Gary Edmond,* Jason M Chin,** Kristy A Martire and Mehera San Roque*****

A WARNING ABOUT JUDICIAL DIRECTIONS AND WARNINGS

Abstract

This article questions our criminal justice system's heavy reliance on judicial directions and warnings. Reviewing a recent case and the directions provided by the trial judge — in a trial where a police officer purported to identify defendants on the basis of listening to intercepted telephone calls — this article explains why orthodox judicial instructions were incapable of assisting the jury with their assessment of the evidence. The analysis in this article explains why judicial directions do not necessarily mediate and therefore justify the admission of opinion evidence. In some cases, judicial directions are incapable of placing decision-makers in a position to rationally evaluate evidence. These conclusions draw on scientific research on voice identification and cognitive bias to illustrate how some judicial directions are not only displaced from scientific knowledge, but sometimes encourage (or expect) jurors to perform impossible feats of cognition.

I Introduction: Trials (and Safeguards) for Show?

The criminal trial on indictment proceeds on the assumption that jurors are true to their oath, that, in the quaint words of the ancient oath, *they hearken to the evidence and that they obey the trial judge's directions. On that assumption, which I regard as fundamental to the criminal jury trial, the common law countries have staked a great deal.* If it was rejected or disregarded, no one — accused, trial judge or member of the public — could have any confidence in any verdict of a criminal jury or in the criminal justice system whenever it involves a jury trial. If it was rejected or disregarded, the pursuit of justice through the jury system would be as much a charade as the show trial of any totalitarian state. Put bluntly, *unless we act on the assumption that criminal*

* Professor, Faculty of Law and Justice, UNSW; Research Professor (fractional), School of Law, Northumbria University; Chair, Evidence-Based Forensics Initiative.

** Senior Lecturer, College of Law, Australian National University.

*** Professor, Faculty of Science, UNSW.

**** Associate Professor, Faculty of Law and Justice, UNSW.

> *juries act on the evidence and in accordance with the directions of the trial judge, there is no point in having criminal jury trials.*[1]

Judicial directions are said to fulfil an important, perhaps essential, role in criminal prosecutions.[2] Our courts rely on directions to capture and convey the law to be used by the trier of fact. With respect to evidence, particularly evidence that might be unreliable or susceptible to misuse, directions are said to bring the collective experience of the judges to assist the trier of fact with evaluation.[3] They draw attention to the limitations and dangers of some kinds of evidence (and reasoning) that are considered by judges to threaten the fairness of proceedings and the rationality of decision-making.[4] Directions are believed to make criminal trials fair by exposing and implicitly protecting against dangers associated with the misunderstanding and misuse of evidence. But what happens if the collective experience and common sense of judges is misguided or simply wrong? What if the experience of courts and the directions provided by trial judges are not readily applied or not actually helpful?[5] This article directly questions the role played by directions in

1 *Gilbert v The Queen* (2000) 201 CLR 414, 425 [31] (McHugh J) (emphasis added). Justice McHugh continues at 426 [32]: 'In my respectful opinion, the fundamental assumption of the criminal jury trial requires us to proceed on the basis that the jury acted in this case on the evidence and in accordance with the trial judge's directions'. See also: *DPP (Vic) v Lyons (Ruling No 3)* [2018] VSC 224, [58]; *Gammage v The Queen* (1969) 122 CLR 444, 463 (Windeyer J): 'A jury in a criminal case … must be assumed to have been faithful to their duty'.

2 In this article we use the terms directions and warnings interchangeably. While some courts draw distinctions — such as directions must be followed (for example, in relation to the *Evidence Act 1995* (Cth) s 95 ('*Evidence Act*') and the prohibited use of evidence for tendency purposes), whereas warnings are intended to assist the trier of fact by drawing attention to a danger — such uses are not consistent. The warnings given in relation to evidence of a kind that might be unreliable at common law and under the uniform evidence law ('UEL') (eg, *Evidence Act* (n 2) s 165) are frequently described, as in our case study, as judicial directions. See, eg: Queensland Law Reform Commission, *A Review of Jury Directions* (Report No 66, December 2009) vol 1, 53 [4.21] ('QLRC, *A Review of Jury Directions*'); New South Wales Law Reform Commission, *Jury Directions* (Report No 136, November 2012) 2 [1.3] ('NSWLRC, *Jury Directions* 2012').

3 See, eg, *Chidiac v The Queen* (1991) 171 CLR 432: 'where part or all of the incriminating evidence against the accused consists of identification evidence, the Court will examine the case in the light of its knowledge, gained from long experience of criminal trials, that identification evidence is a potent source of miscarriages of justice': at 462 (McHugh J), citing *Davies v The King* (1937) 57 CLR 170, 180. Cf *FGC v Western Australia* (2008) 183 A Crim R 313, 317 [4]–[6] (Wheeler JA).

4 Judicial notions of fairness might be understood as emic — an actor's category. For reasons explored in this article, they are not persuasive as etic accounts.

5 We acknowledge the possibility of jury nullification, though we should be anxious about any nullification based on misunderstanding. See generally: Alan Scheflin and Jon Van Dyke, 'Jury Nullification: The Contours of a Controversy' (1980) 43(4) *American Jury* 51; Richard Lorren Jolly, 'Jury Nullification as a Spectrum' (2022) 49(2) *Pepperdine Law Review* 341.

the evaluation of evidence — particularly potentially unreliable opinion evidence. It draws attention to the practical limitations of directions, and the implications of these limitations for the fairness of criminal proceedings.[6] Indirectly, it draws attention to the limits of judicial experience and the distance between legal practice and scientific knowledge.

This article considers the provision and effectiveness of judicial directions through a concrete example. It uses the appellate decision from *Davey v Tasmania* ('*Davey*')[7] as a case study.[8] In *Davey*, the trial judge admitted the voice identification evidence of a police officer who repeatedly listened to lawful telephone intercepts during an investigation. This kind of evidence is now routinely admitted in criminal proceedings even though, as lawyers and courts acknowledge, it is a kind of evidence that may be unreliable.[9] In consequence, in order to make the trial fair, the admission of the police officer's opinion evidence was said to require careful directions to the jury. We contend, for the reasons developed in the ensuing analysis, that epistemic frailties inherent in the opinions of police officers cannot be repaired, or even meaningfully addressed, through judicial directions (or other safeguards).[10] This article is critical of the directions in *Davey*, both their content and effectiveness, and explains how threats to rational decision-making and the fairness of proceedings appear to be misunderstood and radically underestimated by both trial and appellate courts.

II Directions, Warnings and Instructions

Initially, it is useful to consider what courts, law reform commissions and attentive scholars have said about judicial directions. A good place to begin is with their

6 See generally: David Hamer and Gary Edmond, 'Forensic Science Evidence, Wrongful Convictions and Adversarial Process' (2019) 38(2) *University of Queensland Law Journal* 185; Gary Edmond and Andrew Roberts, 'Procedural Fairness, the Criminal Trial and Forensic Science and Medicine' (2011) 33(3) *Sydney Law Review* 359.

7 [2020] TASCCA 12 ('*Davey*').

8 In researching this topic, we sought access to documents, transcripts, submissions and the detective's statements from appellate counsel and the Director of Public Prosecutions (Tasmania). We received no cooperation. The Director did not reply to our request, although his staff assured us in writing that the request had been received. Such attitudes raise questions about open justice, public accountability and public confidence in criminal justice systems. See, eg, New South Wales Law Reform Commission, 'Open Justice Review', *NSW Government Communities & Justice* (Web Page, 29 March 2023) <https://www.lawreform.justice.nsw.gov.au/Pages/lrc/lrc_current_projects/Open-justice/Project_update.aspx>.

9 Similar critiques might be made about the reception of the voice identification and the directions in a number of cases. See, eg: *R v Leung* (1999) 47 NSWLR 405; *Nguyen v The Queen* (2002) 26 WAR 59 ('*Nguyen* 2002'); *Li v The Queen* (2003) 139 A Crim R 281; *R v Riscuta* [2003] NSWCCA 6 ('*Riscuta*'); *Kheir v The Queen* (2014) 43 VR 308 ('*Kheir*'); *Tran v The Queen* [2016] VSCA 79; *Tasmania v Farhat* (2017) 29 Tas R 1 ('*Farhat*'); *R v Phan* (2017) 128 SASR 142 ('*Phan*').

10 This applies regardless of whether it is judges directing jurors or themselves.

function(s). What is the purpose of jury directions given the time and resources they consume in trials and appeals? 'The aim of jury directions is to ensure a fair trial, where the jury's verdict is the result of the application of the law to the facts as found by the jury'.[11] Similarly:

> Ensuring that the defendant receives a fair trial may be seen as the ultimate obligation of a trial judge in presiding over a trial, whether or not the judge is the trier of fact, and as the primary objective of all jury directions and warnings.[12]

The main purpose of jury directions seems to be about making criminal proceedings fair. Ordinarily, the trial judge provides directions to the jury on the legal issues (the law), sometimes explaining the application of the law to the evidence admitted.[13] The trial judge might also direct the jury on specific evidence — providing assistance with some kinds of evidence (for example, eyewitness identification or the testimony of prison informers) thought to raise difficulties or introduce risks — or how they should not reason:

> The fundamental task of a trial judge is, of course, to ensure a fair trial of the accused. … In some cases it will require the judge to warn the jury about how they should *not* reason or about particular care that must be shown before accepting certain kinds of evidence.[14]

---

11 Criminal Law Review, *Jury Directions: A Jury-Centric Approach* (Report, Department of Justice and Regulation (Vic), March 2015) iii ('CLR, *Jury Directions: A Jury-Centric Approach*'). See also: Criminal Law Review, *Jury Directions: A New Approach* (Report, Department of Justice (Vic), January 2013) 4, 18 ('CLR, *Jury Directions: A New Approach*'); New South Wales Law Reform Commission, *Jury Directions* (Consultation Paper 4, December 2008) 4–6 [1.11]–[1.14].

12 QLRC, *A Review of Jury Directions* (n 2) vol 1, 116 [7.5].

13 There may be exceptions, such as where there are few issues or the issues appear straightforward, such that directions are not considered necessary. In some cases the defence may request that one or more directions are not given.

14 *RPS v The Queen* (2000) 199 CLR 620, 637 [41] (Gaudron ACJ, Gummow, Kirby and Hayne JJ) (emphasis in original) (citations omitted), quoted in Mark Weinberg, *Simplification of Jury Directions Project: A Report to the Jury Directions Advisory Group* (Report, Supreme Court of Victoria, August 2012) 277 [5.1]. See also *Dupas v The Queen* (2010) 241 CLR 237, 248–9 [28]–[29] ('*Dupas*'). Cf: *Zoneff v The Queen* (2000) 200 CLR 234, 260–1 [65]–[67] (Kirby J) ('*Zoneff*'); *R v Yasso [No 2]* (2004) 10 VR 466, 482–3 [53]–[60] (Vincent JA); *Wilson v The Queen* (2011) 33 VR 340, 343 [2] (Maxwell P); Murray Gleeson, 'The State of the Judicature' (2007) 14(3) *Australian Journal of Administrative Law* 118, 121; Justice Geoff Eames, 'Tackling the Complexity of Criminal Trial Directions: What Role for Appellate Courts?' (2007) 29(2) *Australian Bar Review* 161; Virginia Bell, 'How to Preserve the Integrity of Jury Trials in a Mass Media Age' (2005) 7(3) *Judicial Review* 311; James Wood, 'Jury Directions' (2007) 16(3) *Journal of Judicial Administration* 151; Justice Peter McClellan, 'Looking Inside the Jury Room' (2011) 10(3) *Judicial Review* 315.

The effectiveness of directions and the fairness of trials is said to be indexed to public confidence in the criminal justice system:

> If directions are not effective, this leaves jurors to navigate the evidence and arguments in a trial on their own, which makes their job harder, and may reduce the community's confidence in the criminal justice system. It is therefore vital to ensure that jury directions are as clear and helpful as possible.[15]

In *Longman v The Queen*[16] and *Bromley v The Queen*,[17] the High Court explained that the trial judge is obliged to give all the directions required to avoid a 'perceptible risk of miscarriage of justice'.[18] In *Carr v The Queen*,[19] Brennan J expanded on this requirement:

> A warning is needed when there is a factor legitimately capable of affecting the assessment of evidence of which the judge has *special knowledge, experience or awareness* and there is a perceptible risk that, unless a warning about that factor is given, the jury will attribute to an important piece of evidence a significance or weight which they might not attribute to it if the warning were given.[20]

This is really just the converse of the requirement that criminal trials should be fair. In the preceding extracts we can observe claims about 'special knowledge' derived from 'judicial experience (actual or inherited)' and the deep institutional investment in the efficacy of directions.[21]

Ironically, directions appear to have contributed to the complexity of criminal proceedings, particularly where judges privilege legal accuracy (or correctness) over simplicity. This sometimes manifests in trial judges rehearsing the technical legal language of appellate courts.[22] However, law reform bodies have placed emphasis on

---

15 CLR, *Jury Directions: A New Approach* (n 11) 15.

16 (1989) 168 CLR 79 ('*Longman*').

17 (1986) 161 CLR 315.

18 *Longman* (n 16) 86 (Brennan, Dawson and Toohey JJ). See also: ibid 324–5 (Brennan J); *RPS v The Queen* (2000) 199 CLR 620, 637 [41] (Gaudron ACJ, Gummow, Kirby and Hayne JJ); *Tully v The Queen* (2006) 230 CLR 234, 252–3 [57] (Kirby J), 259–60 [87], 261 [91] (Hayne J), 274 [132] (Callinan J), 280 [151] (Heydon J), 289 [186] (Crennan J).

19 (1988) 165 CLR 314 ('*Carr*').

20 Ibid 325 (emphasis added). See also *Jenkins v The Queen* (2004) 211 ALR 116, 121–2 [25].

21 *Carr* (n 19) 325 (Brennan J).

22 CLR, *Jury Directions: A New Approach* (n 11) 4, 10, 32. 'Jury directions are too long and too complex': at 10. See also NSWLRC, *Jury Directions* 2012 (n 2) xix.

making 'jury directions as comprehensible … as possible'.[23] This is a reaction to the 'increasing complexity' of directions, the amount of time consumed by directions at trial (particularly in New South Wales and Victoria), and the number of convictions overturned on appeal because of a mistake, irregularity or omission in the provision of directions on the law or evidence.[24]

This article is limited to directions on evidence and the use of evidence; specifically, kinds of evidence that are believed to be vulnerable to misunderstanding or misuse by jurors.[25] In addition to the kinds of evidence enumerated among the various legislative provisions (for example, identification evidence, the evidence of accomplices and bad character evidence), judges are expected (and at common law required) to provide jurors with insights drawn from their collective experience where it is in the interests of justice to do so.[26] Though, the provision of most directions follows a formal request.[27]

Directions tend to be requested (or are required) where

> the court has some *special knowledge or experience* about that kind of evidence *which the jury may not possess* and which may affect its reliability, or because it is the kind of evidence to which a jury may attribute more weight than it really deserves. The risk … may arise because of the nature of the evidence itself

---

23 CLR, *Jury Directions: A Jury-Centric Approach* (n 11) iii. See also: Weinberg (n 14) 7–8; *R v Adomako* [1995] 1 AC 171, 189 (Lord Mackay). Directions on evidence and processes of reasoning, and some directions on law, may be paternalistic: see, eg, QLRC, *A Review of Jury Directions* (n 2) vol 2, 506 [16.4], 509 [16.16]. See also NSWLRC, *Jury Directions* 2012 (n 2) xii, 9, 10, 48. The NSWLRC explained at 9 [1.28]:

> The system of jury directions continues to operate according to a basic premise that jurors will have difficulty in fulfilling their responsibilities without appropriate guidance from the judge. Jury directions aim to help jurors carry out their role of deciding issues of fact in the light of the applicable principles of law.

24 Reform of directions is part of broader system reforms, concerned with (or justified by) efficiencies, particularly the goal of reducing the number of successful appeals and associated retrials. See CLR, *Jury Directions: A Jury-Centric Approach* (n 11) iii, 132.

25 CLR, *Jury Directions: A New Approach* (n 11) 15.

26 *Alford v Magee* (1952) 85 CLR 437, 466; *Longman* (n 16). The common law requirement is more exacting than the *Evidence Act* (n 2). See, eg, ss 165(2)–(3).

27 Legislatures and appellate courts have placed obligations on defendants (through defence counsel) to request directions. Defence counsel are required to identify issues and assist with the content of applicable directions. See: CLR, *Jury Directions: A New Approach* (n 11) 15; Weinberg (n 14) 309, 313, 320. While these developments might seem reasonable from the perspective of appellate courts, and be deemed appropriate by those concerned with institutional efficiencies, in practice they add to the burdens on the most poorly resourced participant(s) in the criminal trial process.

> or because of the significance which may be attached to it by the jury having regard to the evidence in the context of the trial as a whole.[28]

The so-called 'special knowledge or experience' of judges tends to be distilled into bench books, where guidelines or model directions are set out in a form that trial judges are encouraged to adapt to the circumstances of sui generis proceedings.[29]

Directions, along with the commitment to their efficacy, enable judges to admit risky evidence or evidence that might otherwise require exclusion. Our case study considers voice identification evidence, in the shadow of claims about extensive judicial experience with notoriously unreliable identification evidence.[30] The directions in the New South Wales Bench Book ('Bench Book') in relation to voice identification represent a relatively recent addition, supplementing long standing guidance on visual identification by eyewitnesses.[31] The Bench Book does not distinguish between direct and indirect witnesses (for example, those listening in real time compared to those listening to a recording), or draw attention to non-expert investigators expressing their opinions about the identity of speakers.[32] A joint report prepared by the Australian Law Reform Commission ('ALRC'), New South Wales Law Reform Commission ('NSWLRC') and Victorian Law Reform Commission ('VLRC') suggested that

> the most significant difficulty with identification evidence is that — in contrast with other categories of oral testimony — the confidence or apparent credibility

28 *R v Stewart* (2001) 52 NSWLR 301, 322 (Howie J) (emphasis added). See also 'Supreme and District Courts Criminal Directions Benchbook', *Queensland Courts* (Web Page, 14 September 2021) Introduction <https://www.courts.qld.gov.au/court-users/practitioners/benchbooks/supreme-and-district-courts-benchbook>: 'It is the judge's duty to give the jury the benefit of the judge's knowledge of the law and to advise them in the light of the judge's experience as to the significance of the evidence'.

29 A Victorian Law Reform Commission ('VLRC') Report suggested that trial judges 'often face problems in determining when to give directions and in formulating the content of directions': Victorian Law Reform Commission, *Jury Directions* (Final Report No 17, May 2009) 8 ('VLRC, *Jury Directions*').

30 CLR, *Jury Directions: A Jury-Centric Approach* (n 11) xii.

31 Judicial Commission of New South Wales, 'Identification Evidence: Voice Identification', *Criminal Trial Courts Bench Book* (Web Page, October 2012) <https://www.judcom.nsw.gov.au/publications/benchbks/criminal/identification_evidence-voice.html> ('Bench Book'). Drawing primarily from cases such as: *Alexander v The Queen* (1981) 145 CLR 395; *Domican v The Queen* (1992) 173 CLR 555 ('*Domican*'); *Bulejcik v The Queen* (1996) 185 CLR 375 ('*Bulejcik*'); *Festa v The Queen* (2001) 208 CLR 593; *R v Dickman* (2017) 261 CLR 601 ('*Dickman*'); *R v Dupas [No 3]* (2009) 28 VR 380, 462–3 [357] (Weinberg JA).

32 Consider the issues raised with respect to (direct) 'voice identification' suggested in: Bench Book (n 31) [3–100]–[3–120]. See also NSWLRC, *Jury Directions* 2012 (n 2) xi [0.7], 1314 [1.43].

> of an eyewitness [does] not necessarily correlate with the degree of accuracy of this person's identification.[33]

The Queensland Law Reform Commission ('QLRC') drew attention to 'some weaknesses', such as poor light or distance, as matters 'of common sense' whereas 'other potential weaknesses may be "very different from what people expect them to be"'.[34] This focus on the danger of the confident but mistaken eyewitness is consistent with a common law tradition recognising the possibility of wrongful convictions caused by misidentification.[35] The uniform evidence law (UEL)[36] includes a specific section, expressed in apparently mandatory terms, requiring a direction of special caution to be made whenever identification evidence has been admitted against a defendant in a criminal trial.[37]

One of the limitations of the UEL is the narrow scope of the definition of 'identification evidence'. The drafting of the UEL reflects peculiar concerns with eyewitnesses — prioritising the use of formal, live identification parades over other forms of visual identification procedures such as dock identifications, single suspect

---

33 Australian Law Reform Commission, New South Wales Law Reform Commission and Victorian Law Reform Commission, *Uniform Evidence Law* (Report, December 2005) 428 [13.5] ('*Uniform Evidence Law Report*').

34 QLRC, *A Review of Jury Directions* (n 2) vol 2, 526–7 [16.60].

35 *Domican* (n 31). Such concerns are not new, the mistaken identification of Adolf Beck (1841–1909) led to the creation of the English Court of Criminal Appeal in 1907. See also: Hugo Münsterberg, *On the Witness Stand: Essays on Psychology and Crime* (McClure, 1908); *Davies v The King* (1937) 57 CLR 170. More recently, eyewitness error has become a very conspicuous issue in the United States following high profile DNA exonerations. See: Brandon L Garrett, *Convicting the Innocent: Where Criminal Prosecutions Go Wrong* (Harvard University Press, 2011); *New Jersey v Henderson*, 27 A 3d 872 (NJ) (2011); National Research Council et al, *Identifying the Culprit: Assessing Eyewitness Identification* (National Academies Press, 2014).

36 The UEL legislation has been adopted by the Commonwealth, Australian Capital Territory, New South Wales, Northern Territory, Tasmania and Victoria. The relevant acts are: *Evidence Act 2011* (ACT); *Evidence Act 1995* (Cth); *Evidence Act 1995* (NSW) ('*Evidence Act* (NSW)'); *Evidence (National Uniform Legislation) Act 2011* (NT); *Evidence Act 2001* (Tas) ('*Evidence Act* (Tas)'); *Evidence Act 2008* (Vic) ('*Evidence Act* (Vic)'). See also 'Uniform Evidence Acts Comparative Tables', *Attorney-General's Department* (Web Page, 27 August 2015) <https://www.ag.gov.au/legal-system/publications/uniform-evidence-acts-comparative-tables>.

37 *Evidence Act* (n 2) s 116. Victoria has moved most of their warnings from the *Evidence Act* (Vic) (n 36) to the *Jury Directions Act 2015* (Vic) following the formal reviews discussed in this section. It is worth noting that in *Dhanhoa v The Queen* (2003) 217 CLR 1, the High Court limited the effect of the apparently mandatory language in s 116 of the *Evidence Act* (NSW) (n 36), holding instead that such a direction was only required where identification is in issue — where 'in issue' is construed narrowly: at 9 [22] (Gleeson CJ and Hayne J), 16 [53]–[54] (McHugh and Gummow JJ), 26–7 [92]–[94] (Callinan J).

show ups and identifications based on photographs.[38] The UEL addresses identification evidence about the presence of the defendant at a relevant location that is offered by a person who was also present at that location at the same time.[39] Specific sections, designed to regulate the admission of identification evidence, apply only to visual identifications by eyewitnesses and do not regulate image comparison evidence.[40] While identification (or recognition) of a defendant's voice falls within the UEL's definition of 'identification evidence' and thus may require a direction, this does not extend to those listening to voice recordings — ie, indirect voice identification or voice comparisons.[41] Thus, any directions crafted to address the unreliability of voice comparison evidence, or voice identification evidence based on intercepted recordings, sit outside the traditional (albeit often still very limited) formulations designed to address the risks associated with eyewitness identifications. Modern jurisprudence reveals a strong preference for managing admissibility challenges by admitting opinions about identity in conjunction with the provision of a warning about dangers drawn from the experience of the courts.[42]

Given their prominent role in facilitating the admission of many kinds of evidence, and claims about their contribution to the fairness of criminal proceedings, it is informative, at the very least, to touch upon the scholarly study of directions, and characterisations of that research, by law reform bodies. Law reform commissions

---

[38] See, eg, Law Reform Commission, *Evidence* (Interim Report No 26, 1985) vol 1, ch 18. Since the 1980s and 1990s, cameras and recording devices have rapidly proliferated.

[39] See *Evidence Act* (n 2) sch, Dictionary, pt 1 (definition of 'identification evidence'):

> ***identification evidence*** means evidence that is:
>
> (a) an assertion by a person to the effect that a defendant was, or resembles (visually, aurally or otherwise) a person who was, present at or near a place where:
>
> (i) the offence for which the defendant is being prosecuted was committed; or
>
> (ii) an act connected to that offence was done;
>
> at or about the time at which the offence was committed or the act was done, being an assertion that is based wholly or partly on what the person making the assertion saw, heard or otherwise perceived at that place and time; or
>
> (b) a report (whether oral or in writing) of such an assertion.

[40] See, eg, *Evidence Act* (n 2) ss 114–15. Tasmania did not adopt these sections in the *Evidence Act* (Tas) (n 36) but has adopted s 116 regarding directions.

[41] For more detail, consider the issues on (direct) 'voice identification' suggested in Bench Book (n 31) [3–120].

[42] See, eg, *Dickman* (n 31) 610–11 [30], 613 [38], 619 [57]. Cf the preferable approaches adopted by the Victorian Court of Appeal: *Bayley v The Queen* (2016) 260 A Crim R 1, 12–15 [55]–[75], 19 [97]; *Dickman v The Queen* [2015] VSCA 311, [112] (Priest JA and Croucher AJA). Notwithstanding *Smith v The Queen* (2001) 206 CLR 650, a significant limitation on exclusion is the constraint imposed by *IMM v The Queen* (2016) 257 CLR 300 ('*IMM*') — making it difficult to exclude even poor and/or compromised identifications under the UEL (*Evidence (National Uniform Legislation) Act 2011* (NT) s 137). See also *Dupas v The Queen* (2012) 40 VR 182, 235–7 [199]–[206]. Judges and counsel have, on occasion, sought to engage with relevant scientific materials: see, eg, *Winmar v Western Australia* (2007) 35 WAR 159, 166–7 [26]–[30].

and a few attentive judges have acknowledged that, when considered against the results of scientific studies, many of the qualities attributed to directions by appellate courts are overstated or worse.[43] For example, a NSWLRC report on jury directions explained that the report was written 'in the context of a growing concern in Australia and overseas about the problems associated with jury directions'.[44] In terms of the recognition of problems associated with directions, the QLRC reported that

> the status afforded to jury decisions in the criminal justice system, has also led to many assumptions about the way in which juries operate and, importantly for this review, the way in which juries respond to the instructions, directions, comments and warnings given to them by judges. Some of these assumptions do not withstand scrutiny and are challenged by some of the empirical evidence, particularly from psychological and psycho-linguistic sources.[45]

A review conducted by the Victorian Department of Justice accepted that

> [w]hile it is agreed that jurors generally perform their role conscientiously, it is increasingly recognised that what is expected of jurors is unreasonable. This is due to the length and complexity of the issues and material with which they are confronted and, sometimes, the manner in which those issues are presented.[46]

A subsequent review led by Justice Weinberg concluded that 'jury directions are, by and large, unduly complex and in need of reform'.[47]

The NSWLRC report summarised the concerns as follows:

> There is growing awareness that jury directions are not always working well in guiding jurors in their task. There are concerns that jury directions are becoming too complex and uncertain to meet their intended purposes, and that they rely on outmoded communication methods that may confuse rather than assist the jury.[48]
>
> While the [scientific] research … indicates that directions can and do influence juror decision-making, it also reveals that jurors over a number of common law

---

43 See, eg, NSWLRC, *Jury Directions* 2012 (n 2) 23–6. Law reform commissions and reviews reference, and appear to accept, the basic thrust of the scientific research, even though some of the authors (or judicial overseers) occasionally express impressionistic anxiety about research methods. Most of these issues, primarily concerned with the use of mock jurors and other ecological issues, have been substantially addressed. See, eg, Brian H Bornstein et al, 'Mock Juror Sampling Issues in Jury Simulation Research: A Meta-Analysis' (2017) 41(1) *Law and Human Behaviour* 13.

44 NSWLRC, *Jury Directions* 2012 (n 2) 1.

45 QLRC, *A Review of Jury Directions* (n 2) vol 1, 28 [3.11].

46 CLR, *Jury Directions: A New Approach* (n 11) 19.

47 Weinberg (n 14) vi [1.2]. See also VLRC, *Jury Directions* (n 29) 4.

48 NSWLRC, *Jury Directions* 2012 (n 2) xi [0.3]. See also at 23 [1.67], 8 [1.22].

> countries have real difficulties in understanding the directions that they are given.[49]

The executive summary in the NSWLRC report refers to challenges confronting jurors due to the volume of evidence, the complexity of evidence (for example, DNA statistics) and traditional forms of legal presentation.[50] The NSWLRC report also refers to audio and video evidence from surveillance devices, but voice identification and emerging problems with police surveillance recordings are not discussed.[51] A report by the ALRC, NSWLRC and VLRC reviewing the UEL acknowledged concern about directions on issues that are 'new, difficult or counter-intuitive to jurors' commonsense'.[52]

Most of the recent reviews and reports — for example, those from New South Wales, Queensland and Victoria — refer to critical literatures and the need to engage with empirical studies. Somewhat incongruously, reviews tend to favour persisting with prevailing practices. Most recommendations are focused on improving comprehension, reducing length and complexity and limiting scope for appeal by placing greater obligations on defendants (and defence counsel).[53]

Findings and recommendations in the various law reform commission reviews and reports (always judge-led and judge-heavy) tend to be milder than the critical, empirically based contributions of scientists (and legal scholars). Consider, for example, an assessment by researchers James Ogloff and Gordon Rose:

> jurors appear largely incapable of understanding judicial instructions as they are traditionally delivered by the judge. … the overwhelming weight of the evidence is that the instructions are not understood and therefore cannot be helpful.[54]

---

49 Ibid 28 [1.82].

50 Ibid xi [0.5].

51 Ibid. See also ibid 5, 104, 123, 124. However, it should be noted that recommendation 5.6 is directed towards identification from images of the crime scene (eg, closed-circuit television (CCTV)): at xxi, 107–8 [5.119]. The report also discusses giving jurors access to transcripts of audio and video recordings to assist them to listen to or view the evidence: at 124–5.

52 *Uniform Evidence Law Report* (n 33) 593 [18.10].

53 A cynic might note that the initiatives, described as enhancing trial values, are primarily directed toward trial efficiency and reducing the frequency of appeals. There are few modifications that address the effectiveness of directions or the fairness of proceedings.

54 James R P Ogloff and V Gordon Rose, 'The Comprehension of Judicial Instructions' in Neil Brewer and Kipling D Williams (eds), *Psychology and Law: An Empirical Perspective* (Guildford Press, 2005) 407, 425. Jurors process evidence as the trial progresses, influenced by overarching narratives, assumptions and insights that are not restricted to admissible evidence: see Timothy D Wilson and Nancy Brekke, 'Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations' (1994) 116(1) *Psychological Bulletin* 117, 117. While jurors are expected

And, by Lora Levett and colleagues:

> Jurors do seem to have some problems evaluating the reliability of some types of evidence (e.g., eyewitness evidence, confession evidence, expert evidence), and the procedural safeguards intended to assist their discernment of reliability appear to be relatively ineffective. Jurors are also influenced by extra evidentiary factors …[55]

In an influential review of empirical research published in 1997, Joel Lieberman and Bruce Sales concluded that

> it has been consistently shown that jurors do not understand a large portion of the instructions presented to them. It is common to find over half the instructions misunderstood, and even the most optimistic results indicate that roughly 30% of the instructions are not understood.[56]

Citing a study from 1947, Lieberman and Sales noted that the 'strong evidence for a lack of comprehension on the part of jurors … is not new'.[57] They also observed that notwithstanding the longevity of concerns, 'not much has been done by the legal community to address the problem'.[58] A more recent review, focused on identification evidence, summarised the research as indicating 'that jurors often have difficulty understanding and utilizing instructions when determining verdicts'.[59]

While, to various degrees, law reform commissions and commentators recommend changes to directions aimed primarily at improving comprehension (in part by reducing the volume of directions required), there have been relatively few recommendations on the subject of unreliable evidence and continuing reliance on the special knowledge and experience claimed by judges. Typically, law reform bodies seem to be broadly satisfied with the way such evidence is regulated by admissibility rules and directions.[60] Satisfaction is invariably indexed to the availability of

---

to bring their life experience, there are risks from both prejudices (for example, latent racism or beliefs about sexual assault) and the way certain procedures or evidence may induce prejudice (for example, through cognitive biases such as expectations and confirmation).

55 Lora M Levett et al, 'The Psychology of Jury and Juror Decision Making' in Neil Brewer and Kipling D Williams (eds), *Psychology and Law: An Empirical Perspective* (Guildford Press, 2005) 365, 396.

56 Joel D Lieberman and Bruce D Sales, 'What Social Science Teaches Us about the Jury Instruction Process' (1997) 3(4) *Psychology, Public Policy, and Law* 589, 596–7.

57 Ibid 637, citing J Hervey, 'Jurors Look at Our Judges' (1947) 25(1) *Oklahoma Bar Association Journal* 1508.

58 Lieberman and Sales (n 56) 637.

59 Christine M McDermott and Monica K Miller, 'Do Judges' Instructions about Eyewitnesses Really Work?: A 2019 Update' (2019) 55(3) *Court Review* 104, 104.

60 See Weinberg (n 14) 280 [5.14]: 's 165 does not appear to be creating significant problems in practice'.

other trial safeguards, such as scope to cross-examine witnesses.[61] The evidence at the centre of this article, a species of identification evidence, is the kind of evidence that courts and law reformers believe is currently managed reasonably well by admission, cross-examination and directions — provided trial judges 'point out significant matters affecting reliability'.[62]

There are several problems with the reliance on directions in conventional legal proceedings. First, do the directions accurately or adequately capture the law or the dangers with the evidence, and if so, do they convey them clearly and in a manner that is comprehensible and likely to be comprehended?[63] Most of the proposals for reforming directions reflect key recommendations by the VLRC based around directions being 'clear', 'simple', 'brief', 'comprehensible' and 'tailored to the circumstances of the particular case'.[64] These seem to be necessary but hardly sufficient as a foundation for the heavy and continuing reliance on directions as a fundamental safeguard. Law reformers have dedicated limited attention to the content and accuracy of directions pertaining to evidence (rather than law). While there are 'limits to jurors' powers of comprehension', the problems are more profound and extend well beyond effectively communicating content.[65] For, as we shall see, even scientifically informed directions, understood by the jury, may not be capable of addressing or remediating dangers. There would seem to be a need for judges (or legislatures) to have a clear and accurate idea of risks and dangers, such that investigations, trial procedures, admissibility and the use of evidence at trial can be managed appropriately.

There is, in addition, the problem of jury acceptance. The jury may understand a direction but not accept the content or the magnitude of the dangers.[66] The confidence of an identification witness and the impact of their testimony on the assessment of credibility is an example of an issue that might be raised, but is unlikely to be corrected, with directions. For jurors, their assessment of witness credibility is not entirely conscious and not simply managed through cognitive effort. Despite these issues, almost all of the judicial commentary on directions is positive. Much less attention is given to the limits of directions and whether some types of evidence and some threats to cognition and rationality, cannot be — or are not likely to

61 See Gary Edmond et al, 'Forensic Science Evidence and the Limits of Cross-Examination' (2019) 42(3) *Melbourne University Law Review* 858.

62 Weinberg (n 14) ix.

63 NSWLRC, *Jury Directions* 2012 (n 2) xix, for example, refers to the need for directions to be legally accurate but there is no parallel concern with the content being empirically based or scientifically informed. The closest is 'practical advice' but this is based on a concept of practicality steeped in traditional legal practice and judicial experience.

64 VLRC, *Jury Directions* (n 29) 13. See also NSWLRC, *Jury Directions* 2012 (n 2) 29.

65 Weinberg (n 14) 14 [1.47].

66 Lieberman and Sales (n 56) 609; Chantelle M Baguley, Blake M McKimmie and Barbara M Masser, 'Deconstructing the Simplification of Jury Instructions: How Simplifying the Features of Complexity Affects Jurors' Application of Instructions' (2017) 41(3) *Law and Human Behavior* 284.

be — managed or avoided. These are serious issues, for directions 'can only be effective to the extent that they are comprehended by the jury' and, we might add, accurate, accepted and actionable.[67]

This article's challenge extends beyond the clarity and comprehension of directions. Our case study exemplifies an over-reliance on jury directions. It demonstrates how well-intended directions, believed to eliminate substantial unfairness, do not work. It is not restricted to clarity and comprehension, but applies to standard directions that appear incapable of supporting the fundamental aspiration to render jury trials fair and verdicts rational. Our case study illustrates how in some cases it does not matter whether directions are given or understood because they are incapable of conveying or, more importantly, overcoming the threats to rational decision-making introduced by some evidence and some trial procedures. In these circumstances, reliance on directions deceives judges, jurors, lawyers and the public. It also encourages complacency amongst lawyers, judges and police officers.[68] The provision of directions may make claims about trial fairness seem plausible, even persuasive, without attending to the much more difficult question of whether trials are substantially fair.

To be clear, we accept that jurors typically take their task seriously and are conscientious.[69] These criticisms and concerns are not directed at jurors, but rather at the continuing and heavy reliance on directions as meaningful correctives to the misuse of evidence and other risks of unfair prejudice confronting decision-makers. The ongoing heavy reliance on directions and the unrealistic expectations routinely placed upon them seems to reflect an unwillingness to take decades of mainstream scientific research seriously and a failure by the judiciary to rigorously understand (or study) the practices and procedures they routinely preside over, review and defend. The limitations of directions explored in this article are revealing, especially given the number and magnitude of reports on jury directions which have been produced in Australia in the last decade or so. The limitations are all the more revealing because some of the reviews, including the report by the NSWLRC, explicitly referred to challenges that 'derive from the exponential increase in the use of scientific techniques to investigate and prosecute crime'.[70]

---

[67] Lieberman and Sales (n 56) 591.

[68] Additionally, to compound matters, the prevailing confidence in their effectiveness (or adequacy), means that merely giving a direction is likely to satisfy an appellate court that a trial was (formally) fair.

[69] See, eg, QLRC, *A Review of Jury Directions* (n 2) vol 1, 29 [3.16]. This article is not intended as a defence, or critique, of the modern jury. Our criticisms are directed at procedures and structures, based on misguided assumptions and commitments.

[70] NSWLRC, *Jury Directions* 2012 (n 2) 75 [5.2]. Yet, apart from DNA profiling evidence and cursory comments on CCTV, the reports say almost nothing about new problems and new research. The various law reform reports are silent about emerging independent scientific reviews including, reports by the National Academy of Sciences (US), the President's Council of Advisers on Science and Technology (US), the National Institute for Standards and Technology (US) and guidelines issued by the Forensic Science Regulator (UK). See, eg: President's Council of Advisors on Science and Technology, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of*

III Introducing Our Case Study: *Davey v Tasmania* [2020] TASCCA 12

Matthew Davey, David Eaton and Daniel Cure were convicted for stealing firearms in the Supreme Court of Tasmania.[71] Davey and Eaton appealed their convictions in person.[72] Notwithstanding the numerous grounds of appeal, this article will only address the opinion evidence adduced to identify Davey and Eaton as two of the individuals involved in a joint criminal enterprise. Voice identification was central in the circumstantial cases against both appellants and 'was arguably the most significant evidence against the appellant Davey'.[73] Following an expansive investigation (named Operation Oracle) involving covert phone surveillance, the voices of Davey and Eaton were positively identified by one of the investigating police officers.[74] Detective J was called by the prosecutor and allowed 'to give opinion evidence at trial as to the identification of voices on telephone intercepts'.[75]

The admissibility of Detective J's voice identification evidence was unsuccessfully challenged on the voir dire. Following *Kheir v The Queen*,[76] *Nguyen v The Queen*,[77]

Feature-Comparison Methods (Report, September 2016) ('PCAST Report'); National Academy of Sciences, *Strengthening Forensic Science in the United States: A Path Forward* (National Academies Press, 2009) ('NAS, *Strengthening Forensic Science*').

71 *Davey* (n 7) [3], [11]. They were charged with stealing and aggravated burglary. A caveat: before embarking on this critique it is important to make clear that we are agnostic on the question of whether it is Davey (or Eaton) speaking on one or more of the questioned voice recordings. Whether it was Davey speaking or not, the admission of Detective J's opinions and the provision of the recordings to the jury to compare, and the frailties of the jury directions, all contributed to Davey's trial being substantially unfair.

72 This seems significant because they appear to have raised several important points about the voice identification evidence, which might not have been advanced by legal counsel.

73 *Davey* (n 7) [29] (Estcourt JA). Other evidence included: insider information about the absence of the owners of the property; a DNA match with a recovered screwdriver; recovered firearms, including a gun with a fingerprint matched to a relative of Davey's; instructions over the phone; the phone numbers and possession of the phones; and names and other call content: see [15]–[17], [30], [46], [51], [98].

74 We will refer to the police officer as 'Detective J', rather than using the police officer's actual name. It is important to stress that the following discussion is not intended as personal criticism of the investigators, lawyers or judges. Investigators, lawyers and judges who operate in good faith, but are not conversant with scientific methods and knowledge, can unwittingly produce confident opinions that are much more error-prone or controvertible than they appear. We accept that each of the actors might have acted with integrity. In consequence, this critique is directed at rules, procedures and a range of misguided assumptions and beliefs, including misplaced confidence in the ability of directions to overcome serious dangers and biases.

75 *Davey* (n 7) [7].

76 *Kheir* (n 9).

77 (2017) 264 A Crim R 405 ('*Nguyen* 2017').

R v Phan,[78] and *Tasmania v Farhat*,[79] the 'learned trial judge ruled the evidence admissible as lay opinion evidence pursuant to s 78 of the *Evidence Act*'.[80] On appeal, Davey argued that 'Detective [J] was not qualified to give opinion evidence as to the identification of voices heard on telephone intercepts captured by investigating police'.[81] That contention was rejected. The Tasmanian Court of Criminal Appeal ('TASCCA') identified no fewer than four potential admissibility pathways for Detective J's opinions, concluding that 'it is likely to be unnecessary for a trial judge to devote too much time to an analysis of the authorities, as in most cases the evidence will be admissible'.[82]

At trial, Detective J testified that he was able to identify Davey because 'he listened to between 720 and 1200 calls during the operation, and that he had also compared this voice to that of the appellant Davey's in his record of interview with police'.[83] He described Davey's voice as 'consistent in his manner of speech and that his speech was deep and unique'.[84] According to the trial judge, there were two parts to Detective J's evidence. The first related to intercepted calls that 'contained markers' such as the name or nickname of Davey — including 'Matthew', 'Matty' or 'Matthew Davey'.[85] Detective J testified that 'Davey was referred to by name' in about '30 to 50' calls.[86] He used these calls, which he attributed to a single speaker (implicitly Davey), to identify other similar voices among the many intercepted calls. When attributing these calls to the named speaker, Detective J was exposed to the telephone numbers and the locations of calls and had access to transcripts of the calls, as well as other metadata.[87] Detective J was a member of Operation Oracle, understood the grounds on which the warrants had been issued, and knew about other evidence implicating Davey in the crime.[88] The second part of the identification was Detective J's belief that when he compared 'those [named] calls and the calls in question' from among the intercepts it was 'the same voice as in the interview he conducted when he spoke to Mr Davey

---

78 *Phan* (n 9).

79 *Farhat* (n 9).

80 *Davey* (n 7) [54]. See also *Davey* (n 7) [59]–[68].

81 Ibid [53].

82 Ibid [68] (Estcourt JA, Blow CJ agreeing at [2], Geason JA agreeing at [110]).

83 Ibid [55]. This is a large and perhaps revealing range. We are not told how long the recordings were (individually or collectively) or how many of them were alleged to include Davey and Eaton. Depending on the nature of the operation, Davey and Eaton may have been central or peripheral.

84 Ibid.

85 Ibid [56]–[57].

86 Ibid. The broad range suggests that these are estimates or guesses rather than based on an actual count. See also *R v Solomon* (2005) 92 SASR 331, 344 [49] (Doyle CJ).

87 *Davey* (n 7) [57].

88 Detective J presumably also knew things that were inadmissible, such as prior involvement with police and offending, and even subsequent offending. See *Causon v Tasmania* [2021] TASCCA 13.

in person'.[89] Detective J attended Davey's police interview in order to listen to his voice and compare it with the voice(s) he had attributed to Davey on the intercepted recordings. The comparison was said to have confirmed Detective J's voice identification. The many hours spent listening and re-listening to the intercepted calls was the basis advanced and relied upon for the admission of Detective J's lay opinions.[90]

The trial judge provided directions to the jury on Detective J's voice identification evidence in conventional legal terms. These are reproduced in Part IV of this article. The jurors were told that the reliability and significance of Detective J's opinion evidence was a matter for them to determine.[91] While the directions were not challenged on appeal — a fact that might imply they were understood as appropriate or difficult to impugn — we consider their adequacy and effectiveness in detail in Part V.

In response to Davey's challenge to the admission of Detective J's opinions and the fairness of the trial, the directions were said to fulfil an important function. Writing for the TASCCA, Estcourt JA concluded:

> It follows from all that I have said that I accept the submission made by counsel for the State, that, having regard to the state of the authorities, and *in light of the directions given to the jury*, the evidence of Detective [J] was properly admitted and *there was no unfair prejudice from the admission of the evidence*.[92]

The directions facilitated the admission of Detective J's opinion evidence and were said to have removed any associated unfairness.

In addition to Detective J's testimony, the jurors were encouraged to undertake their own assessment of the voice recordings, relying on the intercepted telephone calls and the recording of Davey's police interview.[93] These recordings were said to assist the jurors with the evaluation of Detective J's opinions, but they were simultaneously available for their own voice comparison and identification. Revealingly, the jury comparison does not appear to have been raised as an issue at trial or on appeal.[94] Rather, its propriety and value for fact-finding seem to have been taken

89 *Davey* (n 7) [57].

90 He was assumed to have some advantage over the jury, even though there is no evidence of an ability and ability is not an express requirement of s 78 of the *Evidence Act* (Tas) (n 36). Its significance depends on the way 'necessary' is constructed.

91 *Davey* (n 7) [57].

92 *Davey* (n 7) [74] (Estcourt JA, Blow CJ agreeing at [2], Geason JA agreeing at [110]) (emphasis added).

93 Ibid [57].

94 See Gary Edmond, 'Against Jury Comparisons' (2022) 96(5) *Australian Law Journal* 315. A similar issue arose in *Dickman* where without apparent remark the jury were invited to compare the defendant's voice recorded during a search of his house with a voice captured in the background of an intercepted phone call: see *Dickman* (n 31) 609 [21].

for granted. The jury comparison plays a cameo role in the judicial directions. In describing the jury comparison, the TASCCA in *Davey* drew support from the Full Court of the South Australian Supreme Court in *R v Phan*,[95] quoting the following statement from Hinton J:

> If it is permissible for the jury to undertake voice comparison because '[r]ecognition of a speaker by the sound of the speaker's voice is a commonplace of human experience', it follows that evidence of voice comparison does not fall exclusively within the province of experts and expert opinion evidence.[96]

We will return to differences between recognition and comparison as we consider whether the opinions of those who are not experts should be admitted, whether jurors can be trusted to evaluate their opinions or to make their own comparisons, as well as the ability of directions to identify and reliably mitigate potential problems.

Davey's co-accused, Eaton, was also identified as one of the speakers on the incriminating recordings. The voice identification evidence admitted against Eaton was qualitatively different to the identification of Davey. According to trial testimony, Detective J and Eaton knew one another prior to the theft. There was said to be familiarity between Detective J and Eaton that was unrelated to the investigation.[97] The TASCCA described the pre-existing relationship in the following terms:

- That [Detective J] had known the appellant personally for approximately 10 years.
- That he had regular contact with the appellant in around 2010 in the course of his employment at Bread Café.
- That in this period he communicated with the appellant, they referred to each other by name and the appellant afforded him nicknames.
- That he continued to have contact with the appellant from 2010 until recently. That he spoke to the appellant on approximately 12 occasion and they continued to share a familiarity.
- That he recognised the appellant's voice in some of the telephone intercept material based on his previous dealings.
- That he considered the appellant's voice to be 'quite high pitched with a quite laconic drawl to it' and that assisted in his identification.
- That he listened to approximately 175 calls, on multiple occasions, where he was able to identify the appellant as a speaker.

---

95 *Phan* (n 9).

96 *Davey* (n 7) [61] (Estcourt JA), quoting *Phan* (n 9) 152 [59] (Hinton J). This is a curious statement because it involves ungrounded reasoning.

97 Without much insight into the actual detail of the prior exposure we are taking this at face value because it allows us to make a point. In practice, the degree of familiarity should be an issue that requires more than passing attention.

- That he spoke with the appellant at the conclusion of the investigation and said 'it reaffirmed the belief that I held that it was the voice of Mr Eaton'.[98]

Detective J was said to have recognised Eaton on the intercepted recordings based on prior familiarity with his voice. This evidence was also admitted at trial.[99] In the case against Eaton, jurors were not presented with a recording of his 'no comment' record of interview.[100] Consequently, the jurors were not in a position to undertake their own voice comparison. In Eaton's case, they were required to consider the opinion of Detective J, and listen to the quality of the recordings of the voice attributed to Eaton, in conjunction with the other evidence in the circumstantial case against Eaton.[101] The trial judge instructed the jurors that the question of Detective J's recognition was ultimately a matter for them and they were 'perfectly entitled to have regard to his opinion [and] ultimately … perfectly entitled to accept it *if [they] consider[ed] [it] reliable*'.[102]

Responding to Eaton's appeal, the TASCCA explained that Detective J's 'evidence was even stronger, and was in reality, voice recognition evidence':[103]

> Whilst the basis of the identification of Eaton's voice was different to that of Davey's, the evidence of Detective [J] was admissible and was properly admitted. That Eaton's otherwise inadmissible 'no comment' record of interview with police was not played to the jury, adds nothing to the argument. Nor does the suggestion that Detective [J] had never spoken to Eaton on the telephone.[104]

Curiously, the precise basis for the admission of this opinion evidence was not actually explained.[105]

---

98 *Davey* (n 7) [98]. See generally: *R v Leaney* [1989] 2 SCR 393; United Kingdom Home Office, *Code D: Revised Code of Practice for the Identification of Persons by Police Officers* (Code of Practice, February 2017), brought into operation under the *Police and Criminal Evidence Act 1984 (Codes of Practice) (Revision of Codes C, D and H) Order 2017* (UK) SI 2017/103, ord 2, with respect to police use of video images. See also United Kingdom Home Office, *Advice on the Use of Voice Identification Parades* (Home Office Circular 057/2003, 5 December 2003), with respect to voice identification parades.

99 *Davey* (n 7) [101]. The penultimate dot point (above) is adequate as a description, but it is not evidence of ability. For we do not know if the identifications were accurate.

100 Ibid.

101 Ibid [99]. Allowing jurors to make voice comparisons may impinge on the defendant's decision to testify.

102 Ibid (emphasis added).

103 Ibid [98].

104 Ibid [101].

105 Interestingly, some judges — particularly judges in New South Wales (most now retired) — have characterised such evidence as recognition evidence, said to be a type of fact evidence not caught by the exclusionary opinion rule (*Evidence Act* (n 2) s 76). See, eg: *R v Smith* (1999) 47 NSWLR 419; *Riscuta* (n 9); *Nguyen* 2017 (n 77).

We accept that the trial judge's responses to Detective J's opinions are consistent with practice in Tasmania, as well as New South Wales, Queensland, South Australia, Victoria and Western Australia.[106] However, it is our contention that trial safeguards and appellate review do not adequately regulate the impressions of investigators, especially where they are inexpert (and therefore speculative) and obtained in conditions that are suggestive or likely to confirm expectations. For similar reasons, they are incapable of regulating jury comparisons. Furthermore, as explained in Parts V(A) and V(C), the suggestive context means that it is inappropriate to characterise Detective J's evidence as voice identification evidence at all. It is from this perspective that the heavy reliance on judicial directions and their effectiveness assumes considerable practical significance.

IV Jury Directions in *Davey*

Earwitness or direct voice identifications, as a type of 'identification evidence' under the UEL, tend to attract judicial warnings.[107] Notwithstanding some confusion around the application of s 116 of the *Evidence Act* to the indirect or displaced listening to recordings by police officers (such as Detective J), voice comparison and recognition evidence is typically treated as 'a kind that may be unreliable' such as to require a response to a request under s 165 of the *Evidence Act* and/or residual common law obligations to make sure that criminal proceedings are fair. Section 165 requires the trial judge to: (1) warn the jury that voice identification (or comparison or recognition) evidence might be unreliable; (2) 'inform the jury of the matters that may cause it to be unreliable'; and (3) 'warn the jury of the need for caution in determining whether to accept the evidence and the weight to be given to it' if a party requests they do so.[108]

'[A]fter giving the jury lengthy directions as to the dangers of identification evidence' the trial judge 'directed the jury, specifically as to Davey', in the terms set out below.[109] We reproduce these directions because, notwithstanding their apparently exemplary or orthodox nature, we contend that they are wholly inadequate. For the reasons developed below, these directions are incapable of addressing the dangers

[106] They are consistent with common law practice: see, eg, *Phan* (n 9).

[107] Notwithstanding the way intermediate appellate courts have extended s 78(a) of the UEL to include the opinions of indirect listeners — those displaced in time and space from the matter or event — the definition of 'identification evidence' that guides the application of s 116 is restricted to the identification of a defendant by a person present 'at or about the time at which the offence was committed or the act was done'. It makes little sense to treat the recording (or the voice on the recording) as the matter or event (as Basten JA does in *Nguyen* 2017 (n 77)) because that approach circumvents s 79(1) and enables anyone (with a bit of time on their hands) to form a potentially admissible opinion.

[108] The trial judge need not comply with such a request if there are 'good reasons for not doing so': *Evidence Act* (n 2) s 165(3).

[109] *Davey* (n 7) [57].

they purport to identify, convey and mitigate. Self-evidently they are unlikely to convey or mitigate dangers that are merely alluded to or omitted. In the subsequent analysis we explain why these (and many similar) directions are incapable of placing jurors in a position to rationally evaluate Detective J's opinion evidence or undertake their own comparisons of the recordings.

Specifically, as to Davey, the trial judge directed the jury:[110]

1 I turn now to the specifics of the identification by Detective [J] of Mr Davey's voice in the
2 telephone intercepts so this is part of the same direction about the need for care and caution in
3 relation to accepting identification evidence and what I'm doing is now pointing out the specifics of
4 the identification with respect to Mr Davey so Detective [J]'s evidence and aspects of his evidence
5 which you'll need to carefully consider.
6 In brief Detective [J]'s evidence was that he was not familiar with Matthew Davey's voice prior
7 to Operation Oracle. He gave evidence that he was though able to identify the voice of Matthew
8 Davey in the calls. How was he able to do that? Well, he said he was able to do that based on the
9 fact that there are a number of calls in which Mr Davey identified himself so these were the calls
10 that we listened to on the trial, P6, from the 21 December 2015 to the 11th of February such as
11 Relationships Australia call, the Dave Powell's car yard call, the use of the name Matthew or Matty
12 and he said he heard other calls, other than the ones that were played in Court, in which he heard
13 Matthew, Matty or Matthew Davey, and he said there were approximately 30 to 50 calls where
14 Matthew, Matty or Matthew Davey were used.
15 Other content in the calls also gave rise to his opinion that it was Matthew Davey speaking such as
16 address, locations, the subscriber of other phones. Detective [J] said the voice he attributed to Mr
17 Davey was consistent throughout the calls. Following the end of the operation he conducted an
18 interview with Matthew Davey and his evidence was that he made a comparison of what he heard
19 on the recordings with the voice of Matthew Davey in person and he said when he met with Mr
20 Davey, when he spoke with him, it was the same voice that he'd been hearing all along and which
21 he had attributed to Matthew Davey.
22 He said, 'I was satisfied in my opinion that it was the same voice that I'd heard since the
23 beginning of the operation.' But as you know, as I've explained, *people may be convinced that their*
24 *opinion is strongly grounded but ultimately it's a matter for the jury to carefully scrutinise that* and
25 *I've identified for you the risk of error.* There are a number of matters that have been specifically
26 raised in this case that require your consideration in determining whether the evidence identifying
27 the accused Matthew Davey can be safely acted upon. As I've already said, *bear in mind the*
28 *difficulty of keeping a memory, an imprint of a voice in your mind as opposed to a visual image.*
29 Essentially there are two parts to Detective [J]'s evidence. We've got recognition of the voice
30 as being that voice which in other calls contained markers of identity such as Matty, Matthew Davey
31 or some other marker of identity and what I say to you there is *it's a matter for you as to whether*
32 *that is a reliable marker of identity.* Secondly the evidence of Detective [J] is that when he compared

[110] We have added in line number references to the jury directions for ease of reference in our subsequent analysis.

33 those calls and the calls in question which he identified as the same, it was the same voice as in the
34 interview he conducted when he spoke to Mr Davey in person.
35 Now, as to the identification call to call, *the strength of that evidence as to identification of the*
36 *same voice will depend on the individual call. Obviously if hardly anything is said then the evidence*
37 *of identification is not as strong as a lengthy call with a lot said, which is common sense. At the end*
38 *of the day all this evidence establishes at its highest is that it is the same voice call to call at its*
39 *highest*. As to that identification call to call as being the same voice when he's listening to various
40 voices, *I warn you that mistakes can easily be made even when we're identifying the voice of*
41 *someone close to us, friend or family member, although identifying the voice of a stranger is more*
42 *difficult, much more difficult*.
43 Another question is *what opportunity did Detective [J] have to hear the voice of the person*?
44 Well, *that's self-evident from the calls that you have and you can assess that for yourselves*. As I've
45 said *the reliability of each identification depends on the length of each individual call, how much*
46 *was said and you have that information before you. The reliability of each identification is going to*
47 *vary depending on how long the person speaks for, the nature of the call and so on and so you have*
48 *the calls, you can assess that*.
49 [Defence counsel] has pointed out that *you don't have any contemporaneous notes made by Detective*
50 *[J] about his level of certainty or otherwise that would help you to scrutinise his evidence and it*
51 would seem that there's a risk that somehow he's ultimately globalised, if you like, his opinion so
52 that he's heard a whole lot of calls and ultimately he's decided that they're all so similar, that they're
53 the same calls, but what is his evidence in relation to a particular call which could be significant in
54 this case and *you don't have, if you like, the benefit of contemporaneous notes in relation to a*
55 *specific call which may assist you to scrutinise the reliability of his identification*.
56 Another point which is similar — which is really the same point that I made in relation to Eton, is
57 that *Detective [J] again had the summaries of metadata from TIS, Telephone Intercept Services*,
58 *he's expecting the call to be the voice of Matthew Davey, the risk where in terms of reliability is the*
59 *expectation has influenced his identification of the speaker*. The risk — well, *the risk is that there*
60 *are voices that are similar, but they're not identical in all their characteristics and yet with an*
61 *expectation of the speaker being a certain person, the opinion is filled in with that expectation if you*
62 *like, and so the end opinion is, it is Matthew Davey rather than it sounds like Matthew Davey so, if*
63 *you like, the person's assessment of the voice is shored up by the information they have from TIS*,
64 *another fact that you've got to bear in mind*.
65 *How clearly could the person hear the voice? How was the sound conveyed? Well, they're*
66 *telephone calls, bear in mind the risk that there's some distortion and you have the call so you can*
67 *consider that for yourselves. Was there anything about the voice that would've impressed itself upon*
68 *Detective [J], was there something distinctive about the voice*? Detective [J]'s evidence was that
69 *the voice was difficult to hear, well that's a factor that bears on reliability*. He said, 'It was difficult
70 to hear and understand because it was mumbly, it was deep.' He said, 'To me it was quite unique,'
71 so his evidence was that *to him it was quite a unique voice and that may assist you in assessing the*
72 *reliability of the — that recognition or identification evidence*.
73 *How long did he have to keep the characteristics of the voice in his mind before identifying the*
74 *voice as that of the accused*? Well, ultimately his evidence was, 'I couldn't be certain it was Matthew
75 Davey until I spoke to him in person,' and his evidence was that he interviewed the accused on the

76 26th of April 2016. Now, a final point here is that *you are, yourselves, entitled to compare the voice*
77 *of the accused as you have heard it during the police interview with the voice on the recordings in*
78 *order to assess Detective [J]'s opinion so you have the police interview with Mr Davey, you have*
79 *the recordings and you can undertake that comparison*.
80 Here *bear in mind the risk that you're listening to a recording and the recorded police interview*
81 *may have distorted Mr Davey's voice to some extent*. You need to *consider the risk that the callers*
82 *had a voice similar to that of Mr Davey but were not Mr Davey or was not Mr Davey and that*
83 *Detective [J] was honest but mistaken in his identification of Mr Davey*. In other words *the risk*
84 *that Detective [J] had confused Matthew Davey's voice with another similar voice, a deep,*
85 *mumbly male voice. That's a risk that you need to take into account.*
86 Now, I'm required by law to point out all of these factors because *all of these may bear on the*
87 *reliability of Detective [J]'s opinion and you must give consideration to these matters. Any one of*
88 *these circumstances may possibly lead to error*.[111]

V Evaluating Detective J's Opinion Evidence in Light of the Warnings

These directions draw attention to a range of factors that 'bear on the reliability' (lines 86–7) of Detective J's 'certain' identification of Davey's voice (lines 17–21, 74–5). They highlight the need to 'bear in mind' (lines 27, 64, 66, 69, 80, 86) or 'consider' (lines 5, 26, 67, 81, 87) features of the recordings and Detective J's identification, such as: the length of calls and the quality of the recordings (lines 43–8, 80–1); any distortion or differences in the types of listening or recording (line 80); the distinctiveness of the voice (lines 67–9), noting that Detective J considered it to be unique (lines 70–1); the opportunity Detective J had to hear the voice (line 43); Detective J's exposure to names, phone numbers, locations and metadata when listening and attributing a number of the intercepted conversations to a specific person (identified as Davey, at lines 15–21, 57–8); and that Detective J had to remember the voice from the intercepted calls when comparing (and purportedly recognising) the person speaking in the police interview (lines 27–8).

The directions also draw attention to more general issues and risks with voice comparison and identification that the jury 'need to take into account' (line 85). They distinguish between: unfamiliar and familiar voice identification (or 'recognition', at lines 29, 72) noting that 'mistakes can easily be made' when attempting to identify familiar, let alone unfamiliar voices (lines 40–2); the danger of some voices sounding similar (lines 60–2, 84); and the more general 'possibility' or 'risk of error' (line 25). Notwithstanding Detective J's testimony, the jurors were repeatedly told that the identity of the speaker was a matter for them to decide.[112] They were

---

111 *Davey* (n 7) [57] (emphasis added). There may have been other warnings about the dangers of identification, raised or repeated in relation to Davey's co-accused. There is, however, no suggestion that they provide anything more than the types of 'assistance' exemplified by these directions.

112 See also *Velevski v The Queen* (2002) 187 ALR 233.

instructed to 'carefully consider' or 'carefully scrutinise' Detective J's opinion evidence and testimony, and the way Detective J identified Davey, and listened to the recordings he relied upon, when making their assessment. The jurors were told that they '*must give consideration to these matters*' (line 87).

Here it is important to contemplate what these directions actually do — how do they assist rational evaluation?[113] How is the jury to evaluate Detective J's opinion evidence or approach the voice comparison, either individually or as part of the overall case against Davey? How are jurors to determine the likelihood that Detective J is correct, or assign a weight to his opinion that is not simply a guess or blind speculation? How are they to factor in the general and specific issues summarised in the previous paragraph, let alone Detective J's familiarity with the circumstantial case against Davey (and Eaton)?

In this Part we introduce scientific research on voice comparison and cognitive biases. Not raised by the parties, or perhaps within the experience of trial and appellate judges, this knowledge threatens Detective J's opinion evidence as well as the value of the voice comparison performed by the jurors. It casts durable doubts on the ability of the jurors (and the judges) to understand, let alone rationally evaluate, the voice identification evidence and, ineluctably, the case against Davey (and even Eaton). Mainstream scientific insights raise fundamental issues, both epistemic and procedural, that are not addressed, not adequately addressed, or not capable of being addressed, through directions however cautionary or legally orthodox.

We begin with research on the comparison, recognition and identification of voices.

A *Noise: The Difficulty of Voice Comparison, Familiarity and Detective J's Ability*

It is important to acknowledge, although not surprising, that the directions reproduced in Part IV are not completely mistaken or lacking in insight. However, they reveal little sensitivity to relevant scientific knowledge or the prevalence and magnitude of risks. They mistakenly imply that Detective J has an advantage over the jury. They do not convey the difficulty and frequency of error in voice comparison, and do not respond to the fact that Detective J's identification and the jury comparison are both irreparably contaminated by the circumstances in which the recordings are heard. Importantly, they overlook the fact that Detective J's opinions are not voice identification (or comparison or recognition) in a conventional sense, but rather based on a synthesis of all the information available to him through his participation in the investigation. We will start with error and familiarity before moving to consider the

[113] By way of analogy, what does it mean to say that a bridge or a vaccine might be unsafe or have risks, without clarification, and perhaps quantification to enable rational evaluation — for example, for the bridge, the weight of vehicles that might safely cross it and its anticipated life expectancy, and for the vaccine, its efficacy, the frequency of serious side effects and the length of protection? See Kristy A Martire and Gary Edmond, 'Rethinking Expert Opinion Evidence' (2017) 40(3) *Melbourne University Law Review* 967.

dangers introduced by context and expectations. The actual basis of Detective J's opinion evidence and the implications for admissibility will be considered thereafter.

First, voice identification is more error-prone than most people imagine. Unfamiliar voice identification, based on comparison (rather than recognition), is particularly susceptible to error. Judges and jurors tend to conceive of unfamiliar voice comparison in terms of the way we recognise familiar voices — as 'a commonplace of human experience'.[114] According to attentive scientists, unfamiliar voice (and face) comparison is nothing like familiar voice (or face) recognition.[115] They are fundamentally different cognitive tasks. The conflation leads judges and jurors to believe they are much more accurate at identifying unfamiliar voices (and faces) than they actually are.[116] Simultaneously it generates exaggerated confidence in identifications.[117] Popular beliefs about unfamiliar voice identification are misguided, in part, because of the lack of meaningful feedback we receive in everyday life. Most of us have little idea how error-prone we are when trying to identify speakers we do not know.[118]

In terms of error, the trial judge's comments — 'mistakes *can* easily be made even when we're identifying … [a] friend or family member' (lines 40–1)[119] — do not adequately reflect the known risks. It is not just a matter of people sometimes making mistakes with unfamiliar voices or mistakes being possible. Rather, mistakes are

[114] *Davey* (n 7) [59], citing *Phan* (n 9) 152 [59] (Hinton J).

[115] Gordon E Legge, Carla Grossman and Christina M Pieper, 'Learning Unfamiliar Voices' (1984) 10(2) *Journal of Experimental Psychology* 298, 298; A Daniel Yarmey, 'The Psychology of Speaker Identification and Earwitness Memory' in R C L Lindsay et al (eds), *The Handbook of Eyewitness Psychology* (Psychology Press, 2010) vol 2, 101, 102, 116–18; Jody Kreiman and Diana Sidtis, 'Identifying Unfamiliar Voices in Forensic Contexts' in Jody Kreiman and Diana Sidtis (eds), *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception* (Blackwell, 2011) 237. On faces, see Ahmed M Megreya and Mike Burton, 'Unfamiliar Faces Are Not Faces: Evidence from a Matching Task' (2006) 34(4) *Memory and Cognition* 865.

[116] On faces, and facial comparison, see Gary Edmond et al, 'Facial Recognition and Image Comparison Evidence: Identification by Investigators, Familiars, Experts, Super-Recognisers and Algorithms' (2021) 45(1) *Melbourne University Law Review* 99.

[117] We note that Detective J's opinions were confident and categorical — without qualifications or caveats. Cf PCAST Report (n 70) 96, 26, 74. See also: Simon A Cole, 'Forensics Without Uniqueness, Conclusions Without Individualization: The New Epistemology of Forensic Identification' (2009) 8(3) *Law, Probability and Risk* 233; Jonathan J Koehler and Michael J Saks, 'Individualization Claims in Forensic Science: Still Unwarranted' (2010) 75(4) *Brooklyn Law Review* 1187.

[118] Ironically, it is insensitivity to the different cognitive tasks and the much higher levels of error associated with unfamiliar voice comparison that enables appellate courts to characterise all voice comparison as commonplace.

[119] *Davey* (n 7) [57] (emphasis added).

likely and ubiquitous.[120] Numerous studies confirm that in favourable conditions — with good quality recordings and plenty of time (and without other evidence or suggestion) — ordinary persons make mistakes frequently.[121] Depending on the conditions, error rates can be higher than 50%.[122] But it is not just the prevalence of error that creates problems. Those who purport to identify or recognise voices are surprisingly prone to confidently misattributing a voice to a particular speaker. These types of errors are particularly vulnerable to contextual bias.[123]

The judicial directions refer to the *potential* for familiars (for example, 'friend or family members')[124] to make mistakes and the task of comparison being 'much more difficult' (lines 41–2) for strangers.[125] It then moves to consider the opportunity Detective J had 'to hear the voice of the person' (line 43). The basic premise is correct, but the issue is whether Detective J is genuinely familiar with Davey, and more familiar than the jury.[126] The directions begin by noting that Detective J was 'not familiar with Matthew Davey's voice *prior to* Operation Oracle' (lines 6–7).[127] Implicitly, Detective J became familiar with Davey. According to the testimony and directions, Detective J identified Davey when he recognised the voice (from the telephone intercepts) during Davey's police interview. The issue of familiarity (and Detective J's ability relative to the jurors) is important because unless Detective J had an advantage over the jury and was capable of providing assistance to them, his opinions were, on the basis of the reasoning in *Smith v*

120 And, as we shall see, the other evidence in this case is not independent, but rather constitutive of Detective J's opinions, and so cannot be invoked as independent support or confirmation.

121 See: Harry Hollien, *Forensic Voice Identification* (Academic Press, 2002); Phil Rose, *Forensic Speaker Identification* (Taylor and Francis, 2002); Legge, Grossman and Pieper, 'Learning Unfamiliar Voices' (n 115); Yarmey, 'The Psychology of Speaker Identification and Earwitness Memory' (n 115) 102; Helen Fraser, 'The Reliability of Voice Recognition by "Ear Witnesses": An Overview of Research Findings' (2019) 6(2) *Language and Law* 1; Christopher Sherrin, 'Earwitness Evidence: The Reliability of Voice Identifications' (2016) 52(3) *Osgoode Hall Law Journal* 819.

122 See studies discussed in Gary Edmond, Kristy Martire and Mehera San Roque, 'Unsound Law: Issues with ("Expert") Voice Comparison Evidence' (2011) 35(1) *Melbourne University Law Review* 52, 87–9.

123 See Harriet Mary Jessica Smith and Thom Baguley, 'Unfamiliar Voice Identification: Effect of Post-Event Information on Accuracy and Voice Ratings' (2014) 5(1) *Journal of European Psychology Students* 59.

124 *Davey* (n 7) [57].

125 Ibid. See also *R v Bueti* (1997) 70 SASR 370, 381 (Doyle CJ), quoting directions to the jury by the trial judge: '"You may well think that mistakes in voice recognition do and can happen"'.

126 Allowing the suggestive jury comparison does not make much sense if Detective J was actually better than the jury. In such circumstances, their listening would hardly ever provide useful means for evaluation. Jurors are not encouraged to undertake fingerprint (and other feature) comparisons, so it is curious that they are encouraged to engage in unfamiliar voice comparisons.

127 *Davey* (n 7) [57] (emphasis added).

The Queen, irrelevant and therefore inadmissible.[128] The trial judge implies that Detective J became familiar with Davey's voice through the course of the investigation (lines 6–7) but such familiarity and its significance are left as matters for speculation. To ordinary persons, it might appear that Detective J had become very familiar with the relevant voices through his repeated listening. Moreover, the value of his opinion evidence appears to be supported by his role in the investigation and confirmed by his participation in the trial, along with the way his opinion was said to be formed and strategically aligned with the other evidence adduced against Davey (and Eaton) — see Part V(B) below. However, such descriptions are misguided and very likely to mislead.

Research reveals that listening to a limited set of recordings and repeated listening does not transform a listener into a familiar.[129] This kind of constrained exposure does not appear to significantly improve listener accuracy. In order to become a genuine familiar, it seems important to be exposed to a voice across a range of different settings and moods.[130] The dynamic ways we become familiar with the voices of friends and family members are quite different to the listening associated with intercepted voice recordings. Rather than passively listening to recordings of various quality and duration from a phone or listening device, our interactions with family and friends tend to be varied. We encounter them when they are happy, sad, angry, exuberant, excited, scared, and even intoxicated. We usually interact in person and across a variety of media (for example, via phone). Familiars' exposure usually extends across rather long periods of time where the identity of the interlocutor is rarely in doubt.[131] Research suggests that any limited advantage gained by investigators who are exposed to many hours of recorded speech are rapidly matched by others — such as jurors — when exposed to quite modest amounts of the same recorded materials.[132] Though counter-intuitive, it seems that the jurors would have acquired similar levels of accuracy to Detective J from listening to just an hour or so of the intercepted recordings. Detective J was not a genuine familiar and it is unlikely that he held a meaningful advantage over the jury in terms of his

---

128 (2001) 206 CLR 650, 655 [11] (Gleeson CJ, Gaudron, Gummow, and Hayne JJ), 669–70 [58]–[61] (Kirby J) ('*Smith*'). See also *Phan* (n 9) 152 [59] (Hinton J): 'It … follows that evidence of voice comparison led from a non-expert will be inadmissible unless the non-expert enjoys an advantage over the jury'.

129 Lori R van Wallendael et al, '"Earwitness" Voice Recognition: Factors Affecting Accuracy and Impact on Jurors' (1994) 8(7) *Applied Cognitive Psychology* 661.

130 Niels Schiller and Olaf Köster, 'The Ability of Expert Witnesses to Identify Voices: A Comparison Between Trained and Untrained Listeners' (1998) 5 *Forensic Linguistics* 1.

131 If the identity of Davey was not in doubt that was because his identity was already known or strongly implied.

132 A Daniel Yarmey, 'Earwitness Descriptions and Speaker Identification' (2001) 8(1) *International Journal of Speech, Language and the Law* 113; I Pollack, J M Pickett and W H Sumby, 'On the Identification of Speakers by Voice' (1954) 26(3) *Journal of the Acoustical Society of America* 403.

familiarity and accuracy. There are real doubts about the relevance as well as reliability of Detective J's opinions.

Not only are familiarity and its implications misunderstood and misrepresented — in the admissibility determinations, testimony and directions — the issue of Detective J's (uncertain) abilities and actual ignorance of voice comparison research, methods and limitations were not directly addressed.[133] While the jury was told that Detective J was not an expert in voice comparison, the significance and implications of his lack of specialised knowledge about voices, recordings and comparison methods were not addressed. Detective J does not seem to appreciate his own limitations or the magnitude of dangers.[134] When it comes to describing the voice, all he can say is that: '[i]t was difficult to hear and understand because it was mumbly, it was deep … it was quite unique' (lines 69–70). This does not, however, constrain his identification(s). Drawing upon the loaded idea of uniqueness — though without explaining why or providing any information about the frequency of mumbling or deep voices (which are not unique voice features) — Detective J was willing to categorically identify Davey as the speaker.[135] Reliance on equivocal voice features and assertions about uniqueness are misleading.[136]

There are also the issues of: (1) whether the various voices attributed to Davey on the intercepted recordings are the same person (lines 29–30); and (2) the very limited (though highly suggestive) exposure to Davey's actual voice said to have

[133] Section 78 of the *Evidence Act* (Vic) (n 36) is relied upon in *Kheir* (n 9) because the police officer cannot satisfy s 79, or the common law requirements in *Makita (Australia) Pty Ltd v Sprowles* (2001) 52 NSWLR 705, 743–4 [85] (Heydon JA) ('*Makita (Australia)*').

[134] The jury might have been told that Detective J was not an expert, but he was an employee of the state — a police officer — who was allowed to testify. Like the trial judge, he is not in a position to assist the jury with knowledge and likely to be misunderstood as better than average. See Kathy Pezdek and Daniel Reisberg, 'Psychological Myths about Evidence in the Legal System: How Should Researchers Respond?' (2022) 11(2) *Journal of Applied Research in Memory and Cognition* 143, 144–5.

[135] There is an assumption that all voices are unique, which cannot be tested but may be true (in a trivial way). Even if all speech were unique, it does not follow that all fragments of speech from telephone intercepts are discernibly different, or that an untrained and inexpert individual can reliably distinguish between them in order to identify the speaker. The real question is whether Detective J is capable of distinguishing between and identifying a specific voice on intercepted phone calls. Differences in voices make this possible but reveal nothing about Detective J's ability.

[136] The fact that voice features are difficult to explain should not encourage prosecutors and judges to rely on s 78 of the UEL. Not only are these opinions 'identifications' by those who did not directly perceive the matter or event, but recourse to s 78 rewards ignorance and renders the speculative and biased opinions of investigating police officers largely unaccountable.

confirmed Detective J's identification (lines 74–6).[137] Throughout the investigation, Detective J was exposed to the metadata, telephone numbers, locations, content of the calls (including names), and had access to the transcripts with names attributed to speakers when listening to the voices.[138] He knew the identity of the suspect when purporting to recognise the speaker on the tapes as Davey during Davey's police interview. Detective J claimed that linking the voices on the recordings and identifying those voices as Davey was based on voice comparison and recognition. This was not — despite its representation at trial and in the directions — a voice comparison (or recognition) exercise at all. Rather, Detective J was central to the investigation and was exposed to all of the circumstantial evidence which implicated Davey as the speaker. The impact of this circumstantial evidence is demonstrated by the fact that Detective J was confident about the identity of the speaker before he attended Davey's interview.[139]

The directions refer to the possibility that voices may sound similar, and that Detective J might have mistaken mere similarities for identity (lines 60, 62, 84). The problem with relying on similarities — real or apparent — was discussed by Barack Obama's Presidential Council of Advisers on Science and Technology ('PCAST') in relation to forensic scientists engaged in feature comparisons, such as voice identification:

> The frequency with which a particular pattern or set of features will be observed in different samples, which is an essential element in drawing conclusions, is not a matter of 'judgment.' It is an empirical matter for which only empirical evidence is relevant. Moreover, a forensic examiner's 'experience' from extensive casework is not informative — because the 'right answers' are not typically known in casework and thus examiners cannot accurately know how often they erroneously declare matches and cannot readily hone their accuracy by learning from their mistakes in the course of casework.[140]

This was the advice offered to experienced forensic scientists undertaking feature comparisons — DNA profiles, fingerprints, ballistics, shoeprints and so on. Detective J struggled to find words to describe the voice(s) said to be similar

---

137 *Davey* (n 7) [57]. Detective J's evidence was 'I couldn't be certain it was Matthew Davey until I spoke to him in person'.

138 Ibid. A few limitations with Detective J's opinion evidence (for example, access to names, transcript and metadata) may have been hinted at, but the problems and the level of threat posed, discussed in Part V(B), were not explained.

139 Cf the translator in *Tran v The Queen* [2016] VSCA 79, [34] who made this explicit. See Gary Edmond, 'Investigators, Cognitive Bias and Double-Dipping: Misunderstanding Opinion Evidence in Trials and Appeals' (2023) 97(8) *Australian Law Journal* 543.

140 PCAST Report (n 70) 19, 55. The emphasis is on the construction of a rigorous database and the need to test methods and abilities in conditions where the correct answer is known. Investigations provide neither.

— such as deep, and mumbly.[141] He would not have known whether the telephone or recording contributed to these voice characteristics or how common or interrelated such voice features might be in the community.[142] He was likely dependent on his limited casework experience and what he knew about the case against Davey.[143]

Detective J did precisely what experts are cautioned against. A handful of putative similarities were used to categorically identify Davey as the speaker. Detective J is unfamiliar with scientific research and advice. He is not conversant with appropriate methods for voice comparison, the chance of other voices sounding similar, his ability to distinguish similar voices, or the likelihood that he has or might have made an error.[144] He remained confident, indeed certain, about his identification throughout.[145] This is what happens when non-expert investigators are allowed to proffer opinions at trial. Such non-expert investigators are incapable of helping the jury to evaluate their opinion or method.[146] Compounding the problem, jurors are also very likely to over-estimate their abilities with voices and likely to attribute identity based on expectations (informed by Detective J and exposure to the other evidence) in combination with perceived similarities.[147]

We note that on appeal, Davey (in person) raised the fact that 'there was no testing' of the voice identification process and highlighted the revised English *Criminal Practice Directions* which require 'a sufficiently reliable scientific basis' for such evidence to be admitted in England and Wales.[148] Davey was right to raise these

---

141 *Davey* (n 7) [57].

142 For example, having a deep voice is often correlated with gender.

143 See generally David White et al, 'Passport Officers' Errors in Face Matching' (2014) 9(8) *PLOS One* 1. See also PCAST Report (n 70) 6.

144 Detective J has likely never been tested on his ability in conditions where the actual speaker was known. Empirical evidence suggests that at best, police are no better than ordinary persons in comparison and identification tasks. See: Annelies Vredeveldt and Peter J van Koppen, 'The Thin Blue Line-Up: Comparing Eyewitness Performance by Police and Civilians' (2016) 5(3) *Journal of Applied Research in Memory and Cognition* 252; Anna Lvovsky, 'The Judicial Presumption of Police Expertise' (2017) 130(8) *Harvard Law Review* 1995.

145 Cross-examination is unlikely to shake subjective claims, particularly if sincerely held. It is also unlikely to elicit vulnerabilities because they are not known to most investigators, and are likely to be resisted.

146 See: *HG v The Queen* (1999) 197 CLR 414, 429 [44] (Gleeson CJ); *Makita (Australia)* (n 133).

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

148 Davey (n 7) [73], citing Court of Appeal of England and Wales, *Criminal Practice Directions 2015* [2015] EWCA Crim 1567, 29 September 2015, [19A.4], citing *R v Dlugosz* [2013] EWCA Crim 2 [11]. See also Paul Roberts and Michael Stockdale (eds), *Forensic Science Evidence and Expert Witness Testimony: Reliability through Reform?* (Edward Elgar, 2018).

points, for the validity and reliability of the method is obviously integrally related to the value of derivative (here, Detective J's) opinion. It is the validity and reliability of a method, along with the proficiency of the analyst applying it (and the way subjective interpretation is protected from cognitive bias), that determines the capability of the evidence.[149] Apparently unconcerned by developments in England and Wales (as well as Canada and the United States), Estcourt JA explains that '[n]o such requirement exists in Australia'.[150] The TASCCA appears comfortable with the protections afforded by cross-examination and the fairness inducing qualities it attributes to the directions as given.[151]

B *Great Expectations: Cognitive Bias and Contextual Effects*

> A system of criminal law cannot be attuned to the needs of the people whom it exists to serve unless its administration, proceedings and judgments are comprehensible by both the accused and the general public *and have the appearance, as well as the substance, of being impartial and just*.[152]

There is an expectation, of a fundamental sort, that jurors will not be biased.[153] The 'administration of criminal justice', we are frequently reminded, should have 'the appearance of being, unbiased and detached'.[154] Legal concerns with bias have overwhelmingly focused on the interests and motivations of defendants and witnesses, and the risks to public confidence from judges appearing to lack impartiality.[155]

---

149 See *IMM* (n 42).

150 *Davey* (n 7) [73].

151 This article does not consider cross-examination or rebuttal expertise. We have considered those issues in other articles: Gary Edmond and Mehera San Roque, 'The Cool Crucible: Forensic Science and the Frailty of the Criminal Trial' (2012) 24(1) *Current Issues in Criminal Practice* 51; Gary Edmond et al, 'Forensic Science Evidence and the Limits of Cross-Examination' (2019) 42(3) *Melbourne University Law Review* 858; Gary Edmond, 'Forensic Science and the Myth of Adversarial Testing' (2020) 32(2) *Current Issues in Criminal Justice* 146. However, unless cross-examination is capable of introducing and meaningfully conveying the problems with voice identification it is unlikely to be effective. For the reasons we explain below, cross-examination and directions are incapable of placing a jury in a position where they can reliably or fairly compare voices or assess the biased and speculative opinions of investigators.

152 *Kingswell v The Queen* (1985) 159 CLR 264, 301 (Deane J) (emphasis added). See *QYFM v Minister for Immigration, Citizenship, Migrant Services and Multicultural Affairs* (2023) 66 Fam LR 369.

153 The underlying assumptions — not applied consistently — appear grounded in Lockean metaphysics; that knowledge derived from the senses is independent of bias and judgement. See generally John Locke, *An Essay Concerning Human Understanding* (Oxford University Press, 1689).

154 *Brown v The Queen* (1986) 160 CLR 171, 202 (Deane J).

155 Gary Edmond and Kristy A Martire, 'Just Cognition: Scientific Research on Bias and Some Implications for Legal Procedure and Decision-Making' (2019) 82(4) *Modern Law Review* 633, 633. See also Australian Law Reform Commission, *Without Fear or*

Trial and appellate courts have generally assumed that careful directions will alert jurors to the dangers of partiality so that in most cases they will not misuse evidence or succumb to prejudice and other forms of irrationality.[156] Remarkably, given that the evaluation of evidence is both a routine and core feature of trials (and pleas) and many appeals, our courts have devoted limited attention to the question of whether careful directions actually work. Do directions protect jurors from the biases created by evidence or procedures? In terms of the wide range of cognitive biases that are known to contaminate perception and cognition — sometimes irreparably — there is little evidence in reported jurisprudence of courts recognising dangers or responding in ways that are informed by decades of scientific research or the considered advice of attentive scientists.

Cognitive biases and contextual effects are ubiquitous in human decision-making. When we speak of cognitive biases, we are referring to the systematic ways in which the environment or a person's perspectives (for example, expectations or world view) influences how they perceive and reason.[157] They are a class of effects that automatically influence — sometimes distorting or misleading — judgment. Despite our best efforts they are never fully under conscious control.[158] In this regard, they resemble other reflexive actions, such as sneezing and most movements of our eyelids. Cognitive biases and contextual effects are mental shortcuts or decision-making preferences that help us to process the very large amounts of information produced through our senses to form judgments and make decisions quickly and easily, even automatically.[159] They are generally useful, although in some circumstances — such as where perception or evaluation are unusual or difficult — their effects can become undesirable. They may generate irrational responses to evidence and other stimuli.

Favour: Judicial Impartiality and the Law on Bias (Final Report No 138, December 2021) 21–102.

156 In *Zoneff* (n 14) 261 [67], in an unusually frank discussion of the effectiveness of directions, Kirby J indicated that '[t]he law presumes that triers of fact are able to disregard the prejudicial aspects of testimony and adjust appropriately the weight to be attached to such evidence on the basis of its "probative value"'. Juries gauge and determine the weight of evidence. In Australia, following *IMM* (n 42), probative value is determined by the trial judge taking the capacity of the evidence 'at its highest': at 313 [44], 314 [47] (French CJ, Kiefel, Bell and Keane JJ). Probative value, then, is the maximum value or capacity that can rationally be assigned to some evidence.

157 For an introduction to these and other issues, see generally: Daniel Kahneman, *Thinking, Fast and Slow* (Penguin, 2012); Gerd Gigerenzer, *Simply Rational: Decision Making in the Real World* (Oxford University Press, 2015); Michael J Saks and Barbara A Spellman, *The Psychological Foundations of Evidence Law* (New York University Press, 2016); Dan Simon, *In Doubt: The Psychology of the Criminal Justice Process* (Harvard University Press, 2012); Tom R Tyler, *Advanced Introduction to Law and Psychology* (Edward Elgar, 2022).

158 See generally Jessica Nordell, *The End of Bias: A Beginning* (MacMillan, 2021).

159 Kahneman (n 157); Emma Cunliffe, 'Judging, Fast and Slow: Using Decision-Making Theory to Explore Judicial Fact Determination' (2014) 18(2) *International Journal of Evidence and Proof* 139, 144.

Biased interpretations, responses and attitudes are often unintentional. Individuals striving to act with impartiality and integrity are vulnerable to biases. One conspicuous bias concerns expectancy effects. These arise when our expectations influence perception and interpretation.[160] They may occur in an interpersonal context. If we are told that someone we are going to interact with is angry or mean-spirited, we tend to interpret behaviour, though especially ambiguous behaviours, in ways that are consistent with our expectations.[161] Expectations have been repeatedly demonstrated to inform the way we understand and interpret evidence (or information). Related to expectancy effects is confirmation bias.[162] This is a preference for information that supports rather than contradicts our existing beliefs.[163] Beyond colouring judgment based on our expectations, however, confirmation effects can influence the way we search for information in our environment, how we weigh what we find, and our ability to generate and appropriately evaluate alternative explanations.[164] Confirmation bias has been demonstrated in a variety of contexts, including maintaining beliefs in illusory correlations (for example, that certain weather patterns cause ailments to act up), and belief persistence in the face of strong contradictory evidence.[165] Tunnel vision appears to be a combination of expectation and confirmation biases.[166]

Humans can be biased quickly and often outside their conscious awareness. Researchers refer to the lack of awareness of being biased as a bias blind spot.[167] We

160 D Michael Risinger et al, 'The *Daubert/Kumho* Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion' (2002) 90(1) *California Law Review* 1, 6.

161 E Tory Higgins, William S Rholes and Carl R Jones, 'Category Accessibility and Impression Formation' (1977) 13(2) *Journal of Experimental Social Psychology* 141; Stéphane Doyen et al, 'On the Other Side of the Mirror: Priming in Cognitive and Social Psychology' (2014) 32 (Supplement) *Social Cognition* 12, 16–17. See also Gary Edmond et al, 'Contextual Bias and Cross-Contamination in the Forensic Sciences: The Corrosive Implications for Investigations, Plea Bargains, Trials and Appeals' (2015) 14(1) *Law, Probability and Risk* 1, 4 ('Contextual Bias and Cross-Contamination in the Forensic Sciences').

162 See generally Raymond S Nickerson, 'Confirmation Bias: A Ubiquitous Phenomenon in Many Guises' (1998) 2(2) *Review of General Psychology* 175.

163 Ibid 175.

164 For example, pre-existing attitudes are a robust predictor of how people will search for information on the internet: Dáša Vedejová and Vladimíra Čavojová, 'Confirmation Bias in Information Search, Interpretation, and Memory Recall: Evidence from Reasoning about Four Controversial Topics' (2022) 28(1) *Thinking and Reasoning* 1, 14. See also Joshua Klayman, 'Varieties of Confirmation Bias' (1995) 32(1) *Psychology of Learning and Motivation* 385.

165 Nickerson (n 162) 183.

166 See Keith A Findley and Michael S Scott, 'Multiple Dimensions of Tunnel Vision in Criminal Cases' [2006] (2) *Wisconsin Law Review* 291, 292.

167 Emily Pronin, Daniel Y Lin and Lee Ross, 'The Bias Blind Spot: Perceptions of Bias in Self versus Others' (2002) 28(3) *Personality and Social Psychology Bulletin* 369, 370.

routinely underestimate our vulnerability and the extent to which our thinking is actually biased. This extends to being able to generate sincere, though mistaken and misleading, explanations for (biased) preferences and interpretations.[168] We also tend to believe that other people (including professional peers) are more vulnerable to bias and more likely to be biased.[169] The bias blind spot produces overconfidence, a finding that has been demonstrated across a variety of professions, from architecture to medicine.[170]

Context effects occur when environmental factors influence the way information is perceived and processed.[171] They are most pronounced with unusual or difficult tasks, such as where the information being considered is ambiguous, unclear or vague.[172] In these circumstances decision-makers automatically draw upon other information accessible in the context, or the context itself, to make sense of evidence and attribute meaning. Contextual information is often very useful and can improve decision-making. However, it can also distort and mislead. Contextual effects are particularly concerning where irrelevant, misleading, or unreliable information is accessible to decision-makers, because decision-makers will often unwittingly draw on that information and be affected by it. Exposure to contextual information can lead to overvaluing — or actually, double counting — evidence, especially where the trier of fact is required to consider information that was also available to investigators and experts.[173]

In recent decades, mainstream scientists have advised forensic scientists to manage their exposure to domain irrelevant information.[174] This is information (and context)

---

168 Ibid 374–6.

169 Ibid 376–7.

170 Dunning, Heath and Suls (n 147) 72–3. In one study of forensic scientists, 71% reported that cognitive bias was a concern for forensic science, but only 26% reported that their own judgments were affected by cognitive bias: Jeff Kukucka et al, 'Cognitive Bias and Blindness: A Global Survey of Forensic Science Examiners' (2017) 6(4) *Journal of Applied Research in Memory and Cognition* 452, 454. Many of these respondents had some background knowledge of cognitive bias, through training, yet remained overconfident about their personal ability to control effects that are largely uncontrollable.

171 MJ Saks et al, 'Context Effects in Forensic Science: A Review and Application of the Science of Science to Crime Laboratory Practice in the United States' (2003) 43(2) *Science and Justice* 77, 78.

172 Ibid.

173 Ibid 84.

174 National Commission on Forensic Science, *Ensuring that Forensic Analysis is Based upon Task-Relevant Information* (Report, 2015) 1. Another option is to gradually expose an analyst to contextual information, starting with only the information required to do the specific comparative task. This procedure has been endorsed by many scientists concerned with bias in forensic science: see Adele Quigley-McBride et al, 'A Practical Tool for Information Management in Forensic Decisions: Using Linear Sequential Unmasking-Expanded (LSU-E) in Casework' (2022) 4(1) *Forensic Science International: Synergy* 1.

that is not required to perform a specified task. Those involved in comparing fingerprints, cartridges, shoe and tyre marks, DNA profiles, blood spatter, faces and so on, should, for example, only be provided with the information required to perform the comparison. When undertaking comparisons or reviewing comparisons, the context should not suggest a (preferred or expected) result. Numerous studies have found that forensic scientists are vulnerable to domain irrelevant information.[175] For example, when applying risk assessment algorithms to those convicted of crimes in order to predict their future dangerousness, the conclusions of forensic psychologists are dependent on the party retaining them.[176] Forensic pathologists evaluating human remains are much more likely to interpret the same skull damage as pre-mortem if they are told they are examining skulls from a conflict site, and post-mortem if they are told the remains came from an old cemetery.[177] Providing extraneous information to fingerprint examiners (for example, that the suspect confessed, emotional details of the case, or that other examiners concluded the prints did not match) has been shown to reverse decisions as to whether prints match.[178]

Medical researchers, as just one conspicuous example, employ double blind randomised clinical trials to avoid the corrosive effects of expectation, suggestion and confirmation.[179] Highly trained and experienced biomedical consultants, including those who are conversant with the dangers, are incapable of resisting these invidious influences. Blinding from biasing information is the only effective way to manage the dangers.[180]

Scientific research demonstrates how easily human perception and cognition can be inadvertently biased.[181] Cognitive biases affect not only the general population, but also actors in the legal system. Naturally, this includes witnesses, police officers (and other investigators) and jurors, but it also extends to lawyers and judges as well

---

175 Glinda S Cooper and Vanessa Meterko, 'Cognitive Bias Research in Forensic Science: A Systematic Review' (2019) 297(1) *Forensic Science International* 35, 37–43.

176 Daniel C Murrie et al, 'Are Forensic Experts Biased by the Side That Retained Them?' (2013) 24(10) *Psychological Science* 1889, 1893.

177 Sherry Nakhaeizadeh, Ian Hanson and Nathalie Dozzi, 'The Power of Contextual Effects in Forensic Anthropology: A Study of Biasability in the Visual Interpretations of Trauma Analysis on Skeletal Remains' (2014) 59(5) *Journal of Forensic Sciences* 1177.

178 Cooper and Meterko (n 175) 43; Itiel E Dror, David Charlton and Ailsa E Péron, 'Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications' (2006) 156(1) *Forensic Science International* 74.

179 J Bruce Moseley et al, 'A Controlled Trial of Arthroscopic Surgery for Osteoarthritis of the Knee' (2002) 347(2) *New England Journal of Medicine* 81, 81–2.

180 Well-designed evidence line-ups offer a means to circumvent contextual influences.

181 See, eg, Marcus R Munafò et al, 'A Manifesto for Reproducible Science' (2017) 1(1) *Nature Human Behaviour* 1, 1. Cognitive biases are a major reason for false positives in the scientific literature (and are a reason why reforms to scientific process focus on making science more transparent such that biases are easier to identify and control).

as forensic scientists applying validated methods.[182] In drawing upon the scientific research, we do not mean to imply that humans are incapable of approaching evidence and decision-making critically. Rather, research confirms that the context in which evidence is presented, how evidence is perceived and interpreted, along with our preconceptions (of the evidence or task), can all change the way evidence is understood and weighed.[183] Most prosecutions proceed as though the evaluation of evidence is mechanical, and cognitive biases are peripheral or can be managed through cross-examination, judicial directions, and conscious effort. Such attitudes are inconsistent with decades of scientific research, which demonstrates that even professionals acting with integrity frequently and unwittingly make errors caused by biases.

The directions in *Davey* do not provide insight into, let alone assistance with, context effects and cognitive bias.

C *Context, Contamination and Overvaluing the Circumstantial Evidence*

Starting from the position that risks from cognitive biases are pervasive, the unstructured, impressionistic and suggestive character of Detective J's interpretations, performed as part of an investigation, render them especially susceptible. PCAST reviewed the literature on cognitive bias in forensic sciences and warned about judgments akin to Detective J's: '[s]ubjective methods require particularly careful scrutiny because their heavy reliance on human judgment means they are especially vulnerable to human error, inconsistency across examiners, and cognitive bias'.[184] Cognitive bias looms large in investigative environments. When the source material is unclear (or difficult or ambiguous) and there is no particular methodology for analysing materials, such as the voice recordings, those endeavouring to interpret them will — regardless of any conscious awareness — tend to rely on unintentional

[182] Andrew J Wistrich, Jeffrey J Rachlinski and Chris Guthrie, 'Heart versus Head: Do Judges Follow the Law or Follow Their Feelings' (2014) 93(4) *Texas Law Review* 855. See also: Chris Guthrie, Jeffrey J Rachlinski and Andrew J Wistrich, 'Inside the Judicial Mind' (2001) 86(4) *Cornell Law Review* 777; Birte Englich, Thomas Mussweiler and Fritz Strack, 'Playing Dice with Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making' (2006) 32(2) *Personality and Social Psychology Bulletin* 188. In the experiments, judges made higher damages awards when they rolled a higher number on a die than when they rolled a lower number (ie, they are susceptible to the 'anchoring effect'): 192–5. See also François Rabelais, *Gargantua and Pantagruel, Complete: Five Books of the Lives, Heroic Deeds and Savings of Gargantua and His Son Pantagruel*, tr Sir Thomas Urquhart (Gutenberg, 2004) bk 3, ch 39.

[183] Edmond et al, 'Contextual Bias and Cross-Contamination in the Forensic Sciences' (n 161); Amos Tversky and Daniel Kahneman, 'The Framing of Decisions and the Psychology of Choice' (1981) 211(4481) *Science* 453; Nickerson (n 162).

[184] PCAST Report (n 70) 5.

shortcuts to make sense of it.[185] Detective J's voice identification embodied these problems.[186]

Detective J did not use a reliable method for his voice comparison (and recognition). Apart from the inferences available from the research surveyed in Part V(A), even after cross-examination and directions, his actual ability is uncertain. At no point was Detective J's ability to identify or recognise Davey's voice (or any other voices) tested. We therefore have no reliable insight into Detective J's ability to make accurate voice identifications. Moreover, there were no safeguards in place to protect Detective J's listening and recognition from domain irrelevant information, or the suggestive context of the interview.

Detective J testified that as part of the investigation into the burglary and theft of guns (where many forms of information and other evidence were available to the investigating police officers), he listened to a large number of intercepted calls. The total is stated to be as high as 1200 calls, though the number alleged to involve the person or persons said to be Davey is considerably lower.[187] There were, according to Detective J's evidence, '30 to 50 calls where Matthew, Matty or Matthew Davey were used' during the call (lines 13–14), and an unspecified number of other calls where identity was suggested by the address, locations, subscriber details, other metadata or the subject matter (lines 15–16, 56–7). Detective J testified that he used the calls where the speaker was identified by one of the names shared by the defendant to attribute all the calls featuring a similar sounding speaker to the same source — implicitly, Davey (lines 35, 38).[188] He then (is said to have) confirmed the identity of the speaker as Davey when he 'recognised' (lines 29, 72) the voice during Davey's police interview (lines 74–5). The formation of Detective J's opinions was actually more complicated than this 'just so story'.[189]

---

[185] Amos Tversky and Daniel Kahneman, 'Judgment under Uncertainty: Heuristics and Biases' (1974) 185(4157) *Science* 1124, 1130.

[186] Interestingly, listeners will 'hear' voices and speech in white noise — where there is no speech — when presented with suggestive contextual primes: Michael A Nees and Charlotte Phillips, 'Auditory Pareidolia: Effects of Contextual Priming on Perceptions of Purportedly Paranormal and Ambiguous Auditory Stimuli' (2015) 29(1) *Applied Cognitive Psychology* 129, 130; Paul Lunn and Andy Hunt, 'Phantom Signals: Erroneous Perception Observed During the Audification of Radio Astronomy Data' (Conference Paper, International Conference on Auditory Displays, 6–10 July 2013) 250. For a legal example, see *R v Bain* [2010] 1 NZLR 1, 7 [5] (Elias CJ for Elias CJ and Blanchard J), discussed in Helen Fraser, Bruce Stevenson and Tony Marks, 'Interpretation of a Crisis Call: Persistence of a Primed Perception of a Disputed Utterance' (2011) 18(2) *International Journal of Speech, Language and the Law* 261.

[187] *Davey* (n 7) [5].

[188] We are not told if all of these calls were made using the same phone (number), and whether that played a factor in their selection and the 'identification'.

[189] Derived from Rudyard Kipling's *Just So Stories* (1902), these tend to be narratives featuring frequently untestable and not particularly helpful explanations. See generally Rudyard Kipling, *Just So Stories* (Gutenberg, 2008).

When he sat in the interview room with Davey (and when he listened to the intercepted recordings during the investigation), Detective J was conversant with all of the materials assembled during Operation Oracle.[190] There was a strong circumstantial case against Davey. Davey had, after all, been named on the recordings and arrested. Detective J expected to recognise Davey as the speaker identified on many of the tapes as 'Matthew', 'Matty' or 'Matthew Davey'. Indeed, this seems to have been one of the primary motivations for Detective J attending the interview.[191] The attribution of the recorded voice to Davey is not, in any simple or primary sense, based on Detective J's listening or the development of genuine familiarity leading to recognition. Rather, Detective J's recognition was based on an inextricable combination of non-voice evidence, background information, suggestion, expectation, supposition (or inferences from the circumstantial case), along with his impression of voice similarities at the interview. The manner in which calls were linked, and Davey identified during the investigation and the police interview, render Detective J's purported voice identification evidence practically worthless.[192] Detective J and the other police officers believed the 'speaker' on the intercepted calls was Davey before any comparison or recognition took place.[193]

The comparison from the police interview — presented as some kind of recognition or confirmation of Davey's identification — has little value. This was not an exercise in identification, but a reaffirmation of what Detective J and his colleagues already believed based on the circumstantial evidence they had assembled. Detective J's opinion evidence was based on all of the information — the circumstantial case as well as other background material — to which he had been exposed.[194] This was not

[190] As part of the investigation, Detective J knew the identity of the suspects (for example, from DNA evidence, recovered weapons, associates, and the statements of potential witnesses). He also knew the identities of those being monitored by lawful telephone intercepts, and was provided with metadata, transcripts and location information associated with the very voices he was purporting to compare, associate and identify. The circumstances of the case meant that Detective J was expecting to hear Davey. He knew that Davey was implicated in the offence on the basis of a range of different types of evidence. He knew, even before listening, that Davey was one of the suspects and he knew from the calls that someone called 'Matthew', 'Matty' or 'Matthew Davey', was speaking on a significant number of calls. There were other calls where Davey's names were not used, but Detective J had access to the telephone numbers, metadata, transcripts as well as the content of those calls before he began to listen and recognise.

[191] See generally Ziva Kunda, 'The Case for Motivated Reasoning' (1990) 108(3) *Psychological Bulletin* 480.

[192] If Detective J's opinions on identity were probative (and so relevant), the probative value at its highest was objectively low.

[193] Speaker is in inverted commas because it is uncertain as to whether the same person is speaking across these recordings.

[194] The so-called confirmation through the police interview adds little, if anything, to Detective J's globalised impressions. The trial judge warned the jury 'it would seem that there's a risk that somehow he's ultimately globalised, if you like, his opinion': *Davey* (n 7) [57].

a straightforward case of listening, leading to recognition and identification (lines 21, 74). Detective J's opinions about identity could not be retrieved, or extricated, from what he already knew and the context in which he came to listen. The possibility that Detective J might not have been influenced by the investigation or the other evidence, that he was sincere in his identification(s), or that he might have tried to rise above potential risks, does not matter. Humans are incapable of reliably thinking their way around such powerful contexts and expectations.[195] The conditions in which the listening took place practically guaranteed that Davey would be identified as the speaker already named as 'Matthew', 'Matty' or 'Matthew Davey'.[196]

The prosecutor, trial judge and TASCCA presented Detective J as a police officer who identified Davey on the basis of repeated listening, leading to familiarity with the unique features of his voice.[197] These misunderstandings (and misrepresentations) have serious implications for the admissibility of Detective J's opinions. Relevant exceptions to the exclusionary opinion rule require that opinions must be 'substantially based on [specialised] knowledge' or 'based on what the person saw, heard or otherwise perceived about a matter or event'.[198] Section 79 of the *Evidence Act 2001* (Tas), regarding specialised knowledge, can be quickly dispensed with. Relying on Detective J's experience with the voice recordings is not a solution, because he does not have experience with Davey's voice, unless we assume that it is Davey speaking — and that is what Detective J's opinion is being used to prove. Additionally, s 79 does not provide an admissibility pathway for opinions based on experience. Rather, the opinion must be substantially based on specialised knowledge. Importantly, Detective J has no knowledge of: methods for voice comparison; the difficulty of voice comparison; methods for blinding; the value of documentation; terms for classifying and describing voice features; the frequency of voice features; and statistical models for expressing the strength of opinions. Detective J is not an expert in voice comparison or even Davey's voice, and he is incapable of placing the jury in a position to evaluate his lack of expertise and biased impressions.[199] Section 79 does not provide an admissibility pathway for opinions that are based on biased experience.[200]

---

[195] Kukucka et al (n 170) 452, 456.

[196] *Davey* (n 7) [57].

[197] Cf *R v Harris [No 3]* [1990] VR 310, 322: Justice Ormiston described the listener as having 'engaged in a combination of identification and logic in a way which now cannot be satisfactorily unravelled'.

[198] *Evidence Act* (Tas) (n 37) ss 76(1), 78(a), 79(1). These sections all form part of the UEL.

[199] *Davie v Magistrates of Edinburgh* [1953] SC 34, 39–40 (and *Makita (Australia)* (n 133)) requires the witness to be able to place the fact-finder in a position where they are able to assess the opinion.

[200] There is also the complication that the only experience Detective J has with what is certainly Davey's voice is during the police interview. On ad hoc expert opinion evidence, see Gary Edmond and Mehera San Roque, 'Quasi-Justice: Ad Hoc Expertise and Identification Evidence' (2009) 33(1) *Criminal Law Journal* 8.

As for s 78, there are two problems. The first is the High Court's indication that s 78 is restricted to direct witnesses — this is relevant when considering that Detective J was an indirect witness here (ie, he listened to the recordings, and was not a witness to the relevant events themselves).[201] Intermediate courts of appeal, however, have overlooked this expectation and allowed investigators to opine. The second problem is the basis of Detective J's opinion evidence (also an issue for s 79). Section 78 requires the opinion to be 'based on what the person saw, heard or otherwise perceived about a matter or event' (and s 79 requires the opinion to be substantially based on 'specialised knowledge').[202] Unfortunately, Detective J's opinions about identity are not in any simple sense based upon listening to and comparing the voices.[203] Rather, as we have seen, his opinions about identity are based on his participation in the investigation and knowledge of the suspects, relationships between suspects, the phones being intercepted, the content of messages, including the names of participants, the labelling of speakers on the transcripts, the other evidence in the case (for example, DNA) and so on. To contend that the identification is based on the voice recognition at the police interview is to misconceive the foundation of the opinion evidence and Detective J's vulnerability to what he already knew. It also undermines the second requirement of s 78, that reception of the opinion must be 'necessary to obtain an adequate account or understanding of the person's perception'.[204] Once we recognise that the voice identification was based on Detective J's participation in the investigation, and his exposure to and synthesis of the circumstantial case, we can appreciate that ss 78 and 79(1) do not provide admissibility pathways.[205]

Simultaneously, understanding that Detective J's evidence is inadmissible as opinion evidence exposes fundamental, indeed constitutional, threats to fact-finding. If Detective J's opinion about the speaker is actually based on (or even nontrivially informed by) his exposure to the circumstantial case, then he was trespassing on

201 See: *Lithgow City Council v Jackson* (2011) 244 CLR 352, 370–1 [45]–[46] (French CJ, Heydon and Bell JJ) ('*Lithgow City Council*'); *Smith* (n 128) 669–70 [59]–[60] (Kirby J). See also Gary Edmond, 'Regulating Forensic Science and Medicine Evidence at Trial: It's Time for a Wall, a Gate and Some Gatekeeping' (2020) 94(6) *Australian Law Journal* 427.

202 *Evidence Act* (Tas) (n 36) ss 78, 79. The apparent rejection of the common law basis rule, in *Dasreef Pty Limited v Hawchar* (2011) 243 CLR 588, 604 [37] (French CJ, Gummow, Hayne, Crennan, Kiefel and Bell JJ), does not remove the obligations imposed by the text of ss 78 and 79.

203 On the application of rules following decisions about the meaning of facts, see, eg, Karl N Llewellyn, *The Bramble Bush: The Classic Lectures on the Law and Law School* (Oxford University Press, 2008) 65.

204 *Evidence Act* (Tas) (n 36) s 78(b).

205 We cannot retrospectively claim that it was reducible to the sounds he heard. We have no credible basis for drawing such inferences, especially where they are incompatible with mainstream scientific research and advice.

the prerogative of the jury.[206] The jury was presented with much (though perhaps not quite all) of the information available to Detective J and the police officers.[207] Indeed, this circumstantial prosecution required the jury to consider all of the evidence in order to determine whether Davey was involved. Already, you might have spotted the problem. Because of his exposure to the circumstantial case from his participation in the investigation, Detective J's opinion evidence was both anticipating and reproducing what the jury was expected to do. Detective J improperly trespassed into the realm of jury fact-finding. His opinions, based on a synthesis of the circumstantial case, was redundant. It is the constitutional responsibility of jurors, not investigators, to evaluate and combine the evidence at trial.[208]

The admission of Detective J's opinion evidence thus made it very likely that the jurors would use some of the evidence, likely unwittingly, more than once.[209] Taking a step back, if Detective J's opinions about identity were informed by more than the listening — as it most certainly appears to have been (and the contrary position cannot be credibly established) — and if the jury are allowed to hear and use his opinions as well as the other circumstantial evidence, then there is a serious (and unmanageable) risk that the other evidence will be overvalued (that is, used more than once). This occurs because the context and non-voice evidence contributed to the formation of Detective J's opinions about the speaker being Davey, whether he knew it or not. If the jury relied on Detective J's opinions as well as the evidence which informed (really, contaminated) his opinions, then that evidence is being relied upon, or counted, more than once.[210] Domain irrelevant information was not only available to the jurors, but it was also presented as independent support for Davey's identity and guilt. However, these strands of circumstantial evidence were not independent. They were cross-contaminated and could not be decontaminated. This contamination, as we shall see, has further serious repercussions for the jurors' comparison of the voices and, because Detective J's evidence was incorrectly characterised as independent, there were no directions about the danger of re-using or overvaluing the evidence.[211]

[206] It is the responsibility of the jury to evaluate the evidence. Further, this raises a threshold admissibility question in terms of relevance: see *Smith* (n 128) 655–6 [11]–[12] (Gleeson CJ, Gaudron, Gummow and Hayne JJ).

[207] Police were conversant with the criminal records of those they were investigating. The exposure to inadmissible evidence only compounds the problems, created by the inability to manage the influence of domain-irrelevant information.

[208] There may be occasions when some species of expert (eg, forensic pathologists) might rely on several pieces of evidence in forming an opinion. These witnesses usually possess genuine expertise and are capable of adverting to and explaining their reliance. Such reliance should also be limited to what is technically required.

[209] A piece (or strand) of evidence might be used for more than one purpose in legal proceedings — eg, for tendency and credibility purposes. Where, however, a piece of evidence is used for the same basic purpose over and over — here, by an investigator and jurors to determine the identity of a speaker — the evidence is being misused.

[210] See Edmond, 'Opinion, Bias and Double-Dipping' (n 139).

[211] Such directions would not work; but the dangers do not appear to have been recognised.

These are some of the reasons why investigators should not be invited or allowed to proffer their opinions in criminal proceedings.[212] They are not genuine experts. They are not in a position to provide reliable opinions about identity because their opinions are always irreparably contaminated by exposure to the evidence (and other information). They are not in a position to provide independent or unbiased opinions on identity. Further, Detective J's opinions were not and could not be 'shored up' (line 63) by the other evidence.[213] The same evidence cannot both inform an interpretation and confirm it — that is circular reasoning.[214] Here, the corrosive interactions of context, suggestion, expectation and confirmation cannot be meaningfully addressed or disaggregated. They are irretrievably intertwined in Detective J's irreparably contaminated opinions. It is not possible to extricate Detective J's opinions from his participation in the investigation and his exposure to the circumstantial evidence (and more).

The prosecution case, the judicial directions and appellate decisions, all present (or endorse) Detective J's opinions as identification by voice recognition or comparison. The appeal, after all, is largely concerned with the admission of Detective J's opinion evidence and the adequacy of related directions. In the legally authorised version(s) of provenance, none of the participants appear to be influenced by context or domain irrelevant information, or even particularly confident about the identity of the speaker, until Detective J 'spoke to [Davey] in person'.[215] Indeed, there are repeated references to the identification being based on, and resolved by, the in person meeting (lines 17–21, 32–4, 74–6). The version presented by the prosecutor, repeated in the directions, and accepted (as available) on appeal, is a sanitised reconstruction (or pro-prosecution rationalisation) of events that operates to inoculate against the serious threats posed by context and cognitive biases.[216] This version of events is available, and sustained, because lawyers and judges did not engage with scientific research on voice comparison and do not understand cognitive biases and their insidious effects. Ignorance enables them to accept an undocumented police version of events, even though it would seem, on the basis of studies of voice

---

212 Here we are speaking about detectives, police and translators, rather than those who are trained, certified and demonstrably proficient experts, with the potential to manage their exposure to domain irrelevant information.

213 *Davey* (n 7) [57].

214 See Eric-Jan Wagenmakers et al, 'An Agenda for Purely Confirmatory Research' (2012) 7(6) *Perspectives on Psychological Science* 632. Formal statistical hypothesis testing is invalidated by this very mistake: 'This also means that the interpretation of common statistical tests in terms of [false positives and false negatives] is valid only if the data were used only once and if the statistical test was not chosen on the basis of suggestive patterns in the data. If you carry out a hypothesis test on the very data that inspired that test in the first place then the statistics are invalid (or "wonky", as Ben Goldacre put it)': at 633.

215 *Davey* (n 7) [57].

216 The prosecutor, police and judge recognise there are issues here, but they do not confront them, and their various representations — whether unintentionally or deceptively — downplay serious risks.

comparison and cognitive bias, to be untenable.[217] It bears noting, that if judges are making these oversights (and mistakes) — not recognising the dangers posed by context and cognitive biases and their implications, such as double-dipping — then we can only assume that jurors are responding in similar ways.

We can contrast the insights and evidence-based observations developed in this article, with specific directions given to the jury:

57 that Detective [J] … had the summaries of metadata from TIS, Telephone Intercept Services,
58 he's expecting the call to be the voice of Matthew Davey, the risk where in terms of reliability is the
59 expectation has influenced his identification of the speaker. The risk — well, the risk is that there
60 are voices that are similar, but they're not identical in all their characteristics and yet with an
61 expectation of the speaker being a certain person, the opinion is filled in with that expectation if you
62 like, and so the end opinion is, it is Matthew Davey rather than it sounds like Matthew Davey so, if
63 you like, the person's assessment of the voice is shored up by the information they have from TIS,
64 another fact that you've got to bear in mind.

Although referring to expectations, these directions are not informed by scientific knowledge and do not convey the magnitude of risks. Here we can observe how risks, notorious among attentive scientists, are tentatively raised though ultimately left to the jury. The issue in the extract concerns the strength of Detective J's opinion evidence — positive identification rather than similarity (ie, sounds like). The trial judge's advice is not in consonance with the actual dangers because we regularly mistake voices, and suggestion or expectations can change our perception of whether unfamiliar voices are experienced as similar.[218] Reference to the possibility of Detective J's opinion being 'shored up'[219] by the very evidence that contaminated it, as an issue for the jury to 'bear in mind',[220] reinforces our argument that judges do not appreciate the magnitude of risk created by this type of evidence. Contamination is trivialised — something for the jury to bear in mind — when they come to evaluate Detective J's opinions in conjunction with the very evidence that contaminated them.[221]

217 Again, Detective J might believe it and his testimony may be sincere, but that is part of the problem.

218 Consider the classic card recognition experiment described in Jerome S Bruner and Leo Postman, 'On the Perception of Incongruity: A Paradigm' (1949) 18(2) *Journal of Personality* 206, 209–22. See also Thomas S Kuhn, *The Structure of Scientific Revolutions* (University of Chicago Press, 2nd ed, 1970) vol 2, 62–5.

219 *Davey* (n 7) [57].

220 Ibid.

221 Similarly, the QLRC expressed concern that some directions may lead jurors to 'over-compensate' for bias in arriving at a verdict: QLRC, *A Review of Jury Directions* (n 2) vol 1, 22 [2.21].

VI Evaluating Detective J's Opinion Evidence, the Jury Comparison and Unfair Prejudice

Courts seem to assume that describing what an investigator (or 'expert') did or the procedure(s) used, along with observations on issues to consider or 'bear in mind' (derived from collective judicial experience), provides sufficient assistance to enable a decision-maker to evaluate derivative opinions.[222] This is surely mistaken, especially in relation to feature comparison evidence, such as identification from sound recordings (or the identity of persons of interest in images). Although well-intentioned, the directions in *Davey* were likely to imbue jurors with false confidence — in the possibility of rational fact-finding around the identity of the speaker following the admission of Detective J's contaminated synthetic impressions.[223] This is remarkable, because directions and warnings are the main tools available to trial judges to regulate the evaluation of admissible evidence.[224]

In order to evaluate Detective J's opinions, we need to know quite a bit more than what was available to the jurors. This is a justification for excluding Detective J's speculative opinions, because directions and warnings cannot overcome the absence of information, or place jurors in a position to rationally evaluate them.[225] We need information about, for example: the validity of the methods used; the accuracy of voice comparison (including his own ability); familiarity and whether Detective J had actually become a familiar (discussed above); the quality and quantity of recordings; the frequency of voice attributes among suspect populations; the context(s) in which the comparisons were made; the information available to Detective J when he undertook his listening and identification; and the magnitude of dangers created by cognitive bias.

This kind of information is produced through formal testing and engagement with mainstream scientists and their research.[226] It was not provided by the parties or the Court and so was not available to any of the decision-makers.[227] Directions (and lay testimony) cannot make up for these omissions. The directions did not enable

---

222 The knowledge and experience of the judiciary might be considered mixed. Section 165A of the *Evidence Act* (n 2), for example, documents and proscribes some of what was once presented by judges as collective insight.

223 Trial and appellate courts will tend to assume the directions are sufficient, understood, and followed. Such assumptions, to the extent that they are misguided, appear to threaten the constitutional legitimacy of trials.

224 *Davey* (n 7) [58], [74].

225 See: Gary Edmond, 'Forensic Science Evidence and the Conditions for Rational (Jury) Evaluation' (2015) 39(1) *Melbourne University Law Review* 77, 79–92; Martire and Edmond, 'Rethinking Expert Opinion Evidence' (n 113).

226 See, eg: NAS, *Strengthening Forensic Science* (n 70) 87, 109–10; PCAST Report (n 70) 4–6.

227 This requires formal testing by scientists. It cannot be generated by the parties or elicited through adversarial procedures.

the jury to understand the various deficiencies with Detective J's opinions or to rationally evaluate them.

A very significant additional limitation with the directions, and a profound problem for the fairness of the proceedings, was making the voice recordings available to the jury. The jurors were presented with the contaminated impressions of a police officer actively engaged in Operation Oracle, along with some of the contaminating evidence presented as independent support. They were, in addition, invited to compare the intercepted voice(s) said to be Davey, and sometimes featuring the names 'Matthew', 'Matty' and 'Matthew Davey', with Matthew Davey's police interview.[228] Jurors undertook their comparisons having been primed with Detective J's contaminated opinions along with the circumstantial evidence implicating Davey as the speaker (which also contaminated his interpretations).

Ostensibly, the recordings were provided to the jurors for two purposes. First, to assess Detective J's credibility and the reliability of his voice identification. However, it is unclear how listening to these recordings enabled the jury to evaluate Detective J's opinions.[229] While in some circumstances poor quality recordings might lead the jury to question a witness's ability to make a (reliable) comparison, provision of the recordings does not facilitate meaningful evaluation of Detective J's testimony. Instead, as Davey submitted on appeal, the results of formal (ie, validation) testing provide the appropriate evaluative framework.[230] How — in the absence of insight into the difficulty of the task and Detective J's ability (or even the ability of other, ordinary persons), and other problems — jurors could make a sensible evaluation of Detective J's opinions is anyone's guess. Assessment of Detective J's opinions was necessarily speculative, and inextricably linked to their own comparisons. How, and more importantly why, were the jurors expected to evaluate the opinions of a police officer that were based on, and perhaps determined by, much of the same evidence they heard in the case against Davey?

Second, the recordings were played to enable the jurors to make their own voice comparisons. However, this was not a comparison of similarities in the voices. Rather, it was a holistic evaluation of all the evidence masquerading (and being presented by the judges) as voice comparison and identification. The jurors were invited to compare the voices in the shadow of the circumstantial case against Davey (and Eaton).[231] When they embarked on this difficult comparison, in the context of an

[228] *Davey* (n 7) [57].

[229] Cf *R v Hawat [No 5]* [2019] NSWSC 1727, [30].

[230] See, eg: Forensic Science Regulator (UK), *Validation* (Guidance, 22 September 2020); Forensic Science Regulator (UK), *Cognitive Bias Effects Relevant to Forensic Science Examinations* (Guidance, 22 July 2020); Forensic Science Regulator (UK), *Appendix: Speech and Audio Forensic Services* (Codes of Practice and Conduct, September 2020).

[231] *Davey* (n 7) [57]. Unwittingly, the Court embarked on an exercise that was very likely to lead the jurors to overvalue the evidence and agree with Detective J even if he was wrong.

accusatorial trial, the evidence was pointing to Davey being the speaker. Here we can observe how evidence which irreparably contaminated Detective J's opinions was (again) available to contaminate the juror comparisons. To the extent that Detective J's opinion evidence was contaminated, or that jurors relied upon it or other evidence to inform their listening, this evidence was being used more than once. In other words, evidence which informed Detective J's interpretation was very likely to have informed (ie, contaminated) the juror comparisons. The suggestive context of the accusatorial trial was simply ignored. Compounding these problems, the contaminating non-voice evidence was said to be also available as independent evidence implicating Davey as the speaker. In these ways the non-voice evidence might have been unwittingly used on as many as three separate occasions. The voice comparison was likely to result in jurors concluding that it was Davey speaking — even if it was not — and simultaneously, and inextricably, that Detective J was a credible and reliable witness.

Judicial directions are incapable of preventing or repairing contaminated perception and evaluation. They are incapable of preventing or repairing cognitive bias. This applies to impressionistic directions, such as those provided in *Davey*, but would also apply to scientifically informed assistance. Directions are incapable of enabling the jury to avoid unconscious influences on their perception and cognition and the double and triple counting of evidence which could not be prevented following the admission of Detective J's synthetic opinions.[232] Jurors are very likely to be persuaded by the sincere, confident, and resilient opinions proffered by police officers.[233] They are extremely vulnerable to the context (a trial with Davey sitting in the dock) and other suggestive evidence — such as the matching DNA profile and the opinion evidence of Detective J — unconsciously and irreparably contaminating their perception, interpretation and evaluation of the voices.[234] Jurors

---

232 See also: *Nguyen* 2002 (n 9) 90 [138] (Anderson J); *Neville v The Queen* (2004) 145 A Crim R 108, 125–6 [66]–[72] (Miller J). In *Bulejcik* (n 31) Toohey and Gaudron JJ stated at 398–9:

> Where the jury is itself asked to make a comparison of voices in a situation such as this one, very careful directions are called for … it is unsafe to leave that matter to the jury without very careful directions as to those considerations which would make a comparison difficult and without a strong warning as to the dangers involved in making a comparison.

Chief Justice Brennan also insisted on the need for 'a satisfactory warning': at 383.

233 See generally: Jerry W Kim and Brayden G King, 'Seeing Stars: Matthew Effects and Status Bias in Major League Baseball Umpiring' (2014) 60(11) *Management Science* 2619; Kanu Okike, 'Single-Blind vs Double-Blind Peer Review in the Setting of Author Prestige' (2016) 316(12) *Journal of the American Medical Association* 1315, 1316; Andrew Tomkins, Min Zhang and William D Heavlin, 'Reviewer Bias in Single-Versus Double-Blind Peer Review' (2017) 114(48) *Proceedings of the National Academy of Sciences of the United States of America* 12708; Simine Vazire, 'Our Obsession with Eminence Warps Research' (2017) 547(7661) *Nature* 7.

234 A simple example might be considering the DNA evidence as implicating Davey, even though the same evidence was part of the matrix of contextual information that unwittingly informed Detective J's opinion evidence, and also unwittingly informed

should not be tasked with voice (or other) comparisons in the context of accusatorial proceedings.[235]

Interestingly, the trial and appellate judges in *Davey* — like many of the trial and appellate judges who have considered the opinion evidence of investigators in recent years[236] — could not discern much unfairness attending the admission and reliance on Detective J's opinions or in allowing the jurors to engage in voice comparisons. Having deemed Detective J's contaminated and impressionistic opinion evidence admissible, according to our remarkably lax admissibility frameworks, the trial judge and TASCCA were satisfied that there was 'no unfair prejudice' to Davey because of the work attributed to the directions.[237] Consider the formulaic assessment offered by the TASCCA: 'Once the required warnings are given, no question of unfair prejudice within the meaning of s 137 of the *Evidence Act* is likely to be discernible'.[238] With the benefit of knowledge, this assessment seems coldly indifferent.

After *R v Tang*,[239] *Honeysett v The Queen*[240] and *IMM v The Queen*,[241] scope for exclusion on the ground of cognitive bias or the reliability (or accuracy) of opinions, like those of Detective J, seem to be largely relegated to unfair prejudice when balancing the probative value of biased and speculative opinions against the danger of unfair prejudice to the defendant.[242] Ignorant of relevant scientific knowledge,

any jury comparison. In this case, the DNA evidence is being triple counted because even though it could be independent of the comparisons, exposing those engaged in comparisons means it is no longer independent. Unwittingly, the evidence might be counted three times: (i) the DNA evidence as an implicitly independent strand of circumstantial evidence; (ii) the DNA evidence influencing Detective J's interpretation; and (iii) the DNA evidence influencing the jurors' comparisons. Such exposure is not only detrimental to the comparisons, but it threatens the rational evaluation of the entire case.

235 Edmond, 'Against Jury Comparisons' (n 94). The fact of admission and the implied value of Detective J's listening and the voice comparison exercise appear to be far stronger signals of their potential, or assumed value, than the limited insight offered through the orthodox directions.

236 See above n 9.

237 *Davey* (n 7) [74].

238 Ibid [58]. See also ibid [74]. This is a peculiar form of words.

239 (2006) 65 NSWLR 681.

240 (2014) 253 CLR 122.

241 *IMM* (n 42).

242 According to *Xie v The Queen* (2021) 386 ALR 371, 457 [301]: '*IMM* left open the possibility that an assessment of the "reliability" of evidence may be permissible as part of an inquiry into the "danger of unfair prejudice"'. This produces the absurdity that the capability of evidence (which inexorably requires consideration of reliability 'at its highest') is to be balanced against actual reliability. For a review of the prejudice associated with expert evidence, see Jason M Chin, Hayley J Cullen and Beth Clarke, 'The Prejudices of Expert Evidence' (2023) 48(2) *Monash University Law Review* 59.

the TASCCA assumed that Detective J's opinions were quite probative (because an uninformed jury might treat them as such) and balanced that impression (and possibility) against an equally naïve response to the dangers.[243] The lawyers and judges appear to have been largely oblivious to dangers outlined in this article. According to the TASCCA there was no actual need to engage in a balancing exercise, for once the warnings were given 'there was no unfair prejudice from the admission of the evidence'.[244]

Here, we can observe adjectival law — and specifically a safeguard intended to protect defendants — operating asymmetrically, in favour of the Crown. Against decades of scientific research, criminal trials and appeals proceed on the basis that risks to the defendant from speculative opinions and cognitive bias can and will be managed by directions.

VII Was Detective J's Opinion Evidence Necessary?

Police, prosecutors and judges seem to assume that voice identification is important, and perhaps necessary, in many cases. But is it? In principle, courts should aim to admit all relevant evidence that can be rationally evaluated by the trier of fact. There are, however, limits in relation to what ought to be admitted where there are significant costs, serious threats to fairness, or manifest risks of error and irrationality. Voice comparison and identification evidence is not only costly to produce (it consumes a great deal of police time), and expensive to hear and contest (consuming time preparing for and presenting it in court), but it is also extremely difficult to manage (through directions and warnings), and very likely to mislead lawyers, trial judges, jurors, and appellate courts.

The question that prosecutors and judges have not adequately grappled with is whether voice identification evidence and jury comparisons are actually required. We contend that in many cases, perhaps most cases, there is no need for voice identification evidence. The number of cases where jury comparisons are helpful is likely to be vanishingly small. Consider the case against Davey. Apart from witnesses placing Davey's co-accused (and friends, Eaton and Cure) in the vicinity of the crime, there was a match with Davey's DNA profile on a recovered screwdriver, guns and ammunition.[245] A stolen gun was recovered close to Davey's house

---

243 *Davey* (n 7) [103]. See also *Davey* (n 7) [99]. Cf *Bulejcik* (n 31) 382 (Brennan CJ):

> the ordinary rules of evidence confer on a judge a discretion to exclude evidence that is unduly prejudicial, albeit the evidence is otherwise admissible. The exercise of that discretion is designed to avoid a significant risk that the evidence will be misused by the jury in a way that cannot be guarded against by an appropriate warning.

244 *Davey* (n 7) [58], [74]. See Jason M Chin, Gary Edmond and Andrew Roberts, 'Simply Unconvincing: The High Court on Probative Value and Reliability in the *Uniform Evidence Law*' (2022) 50(1) *Federal Law Review* 104, 122–3.

245 *Davey* (n 7) [58], [74]. Of course, transfer was a possibility, but that was to be considered in the context of the case as a whole.

with a fingerprint identified to one of his relatives.[246] There were recorded conversations between the burglars and a leader. Phones and phone numbers used by the burglars to communicate during the weekend of the burglary were in the vicinity of the crime scene. They were used at other times by a person calling himself or responding to the names 'Matthew', 'Matty' or 'Matthew Davey'.[247] This evidence could be used as circumstantial evidence implicating Davey in a joint criminal enterprise.[248] There was an abundance of evidence with which to prosecute and convict without recourse to the partial opinions of those who participated in the investigation. Perceived insufficiency, or the desire to prop up a case, should not be addressed by allowing investigators to proffer speculative opinions, however well-intentioned, sincere or plausible.

If there is a need to identify a speaker — whether Davey or some other suspect — then all is not lost. There are a range of valid and reliable approaches to voice comparison available.[249] Rather than call on the impressions of police (or translators) embedded in the investigation, prosecutors should obtain the services of genuine experts. These are individuals with: (1) validated methods; (2) a clear idea of their abilities with different types of voice recordings; (3) a deep understanding of speech, voice comparisons and their limitations; and (4) procedures to manage the corrosive effects of context and cognitive bias. Experts are able to provide reliable opinions on identity or the significance of similarities informed by statistics and/or the risk of error. Opinions obtained from experts using validated methods are opinions based on specialised knowledge. They facilitate the goals of accuracy and fairness by providing relevant evidence in a form that enables rational assessment of their probative value and weight. By managing exposure to domain irrelevant information, genuine experts can provide independent evidence. This helps to prevent double counting evidence and enables courts to avoid asking jurors to compare voices in the very suggestive conditions attending all prosecutions.[250]

---

[246] Ibid [15]–[17].

[247] Ibid [57].

[248] The jury might hear incriminating calls intercepted from the phone linked to Davey, but there was no need to engage them in a biased and speculative comparison exercise: ibid [29].

[249] See: Geoffrey Stewart Morrison and William C Thompson, 'Assessing the Admissibility of a New Generation of Forensic Voice Comparison Testimony' (2017) 18(2) *Columbia Science and Technology Law Review* 326; Geoffrey Stewart Morrison et al, 'Consensus on Validation of Forensic Voice Comparison' (2021) 61(3) *Science and Justice* 299.

[250] Proscribing jury comparisons and excluding the impressions of biased investigators actually simplifies the provision of warnings and will tend to make trials fairer and decision-making less likely to be biased and irrational.

VIII Detective J's Recognition of Eaton

We accept that, in principle, Detective J might express an opinion about Eaton's voice based on his longstanding acquaintance — his apparently genuine (ie, pre-investigative) familiarity.[251] This was qualitatively different to his purported recognition of Davey's voice at the police interview. Our evidence law should provide a clear mechanism to admit the opinions of those who are genuine (ie, non-investigative) familiars. This is because genuine familiarity can rationally assist fact-finding.

There were, however, problems with Detective J's recognition of Eaton and the directions provided.[252] The opinions of familiars are vulnerable to contamination by the conditions in which they are obtained, such that admission and a warning might not repair the threat to probative value or redress the magnitude of unfair prejudice to the defendant. As with eyewitness identifications, the identity of the suspect should not be implied or suggested by the request or the procedure structuring the identification or recognition.[253] In *Davey*, it seems likely that Eaton's involvement was known to the investigators before Detective J purported to identify him on the intercepted recordings. These were, after all, intercepts obtained on the suspicion required to secure legal permission in the form of a warrant. If the recognition was produced in suggestive circumstances — such as where Detective J was listening, already aware that Eaton was a prime suspect or warrants had been obtained to intercept his phone — then it should have been excluded. If it was recognition evidence, then Detective J was primed, expecting to hear Eaton. Similarly, the putative confirmation obtained by speaking to Eaton (at the police interview) should have been excluded because it contributed nothing to the claimed recognition but was likely to mislead.[254] The prosecution's narrative suggested a level of caution and confirmation that was not compatible with the suggestive conditions in which the listening and identification appear to have occurred.

---

[251] Precisely where such opinions sit within the UEL is unclear. We cannot, as Basten JA seems to contend in *Nguyen* 2017 (n 77), just admit the evidence because it was admissible at common law. Section 76 appears to proscribe the admission of such opinions: at 411–12 [20]–[21]. Section 78 is not obviously suited to displaced perceivers and should, consistent with High Court authority, be limited to direct witnesses: *Lithgow City Council* (n 201) 370–1 [45]–[46] (French CJ, Heydon and Bell JJ).

[252] *Davey* (n 7) [99].

[253] Best practices for eyewitness parades (ie, line-ups) require that the suspect not stand out among the foils and that an individual not involved in the investigation, and not aware of the suspect, conduct the parade: Gary L Wells et al, 'Policy and Procedure Recommendations for the Collection and Preservation of Eyewitness Identification Evidence' (2020) 44(1) *Law and Human Behavior* 3, 8. These prescriptions are reflected in the *Crimes Act 1914* (Cth) s 3ZM(6).

[254] *Davey* (n 7) [98].

The directions did not draw known risks to the attention of the jury or place jurors in a position to make sense of Detective J's recognition and attendant problems.[255]

IX Conclusion: Lacking Direction

> What … is vital to the criminal justice system is the capacity of jurors, when properly directed by trial judges, to decide cases in accordance with the law, that is, by reference only to admissible evidence led in court and relevant submissions, uninfluenced by extraneous considerations. That capacity is critical to ensuring that criminal proceedings are fair to an accused.[256]

> If directions are ineffective, then there is little or no point giving them. They add to the length of a trial without any real benefit. Giving directions that may backfire (ie result in jurors reasoning in the opposite way than is intended) is even more problematic, as such directions can be detrimental to the party which the direction is meant to benefit.[257]

We have identified and endeavoured to explain some of the problems with the evidence characterised as voice identification or recognition in *Davey*. In reviewing the conventional judicial directions attending the admission of a police officer's opinion evidence, we have observed how the evidence was mischaracterised in ways that helped to rationalise admission. Serious dangers were overlooked, treated superficially, or addressed in ways that were not merely misguided but misleading and unfair. The fact that the police officer's opinions were a synthesis of the entire circumstantial case was overlooked, even though it makes the opinion irrelevant and, more fundamentally, redundant — trespassing on the constitutional prerogatives of the jury. Detective J's opinions and the jury comparison were both irreparably contaminated such that they had limited, if any, value but great potential for unfair prejudice and irrationality.

Directions provided with the authority and experience of the trial court and endorsed by the TASCCA (and by extension, appellate courts in other Australian jurisdictions) did not and could not have assisted decision-making. They alluded to a few of the risks (or possibilities) for the jury to consider but did not provide practical assistance. Overlooking scientific knowledge and advice, they treated voice identification as mundane — based on the experience of the judges and the common sense of jurors — and effectively trivialised cognitive contamination. The directions in *Davey* (and many other cases) present the reader with complacent assumptions, ignorance and popular misnomers masquerading as common sense, legal experience and even wisdom. They facilitate the admission of the opinions of

---

255 Ibid [99]. Nothing a judge can say is likely to persuade the jury of the reality or the seriousness of the risk of error or the dangers posed by suggestion and expectation when a police officer purports to recognise an acquaintance.

256 *Dupas* (n 14) 248–9 [29].

257 CLR, *Jury Directions: A New Approach* (n 11) 24.

police officers, privileging the case advanced by the prosecutor, while insinuating that any problems could be (or should have been) effectively addressed at trial by competent defence counsel. According to the TASCCA, 'in light of the directions given to the jury … there was no unfair prejudice'.[258] Proceeding as though words are magical — this was, after all, Operation Oracle — directions were said to have removed all unfair prejudice attending the admission of Detective J's opinions.[259]

Convictions obtained through reliance on the impressions of investigators or biased jurors are incompatible with fairness and justice. Safe and socially legitimate convictions are not obtained through cognitive traps. Courts should not pretend that directions eliminate scientifically notorious dangers. Where directions are unlikely to work or unlikely to place the trier of fact in a position to rationally evaluate evidence, judges must exclude the evidence. If they do not, then some trials and some of our trial processes would seem to be primarily for show.

258 *Davey* (n 7) [74].

259 Note that Estcourt JA in *Davey* (n 7) at [58] referred to 'discernible' unfair prejudice, but risks and dangers that are not known to judges, from their (remarkably) limited and non-systematic experiences at the bench and bar, will not be discernible. See Justice Marcia Neave, 'Jury Directions in Criminal Trials: Legal Fiction or the Power of Magical Thinking?' (2012) (Speech, Supreme and Federal Court Judges' Conference, 23 January 2012) 6. See also John L Austin, *How to Do Things with Words: The William James Lectures Delivered at Harvard University in 1955* (Oxford University Press, 1962).