

# ICARUS AND THE EVIDENCE ACT: SECTION 137, PROBATIVE VALUE AND TAKING FORENSIC SCIENCE EVIDENCE ‘AT ITS HIGHEST’

GARY EDMOND\*

*When determining probative value for the purposes of balancing the probative value of evidence against the danger of unfair prejudice to the accused, the High Court has favoured an approach that appears incompatible with a rational response to the treatment of opinions based on specialised knowledge. A majority in IMM v The Queen concluded that when determining the probative value of evidence for the purpose of s 137 the trial judge should take the contested evidence at its highest. In doing so the trial judge is prevented from considering the reliability of the evidence or the credibility of the witness. This article explains why, whatever the merits of such an approach might be in relation to other types of evidence and witness, trial judges are incapable of assessing the probative value of most forms of scientific, medical and technical opinion evidence — particularly forensic science and forensic medicine evidence — without insight into reliability or trustworthiness demonstrated through formal evaluation. Using examples, the article explains why blinding trial judges to reliability and validity obliges them to speculate about the value of procedures and opinions. Not only are trial judges obliged to speculate, but inattention to the results of formal evaluation has detrimental system effects, elides a range of serious dangers to the accused and denudes s 137 of value as a trial safeguard.*

## CONTENTS

I	Introduction: Escaping the Cretans.....	107
II	IMM v The Queen and Determining Probative Value for the Purpose of Section 137 .....	112
	A Too Close to the Sun: Probative Value at Some Imagined ‘Highest’ ....	112

\* Professor, School of Law, The University of New South Wales; Research Professor (fractional), Northumbria Law School, Northumbria University; Chair of the Evidence-Based Forensics Initiative. The author would like to thank David Hamer, Andrew Ligertwood, Kristy Martire, Mehera San Roque, Kaye Ballantyne and several anonymous referees for comments. The research was supported by the Australian Research Council (LP16010000).

B	Between the Sun and the Sea: Probative Value among the Dissentients .....	115
1	Probative Value at its Actual 'Highest' .....	115
2	(Actual) Probative Value .....	117
C	Seeing in the Fog .....	118
III	The Probative Value of Scientific, Medical and Technical Evidence.....	121
A	Determining the Probative Value of Opinions Based on Specialised Knowledge .....	121
B	States of Knowledge and Ignorance.....	125
C	Carried Away with Phaethon: What 'at its Highest' Cannot Mean.....	128
D	Reliability and Credibility and Probative Value.....	130
E	Standing by Itself.....	133
F	Free Flying? Limits on the Jury's Prerogative .....	134
IV	Seeing through Heydon's Fog: Interpreting Images after <i>Honeysett</i> .....	135
A	The Opinion of a Passport Officer .....	135
B	Can Trial Safeguards Identify Problems and Convey Them to Jurors? 140	
V	'Danger of Unfair Prejudice to the Defendant' and the Balancing Exercise....	142
A	The Danger of Unfair Prejudice .....	143
B	The Balancing Exercise and Mitigation of the Risk.....	147
VI	Re-Imagining Probative Value for Opinions Based on Specialised Knowledge .....	149

## I INTRODUCTION: ESCAPING THE CRETANS

In a criminal proceeding, the court must refuse to admit evidence adduced by the prosecutor if its probative value is outweighed by the danger of unfair prejudice to the defendant.<sup>1</sup>

'[P]robative value' of evidence means the extent to which the evidence could rationally affect the assessment of the probability of the existence of a fact in issue.<sup>2</sup>

In the classical myth, Icarus enjoyed the freedom flight afforded. Inattentive to knowledgeable admonition, he pushed his new-found liberty, falling 'in love

<sup>1</sup> *Uniform Evidence Law* s 137 ('UEL'). The UEL has been implemented in most Australian jurisdictions: *Evidence Act 1995* (Cth); *Evidence Act 2011* (ACT); *Evidence Act 1995* (NSW); *Evidence (National Uniform Legislation) Act* (NT); *Evidence Act 2001* (Tas); *Evidence Act 2008* (Vic).

<sup>2</sup> UEL (n 1) s 3 (definition of 'probative value').

with the sky.<sup>3</sup> Icarus flew too close to the sun. Wax securing the feathers on the wings crafted by his father melted and he fell catastrophically. The myth of Icarus offers a salutary lesson about the value of attending to knowledge; in the myth, the advice of Icarus' father. Daedalus warned his son against complacency (flying too low) and hubris (flying too high). This article endeavours to explain the related dangers of disregarding reliability and soaring unrestrained when assessing the probative value of opinions based on specialised knowledge in criminal proceedings.<sup>4</sup>

With the exception of DNA profiling, Australian courts have been remarkably resistant to engaging with the reliability of scientific, medical and technical evidence adduced by prosecutors in criminal proceedings.<sup>5</sup> No Australian court requires proponents of opinions purportedly based on specialised knowledge to provide evidence of the underlying procedure's validity (ie that is actually works), its level of error, or the proficiency of the 'expert'.<sup>6</sup> This is unfortunate because the appropriate mechanism for assessment of reliability (and validity) exists in s 79(1), which provides that

[i]f a person has specialised knowledge based on the person's training, study or experience, the opinion rule does not apply to evidence of the person's opinion that is completely or substantially based on that knowledge.<sup>7</sup>

That contention finds support in *Honeysett v The Queen* where the High Court unanimously endorsed the following definitions of 'knowledge', though without addressing the reliability issue:

The *Macquarie Dictionary* defines 'knowledge' as 'acquaintance with *facts, truths, or principles*, as from study or investigation' (emphasis added) 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

<sup>3</sup> Ovid, *Metamorphoses*, tr David Raeburn (Penguin Books, 2004) 305.

<sup>4</sup> This article is primarily focused on the comparison or identification sciences, but has broader implications.

<sup>5</sup> See, eg, *R v Gallagher* [2001] NSWSC 462; *R v Karger* (2002) 83 SASR 135. As explained in Part VI, the Victorian Court of Appeal attempted to impose a reliability threshold for forensic science evidence in *Tuite v The Queen* [2015] VSCA 148.

<sup>6</sup> The word 'expert' is sometimes emphasised to reinforce the fact that in many cases it is uncertain whether those claiming expertise or recognised as experts actually possess heightened abilities.

<sup>7</sup> Unless otherwise indicated, all references to sections are to the *UEL*. The proponent of the opinion should support the admissibility decision by explaining the two criteria in s 79(1): *Dasreef Pty Ltd v Hawchar* (2011) 243 CLR 588, 602–3 [32]. See also *HG v The Queen* (1999) 197 CLR 414, 427 [39].

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>8</sup>

It bears noting that in *Daubert* the Supreme Court of the United States interpreted the term ‘knowledge’, from the phrase ‘scientific, technical, or other specialized knowledge’ in r 702 of the *Federal Rules of Evidence*, to require federal judges to attend to validity and reliability.<sup>9</sup> This approach to ‘scientific knowledge’ was extended to ‘technical, or other specialized knowledge’ in *Kumho Tire Co Ltd v Carmichael* and eventually led to revision of the *Federal Rules of Evidence* in 2000 and 2011.<sup>10</sup> Subsequently, the Supreme Court of Canada — following the US lead — interpreted its common law to require trial judges to consider the reliability of expert opinion evidence in the aftermath of a series of wrongful convictions.<sup>11</sup> More recently, a rules committee chaired by the Lord Chief Justice of England, responding to the thrust of recommendations (and a Bill drafted) by the Law Commission of England and Wales,<sup>12</sup> embedded the need for reliability in the *Criminal Procedure Rules 2015* (England and Wales).<sup>13</sup>

<sup>8</sup> (2014) 253 CLR 122, 131–2 [23], quoting *Daubert v Merrell Dow Pharmaceuticals Inc*, 509 US 579, 590 (1993). No Australian court has interpreted s 79(1) or the common law equivalents to require that the opinion be reliable. I have argued in favour of a reliability standard at length: Gary Edmond, ‘Specialised Knowledge, the Exclusionary Discretions and Reliability: Reassessing Incriminating Expert Opinion Evidence’ (2008) 31 *University of New South Wales Law Journal* 1; Gary Edmond, ‘The Admissibility of Forensic Science and Medicine Evidence under the Uniform Evidence Law’ (2014) 38 *Criminal Law Journal* 136; Gary Edmond, ‘A Closer Look at *Honeysett*: Enhancing Our Forensic Science and Medicine Jurisprudence’ (2015) 17 *Flinders Law Journal* 287.

<sup>9</sup> *Daubert* (n 8) 590. The majority explained that ‘in order to qualify as “scientific knowledge,” an inference or assertion must be derived by the scientific method. Proposed testimony must be supported by appropriate validation — ie, “good grounds,” based on what is known. In short, the requirement that an expert’s testimony pertain to “scientific knowledge” establishes a standard of evidentiary reliability.’

<sup>10</sup> 526 US 137, 147 (1999). Rule 702 now requires the testimony to be ‘the product of reliable principles and methods ... reliably applied ... to the facts of the case.’

<sup>11</sup> *R v DD* [2000] 2 SCR 275; *R v J-LJ* [2000] 2 SCR 600; *R v Trochym* [2007] 1 SCR 239. In *J-LJ*, Binnie J wrote that ‘[t]he admissibility of the expert evidence should be scrutinized at the time it is proffered, and not allowed too easy an entry on the basis that all of the frailties could go at the end of the day to weight rather than admissibility’: at 613 [28]. See also Justice W Ian C Binnie, ‘Science in the Courtroom: The Mouse that Roared’ (2008) 27(2) *Advocates’ Society Journal* 11.

<sup>12</sup> Law Commission (England and Wales), *Expert Evidence in Criminal Proceedings in England and Wales* (Report No 325, 2011).

<sup>13</sup> *Criminal Procedure Rules 2015* (England and Wales) pt 19; *Criminal Practice Directions 2015* (England and Wales) div V.

With Australian courts reluctant, thus far, to consider reliability as part of s 79(1),<sup>14</sup> the primary protections against the admission and misuse of weak, speculative and unreliable opinion evidence are ss 135 and 137, and judicial warnings.<sup>15</sup> This article explains why judges must attend in their admissibility jurisprudence and practice to reliability — really validity and scientific reliability.<sup>16</sup> Without abandoning the need to consider reliability as part of s 79(1), it responds to *IMM v The Queen*, where, considering probative value in relation to the admission of tendency and context evidence in a sexual assault prosecution, a bare majority of the High Court insisted that issues of reliability and credibility should play no part in the trial judge's assessment of probative value for the purposes of ss 135 and 137.<sup>17</sup>

In their attempts to determine 'the extent to which' evidence can 'rationally affect the assessment of the probability of the existence of a fact in issue', trial judges cannot avoid issues of reliability and credibility. This article explains why trial judges should consider both the reliability of procedures and the proficiency of the witness when determining the probative value of opinions based on specialised knowledge — so-called 'expert evidence'.<sup>18</sup> When it comes to scientific, medical and technical evidence, there are very few means of gauging probative value (and weight) without determining whether the procedure works, how well and in what conditions, and whether the forensic

<sup>14</sup> See, eg, *Velevski v The Queen* (2002) 187 ALR 233; *R v Tang* (2006) 65 NSWLR 681; *Dasreef* (n 7).

<sup>15</sup> See especially *Aytugrul v The Queen* (2012) 247 CLR 170, 176 [5]–[6]; *Tuite* (n 5) [42]–[114].

<sup>16</sup> Reliability has both a common and a technical meaning. In this essay, where 'reliability' is used by itself it generally refers to its everyday meaning, namely trustworthiness. When collocated with validity or the qualifier 'scientific' it refers to the consistency of a measurement. This is sometimes captured in the terms repeatability, reproducibility and accuracy.

<sup>17</sup> (2016) 257 CLR 300. For general commentary on *IMM*, see Stephen Odgers, 'Stephen Odgers SC on Probative Evidence after *IMM v The Queen*' [2016] (Winter) *Bar News* 36; Richard Lancaster, '*IMM v The Queen*: A Response from Richard Lancaster SC' [2016] (Winter) *Bar News* 40. For an attempt to redeem the majority's position, see Stephen Odgers, *Uniform Evidence Law* (Lawbook, 12<sup>th</sup> ed, 2016) 1184–6. See also David Hamer, 'The Unstable Province of Jury Fact-Finding: Evidence Exclusion, Probative Value and Judicial Restraint after *IMM v The Queen*' (2017) 41(2) *Melbourne University Law Review* (forthcoming).

<sup>18</sup> It bears noting that none of the judgments in *IMM* considers the implications of the various approaches to s 137 for scientific, medical and technical evidence adduced via s 79(1). In terms of the decisions cited in *IMM*, apart from *Dupas v The Queen* (2012) 40 VR 182, none was concerned with, or directed attention toward, opinions based on specialised knowledge: *R v Christie* [1914] AC 545; *Rozenes v Beljajev* [1995] 1 VR 533; *Carusi* (1997) 92 A Crim R 52; *Papakosmas v The Queen* (1999) 196 CLR 297; *Adam v The Queen* (2001) 207 CLR 96; *Festa v The Queen* (2001) 208 CLR 593; *R v Cook* [2004] NSWCCA 52; *R v Shamouil* (2006) 66 NSWLR 228; *Phillips v The Queen* (2006) 225 CLR 303; *R v XY* (2013) 84 NSWLR 363.

practitioner is proficient with procedures known to work.<sup>19</sup> Opinions produced using scientific, medical and technical procedures cannot be rationally assessed unless the procedures have been subjected to formal evaluation. Trial safeguards, such as s 137 (and *Christie* in common law jurisdictions), intended to prevent unfair prejudice to the accused, are rendered impotent when prosecutors and trial judges do not engage with actual probative value derived through formal studies.<sup>20</sup> In such cases, issues of validity and reliability fall to be contested by lawyers and evaluated by laypersons as part of an adversarial proceeding. Our accusatory system — and here we should not overlook the system's heavy reliance on plea (and charge) bargains and the limited resourcing available to most defendants — has not proven capable of consistently identifying and conveying fundamental problems with new procedures (eg facial mapping, voice comparison and forensic gait analysis) let alone the many untested, or insufficiently tested, procedures (eg firearm, shoeprint, tyre, fibre and document comparisons) routinely used by investigators and adduced by prosecutors.<sup>21</sup>

To the extent that trial judges are concerned with the ability of opinion evidence to 'rationally affect the assessment of the probability of the existence of a fact in issue', they must be constrained by the availability of *knowledge* derived through the formal evaluation of procedures. Whether they like it or not, and regardless of whether they understand the issues, when it comes to the rational determination of the probative value of forensic science evidence, trial judges are dependent on evidence of validity, scientific reliability and proficiency.<sup>22</sup>

In the fraught tradition of Daedalus, this article constitutes a warning. Australian courts must focus attention on the reliability of opinions based on specialised knowledge. Attempts to determine the probative value of scientific, medical and technical evidence without considering reliability risk flying too close to the sun.

<sup>19</sup> The general insights and experiences available to the tribunal of fact, useful for assessing much ordinary evidence, do not enable the tribunal to make appropriate assessments of opinions based on specialised knowledge.

<sup>20</sup> *Christie* (n 18) is the common law forebear to s 137. See also *R v Fletcher* (1953) 53 SR (NSW) 70, 76; *Driscoll v The Queen* (1977) 137 CLR 517, 541; *Bunning v Cross* (1978) 141 CLR 54; *Alexander v The Queen* (1981) 145 CLR 395; *Cleland v The Queen* (1982) 151 CLR 1; *Harriman v The Queen* (1989) 167 CLR 590; *Doney v The Queen* (1990) 171 CLR 207, 212; *Sandford* (1994) 72 A Crim R 160, 178.

<sup>21</sup> See, eg, *Morgan v The Queen* [2016] NSWCCA 25; *Tran v The Queen* [2016] VSCA 79.

<sup>22</sup> See Gary Edmond, 'Forensic Science Evidence and the Conditions for Rational (Jury) Evaluation' (2015) 39 *Melbourne University Law Review* 77.

## II *IMM* V *THE QUEEN* AND DETERMINING PROBATIVE VALUE FOR THE PURPOSE OF SECTION 137

The vexed question of how trial judges should approach probative value was recently considered by the High Court in *IMM*.<sup>23</sup> While *IMM* was not concerned with scientific, medical or technical evidence, the decision appears set to structure the way s 137 applies to opinions based wholly or substantially on specialised knowledge (admitted via s 79). Rather than repeat the divergent approaches to statute and the common law, this section endeavours to capture what the various judgments require of trial judges.

The basic issue at the centre of *IMM* was summarised as follows:

Where the tribunal of fact is a jury, is a judge determining probative value required to assume that the jury would find the evidence to be credible and otherwise reliable and to assess, on that assumption, the extent to which the jury could rationally infer from the evidence that a fact in issue is more or less probable? Alternatively, is the judge required to examine whether the jury could rationally find evidence to be credible and otherwise reliable as a step in determining the extent to which the jury could rationally infer from the evidence that the fact in issue is more or less probable?<sup>24</sup>

### *A Too Close to the Sun: Probative Value at Some Imagined 'Highest'*

According to French CJ, Kiefel, Bell and Keane JJ (the majority), determining the probative value of evidence for the purposes of ss 97 and 137 requires 'an assessment of the probative value of the evidence tendered'.<sup>25</sup> However, for the majority, the trial judge's task does not end with that (preliminary) assessment. In order to identify a probative value for the balancing exercise mandated by s 137, the trial judge must identify and use the highest probative value that the contested evidence can support. This expectation is repeated throughout the joint judgment:

The assessment of 'the extent to which the evidence could rationally affect the assessment of the probability of the existence of a fact in issue' requires that the possible use to which the evidence might be put, which is to say how it might be used, be taken at its highest. ...

<sup>23</sup> *IMM* (n 17). Cf *Dupas* (n 18); *XY* (n 18).

<sup>24</sup> *IMM* (n 17) 322 [84] (Gageler J).

<sup>25</sup> *Ibid* 313 [42]. This also applies to 'probative value' in ss 98, 101 and 137.

[T]he requisite probative value of the evidence is not spelled out in s 137. It requires the ‘probative value’ of the evidence to be weighed against the danger of unfair prejudice to the defendant. This again requires that the evidence be taken at its highest in the effect it could achieve on the assessment of the probability of the existence of the facts in issue.<sup>26</sup>

The *Uniform Evidence Law* defines the ‘credibility’ of a witness as ‘the credibility of any part or all of the evidence of the witness, and includes the witness’s ability to observe or remember facts and events about which the witness has given, is giving or is to give evidence.’<sup>27</sup> The reliability of evidence is not defined, but Gageler J equates reliability with trustworthiness: ‘evidence that is trustworthy is evidence that is “reliable”’.<sup>28</sup> For the majority, the commitment to taking probative value at its highest appears to preclude any inquiry into credibility or reliability:

Once it is understood that an assumption as to the jury’s acceptance of the evidence must be made, it follows that no question as to credibility of the evidence, or the witness giving it, can arise. For the same reason, no question as to the reliability of the evidence can arise.<sup>29</sup>

The majority also question the ability to clearly demarcate reliability and credibility:

The *Evidence Act* itself creates a difficulty in separating reliability from credibility. The definition of ‘credibility’, which concerns both a person who has made a representation that has been admitted into evidence and a witness, includes the person’s or witness’s ‘ability to observe or remember facts and events’ relevant to the representation or their evidence. These are matters which go to the reliability of the evidence.<sup>30</sup>

For the majority, taking probative value at its highest is intended to prevent the trial judge from trespassing on the fact-finding prerogative of the jury by excluding relevant evidence.<sup>31</sup> By assuming that the evidence is reliable and

<sup>26</sup> Ibid 313 [44], 317 [47]. See also at 314–15 [50]–[51].

<sup>27</sup> *UEL* (n 1) s 3 (definition of ‘credibility’).

<sup>28</sup> *IMM* (n 17) 321 [82]; see also at 330 [114], 343 [152] (Nettle and Gordon JJ); *Dupas* (n 18) 253–5 [260]–[266].

<sup>29</sup> *IMM* (n 17) 315 [52].

<sup>30</sup> Ibid 315–16 [53]; see also at 347–8 [164] (Nettle and Gordon JJ). Cf *Dupas* (n 18) 253–5 [257]–[265].

<sup>31</sup> See also *Shamouil* (n 18) 237–8 [63]–[65]; *XY* (n 18) 400 [167]–[171].

the witness credible, any potential difference in assessment between the trial judge's impression and the weight that a jury might attribute will not necessarily lead to the exclusion of evidence before the jury has had the chance to consider it. Taking the evidence at its highest is intended to prevent a trial judge from pre-emptively excluding evidence on the basis of personal doubts about the value of the evidence where those doubts might not be shared by a jury.

For the majority, it is fundamental to determine whether the contested evidence, taken at its highest, could rationally affect fact-finding:

This assumption necessarily denies to the trial judge any consideration as to whether the evidence is credible. Nor will it be necessary for a trial judge to determine whether the evidence is reliable, because *the only question is whether it has the capability, rationally, to affect findings of fact.*<sup>32</sup>

This is important, and perhaps vitally important with respect to opinions based on specialised knowledge. It raises the question of how a trial judge should determine the probative value of such opinions without knowing about the reliability (or trustworthiness) of the evidence. Closely related is the question of how a trial judge should determine the highest probative value of scientific, medical and technical forms of evidence.

In rationalising its approach to probative value and the prohibition on trial judges encroaching on fact-finding, the majority attach significance to the fact that s 137 refers to neither credibility nor reliability:

The *Evidence Act* contains no warrant for the application of tests of reliability or credibility in connection with ss 97(1)(b) and 137. The only occasion for a trial judge to consider the reliability of evidence, in connection with the admissibility of evidence, is provided by ss 65(2)(c) and (d) and 85. It is the evident policy of the Act that, generally speaking, questions as to the reliability or otherwise of evidence are matters for a jury, albeit that a jury would need to be warned by the trial judge about evidence which may be unreliable pursuant to s 165.<sup>33</sup>

Though accurate as a literal description of the text of the Act, the Court has yet to determine whether s 79(1) requires trial judges to consider the reliability of opinions based on 'specialised *knowledge*' in criminal proceedings.<sup>34</sup> This remains a significant live issue because opinions based on specialised

<sup>32</sup> *IMM* (n 17) 312 [39] (emphasis added); see also at 307 [18].

<sup>33</sup> *Ibid* 316 [54]; see also at 306–7 [17].

<sup>34</sup> Obiter in *IMM* might not be promising, but the High Court has yet to decide on whether s 79(1) requires the trial judge to consider reliability. Cf *Daubert* (n 8) 590.

knowledge need to satisfy the terms of s 79 before s 137 can be engaged. Section 79(1) requires the proponent of opinion evidence to identify the ‘knowledge’ underpinning the proffered opinion. Requiring trial judges to engage with knowledge (and therefore reliability) in s 79(1) is not only consistent with the scheme of the uniform legislation; simultaneously it will enable lawyers and trial judges to assess probative value (at its highest).

With respect to ‘the danger of unfair prejudice to the defendant’<sup>35</sup> — the other side of the ‘balance’ — the majority has much less to say. The joint judgment offers limited insight into the danger of unfair prejudice or how a trial judge should balance the putative ‘incommensurables.’<sup>36</sup>

### B *Between the Sun and the Sea: Probative Value among the Dissentients*

Three judges disagreed with the interpretation of s 137 advanced by the majority. Gageler J wrote a single judgment and Nettle and Gordon JJ joined in a separate judgment. These judges are all open to the trial judge taking reliability and credibility into account when determining the probative value of contested evidence.<sup>37</sup>

#### 1 *Probative Value at its Actual ‘Highest’*

Gageler J summarises the two basic approaches and their implications:

On neither approach is the judge required to do more than make an assessment of the extent to which the jury ‘could’ rationally infer from the evidence that a fact in issue was more or less probable.

... The judge’s assessment of probative value is an assessment of the maximum potential for the evidence rationally to affect the jury’s assessment of the probability of the existence of a fact in issue. The judge has to ask: how much is the evidence rationally capable of contributing to the jury’s assessment that the existence of that fact is more or less probable?

The difference between the two approaches concerns what is or can be involved in assessing *the highest use* to which the evidence is rationally capable of being put by the jury. On one approach, the reliability of the evidence must be taken as given. On the other approach, the reliability of the evidence forms part of the assessment. But on either approach, the assessment to be made by the

<sup>35</sup> *UEL* (n 1) s 137.

<sup>36</sup> See *Bendix Autolite Corp v Midwesco Enterprises Inc*, 486 US 888, 897 (1988); *Pfennig v The Queen* (1995) 182 CLR 461, 528; *IMM* (n 17) 329 [109] (Gageler J).

<sup>37</sup> *IMM* (n 17) 325 [95]–[96] (Gageler J), 336–7 [139]–[140] (Nettle and Gordon JJ).

judge remains an assessment of how much the evidence is rationally capable of contributing to the jury's assessment that the existence of a fact in issue is more or less probable.<sup>38</sup>

The entire Court agrees that the question of the *weight* a jury might attach to the contested evidence is irrelevant. Rather, the concern of the trial judge is with the ability of the evidence to rationally influence the assessment of facts in issue. Both of the 'approaches' in the extracted passage require the trial judge to determine 'how much the evidence is rationally capable of contributing to the jury's assessment that the existence of a fact in issue is more or less probable'.<sup>39</sup> This requires the trial judges to ask:

[H]ow much is that testimony rationally capable of contributing to the jury's assessment that the existence of a fact in issue is more or less probable? Performance of that assessment necessitates identification of the fact in issue and of the steps by which it would be open to the jury to reason from the testimony to a conclusion that the existence of that fact is more or less probable.<sup>40</sup>

Undertaking that assessment entails the trial judge determining what a jury might legitimately do with the evidence.

Gageler J's preferred approach — 'the other approach' in the extracted passage — requires the trial judge to consider the reliability of the evidence:

The conceptual framework which the statutory language erects therefore admits of the possibility that relevant evidence will lack probative value because it is not reliable.

... The legislative design was that probative value would involve an assessment of reliability and that relevance would not.<sup>41</sup>

In seeking to determine the highest probative value the evidence can support, according to Gageler J a trial judge should consider the reliability of the evidence and the credibility of the witness. For Gageler J, the determination of probative value at its highest is constrained by consideration of reliability and credibility.<sup>42</sup>

<sup>38</sup> Ibid 323–4 [89]–[91] (emphasis added) (citations omitted).

<sup>39</sup> Ibid 323 [91].

<sup>40</sup> Ibid 326 [99].

<sup>41</sup> Ibid 326 [96]–[97] (citations omitted). Gageler J adopts the reasoning of McHugh J from *Papakosmas* (n 18) 323 [86]: at 325–6 [96]. The majority, following Gaudron J in *Adam* (n 18) 115 [60] and Spigelman CJ in *Shamouil* (n 18) 237–8 [63]–[64], adopts a different course: at 309 [27].

<sup>42</sup> *IMM* (n 17) 323–4 [90]–[91], 324 [93], 326–7 [99].

In thinking about the practical implications of the two approaches, Gager J suggests that differences are only likely to emerge in ‘an extreme case’:

Once it is borne in mind that the judge’s assessment concerns *the highest use* to which the evidence is capable of being put by the jury, it is difficult to see significance in the difference between the two approaches other than in an extreme case where the judge is able to determine at the time evidence is sought to be adduced that it would not be open to the jury rationally to find that evidence to be reliable.<sup>43</sup>

## 2 (Actual) Probative Value

Also favouring ‘the other approach’, Nettle and Gordon JJ are committed to the trial judge engaging with reliability and credibility when determining the probative value of the evidence. They explain:

Evidence cannot affect the assessment of the probability of the existence of a fact in issue unless the evidence is rationally capable of being accepted. Hence, to determine whether evidence has the capacity rationally to affect the assessment of the probability of the existence of a fact in issue requires a determination of whether the evidence is rationally capable of acceptance. And for the court to determine whether it thinks that evidence is rationally capable of acceptance requires the court, among other things, to determine whether it thinks that the degree of reliability which it would be open to the jury rationally to attribute to the evidence is such that it will be open to the jury rationally to accept the evidence. ...

[B]oth ss 97 and 137 should be construed such that both credibility and reliability are relevant considerations in determining whether evidence is of such probative value as not to be outweighed by the danger of unfair prejudice to the defendant.<sup>44</sup>

A potential point of departure from the position advocated by Gager J is that Nettle and Gordon JJ do not refer to taking the evidence at its highest. Nettle and Gordon JJ probably support taking the evidence at its highest in a process that incorporates consideration of reliability and credibility, but their decision might also be read as expecting trial judges to make an assessment

<sup>43</sup> Ibid 324 [93] (emphasis added); see also at 343 [152], 343–4 [154] (Nettle and Gordon JJ).

<sup>44</sup> Ibid 337 [140], 348 [165].

of actual probative value and to use that value when undertaking the balancing exercise.<sup>45</sup>

In explaining their approach to probative value, Nettle and Gordon JJ refer to the role played by s 137 in ensuring that the accused receives a fair trial:

In view of the critical importance of s 137 in ensuring that an accused receives a fair trial, such difficulties as might attend those procedures are insufficient to adopt a construction of s 137 that excludes consideration of the reliability of evidence.<sup>46</sup>

Nettle and Gordon JJ dismiss the contention, advanced by the majority, that engaging with reliability represented an illegitimate encroachment on jury prerogatives:

Such an assessment is not in any sense a usurpation of the jury's function. It is the discharge of the long recognised duty of a trial judge to exclude evidence that, because of its nature or inherent frailties, could cause a jury to act irrationally either in the sense of attributing greater weight to the evidence than it is rationally capable of bearing or because its admission would otherwise be productive of unfair prejudice which exceeds its probative value.<sup>47</sup>

### *C Seeing in the Fog*

French CJ, Kiefel, Bell and Keane JJ and Gageler J require the trial judge to take the evidence 'at its highest' when determining probative value. However, the majority proscribed consideration of reliability and credibility.<sup>48</sup> Gageler J, in contrast, favours the trial judge assessing reliability and credibility when determining 'the highest use to which the jury could rationally put the testimony.'<sup>49</sup> Nettle and Gordon JJ expect the trial judge to determine the probative value of the evidence (probably at its highest), informed by consid-

<sup>45</sup> For reasons developed in this article, little turns on this difference. I have retained reference to actual probative value in order to help with explanations acknowledging that these dissentients are probably committed, like Gageler J, to taking the evidence 'at its highest' once reliability and credibility have been considered.

<sup>46</sup> *IMM* (n 17) 345 [156] (citations omitted).

<sup>47</sup> *Ibid* 346 [161].

<sup>48</sup> The majority notes the possibility of exceptions: *ibid* 316–17 [57]–[58]. However, saying that 'evidence which is inherently incredible or fanciful or preposterous would not appear to meet the threshold requirement of relevance', the majority suggests that these unspecified exceptions might be narrowly conceived; see also at 312 [39].

<sup>49</sup> *Ibid* 327 [99].

eration of reliability and credibility. This approach is presented as integral to s 137 functioning as a trial safeguard.

*Table 1: Summary of the Various Approaches to Probative Value, Reliability and Credibility*

Approaches to the contested evidence	French CJ, Kiefel, Bell and Keane JJ	Gageler J	Nettle and Gordon JJ
Try to determine the probative value that the jury might attribute to the evidence	No: [18], [28], [30], [39]	No: [88]	No: [166]
Determine the extent to which the evidence could rationally affect the assessment of the probability of the existence of a fact in issue (actual probative value)	Yes: [42], [48]	Yes: [89], [90], [99]	Yes: [140], [160], [162], [164], [165], [172]
Take the probative value at its highest	Yes: [44]	Yes: [90], [93]	No mention, but see [176]
Consider the reliability and credibility of the evidence	No: [17], [52], [54] (except in exceptional cases: [39], [57]–[58])	Yes: [96], [97]	Yes: [139], [140], [160]
Consider whether the evidence is weak or unconvincing	Yes: [50] (may form part the determination)	See previous answers and [92]	See previous answers

The clarity of this summary (see also Table 1), and the deceptive simplicity of the majority's approach, is jeopardised by a beguiling example incorporated within the majority's judgment. The majority deploys an example involving eyewitness evidence, adopted from a speech by Dyson Heydon.<sup>50</sup> That example purports to exemplify how probative value at its highest might be evaluated. It refers to factors that a trial judge might consider in relation to the

<sup>50</sup> Ibid 315 [50]–[51], citing JD Heydon, 'Is the Weight of Evidence Material to Its Admissibility?' (2014) 26 *Current Issues in Criminal Justice* 219, 234.

assessment of the probative value of the identification evidence of an eyewitness. The example is concerned with an identification in sub-optimal conditions. It raises (or perhaps begs) questions about reliability and the kinds of insights and 'knowledge' that a judge might entertain about some kinds of evidence. Following Heydon's lead, in their attempt to explain the probative value of the eyewitness evidence at its highest, the majority consider the identification to be 'unconvincing'.<sup>51</sup> The circumstances in which the observation was made (ie in foggy conditions, in bad light, by a stranger) led them to find the identification to be 'weak'.<sup>52</sup>

In coming to that conclusion, both Heydon and the majority incorporate factors that bear directly upon the reliability of the eyewitness evidence.<sup>53</sup> The eyewitness evidence is assessed subject to the specific environmental conditions, judicial 'common sense' and epistemic threats (such as scientific research identifying the increased risk of error associated with cross-racial identifications).<sup>54</sup> On its face, determining that the eyewitness identification was weak or unconvincing, by attending to a number of context-related factors, is not easily reconciled with the majority's explicit rejection of reliability (and credibility). This is because reliability factors conspicuously intrude into the assessment of probative value. In finding the eyewitness evidence to be, 'at its highest', 'weak', Heydon and the majority unilaterally and opaquely invoke and apply reliability-based considerations.<sup>55</sup>

The discussion and analysis following this section explains why trial judges, to the extent that they are operating in the 'rationalist tradition', must attend to the reliability (and validity) of contested forensic science and medicine evidence.<sup>56</sup> Now we turn to consider how to determine the probative

<sup>51</sup> Ibid 315 [50].

<sup>52</sup> Ibid. Neither Heydon nor the majority specify or seek to specify a highest probative value beyond 'weak' or 'unconvincing'. The original example was slightly more detailed and included a racial dimension: Heydon (n 50) 234.

<sup>53</sup> *IMM* (n 17) 315 [50].

<sup>54</sup> Significantly, this was not all derived through endogenous legal awareness. The error-prone nature of strangers and cross-racial identifications was revealed by scientific research. We might say the same about the deleterious impact of stress, weapon focus and so forth. The invocation of insights, whether as common sense or science-based, illustrates the problems with *Aytugrul* (n 15). These, after all, are adjudicative facts, at least.

<sup>55</sup> *IMM* (n 17) 324 [92] (Gageler J).

<sup>56</sup> See generally William Twining, 'The Rationalist Tradition of Evidence Scholarship' in William Twining, *Rethinking Evidence: Exploratory Essays* (Cambridge University Press, 2<sup>nd</sup> ed, 2006) 35.

value of opinions based on specialised knowledge, particularly forensic science evidence.

### III THE PROBATIVE VALUE OF SCIENTIFIC, MEDICAL AND TECHNICAL EVIDENCE

Whatever might be thought about the majority's interpretation of s 137 and its application to ordinary witnesses, the strict proscription against considering reliability and credibility is not suited to attempts to determine the probative value of opinions based on specialised knowledge.<sup>57</sup> This section explains what is required to gauge the probative value of scientific, medical and technical forms of evidence. It also aims to illustrate why interest in validity and reliability is unavoidable if we intend to rationally engage with the probative value of opinions based on specialised knowledge in criminal proceedings.

#### *A Determining the Probative Value of Opinions Based on Specialised Knowledge*

There are two very important questions that *should* underlie the law's admission of and reliance upon forensic evidence in criminal trials: (1) the extent to which a particular forensic discipline is founded on a reliable scientific methodology that gives it the capacity to accurately analyze evidence and report findings and (2) the extent to which practitioners in a particular forensic discipline rely on human interpretation that could be tainted by error, the threat of bias, or the absence of sound operational procedures and robust performance standards.<sup>58</sup>

<sup>57</sup> The Supreme Court of the United States suggested that expert opinion evidence should be distinguished from other forms of evidence. In *Daubert*, the majority endorsed the position of Jack B Weinstein, the veteran judge and evidence scholar who oversaw the Agent Orange litigation. Weinstein stated that '[e]xpert evidence can be both powerful and quite misleading because of the difficulty in evaluating it. Because of this risk, the judge in weighing possible prejudice against probative force under Rule 403 of the present rules exercises more control over experts than over lay witnesses': *Daubert* (n 8) 595, quoting Jack B Weinstein, 'Rule 702 of the Federal Rules of Evidence Is Sound; It Should Not Be Amended', 138 FRD 631, 632 (1991). Even if not considered epistemologically exceptional, there may be compelling practical reasons to distinguish scientific, medical and technical opinions from other types of evidence: see generally Geoffrey Bowker and Susan Star, *Sorting Things Out: Classification and Its Consequences* (MIT Press, 1999).

<sup>58</sup> Committee on Identifying the Needs of the Forensic Science Community, National Research Council et al, *Strengthening Forensic Science in the United States: A Path Forward* (National Academies Press, 2009) 87 (emphasis in original).

[V]alid scientific knowledge can *only* be gained through *empirical* testing of specific propositions.<sup>59</sup>

With these epigraphs in mind, it is useful to begin by listing factors that are not reliable guides to the assessment of the probative value of opinions based on specialised knowledge. Probative value is not determined by speculation, the impressions of lawyers and judges, institutional ‘common sense’, long use of a procedure, a witness’s confidence and demeanour, the ability to survive cross-examination, formal training, study and experience, certification and accreditation, or the apparent strength of the case.<sup>60</sup> The probative value of opinions based on specialised knowledge, at its highest or otherwise, is not illuminated by a trial judge finding the evidence plausible, convincing or compelling. Similarly, the fact that other judges have found an opinion (or procedure) to be probative by itself reveals nothing about probative value or ‘the extent to which the evidence could rationally affect the assessment of the probability of the existence of a fact in issue’. Reliance on such factors — both individually and even in combination — is misguided and may produce misleading assessments of probative value in circumstances that are not, to adopt Gageler J’s terminology, ‘extreme’.<sup>61</sup>

The probative value of scientific, medical and technical procedures can only be rationally gauged through processes of formal evaluation<sup>62</sup> — that is, through studies designed to test procedures and/or the abilities of forensic practitioners in conditions where the correct answer is known.<sup>63</sup> In the absence of such studies, claims about the probative value of opinions based on specialised knowledge are speculative. In order to determine the probative value of an opinion based on specialised knowledge we need to *know* whether the procedure does what it is supposed to do, how well, and in what conditions. Such testing is often described by scientists as *validation*. Validation provides the basis for standardised procedures and protocols. It can also provide an indication of accuracy, error and limitations, and these contribute to empirically-based protocols and forms of expression for reporting. To put

<sup>59</sup> President’s Council of Advisors on Science and Technology, *Forensic Science in Criminal Proceedings: Ensuring Scientific Validity of Feature Comparison Methods* (Report to the President, September 2016) 46 (emphasis in original) (citations omitted). For an accessible overview of the report, see Gary Edmond and Kristy Martire, ‘Forensic Science in Criminal Courts: The Latest Scientific Insights’ (2016) 42 *Australian Bar Review* 367.

<sup>60</sup> See *Forensic Science in Criminal Proceedings* (n 59) 32–3.

<sup>61</sup> *IMM* (n 17) 324 [93].

<sup>62</sup> Cf *R v Jung* [2006] NSWSC 658, [78].

<sup>63</sup> *Forensic Science in Criminal Proceedings* (n 59) 32–3.

this another way, in order to determine probative value, and probative value at its highest, there is a need to attend to the ‘knowledge’ derived through formal processes of evaluation. To adopt the terminology of the High Court in *Honeysett*, that involves an ‘acquaintance with facts, truths, or principles, as from study or investigation’.<sup>64</sup>

In addition, we should *know* if the individual proffering the particular opinion is proficient with the validated procedure or has demonstrated the claimed ability in controlled conditions.<sup>65</sup> Scientists are typically less concerned with formal qualifications and experience than evidence of proficiency with a validated procedure or a demonstrable ability.<sup>66</sup> Ability is typically demonstrated through accurate performance (against answers known to be correct) or a heightened level of performance relative to non-experts. Study, training, formal qualifications and even experience with a specific procedure are not substitutes for formal evaluation and do not guarantee expertise, even where a person is experienced with a specific procedure.<sup>67</sup>

Only formal evaluation produces the kind of knowledge required to make sense of procedures and claimed abilities. This is why biomedical researchers and engineers test pharmaceuticals, therapeutics, materials and designs. If the impressions of judges provided a useful surrogate for formal evaluation, researchers might ask their opinions. Predictably, they do not. Biomedical researchers, engineers and other scientists do not even rely on their own experience, expectations or beliefs to assess efficacy and safety.<sup>68</sup> In the same vein, we do not typically rely on extrapolation from similar (but different) drugs or similar (but different) designs.<sup>69</sup> Rather, scientists and engineers are systematically engaged in elaborate and costly experiments. Methodologically

<sup>64</sup> *Honeysett* (n 8) 131 [23] (emphasis altered), quoting *Macquarie Dictionary* (rev 3<sup>rd</sup> ed, 2001), ‘knowledge’ (def 1).

<sup>65</sup> *Forensic Science in Criminal Proceedings* (n 59) 57–9.

<sup>66</sup> Kristy Martire and Gary Edmond, ‘Rethinking Expert Opinion Evidence’ (2017) 40 *Melbourne University Law Review* 967, 984.

<sup>67</sup> *Ibid.* Opinions based *primarily* on ‘training, study or experience’ are not admissible via the exception to opinion evidence provided by s 79(1); see K Anders Ericsson, ‘The Influence of Experience and Deliberate Practice on the Development of Superior Expert Performance’ in K Ericsson et al (eds), *The Cambridge Handbook of Expertise and Expert Performance* (Cambridge University Press, 2006) 683, 685–705.

<sup>68</sup> Hypotheses and the imagination are acceptable for formulating research questions, but not for answering them.

<sup>69</sup> In relation to biomedical research, there has been a shift in assumptions about the applicability of the results of clinical trials to those who were not historically studied, such as women, children, the aged and non-Europeans: see generally Steven Epstein, *Inclusion: The Politics of Difference in Medical Research* (University of Chicago Press, 2007).

complicated, these studies typically involved elaborate efforts to keep suggestive (ie potentially biasing) information away from human participants — as in double-blind clinical trials.<sup>70</sup>

Legal assessment of the probative value of opinion based on specialised knowledge — whether by judges or juries — cannot be based on a guess or impression, however reasonable such a guess or impression might appear or be made to appear. Following *IMM*, the High Court requires trial judges to determine probative value at its highest when applying ss 135 and 137. Legal assessment of the probative value of opinions based on specialised knowledge must be informed by relevant evidence (ie knowledge) rather than judicial impressions or past practices. Formal evaluation and the knowledge it produces marks the boundaries of probative value. Assigning values significantly beyond the scope of what formal evaluation establishes, or might establish if it was actually conducted, is illegitimate.<sup>71</sup> It is unavoidably speculative. Where the value of procedures and abilities is unknown (ie has not been formally assessed), the risks of overvaluation, misuse and misunderstanding are legion. Without insight into the value of procedures derived through formal evaluation or demonstrable evidence of ability, the link to rationality is sacrificed.

Most of the procedures used in forensic science and forensic medicine can be formally evaluated; however, '[l]ittle rigorous systematic research has been done to validate the basic premises and techniques in a number of forensic science disciplines.'<sup>72</sup> There are few excuses for the lack of formal studies and inattention to evaluation and 'knowledge'. Where procedures are in routine or widespread use (eg ballistics and tool marks, latent fingerprints, foot and shoeprints, tyre marks, image and voice comparison, blood spatter analysis, DNA profiling, document comparison, crash reconstruction and arson investigation) they must be formally evaluated. Procedures and protocols should avoid notorious forms of contamination and bias. Reports and testimony should include error rates, uncertainties and limitations.

<sup>70</sup> See, eg, the concerns about testing and the placebo effect in R Barker Bausell, *Snake Oil Science: The Truth about Complementary and Alternative Medicine* (Oxford University Press, 2007). For a revealing example from gravitational wave research, see Harry Collins, *Gravity's Ghost: Scientific Discovery in the Twenty-First Century* (University of Chicago Press, 2010).

<sup>71</sup> *Forensic Science in Criminal Proceedings* (n 59) 3.

<sup>72</sup> *Strengthening Forensic Science in the United States* (n 58) 189.

### B *States of Knowledge and Ignorance*

Directing attention to the underlying knowledge enables us to gauge the probative value of forensic science and medicine evidence. When it comes to considering the evidence supporting a procedure or claimed ability, there are three basic categories — really ideal types.

First, where a procedure has been formally evaluated and the opinion is based squarely within the standard operation or protocol, we have a good idea of its probative value. Where a procedure has been validated, for example, we know how well it performs in specific conditions. Where individuals are proficient with validated procedures, we have a reasonable idea of accuracy based on standardised testing. That is, we know about strengths and some weaknesses and this assists with the presentation and evaluation of the opinion evidence. It may be that a particular forensic practitioner is, in addition, very experienced with the specific procedure and possesses apparently relevant formal qualifications. However, unless there is evidence demonstrating that these *supplementary factors* translate into enhanced performance, claims about superior performance and a heightened probative value remain speculative.<sup>73</sup>

Secondly, where a procedure has been formally evaluated, but the opinion extends beyond the results of the evaluation or practices associated with validation, the probative value of the opinion becomes uncertain. Similarly, where a procedure or ability has been evaluated, but the specific way the procedure was used in the case has not, the probative value of the resulting opinion may be uncertain. In this case, the trial judge must carefully consider the significance of any differences and whether the procedure has been extended too far (beyond the conditions of known validity). Unfortunately, this process is more complicated and vulnerable to error than it might (at first) appear,<sup>74</sup> for expertise is not generally transferable and even activities and abilities that appear similar might in reality be quite different.<sup>75</sup> There may be embarrassing surprises when those claiming special abilities are

<sup>73</sup> See Part III(B). The claim that the forensic practitioner has never made a mistake is unhelpful as it is misguided and misleading. How would they know? What procedures are in place to actually 'catch' errors? Assertions that there have been few or no errors tend to reveal more about the culture and (lack of) methodological sophistication than the accuracy of specific opinions.

<sup>74</sup> This process may resemble 'articulation work': see Anselm Strauss, 'The Articulation of Project Work: An Organizational Process' (1988) 29 *Sociological Quarterly* 163.

<sup>75</sup> Jean Bédard and Michelene TH Chi, 'Expertise' (1992) 1 *Current Directions in Psychological Science* 135, 138–9.

rigorously tested in controlled conditions.<sup>76</sup> The point to take away is that any attribution of validity or ability (or credibility) that is not based on knowledge derived through appropriate forms of evaluation dramatically increases the risk that the ability does not exist or that the opinion will be exaggerated or mistaken and consequently overvalued or misused.

The third category is composed of opinions based on procedures or abilities that have not been formally evaluated. These are the most difficult because, notwithstanding their potential appeal, in the absence of formal evaluation we do not *know* if the procedure works, nor how well, nor in what circumstances. We do not *know* if the evidence is capable of ‘rationally affect[ing] the assessment of the probability of the existence of a fact in issue,’ nor ‘the extent’ of any influence.<sup>77</sup> This may strike the reader as alarmist. Yet only formal evaluation provides the kinds of information that enable judges (and others) to determine probative value — recall the epigraphs.<sup>78</sup> Long institutional use, legal recognition and reliance, the experience of the ‘expert’, previous convictions, the existence of a field, plausibility and even standards, accreditation, disinterest and apparent impartiality do not provide direct insight into probative value and should not be substituted for formal evaluation.<sup>79</sup> They do not enable a scientist or a trial judge to rationally determine if, or the extent to which, an opinion might rationally influence the assessment of facts in issue. Traditional legal proxies do not provide appropriate insight into the probative value of procedures and abilities. Rather, their use tends to disguise judicial recourse to impressions and speculation.

Lest these oversights be considered hypothetical or exaggerated, it is helpful to list a few of the growing number of procedures once admitted and relied upon but now discarded by various common law courts.<sup>80</sup> After formal

<sup>76</sup> See, eg, Michael J Saks et al, ‘Forensic Bitemark Identification: Weak Foundations, Exaggerated Claims’ (2016) 3 *Journal of Law and the Biosciences* 538, 553–4.

<sup>77</sup> It may influence beliefs, but the rational decision-maker ought to take the lack of evidence about validity and reliability to be a significant reason that goes against believing the opinion to reflect what Andrew Roberts would describe as ‘[t]he truth of the matter’: Andrew Roberts, ‘Expert Evidence on the Reliability of Eyewitness Identification: Some Observations on the Justifications for Exclusion’ (2012) 16 *International Journal of Evidence and Proof* 93, 99. Without this evidence there may not be very good grounds for believing the opinion. The risk that the jury will not take the absence of this evidence to be a significant reason against believing the opinion goes to unfair prejudice.

<sup>78</sup> *Forensic Science in Criminal Proceedings* (n 59) 6.

<sup>79</sup> *Ibid* 55.

<sup>80</sup> For the most recent authoritative expression of concern about longstanding forensic procedures, see *ibid* 25–39.

evaluation or wrongful convictions, the following practices have been abandoned or curtailed: voice spectroscopy;<sup>81</sup> bite mark comparisons;<sup>82</sup> ear print identification;<sup>83</sup> microscopic hair comparison;<sup>84</sup> bullet lead analysis (to match a bullet to a batch of bullets);<sup>85</sup> facial and body mapping;<sup>86</sup> the triad (to prove that harm has been deliberately inflicted on an infant);<sup>87</sup> and features such as V-shapes, pour patterns and puddling to identify arson.<sup>88</sup> Each of these procedures was relied upon in serious prosecutions notwithstanding the absence of support through appropriate formal evaluation. Each of them contributed to exaggerated or misleading opinions, unfair prosecutions and wrongful convictions.<sup>89</sup>

In the absence of formal studies there is no way for a scientist, lawyer, trial judge or jury to determine whether an anatomist can accurately identify persons of interests in CCTV images, tool and firearm specialists can accurately match shell casings to firearms, fingerprint examiners can accurately match latent fingerprints to reference prints, linguists or interpreters can accurately match a voice recording to a specific individual, and so on and so

<sup>81</sup> See Committee on Evaluation of Sound Spectrograms, National Research Council, *On the Theory and Practice of Voice Identification* (National Academy of Sciences, 1979).

<sup>82</sup> See Erica Beecher-Monas, 'Reality Bites: The Illusion of Science in Bite-Mark Evidence' (2009) 30 *Cardozo Law Review* 1369; Mary A Bush, Howard I Cooper and Robert BJ Dorion, 'Inquiry into the Scientific Basis for Bitemark Profiling and Arbitrary Distortion Compensation' (2010) 55 *Journal of Forensic Sciences* 976; Mark Page et al, 'Expert Interpretation of Bitemark Injuries: A Contemporary Qualitative Study' (2013) 58 *Journal of Forensic Sciences* 664.

<sup>83</sup> See *Expert Evidence in Criminal Proceedings in England and Wales* (n 12) 44–5 [3.118]–[3.124].

<sup>84</sup> See Federal Bureau of Investigation, 'FBI Testimony on Microscopic Hair Analysis Contained Errors in at Least 90 Percent of Cases in Ongoing Review' (Press Release, 20 April 2015) <[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](http://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)>, archived at <<https://perma.cc/J3GQ-9S5L>>.

<sup>85</sup> See Committee on Scientific Assessment of Bullet Lead Elemental Composition Comparison and Board on Chemical Sciences and Technology, National Research Council, *Forensic Analysis: Weighing Bullet Lead Evidence* (National Academies Press, 2004).

<sup>86</sup> See Gary Edmond et al, 'Law's Looking Glass: Expert Identification Evidence Derived from Photographic and Video Images' (2009) 20 *Current Issues in Criminal Justice* 337.

<sup>87</sup> See Deborah Tuerkheimer, *Flawed Convictions: 'Shaken Baby Syndrome' and the Inertia of Injustice* (Oxford University Press, 2014) ch 2. See generally Emma Cunliffe, *Murder, Medicine and Motherhood* (Hart Publishing, 2011).

<sup>88</sup> See John J Lentini, *Scientific Protocols for Fire Investigation* (CRC Press, 2<sup>nd</sup> ed, 2012) ch 8.

<sup>89</sup> See Brandon L Garrett, *Convicting the Innocent: Where Criminal Prosecutions Go Wrong* (Harvard University Press, 2011) ch 4.

forth. Complicating assessments, opinions might be presented in confident terms by a credentialed and experienced witness, situated in a specialist group operating within a prestigious investigative institution run by the state, alongside references to the witness's prior involvement in investigations and successful prosecutions. None of this, however, addresses the probative value of the opinion — whether, for example, microscopic hair comparison is useful for the purposes of identification.<sup>90</sup> It provides no insight into what a microscopist could legitimately opine, how confident she should be, or the number of times she might be mistaken while engaged in similar comparisons.

Real world examples will often be more refractory than the ideal types presented here. There may, for example, be multiple studies with inconsistent results.<sup>91</sup> Nevertheless, the question for lawyers, judges and juries is: What evidence supports the procedure or ability? This will almost always be answered by reference to *knowledge* — independent of the witness and identifiable — derived through some kind of validation or performance study. This is what prosecutors must assemble and trial judges consider when the defence objects to the admission of opinions purportedly based on specialised knowledge under s 137.

### *C Carried Away with Phaethon: What 'at its Highest' Cannot Mean*

When it comes to determining probative value at its highest we cannot assume that an opinion putatively based on specialised knowledge is correct or error-free. Such an approach negates our interest in knowledge and the extent to which an opinion can rationally influence decision-making. Moreover, it would have the tendency of overwhelming the balancing exercise in s 137. Where the opinion of a forensic scientist (such as a positive identification of a person) is assumed to be correct, then many of the dangers associated with exaggeration, overvaluation and error will be trivialised when it comes to considering the danger of unfair prejudice.<sup>92</sup> How do we balance the ubiquitous possibility of error, or the lack of formal evaluation, against an opinion assumed to be correct or error-free?

'At its highest' cannot mean that we assume that an untested procedure works or that an opinion is free from error. Though unreasonable, the claim

<sup>90</sup> Hair comparison might be useful for *excluding* persons of interest, but that is not how most hair comparison evidence has been used in common law courts.

<sup>91</sup> See *Forensic Science in Criminal Proceedings* (n 59) 67–123, where several of the studies traditionally relied upon by forensic practitioners were criticised for their poor design.

<sup>92</sup> See *Dupas* (n 18) 198–9 [72]–[73], 216 [131], 237 [206].

that we can make that assumption is not adequately denounced in jurisprudence around taking evidence ‘at its highest’.<sup>93</sup> Assuming that procedures work or are error-free would mean that untested procedures that a trial judge finds convincing might enter the balancing exercise on the basis that they are correct. Procedures that have never been formally evaluated might be accorded higher probative values than demonstrably reliable procedures. The probabilistic results produced by DNA profiling might, for example, be treated as less probative — because we know about its limitations — than a positive identification based on untested forensic gait analysis<sup>94</sup> or a cross-lingual voice comparison by an investigator.<sup>95</sup> Where the court is ignorant of scientific methods or the results of scientific research, a positive (or categorical) identification based on a discredited procedure, such as microscopic hair or bite mark comparison, might also be treated as correct. Inattention to reliability also increases inconsistency, placing trial judges and appellate courts at the mercy of the abilities, performances and resources available to the parties in specific proceedings.

Limitations and the ubiquitous risk of error are directly related to ‘the extent to which the evidence could rationally affect the assessment of the probability of the existence of a fact in issue’. Regardless of whether errors are (also) considered in relation to the danger of unfair prejudice to the defendant, they should be incorporated in the attempt to determine probative value at its highest.<sup>96</sup> The highest probative value of any opinion depends on consideration of limitations, whether caused by environment, procedure or personnel, or some combination of these. If we recall the *IMM* majority’s example of identification in foggy conditions, the highest probative value that the eyewitness evidence could support was not said to be a correct identification.<sup>97</sup> Probative value at its highest was not equated with error-free performance.

In a recent review of the forensic sciences, the US National Research Council (‘NRC’) directed forensic scientists to determine and disclose uncertainty, accuracy and error in the following terms:

<sup>93</sup> See, eg, *Tran* (n 21) [61].

<sup>94</sup> See, eg, *Otway v The Queen* [2011] EWCA Crim 3.

<sup>95</sup> See, eg, *Kheir v The Queen* (2014) 43 VR 308; *Tran* (n 21); *Nguyen v The Queen* [2017] NSWCCA 4.

<sup>96</sup> *IMM* (n 17) 314 [47] (French CJ, Kiefel, Bell and Keane JJ).

<sup>97</sup> *Ibid* 314–15 [50].

All results for every forensic science method [ie procedure] should indicate the uncertainty in the measurements that are made, and studies must be conducted that enable the estimation of those values. ...

[T]he accuracy of forensic methods resulting in classification or individualization conclusions needs to be evaluated in well-designed and rigorously conducted studies. The level of accuracy of an analysis is likely to be a key determinant of its ultimate probative value.<sup>98</sup>

Foot, shoe and tyre prints, latent fingerprints, ballistics, and some voice identifications are frequently presented as effectively error-free,<sup>99</sup> yet these and other procedures have been subject to sustained criticism by the NRC and numerous commentators.<sup>100</sup> Relevantly, the NRC indicated that claims about error-free performance are not sustainable. Commenting on latent fingerprint evidence (ie friction ridge analysis), the NRC explained:

Although there is limited information about the accuracy and reliability of friction ridge analyses, claims that these analyses have zero error rates are not scientifically plausible.<sup>101</sup>

Subsequent studies, led by Ulery and Tangen, confirmed the NRC's concerns.<sup>102</sup>

#### D *Reliability and Credibility and Probative Value*

It is useful to say a few words about the way reliability and credibility bear upon the assessment of probative value in this context. Reliability, in the sense of 'trustworthiness', should predominate in any assessment of the probative value of opinion based on specialised knowledge.<sup>103</sup> Logically, we should consider whether a procedure is known to work, or whether an individual

<sup>98</sup> *Strengthening Forensic Science in the United States* (n 58) 184.

<sup>99</sup> Simon A Cole, 'More than Zero: Accounting for Error in Latent Fingerprint Identification' (2005) 95 *Journal of Criminal Law and Criminology* 985, 1043.

<sup>100</sup> Michael J Saks and Jonathan J Koehler, 'The Coming Paradigm Shift in Forensic Identification Science' (2005) 309 *Science* 892, 895; *Strengthening Forensic Science in the United States* (n 58).

<sup>101</sup> *Strengthening Forensic Science in the United States* (n 58) 142.

<sup>102</sup> Bradford T Ulery et al, 'Accuracy and Reliability of Forensic Latent Fingerprint Decisions' (2011) 108 *Proceedings of the National Academy of Sciences* 7733; Jason M Tangen, Matthew B Thompson and Duncan J McCarthy, 'Identifying Fingerprint Expertise' (2011) 22 *Psychological Science* 995. See also *Forensic Science in Criminal Proceedings* (n 59) 87–103.

<sup>103</sup> *IMM* (n 17) 321 [82] (Gageler J).

actually possesses an enhanced ability (ie demonstrable expertise) relative to the tribunal of fact, before we attend to credibility. In determining probative value, we should be concerned with whether the procedure works, and if so, how well and in what conditions. We should also be concerned with the competence (or proficiency) of the practitioner with the specific procedure. In the case of DNA profiling, for example, before attending to whether the practitioner is a qualified biologist with a good reputation we would want to know that the procedure they used has been validated and that they are demonstrably proficient in the use of the specific protocol. These are reliability issues.

As the majority in *IMM* recognise, reliability issues are not always clearly demarcated from the credibility of a witness.<sup>104</sup> Consequently, although ordinarily subservient to questions of validity, scientific reliability and proficiency, the credibility of a witness may not be inconsequential in the assessment of probative value. Notwithstanding the primacy of reliability in relation to opinions based on specialised knowledge, assessment of the credibility of a witness might occasionally trump reliability. Where, for example, a witness is extremely partisan<sup>105</sup> or has had their performance questioned,<sup>106</sup> such considerations might be used to discount the probative value of opinion evidence (at its highest) even where the opinion appears to be based on the application of a sound procedure.<sup>107</sup> In determining probative value under s 137, just as we might want to know about the vision of an eyewitness, so too we might want to know about the reputation of an expert, particularly where they have been disciplined or censured by a professional body.

Assessment of credibility might be used to *decrease* probative value. However, credibility cannot be used to overcome the lack of formal evaluation. The

<sup>104</sup> Ibid 315–16 [53].

<sup>105</sup> See, eg, Associate Professor Cross in *Wood v The Queen* (2012) 84 NSWLR 581. McClellan CJ at CL indicated that the issue of an extremely partisan witness might be considered under s 137: at 620 [729].

<sup>106</sup> See, eg, Sergeant Cocks in *Royal Commission of Inquiry in Respect to the Case of Edward Charles Splatt* (Report, 1984) 338–41; Mrs Kuhl, Mr Brown and Professor Cameron in *Royal Commission of Inquiry into Chamberlain Convictions* (Report, 1987) 312–14, 324–6; Dr Sutisno in *Tang* (n 14); Professor Henneberg in *Morgan v The Queen* (2011) 215 A Crim R 33, 44–61 [71]–[146] (Hidden J); Dr Lawrence in *Gilham v The Queen* [2012] NSWCCA 131, [616]–[621] (McClellan CJ at CL, Fullerton and Garling JJ); Dr Manock in *R v Keogh [No 2]* (2014) 121 SASR 307.

<sup>107</sup> See generally Bibi Sangha and Robert N Moles, *Miscarriages of Justice: Criminal Appeals and the Rule of Law in Australia* (LexisNexis Butterworths, 2015) ch 9.

apparent believability, sincerity, confidence or reputation of a witness (eg Sir Roy Meadow or Sir Bernard Spilsbury) cannot overcome the absence of formal evaluation.<sup>108</sup> None of these are correlated (or they are only weakly correlated) to the probative value of the opinion. In the absence of formally evaluated procedures, imputed credibility cannot fill reliability gaps. We should not infer that a procedure is valid or that an individual has abilities on the basis of imputed credibility.

It bears stating that credibility tends to be impressionistic. It is more susceptible to subjective impressions than formal evaluation and is of limited utility in understanding the value of procedures. Many of the forensic practitioners discredited in the wake of wrongful convictions and innocence projects were charismatic, confident and impressive witnesses. These witnesses appeared credible but their opinions were unreliable — exaggerated, misleading, mistaken and in a few cases fraudulent.<sup>109</sup>

Some issues, such as bias, might be indexed to the reliability of the evidence or the credibility of the witness. An interest or association, for example, might be used to question a conclusion or to impugn the integrity of the witness. Threats to cognition in the forensic sciences have, for example, been demonstrated to threaten, and even alter, expert opinions.<sup>110</sup> Given the prominence of criticism and concerns about cognitive biases expressed by numerous authoritative bodies and commentators over the past decade, insensitivity or indifference to threats to cognition might impugn witness credibility. Moreover, is the opinion of an expert witness who has not disclosed the lack of validation, or sought to avoid well-known risks to cognition, credible (or reliable)? Returning to the terms of the *IMM* majority's example, once judges are more conversant with the dangers posed by human factors — particularly threats to cognition from context — will they find expert opinions developed in conditions that were inattentive to known dangers 'weak' or 'unconvincing'?<sup>111</sup>

<sup>108</sup> Such reputations do not always stand the test of time: see, eg, Andrew Rose, *Lethal Witness: Sir Bernard Spilsbury, Honorary Pathologist* (Kent State University Press, 2007).

<sup>109</sup> See, eg, *Inquiry into Pediatric Forensic Pathology in Ontario* (Report, 30 September 2008) vol 2, 115–204.

<sup>110</sup> Itiel E Dror, David Charlton and Ailsa E Péron, 'Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications' (2006) 156 *Forensic Science International* 74.

<sup>111</sup> Judges are not immune to cognitive biases: Chris Guthrie, Jeff J Rachlinski and Andrew J Wistrich, 'Inside the Judicial Mind' (2001) 86 *Cornell Law Review* 777; Dan Simon, *In Doubt: The Psychology of the Criminal Justice Process* (Harvard University Press, 2012) 150–60.

### E *Standing by Itself*

This brings me to a refractory issue. Various appellate judges have suggested that in determining probative value (at its highest) a trial judge might consider other evidence or the case as a whole.<sup>112</sup> When it comes to scientific, medical and technical evidence, apart from evidence relevant to the collection and transportation of samples (and perhaps other details relating to continuity or how the evidence articulates with the prosecution case) it is difficult to imagine why a trial judge would look beyond the results of formal evaluation (of the procedure or ability) when determining the probative value of an opinion based on specialised knowledge adduced by the prosecutor.

Nothing is gained by considering other (independent) evidence or the strength of a case for the purposes of determining the admission of opinions based on specialised knowledge or the probative value of such opinions. Rather, the consideration of other evidence (or the overall case) is likely to distract or mislead, especially where a procedure or ability has not passed formal evaluation. Other evidence and the strength of the case reveal nothing conclusive about whether or how well a procedure works.<sup>113</sup> Other evidence reveals nothing about the ability of the forensic practitioner and in many cases unnecessarily introduces the risk of double counting, biasing the practitioner, and misleading judge and jury.<sup>114</sup> Moreover, where the strength of a case is perceived as strong and taken into account, s 137 will have little or no role to play even if the opinion is weak, unreliable or mistaken.

<sup>112</sup> *IMM* (n 17) 310 [30], 315 [51] (French CJ, Kiefel, Bell and Keane JJ), citing *XY* (n 18) 400 [167], [170]; cf at 344–5 [156] (Kiefel and Nettle JJ). See also *Old Chief v United States*, 519 US 172, 182–5 (1997). It might be useful to compare s 137 with the text of other parts of the statute. Sections 97(1)(b) and 98(1)(b) provide that tendency and coincidence evidence is not admissible unless ‘the court thinks that the evidence will, *either by itself or having regard to other evidence* adduced or to be adduced by the party seeking to adduce the evidence, have significant probative value’ (emphasis added). Section 137 features no equivalent text. Section 137 refers to the probative value of the evidence and, unlike provisions regulating tendency and coincidence evidence, it does not encourage the court to consider probative value ‘by itself’ or ‘having regard to other evidence to be adduced’ by the prosecutor.

<sup>113</sup> Like conviction, it does not validate a procedure because we do not know if the correct answer was reached or why the jury convicted.

<sup>114</sup> ‘Double counting’ may occur when decision-makers treat two strands of evidence as independent when they are not independent. Where, for example, a forensic practitioner is unnecessarily exposed to suggestive information — such as an admission — and the jury treats the admission and the forensic practitioner’s opinion as independent, that is a misunderstanding that may lead to the evidence being overvalued, even double-counted: Emma Cunliffe, ‘Judging, Fast and Slow: Using Decision-Making Theory to Explore Judicial Fact Determination’ (2014) 18 *International Journal of Evidence and Proof* 139, 156–7.

The *tribunal of fact* should be able to combine and trade *admissible* evidence, but opinion based on specialised knowledge should not be admitted unless there is knowledge and, where contested, the prosecutor can demonstrate that the probative value (at its highest) outweighs the danger of unfair prejudice to the defendant. If the procedure works and the forensic practitioner is proficient, then the trial judge needs to determine the probative value of the opinion and undertake a balancing exercise. If the procedure does not work or is not known to work, there is little the judge can do that is indexed to knowledge or rationality. Strength of the case and independent strands of evidence are not relevant to the determination of the probative value of an opinion purportedly based on specialised knowledge.

#### F *Free Flying? Limits on the Jury's Prerogative*

Another issue arises where the prosecutor and trial judge do not consider reliability (and credibility). This concerns the jury's evaluation of the evidence at trial. Once opinion based on specialised knowledge is admitted, a rational jury cannot assign to it any value they desire. If a jury deliberately or unwittingly attributes more weight to the evidence than it can support, that is unfairly prejudicial and may result in a miscarriage of justice. Interpretations of opinion evidence must be disciplined by *known* abilities and limitations.

An evidence-based approach to opinion based on specialised knowledge not only is consistent with the need for 'knowledge' (from s 79) but also addresses appellate courts' concerns about trial judges usurping the province of juries. Jurors cannot assign more weight to an opinion than formal evaluation supports.<sup>115</sup> To do so liberates legal decision-making from the constraints imposed by knowledge and rationality. In the absence of formal evaluation, many opinions should be excluded via ss 79(1) or 137 or, in the alternative, read down. It is not trespassing on the prerogative of the jury to prevent them from speculating about opinion evidence when they could be provided with knowledge from formal studies that would enable them to make sense of it. Indeed, this is one of the prime reasons for s 137 (and s 79). In circumstances where the probative value of the evidence is unknown but knowable, testable

<sup>115</sup> Jurors are entitled to combine different strands of evidence to be satisfied beyond reasonable doubt, even where validated procedures generate opinions that have non-trivial levels of error. Problems may arise where prosecutions are based on a single strand of forensic science or medicine evidence — eg DNA or fingerprint-only cases. In such circumstances the known error rate (in conjunction with any other limitations) is as high as the probative value can ascend.

but untested, we should not encourage (or allow) the tribunal of fact to assume that the procedure works and that the forensic practitioner is proficient. We should not ignore oversights or invariably give the benefit of uncertainty to the state. The state should not be rewarded for its failure to evaluate the techniques it relies upon. Excluding opinion evidence of unknown probative value may be preferable to admitting the evidence and allowing, or requiring, the jury to speculate about its value or use other evidence as a makeweight. That approach to fact-finding frequently represents a form of unfair prejudice to the accused.<sup>116</sup>

Concerns about trial judges trespassing on the prerogatives of the jury do not arise when they are engaged in determining whether procedures work and whether forensic practitioners are proficient. Lack of formal evaluation raises more serious threats to the administration of justice than denying the jury an opportunity to speculate about what an opinion might be worth in the absence of relevant knowledge or other indicia that would enable them to rationally assess it. There is no judicial usurpation if the jury cannot be placed in a position conducive to the evaluation of opinion evidence.

#### IV SEEING THROUGH HEYDON'S FOG: INTERPRETING IMAGES AFTER *HONEYSETT*

This section develops an example of the issue at the centre of this article. It illustrates why legal institutions must attend to 'knowledge' in s 79(1), and why reliability is essential to any assessment of the probative value of opinions based on specialised knowledge for the purposes of s 137.

##### *A The Opinion of a Passport Officer*

Imagine, in the aftermath of *Honeysett*, that a prosecutor sought to adduce the evidence of a passport examiner of 10 years' experience to proffer an opinion about the identity of a person of interest captured in CCTV images of a robbery. On the basis of *Honeysett* and conventional legal practice, it seems likely that a passport examiner would be allowed to testify.<sup>117</sup> After all, passport officers spend their days comparing images of persons, or persons and images, for the purposes of identification. In contrast to the treatment of the professor of anatomy in *Honeysett*, who it was considered lacked expertise

<sup>116</sup> See Part V.

<sup>117</sup> While passport officers might not ordinarily examine CCTV images, courts are likely to overlook such potential issues, even in the absence of formal evaluation.

in image interpretation and the comparison of features for the purposes of identification,<sup>118</sup> our passport examiner would appear to possess these skills on the basis of her training (as a passport officer) and years of experience (comparing images). A trial judge (and appellate court) might accept that the opinion of our passport officer was based on training, study or experience, or even that the ‘specialised knowledge’ was an ability to interpret and compare facial features derived from training and experience. Further, think about what the passport examiner might be allowed to say. She might identify a person as being the person of interest, although earlier court decisions might be invoked by the trial judge to restrict the opinion to describing similarities between features in the images.<sup>119</sup>

Assume that the passport officer’s evidence is deemed relevant and admissible according to s 79(1). This should not require a great feat of imagination given that facial mapping evidence was admissible in Australian courts for more than a decade, and may still be — if the anatomist spends more time looking at the images than the professor did in *Honeysett* — as ad hoc expertise.<sup>120</sup> Now, what happens if the defence object to this otherwise admissible opinion evidence on the basis of s 137? Upon objection, the trial judge is required to determine the probative value of the evidence at its highest, determine the danger of unfair prejudice to the defendant and then to balance the highest probative value a jury could rationally assign against the danger of unfair prejudice.<sup>121</sup>

Before proceeding, I challenge you to make an assessment of the probative value of the passport examiner’s identification evidence, and its probative value taken at its highest. I would, in addition, encourage you to reduce your thoughts and reasons to writing.

Strictly applying the approach proposed by the majority in *IMM* brings practical limitations to the fore. Prevented from considering the reliability of the identification evidence or the credibility of the witness, we must determine not only the probative value, but the probative value at its highest. In the absence of information about whether the procedures used by the passport

<sup>118</sup> *Honeysett* (n 8) 137–8 [42]–[46]. This was declaratory, as there was no assessment of the professor’s procedure or ability.

<sup>119</sup> See *Tang* (n 14) 709 [120], 716 [157]; *Murdoch v The Queen* (2007) 167 A Crim R 329, 356 [298]–[300].

<sup>120</sup> Recognition of ad hoc expertise provides an unsatisfactory means of circumventing the statutory need for knowledge: Gary Edmond and Mehera San Roque, ‘Quasi-Justice: Ad Hoc Experts and Identification Evidence’ (2009) 33 *Criminal Law Journal* 8, 22.

<sup>121</sup> The danger of unfair prejudice is considered in Part V.

examiner are valid and reliable, and without insight into the performance of this passport officer or passport officers in general, any attribution of probative value is unavoidably speculative. It involves speculating about ability and performance. It involves imagining a value on the basis of factors (heretofore proxies) that may or may not provide insight into 'the extent to which the evidence could rationally affect the assessment of the probability of the existence of a fact in issue'. To be clear, trial and appellate judges may attribute probative value and seek to justify the value they assign. These are interesting epiphenomena that, in the absence of formal evaluation, do not afford useful insight into the actual probative value or probative value at its highest. Such attitudes and rationalisations are decoupled from knowledge.

Fortunately, there is knowledge in this domain. Scientific studies have found that training, employment and experience as a passport examiner make little difference to performance comparing and identifying persons in images. When tested, passport examiners were found to 'show no performance advantage over the general population'.<sup>122</sup> White and colleagues summarised their research findings as follows:

Consistent with previous research, our results emphasise that unfamiliar face matching is a difficult and error-prone task. Further, we show that this is not merely a laboratory phenomenon that is limited to novice participants. Trained passport officers also perform poorly when matching unfamiliar faces. High error rates were consistent across three tests, each of which was designed to emulate face matching in occupational settings. Further, across all experiments, length of time employed as a passport officer did not predict accuracy. Given the many face matching decisions made by passport officers as part of their daily workflow, we interpret this as strong evidence that experience alone does not improve accuracy on face matching tasks.<sup>123</sup>

This research is significant because it should inform attribution of probative value. Regardless of whether it is located by lawyers, accepted or admitted by the trial judge, the available scientific research suggests that the opinion of the passport officer is not relevant. According to the majority in *Smith v The Queen*, the opinion of the passport officer is incapable of rationally affecting the assessment of the probability of the existence of a fact in issue.<sup>124</sup> Given general legal complacency in response to opinions purportedly based

<sup>122</sup> David White et al, 'Passport Officers' Errors in Face Matching' (2014) 9(8) *PLOS ONE* 1, 1.

<sup>123</sup> *Ibid* 5.

<sup>124</sup> (2001) 206 CLR 650, 655 [11].

on specialised knowledge and adduced by the state, the relevance point might be considered disconcerting. It might encourage judges to reflect on the relevance and probative value of many other types of forensic science and forensic medicine evidence, and their indifference to knowledge. Our primary concern, however, is with probative value. Scientific research suggests that the opinion of a passport officer has no probative value, because passport officers can do no better than what the jury might do in a similarly error-prone fashion.

This example illustrates the difficulties and dangers confronting trial judges who are not provided with the results of formal evaluation. How is a trial judge who is not provided with relevant scientific literature (or aware of the implications of its absence) to determine the probative value of the passport officer's opinion evidence? Such a trial judge might assign any of the following probative values:

- 1 accept the opinion about identity as correct (ie categorical);
- 2 accept the opinion as likely to be correct (ie probabilistic);
- 3 accept the opinion as correct or likely to be correct but require the passport officer to limit the testimony to describing similarities and differences (ie an implicitly conservative compromise); or
- 4 some alternative that treats the passport officer's opinion as probative (ie relevant).

None of these responses embodies what is known about the procedure and the ability of passport officers (see Table 2). None is likely to assist with fact-finding. Only 1 and 2 seem consistent with the strict proscription on reliability and credibility. Option 3, and perhaps 2, might be consistent with an approach that attends to how convincing the evidence appears.

*Table 2: Assessing the Probative Value of the Opinion of a Passport Officer with and without Access to ‘Specialised Knowledge’*

Probative value and relevance	Impression without (attending to) validity or performance study	Assessment sensitive to validation or performance study
Actual probative value	Strong	Not probative
Probative value (at its highest) not considering reliability <i>or</i> credibility	Strong (perhaps assume identification <i>correct</i> )	Not probative
Probative value (at its highest) not considering reliability <i>and</i> credibility	Strong (no reason to think otherwise)	Not probative
Is the evidence convincing?	Yes (appears persuasive)	No
Is the evidence relevant?	Yes (perhaps self-evidently)	No

Ignorant of the scientific research, it is difficult to imagine any Australian trial or appellate judge finding the opinion of the passport officer not convincing and inadmissible. (It is also hard to imagine a trial judge finding unfair prejudice as a basis for excluding this evidence.)<sup>125</sup> And yet, all of these approaches are clearly misguided and likely to create unfair prejudice.<sup>126</sup> Notwithstanding what a judge might declare (in ignorance), the passport officer’s opinion is not based on ‘specialised knowledge’. It is not probative (even if accepted). Any judicial ascription is likely to be misguided and any justification or instruction to the jury, in consequence, misleading. Any use seems inappropriate and likely to compromise the burden and standard of proof.

<sup>125</sup> There are remarkably few reported examples after 1995 of trial judges excluding opinion based on specialised knowledge under s 137. A few of the early challenges to DNA evidence led to exclusion: see, eg, *R v Elliott* (Supreme Court of New South Wales, Hunt J, 6 April 1990) 10; *Tran* (1990) 50 A Crim R 233; *R v Lucas* [1992] 2 VR 109; *R v Green* (Supreme Court of New South Wales Court of Criminal Appeal, Gleeson CJ, Cripps JA and Abadee J, 26 March 1993); *Pantoja* (1996) 88 A Crim R 554.

<sup>126</sup> They also represent a waste of time and resources (under s 135), given the doubts about probative value.

Trial judges should not try to muddle through without reference to appropriate scientific research. Trial judges should not assume that a procedure works or that an untested opinion is reliable. Trial judges should be given, or ask for, evidence of validity and/or proficiency.

### B *Can Trial Safeguards Identify Problems and Convey Them to Jurors?*

The answer is that they might on occasion, though in most cases they are unlikely to place decision-makers in a position to rationally evaluate contested opinion evidence. Significantly, trial safeguards afford little protection in relation to the quality of opinions purportedly based on specialised knowledge used in plea bargains and charge negotiations.

Prosecutors, largely inattentive to the probative value of even novel procedures, routinely leave issues of validity, uncertainty, error and proficiency to the defence.<sup>127</sup> It might be considered ironic that many prosecutors present opinion evidence at its highest, or at some level of probative value beyond what is known or can be supported. Cross-examination might expose problems, such as lack of formal evaluation or performance testing. However, to be effective, cross-examiners and decision-makers must appreciate the significance of such omissions and their implications.<sup>128</sup> Where the witness is an apparently disinterested state employee (such as our passport officer), employed to perform a specific task and confident in her ability to perform that task, the chances of *the defence* persuading lay decision-makers that the employee has no special ability and that her experience may not matter are remote. Perversely, the passport officer's lack of knowledge about relevant studies, scientific research methods and notorious dangers with image comparison and facial identification may make it difficult to cross-examine her in ways that engage with knowledge or secure concessions.<sup>129</sup>

Rebuttal experts, including those offering methodological criticisms, are relatively uncommon and, notwithstanding claims about 'equality of arms', they struggle to compete with the positive (sometimes categorical) opinion and the apparently disinterested appearance of the state's employees and

<sup>127</sup> See, eg, FHR Vincent, *Inquiry into the Circumstances that Led to the Conviction of Mr Farah Abdulkadir Jama* (Report, May 2010) 32–8.

<sup>128</sup> A great deal of cross-examination focuses on legal proxies, folk knowledge, common prejudices and credibility issues.

<sup>129</sup> This is an acute problem with ad hoc experts who may not possess relevant disciplinary expertise or knowledge.

consultants.<sup>130</sup> Defence witnesses may be represented as idealistic, partisan or 'out of touch',<sup>131</sup> even when presenting mainstream scientific perspectives. To compound these difficulties, judges occasionally question the fact that rebuttal witnesses have not produced positive evidence, where the only procedures available are the untested procedures they are questioning.<sup>132</sup>

Directions and warnings are also fundamentally compromised where there is insufficient attention to relevant knowledge. To the extent that a trial judge is not provided with the results of formal evaluation or relevant scientific literature, any guidance they provide to the jury is likely to be inadequate or misleading. We might reflect on what a trial judge who had found the opinion of the passport officer to be admissible could say that would assist the jury. Such a trial judge has not engaged with knowledge and is unlikely to appreciate the level of error or the significance of not obtaining the results of formal evaluation. Attempts to draw attention to potential dangers cannot overcome the fact of admission, inadequate means of evaluation, or the fact that any identification evidence (even in a restricted form) would not be probative.

Compounding these problems is an issue that trial and appellate judges have yet to consider. Some types of evidence, such as identification from images and voice comparisons (and for some courts, fingerprints), are routinely presented to the tribunal of fact. In many cases the tribunal is entitled, and perhaps encouraged, to undertake its own comparisons following suggestive 'expert' opinion.<sup>133</sup> Courts continue to facilitate and endorse such practices notwithstanding the biasing conditions of the trial and the notoriously high error rates associated with many of these comparisons. When asked to compare the accused with a person in an image (or video) for example, the trial judge and jury are usually in a situation where the accused is both nominated and sitting conspicuously in front of them. The accused may have been selected because he resembles the person in the image.

<sup>130</sup> See generally Jenny McEwan, 'Ritual, Fairness and Truth: The Adversarial and Inquisitorial Models of Criminal Trial' in Antony Duff et al (eds), *The Trial on Trial: Truth and Due Process* (Hart Publishing, 2004–7) vol 1, 51.

<sup>131</sup> For example, the defence might only have access to retired examiners where the state has an effective monopoly on a type of 'expertise', such as ballistics.

<sup>132</sup> See, eg, *R v Madigan* [2005] NSWCCA 170, [90]; *Otway* (n 94) [21]; *JP v DPP* [2015] NSWSC 1669, [77]. For accounts of the experiences of critics, see Michael Lynch and Simon Cole, 'Science and Technology Studies on Trial: Dilemmas of Expertise' (2005) 35 *Social Studies of Science* 269; Simon A Cole, 'A Cautionary Tale about Cautionary Tales about Intervention' (2009) 16 *Organization* 121.

<sup>133</sup> Cf *Bennett v Police* [2005] SASC 415, [4]–[6]; *R v Murdoch [No 4]* (2005) 195 FLR 421, 440 [108].

Evidence presented — such as reference to alleged similarities or even more indirect forms, such as convictions for similar offences — may (unconsciously) contaminate the judge's and jury's interpretations.<sup>134</sup> These are influences that the human brain cannot overcome. We cannot think our way out of such highly suggestive environments and the number of jurors does not assist in these endeavours. Such conditions tend to make unfamiliar face comparison even less reliable, while misleading the tribunal of fact (and judges) into being overconfident about their abilities.<sup>135</sup>

The ability of trial safeguards to afford protection to the accused should inform the trial judge's application of s 137. Trial and appellate judges should be reluctant to glibly rehearse the efficacy of trial safeguards based on (collective) judicial experience while being inattentive to relevant studies and literature.<sup>136</sup>

#### V 'DANGER OF UNFAIR PREJUDICE TO THE DEFENDANT' AND THE BALANCING EXERCISE

A great deal of attention has been lavished on probative value and how it should be assessed. Much less attention has been directed to determining the probative value of opinions based on specialised knowledge. Perhaps even more revealing is how little attention has been dedicated to the danger of unfair prejudice to the defendant.

Building upon the example of the passport examiner, this section explores some of the dangers of unfair prejudice associated with opinion based on specialised knowledge. This section may surprise the reader because trial and appellate judges so rarely engage with the dangers in detail, perhaps because the parties do not provide adequate assistance. This section reinforces the need to attend to formal evaluation and reliability. Formal evaluation helps the trial judge (and eventually the decision-maker) to determine probative value as well as some of the dangers of unfair prejudice from expert opinion evidence. Without insight into actual probative value, when it comes to opinion based on specialised knowledge many of the dangers of unfair prejudice are acute and the balancing exercise becomes a sham. The 'scales' are, in effect, fixed in favour of admission.

<sup>134</sup> Cf Dror, Charlton and Péron (n 110); Gary Edmond et al, 'Thinking Forensics: Cognitive Science for Forensic Practitioners' (2017) 57 *Science and Justice* 144, 146–7.

<sup>135</sup> See generally Daniel Kahneman, *Thinking, Fast and Slow* (Farrar, Straus and Giroux, 2011) pt 3.

<sup>136</sup> *Expert Evidence in Criminal Proceedings in England and Wales* (n 12) 5 [1.20].

*A The Danger of Unfair Prejudice*

The danger of unfair prejudice is the risk that the tribunal of fact may use the evidence on an improper basis. The danger of unfair prejudice includes the risk that the jury will misuse or misunderstand the evidence, its limitations and uncertainties, as well as the risk that the tribunal of fact will not be placed in a position conducive to the rational determination of its weight. It also includes the danger that ‘on hearing the evidence the fact-finder may be satisfied with a lower degree of probability than would otherwise be required’.<sup>137</sup> Unfair prejudice may extend to procedural disadvantages flowing from the admission of forensic science evidence. It may be difficult to effectively cross-examine some witnesses even though their opinions are weak or not probative.<sup>138</sup> Increasingly, the scarcity of resources (and the state’s monopoly in many areas of forensic science and medicine) may make a defendant dependent on the state, the ability and impartiality of the witness, the adequacy of disclosure, and the transparency of reporting and testimony.

As the Victorian Court of Appeal has said, ‘[t]he obvious risk in a criminal trial when expert evidence is led from a forensic scientist is that a jury will give the evidence more weight than it deserves’.<sup>139</sup> For opinions based on specialised knowledge, the risk of unfair prejudice to the defendant is acute where the procedure being relied upon has not been formally evaluated or is used in a manner that is inconsistent with testing and protocols. The risk is compounded when the forensic practitioner is not demonstrably proficient, testifies in terms that are not scientifically supportable, or is inattentive to risks from contextual and other biases. Some of the main risks associated with forensic science evidence are considered below.

First, forensic science evidence must be presented in a manner that enables the trial judge and tribunal of fact to rationally evaluate it.<sup>140</sup> Similar difficulties to those confronting a trial judge attempting to determine probative value

<sup>137</sup> Law Reform Commission (Cth), *Evidence* (Report No 26, Interim, 1985) vol 1, 352 [644]. See also Australian Law Reform Commission, New South Wales Law Reform Commission and Victorian Law Reform Commission, *Uniform Evidence Law* (Report, December 2005) 558–9 [16.23]–[16.26] (*‘UEL Report’*).

<sup>138</sup> ‘Weakness’, per se, is not a reason for excluding relevant evidence: *Festa* (n 18) 609 [51] (McHugh J). Although, where the evidence is weak or uncertain, many of the dangers with opinions based on specialised knowledge are heightened.

<sup>139</sup> *Tuite* (n 5) [11].

<sup>140</sup> *Davie v Magistrates of Edinburgh* [1953] SC 34, 40; *Makita (Australia) Pty Ltd v Sprowles* (2001) 52 NSWLR 705, 741 [81]. See also Supreme Court of Victoria, *Practice Note SC CR 3: Expert Evidence in Criminal Trials*, 30 January 2017. Cf *Criminal Procedure Rules 2015* (England and Wales) pt 19.

(at its highest) confront the tribunal of fact when trying to determine the weight for the purposes of proof. The failure to provide information about validity and scientific reliability of evidence and the ability of the forensic practitioner means that the trial judge and tribunal of fact might not be able to rationally attach a weight to the forensic science evidence.<sup>141</sup> Lacking information about validity, scientific reliability and so forth, the tribunal of fact (and trial judges) will be obliged to speculate about weight or to rely on proxies (such as demeanour, experience, apparent plausibility and the fact of admission) which are of more limited utility in determining the probative value of scientific, medical and technical evidence. Inability to rationally assess opinion evidence is not only a danger of unfair prejudice to the defendant; it simultaneously threatens both rectitude and the legitimacy of accusatorial proceedings.

Second, there is a danger that the tribunal of fact will misunderstand or overvalue forensic science evidence. Where forensic science evidence is not supported by formal evaluation and evidence of proficiency, there is an acute risk that the tribunal of fact will misunderstand or overvalue the forensic science evidence. The undervaluation of forensic science evidence adduced by the prosecutor may be a threat to rational fact-finding but is not relevant to the danger of unfair prejudice to the defendant.<sup>142</sup>

Third, there is a danger that the tribunal of fact may trivialise error rates, limitations and uncertainty. All forensic science and medicine evidence is subject to limitations, uncertainty and error. However, some procedures have quite low error rates, whereas other procedures have surprisingly high rates of error. Consequently, it is very important to know about limitations, uncertainties and error rates when trying to determine both the probative value and whether that value might be understood by a tribunal of fact.<sup>143</sup> Where a procedure has a substantial error rate, there are dangers that the error rate will not be identified, explained or understood.<sup>144</sup> There are also risks that the

<sup>141</sup> *Forensic Science in Criminal Proceedings* (n 59) 45.

<sup>142</sup> Kristy A Martire et al, 'The Expression and Interpretation of Uncertain Forensic Science Evidence: Verbal Equivalence, Evidence Strength, and the Weak Evidence Effect' (2013) 37 *Law and Human Behavior* 197, 202.

<sup>143</sup> We can see here that probative value and the danger of unfair prejudice will not always be incommensurable. This is one of the reasons some judges contemplate incorporating limits to probative value (eg reliability and credibility) on the unfair prejudice side of the scales.

<sup>144</sup> Examples might include unfamiliar face matching: see Vicki Bruce et al, 'Verification of Face Identities from Images Captured on Video' (1999) 5 *Journal of Experimental Psychology: Applied* 339, 358; unfamiliar voice matching: see A Daniel Yarmey et al, 'Commonsense Beliefs and the Identification of Familiar Voices' (2001) 15 *Applied Cognitive Psychology* 283;

tribunal of fact will discount error because of factors that may not be pertinent — such as their impression of witness experience and demeanour or unwarranted confidence in their own abilities (relative to others).<sup>145</sup>

Fourth, there is a danger of the tribunal of fact treating the evidence as (basically) correct or error-free. Particularly detrimental to the general operation of s 137 is the danger of a trial judge equating probative value at its highest with the opinion being correct or error-free. If the highest probative value is equated with the opinion being correct, then the balancing exercise will invariably favour the admission of forensic science and forensic medicine evidence. Real dangers and ubiquitous human error tend to be discounted when balanced against putatively correct or error-free opinions.

Fifth, there is a danger of improperly deferring to the forensic practitioner. There is a risk that expert opinions may be invested ‘with a spurious appearance of authority, and [that] legitimate processes of fact-finding may be subverted’.<sup>146</sup> Where forensic science evidence is not supported by formal evaluation and evidence of proficiency, the tribunal of fact may be obliged to defer to the forensic practitioner or to rely upon alternative information of more limited utility.<sup>147</sup>

Sixth, there is a danger that the tribunal of fact may fail to appreciate the significance of formal evaluation or its absence. The vast majority of procedures used by forensic scientists are amenable to formal evaluation. Where forensic science evidence is admitted not having been formally evaluated, there is a risk that the tribunal of fact will not appreciate how fundamental formal evaluation is to conventional scientific practice and the generation of knowledge. The tribunal of fact may mistake admission (and reliance by investigators) as implicit legal endorsement. Admission might be interpreted to mean that the procedure and the opinion are sufficiently reliable for use in

Philip Rose, *Forensic Speaker Identification* (Taylor & Francis, 2002); and unfamiliar gait-matching from images: see Ivan Birch et al, ‘The Identification of Individuals by Observational Gait Analysis Using Closed Circuit Television Footage’ (2013) 53 *Science and Justice* 339.

<sup>145</sup> See Kay L Ritchie et al, ‘Viewers Base Estimates of Face Matching Accuracy on Their Own Familiarity: Explaining the Photo-ID Paradox’ (2015) 141 *Cognition* 161. How judges should determine the kinds of error rates they are willing to tolerate when admitting opinion based on specialised knowledge is an issue that has received almost no attention.

<sup>146</sup> *HG* (n 7) 429 [44].

<sup>147</sup> *Honeysett* (n 8) 131 [23].

serious criminal proceedings even where, as the study of the passport officers illustrates, they have little or no probative value.<sup>148</sup>

Importantly, being told (even by a trial judge) that appropriate evaluation has not been performed does not enable a decision-maker to gauge probative value. Rather, it identifies a serious omission (that should inform admissibility determinations and assessment of the credibility of a witness) and obliges a decision-maker to speculate.<sup>149</sup>

Seventh, there is a danger that the tribunal of fact may rely on general acceptance, longstanding use and previous admission. General acceptance of a procedure or practice within forensic science communities or legal institutions is not a substitute for formal evaluation.<sup>150</sup> There is a danger that tribunals of fact may substitute general acceptance, longstanding use, or previous admission for evidence of validity and scientific reliability.<sup>151</sup>

Eighth, there is a danger than the tribunal of fact may rely on witness experience, confidence or demeanour, or use the perceived strength of the case as a makeweight. Where procedures and the ability of the forensic practitioner have not been formally evaluated, there is a risk that the tribunal of fact will attribute too much weight to criteria of more limited utility, such as formal qualifications, training, study, experience, confidence and demeanour. Alternatively, the tribunal of fact may use the strength of the case as a makeweight.

Longstanding use, previous admission, experience, confidence and demeanour, and the apparent strength of the case reveal neither whether a procedure works, nor how well it works. They tell us nothing direct about the forensic practitioner's proficiency. Importantly, they do not provide insight into what the practitioner might legitimately opine.

Ninth, there is a danger that the tribunal of fact may not appreciate the corrosive potential of contextual information and other cognitive biases. If forensic science evidence is produced in conditions where the forensic practitioner is unnecessarily exposed to suggestive processes, gratuitous information or other threats to analysis and interpretation, there is a risk that

<sup>148</sup> Even references to use in investigations might convey a similar message or be invoked to insinuate reliability. Impressions are difficult to manage even when putative expertise is constrained.

<sup>149</sup> *Cf R v Atkins* [2010] 1 Cr App R 8, 126–7 [23].

<sup>150</sup> *Tuite* (n 5) [104]. *Cf Daubert* (n 8) 594.

<sup>151</sup> This is ironic because liberal admission has tended to discourage formal evaluation. Forensic practitioners have tended to look to courts rather than scientists and formal evaluation for epistemic legitimacy.

the resulting opinion will be contaminated or mistaken. In these circumstances, it may be difficult to convey to the tribunal how great the risk is, and the defendant may be procedurally disadvantaged in having to attempt to identify and explain subtle though potentially corrosive psychological influences. How, for example, do you effectively cross-examine a confident and experienced witness about unconscious influences?

Lack of formal evaluation of procedures and lack of evidence of forensic practitioners' abilities may produce procedural disadvantages.<sup>152</sup> The absence of relevant information (or the lack of disclosure) may make it difficult to effectively cross-examine the forensic practitioner or convey the seriousness of limitations and omissions. In such circumstances, it may be difficult to persuade the tribunal of fact of the fundamental importance of validation, the significance of its absence and the real risk of error, even though the defence has no such formal legal obligation.

In concluding this discussion, we should acknowledge that the reliability of evidence and the credibility of the witness could be considered on the 'danger of unfair prejudice' side of the scale rather than as part of probative value. That possibility, recognised in several decisions, is inconsistent with the conventional common law commitment to probative value going to proof and unfair prejudice being concerned with fairness.<sup>153</sup> While this might not prevent the inclusion of reliability and credibility among the dangers of unfair prejudice, that approach will tend to inflate the probative value of opinion putatively based on knowledge, even where no knowledge is identified. Moreover, the reliability of evidence is directly related to probative value, whereas it tends to merely inform some of the dangers of unfair prejudice raised by scientific, medical and technical evidence.

### B *The Balancing Exercise and Mitigation of the Risk*

The final stage in the application of s 137 is the 'balancing exercise'. This requires the trial judge to determine whether the probative value of the evidence outweighs the danger of unfair prejudice to the defendant. The trial judge must evaluate the probative value of forensic science evidence, drawing on information provided by the prosecutor (though ideally contained in the

<sup>152</sup> *UEL Report* (n 137) 559–61 [16.27]–[16.32], 564 [16.45].

<sup>153</sup> *XY* (n 18) 376–7 [48]. Cf *Pfennig* (n 36) 482; *Shamouil* (n 18) 236–7 [56], quoting *Cook* (n 18) [43]; *Police v Dunstall* (2015) 256 CLR 403, 418–20 [31]–[33].

expert report or certificate) along with any insights provided by the defence.<sup>154</sup> In most cases involving forensic science and medicine evidence, this will require the trial judge to consider the procedure's validity, scientific reliability and error rate, as well as the ability of the forensic practitioner. The trial judge is then obliged to consider the danger of unfair prejudice to the defendant arising from risks associated with the admission of the forensic science evidence. In undertaking this assessment, the trial judge should consider any directions that will ameliorate risks to the defendant.<sup>155</sup>

In undertaking the balancing exercise, the trial judge should consider the ability of directions to identify and convey the significance to the rational evaluation of forensic science evidence of issues such as validation, scientific reliability, uncertainty, limitations, error rates and contextual bias. In the absence of information about validation and scientific reliability, it may be difficult to prevent juries attributing too much weight to opinions that are presented as, or may appear to be, scientifically or technically predicated. Explaining to the jury that a procedure has not been formally evaluated will rarely place the jury in a position to rationally evaluate related forensic science and medicine evidence. In the absence of formal evaluation, the tribunal of fact should be cautious, perhaps even sceptical. There is a particular need for caution where the trial judge proposes to manage the absence of formal evaluation by moderating the strength of the expression used by the forensic practitioner (using s 136 or on some other basis). Untested procedures might not support even weak conclusions. For reasons that should now be obvious, restricting our passport officer to expressing opinions about features in the images and any similarities or differences is not a credible response to admissibility decision-making.

Attempts to ameliorate unfair prejudice by characterising the evidence of forensic practitioners, or those historically recognised as forensic practitioners (or scientists), as non-scientific, non-technical or experience-based may have little practical effect, especially if the procedure has been in longstanding use and might be popularly perceived as scientific, technical or otherwise trustworthy. The issue here is not one of classification or nomenclature, but rather

<sup>154</sup> On the limits of certificates and reports, particularly non-compliance with codes of conduct and practice directions for expert witnesses, see Gary Edmond, Kristy Martire and Mehera San Roque, 'Expert Reports and the Forensic Sciences' (2017) 40 *University of New South Wales Law Journal* 590.

<sup>155</sup> Where there is 'a real risk that the jury would attach more weight to [the evidence] than it deserved, and that risk could not be overcome by strong directions from the trial judge, the evidence would be excluded': *Dupas* (n 18) 219 [142].

a problem of proof and fairness that requires appropriate evidence of probative value.

## VI RE-IMAGINING PROBATIVE VALUE FOR OPINIONS BASED ON SPECIALISED KNOWLEDGE

This article has endeavoured to explain the need to consider reliability when determining the probative value of opinions based on specialised knowledge for the purpose of s 137. Regardless of any stance trial and appellate judges take in relation to ordinary witnesses or the way the meaning of ‘specialised knowledge’ is developed in the aftermath of *Honeysett*, they cannot rationally determine the probative value of scientific, medical or technical evidence (at its highest or otherwise) without knowing whether the underlying procedure works and, if so, the conditions in which it is known to work. Where the opinion is dependent upon some putative ability, there should, in addition, be evidence of the witnesses’ competence or level of proficiency. Without insight into validity, reliability and proficiency, we do not know the extent to which an opinion might rationally influence the probability of facts in issue. Without this knowledge we are ignorant. We cannot be confident that opinions presented as ‘expert’ are, notwithstanding appearances, even relevant.

To the extent that any of this is inconsistent with *IMM*, courts must develop an exception for opinions based on specialised knowledge adduced by the prosecutor.<sup>156</sup> The foundations of such an approach are already embodied in the various decisions. The majority recognises that trial judges need to determine the ‘extent to which the evidence could rationally affect the assessment of the probability of the existence of a fact in issue’. Moreover, notwithstanding the strident approach to reliability, their eyewitness example engages with reliability, for the example confirms, albeit indirectly, that information about the reliability of an eyewitness identification should inform judicial interpretations of how *convincing* testimony is for the purpose of determining probative value. Incorporating contextual considerations the majority considers the eyewitness testimony weak.<sup>157</sup> Of more direct utility are the dissenting judgments. Gageler J and Nettle and Gordon JJ expressly

<sup>156</sup> The majority left some limited scope for exceptions: *IMM* (n 17) 316–17 [57]–[58].

<sup>157</sup> It is no coincidence that the majority’s impression is consistent with — really informally informed by — decades of scientific research on eyewitness identification evidence: see Committee on Scientific Approaches to Understanding and Maximizing the Validity and Reliability of Eyewitness Identification in Law Enforcement and the Courts, *Identifying the Culprit* (National Academies Press, 2014).

favour trial judges considering reliability (and credibility). Their approaches are consistent with the majority's example and are well suited to the evaluation of opinion based on specialised knowledge under s 137. They also sit more comfortably with the text of the uniform legislation and s 137 operating specifically as a trial safeguard. These approaches enable the trial judge to assess the extent to which the opinion can rationally influence the jury, while facilitating a more transparent engagement with actual limitations and dangers.

If judges do not attend to the reliability of opinions based on specialised knowledge for the purposes of s 137, then, as things stand, prosecutors and trial judges are not required to consider the trustworthiness of expert opinion evidence at any stage of their admissibility decision-making. Following *Tang* in New South Wales and *Tuite* in Victoria, trial judges are not required to consider reliability (or validity) as part of the assessment of 'knowledge' under s 79(1).<sup>158</sup> If the stringent approach proposed by the majority in *IMM* were applied to opinions based on specialised knowledge challenged via s 137 (or s 135) then reliability (and validity) have no role in contemporary admissibility practice. Questions of validity and reliability will be left exclusively for the tribunal of fact. In such circumstances, a forensic procedure might be relied on over and over without prosecutors ever producing evidence that the procedure is valid or affording insight into the witness's actual ability. This is perverse. The very information that would enable lawyers, judges and jurors to evaluate the opinion evidence should be requested and provided. Rather than pay lawyers and 'experts' to naively conjecture about whether some procedure works, we should require *knowledge* derived through formal evaluation. That is, 'from study or investigation' to repeat the formulation advanced in *Honeysett*.<sup>159</sup>

There is one appellate decision that makes precisely this case for s 137. It is well suited to the determination of probative value — including probative value at its highest — and the question of whether an opinion based on

<sup>158</sup> Spigelman CJ's contention in *Tang* (n 14) 712 [137] that '[t]he focus of attention must be on the words "specialised knowledge", not on the introduction of an extraneous idea such as "reliability"' is misguided and unhelpful. We should not overlook the fact that *Tang* was decided before *Strengthening Forensic Science in the United States* (n 58) and *Forensic Science in Criminal Proceedings* (n 59) exposed serious problems with many forensic sciences. In *Tuite* (n 5) [58]–[59], Maxwell ACJ, Redlich and Weinberg JJA overemphasised both the passing references to *Tang* in *Honeysett* and the need for comity where the New South Wales Court of Criminal Appeal was mistaken.

<sup>159</sup> Studies are cheaper and more informative than contesting 'expert' evidence in trials and appeals.

specialised knowledge is ‘convincing’. Its origins, in the contest around *XY* and *Dupas* are less important than the provision of a practical means of determining if and ‘the extent to which’, opinion evidence ‘could rationally affect the assessment of the probability of the existence of a fact in issue’. In *Tuite*, the Victorian Court of Appeal explained why there is a need to consider the reliability (and validity) of forensic science evidence for s 137 and set out a basic framework for determining the probative value of forensic science and medicine evidence.<sup>160</sup> The need was demonstrated by concern about serious deficiencies with forensic science and forensic medicine evidence expressed by superior courts, authoritative scientific organisations and law reform bodies from around the common law world.

Though characterised as a means of assessing ‘the reliability of scientific evidence’,<sup>161</sup> the Court of Appeal’s approach is really a means — the only viable means — of determining probative value. Logically, the highest probative value must be predicated upon what is currently *known* rather than what is possible or what is imaginable. So, in determining the probative value of forensic science evidence, a trial judge must consider (the legal idea of) reliability. The Court in *Tuite* explained:

In our view, the touchstone of reliability for scientific evidence must be trustworthiness, and trustworthiness depends on validation. ...

[T]he focus on proven validation has a number of advantages. First, and most importantly, it means that the scrutiny of scientific evidence in the judicial process will apply the rigour which the discipline of science itself demands. As it was put in *Daubert*, evidentiary reliability will be based on scientific validity. Secondly, the trial judge considering scientific evidence will ordinarily be able to assess the sufficiency of validation — based on the published results of validation tests — without needing to acquire particular expertise in the relevant field of science.

Thirdly, validation studies provide a framework which assists the judge — and, ultimately, the jury — to evaluate the evidence. Fourthly, this approach avoids what we consider to be the unworkable imprecision of a ‘general acceptance’ test, and will ensure that new developments and novel techniques are not excluded, provided always that their scientific validity is established to the satisfaction of the court.<sup>162</sup>

<sup>160</sup> *Tuite* (n 5) [85]–[114].

<sup>161</sup> *Ibid* [88].

<sup>162</sup> *Ibid* [101]–[104] (citations omitted).

For the Court, the need to attend to reliability was pressing where opinion based on specialised knowledge is novel:

Special care must be taken, of course, in a case where the proposed expert evidence is based on ‘new science’ properly so-called. In such a case, the party proposing to rely on the expert evidence will need ... to establish that the underlying science is ‘sufficiently reliable to be admitted in a court of law.’<sup>163</sup>

Concern with probative value was not, however, restricted to novel procedures.<sup>164</sup>

Following *IMM*, there is no requirement for forensic scientists and prosecutors to provide evidence about validity, reliability, and proficiency in any Australian jurisdiction. These subjects might be disclosed in reports and/or explored during trial, but they are not required for opinions characterised as expert to be adduced, admitted and relied upon by the state.<sup>165</sup> For reasons made clear by the example of the passport officer, this is unacceptable. Opinions not known to be probative, and opinions that might be presented and accepted as more probative than they are known to be, are routinely adduced, admitted and relied upon in criminal proceedings.<sup>166</sup> This is not only inconsistent with our statutory arrangements; it is dangerous. Indifference to the actual probative value of scientific, medical and technical evidence threatens ‘the integrity and fairness of the criminal justice system.’<sup>167</sup> Moreover, our historical lack of interest in reliability in admissibility decision-making has had the unfortunate effect of discouraging research and formal scientific evaluation. Many forensic scientists have sought and prematurely received legal recognition and reliance.<sup>168</sup>

In its recent and sobering review of seven feature comparison forensic procedures — including DNA profiling, latent fingerprints, ballistics, bite marks, shoeprints and hair — the two dozen scientists, engineers and statisticians composing the President’s Council of Advisors on Science and

<sup>163</sup> Ibid [106], quoting *Trochym* (n 11) 262 [33]. Cf *Clarke* [1995] 2 Cr App R 425, 429–30.

<sup>164</sup> *Tuite* (n 5) [104].

<sup>165</sup> Again, where pleas are negotiated, the reliability of ‘expert’ opinions may not be considered.

<sup>166</sup> See *Forensic Science in Criminal Proceedings* (n 59) 67–123.

<sup>167</sup> *Tuite* (n 5) [82].

<sup>168</sup> See, eg, Gary Edmond and Emma Cunliffe, ‘Cinderella Story? The Social Production of a Forensic “Science”’ (2016) 106 *Journal of Criminal Law and Criminology* 219, 264. See generally Jennifer L Mnookin et al, ‘The Need for a Research Culture in the Forensic Sciences’ (2011) 58 *UCLA Law Review* 725; David A Harris, *Failed Evidence: Why Law Enforcement Resists Science* (New York University Press, 2012).

Technology ('PCAST') offered the following insight to President Obama, the Department of Justice and the federal judiciary:

Scientific validity and reliability require that a method has been subjected to empirical testing, under conditions appropriate to its intended use, that provides valid estimates of how often the method reaches an incorrect conclusion. ... 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.<sup>169</sup>

Salutary, independent and unquestionably authoritative, this advice is in no way limited to the assessment of probative value in US federal proceedings.<sup>170</sup> PCAST confirms that opinions based on procedures and abilities that have not been formally evaluated are not (known to be) reliable. They are, to apply the words of the *IMM* majority, weak and unconvincing. Such opinions introduce 'considerable potential' for unfair prejudice.<sup>171</sup> PCAST recommended that the Department of Justice should not offer such testimony.

However they choose to do it, Australian trial judges must consider evidence of reliability for opinions based on specialised knowledge at some stage in their admissibility decision-making. 'Reliability' should be read into s 79(1) in a manner that is consistent with the emerging jurisprudence around 'knowledge' in *Honeysett*. In principle, the section regulating the admission of expert opinion should require the proponent to identify 'knowledge' and the means to evaluate the opinion.<sup>172</sup> Clearly, the value of expert opinion evidence might also be addressed by requiring trial judges to attend to reliability when determining the probative value of opinions based on specialised knowledge for the purposes of s 137.<sup>173</sup> There are few alternatives.<sup>174</sup> If our criminal

<sup>169</sup> *Forensic Science in Criminal Proceedings* (n 59) 143.

<sup>170</sup> Note the comments by PCAST's co-chair on the implications for Australia: Eric S Lander, 'Response to the ANZFSS Council Statement on the President's Council of Advisors on Science and Technology Report' (2017) 49 *Australian Journal of Forensic Sciences* 366.

<sup>171</sup> *Forensic Science in Criminal Proceedings* (n 59) 32. PCAST was unwilling to speculate about probative value in the absence of formal evaluation. According to the report, 'methods must be presumed to be unreliable until their foundational validity has been established based on empirical evidence and ... even then, scientific questioning and review of methods must continue on an ongoing basis'.

<sup>172</sup> *Davie* (n 140) 40; *Makita* (n 140) 741 [81].

<sup>173</sup> Though here, it is the defence rather than the proponent (ie the prosecutor) who carries the burden. Nevertheless, s 137 should be used to regulate the admission of investigators' opinions about the identity of speakers (and the words spoken) via s 78 of the *UEL*. Consider, for example, *Tran* (n 21) and *Nguyen* (n 95).

justice system is to benefit from scientific, medical and technical opinions, then opinions must be ‘wholly or substantially based on’ knowledge so that decision-makers have rational means of evaluating them. The alternative is the constant risk of a spectacular and unedifying fall.

Icarus, according to the legend, flapped his featherless arms as he fell.

<sup>174</sup> Any potential recognised by the Victorian Court of Appeal, in *Haddara v The Queen* [2014] VSCA 100, does not provide an appropriate means of regulating the admission and evaluation of opinions based on specialised knowledge.