempering the strictness of requirements around admissibility in some circumstances, but these do not apply to opinion evidence adduced by the state.

In conclusion, we would make two emphatic points. First, when it comes to expert opinion evidence adduced by the state in criminal proceedings, there is a need to attend to indicia of *expert performance*. Where the evidence is of a scientific, medical or technical nature there appear to be very few credible reasons for exempting experts from the need to identify the scientific research supporting their practices and claims and, where appropriate, evidence of their own proficiency or ability in the specific domain. Currently, no jurisdiction in Australia requires such information as part of its admissibil-

<sup>132</sup> See, eg, *Thorn v Worthing Skating Rink Co* (1876) 4 Ch D 415, 416. For a more expansive discussion, see Golan, above n 6.

<sup>133</sup> See Gary Edmond, Kristy Martire and Mehera San Roque, 'Expert Reports and the Forensic Sciences' (2017) 40 *University of New South Wales Law Journal* (forthcoming).

ity practice.<sup>134</sup> Secondly, as a general principle, courts should not allow or persist with practices that are inconsistent with mainstream scientific knowledge or ways of knowing used by scientists. If there are compelling reasons to do so, whether institutional reasons or issues pertinent to the specific case, then these should be identified and explained. Fact-finding in law should, as far as possible, resemble fact-finding elsewhere. When it comes to establishing expertise that is task-related, demonstrably superior performance is fundamental.

<sup>134</sup> Cf *Tuite* [2015] VSCA 148 (12 June 2015). *Tuite* was the only appellate decision in Australia stipulating that forensic science evidence should be produced using validated procedures, albeit by virtue of s 137 of the uniform evidence legislation. In the aftermath of *IMM* (2016) 257 CLR 300, 306 [17], 314 [50], the status of *Tuite* and the requirement that forensic science be valid and reliable is uncertain. As this article explains, it is certainly arguable that forensic science evidence, most conspicuously opinions derived via the feature comparison methods, is 'weak' or 'unconvincing' where the procedures have not been formally validated and actual expertise has not been demonstrated. See also Edmond, 'Icarus and the *Evidence Act*', above n 4.