UNSOUND LAW: ISSUES WITH ('EXPERT') VOICE COMPARISON EVIDENCE

GARY EDMOND,* KRISTY MARTIRE† AND
MEHERA SAN ROQUE‡

[*Since the 1980s the volume of identification evidence derived from surveillance devices and telephones has increased dramatically. This article offers a critical analysis of the forensic use of voice comparison and identification evidence. First, it reviews the contemporary jurisprudence in common law and uniform* Evidence Act *jurisdictions, then explains some of the limitations with our current responses to voice evidence, particularly the dramatic rise in the reliance placed upon the opinions of investigators, interpreters and (other ad hoc) 'experts' as well as the willingness to leave voice comparison evidence (and exercises) to juries. Employing an original multi-disciplinary methodology, the article then problematises legal practice through the introduction of relevant social science research on voice comparison (and recognition). As the authors explain, relevant scientific research and opinions are rarely adduced by lawyers or referred to by trial judges when instructing or cautioning juries. In consequence, it is suggested that current legal rules and procedures do not adequately represent what is known beyond the courts and thereby fail to embody fundamental criminal justice principles concerned with truth and fairness.]*

CONTENTS

I Introduction

In recent years most Australian courts have become remarkably receptive to comparison evidence derived from audio surveillance technologies. In most cases the courts are considering whether to allow witnesses to give evidence of their opinion as to whether a voice captured on a surveillance tape is the same as the voice of the accused. These witnesses are often, though not always, characterised as 'experts',[1] sometimes by virtue of formal training, but mostly by virtue of 'displaced' exposure — ie remote listening, usually repeatedly — to the tapes in question. Often characterised as 'identification' evidence, displaced comparison evidence is situated awkwardly at common law and does not come within the definition of 'identification evidence' under the uniform *Evidence Acts* ('UEAs').[2] Australian courts have become reluctant to impose specific conditions on the admission of voice comparison evidence. Indeed, they have demonstrated a willingness to allow juries to make their own assessments of direct and displaced witness testimony and, where tape recordings (or voices) are available, to undertake their own voice comparisons.

This article aims to examine recent trends in voice comparison and identification evidence, focusing primarily upon the evidence of 'displaced non-familiars' and the use of voice recordings.[3] It is our contention that decisions on the admissibility of voice comparison evidence display a troubling readiness to admit incriminating opinion evidence of unknown probative value, an over-reliance on the capacity of traditional features of the adversarial trial — such as cross-examination and warnings to juries — to expose and convey weaknesses, and a hostility towards attempts to require some assessment of the methods used by displaced non-familiars to provide opinions about identity.

Judicial confidence in traditional adversarial mechanisms appears misplaced when set against empirical research concerned with the validity and reliability of

---

[1] We use scare quotes because the ability of many witnesses, including those qualified legally as experts, to provide reliable opinions about identity is in genuine doubt. Many of these 'experts' have no experience or, more importantly, expertise in voice comparisons.

[2] The UEAs are *Evidence Act 1995* (Cth); *Evidence Act 2011* (ACT); *Evidence Act 1995* (NSW); *Evidence Act 2001* (Tas); *Evidence Act 2008* (Vic). According to the Acts' Dictionaries, 'identification evidence' 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.

[3] 'Displaced non-familiars' are those who are not conversant with the suspect (or person of interest) and were not present at the crime scene or its aftermath so as to directly perceive a voice (or sound). On the special dangers arising with respect to strangers and identifications, see, eg, *Kelleher v The Queen* (1974) 131 CLR 534, 550–1 (Gibbs J).

voice comparison, and the efficacy of rules of evidence, procedural safeguards, and appellate review.[4] Engaging with experimental studies and scientific research can help courts to make more appropriate decisions on admissibility (and weight). Remarkably, Australian courts are yet to engage with the considerable scientific literature on these subjects. Rather, judges have preferred to rely upon their own impressions and experiences, assessed against past practice and new statutory arrangements, and subject to the vagaries of prosecution and defence interest and ability.

In this article, we provide a general overview of modern jurisprudence on voice identification and comparison evidence before turning to consider the increasingly prominent role of displaced non-familiar listeners. After describing several recent cases we review some of the relevant scientific research that, we suggest, should be used by courts in their response to voice evidence in order to improve the accuracy of decisions and reduce the number of substantially unfair trials and appeals. Courts, to the extent that they claim to operate in a rational tradition (or capacity),[5] cannot afford to ignore — or have procedures and rules that do not require reference to — relevant scientific studies that bear directly on incriminating evidence.

II OVERVIEW OF THE AUSTRALIAN LAW ON VOICE COMPARISON EVIDENCE

The admissibility and treatment of voice identification evidence can be contrasted with the legal approach to *visual* identification evidence (and images). It is accepted, both at common law and under the UEA, that because of notorious dangers, visual identification evidence is a type of evidence requiring special attention and caution in terms of both admissibility and warnings to the jury.[6] There are extensive statutory arrangements governing the use of eyewitness testimony, identification parades, photo arrays, and visual and image comparison evidence.[7] In addition, where 'expert' witnesses are called to testify based on their interpretations of (often low quality) CCTV images, they are prohibited, both at common law and under the UEA, from expressing opinions about identity (ie positive identification or 'individualisation').[8] Their interpretations are usually restricted to descriptions of similarities (and differences).[9] It is not

---

[4] See Gary Edmond and Kent Roach, 'A Contextual Approach to the Admissibility of the State's Forensic Science and Medical Evidence' (2011) 61 *University of Toronto Law Journal* 343.

[5] On the rationalist tradition, see William Twining, *Rethinking Evidence: Exploratory Essays* (Cambridge University Press, 2nd ed, 2006) ch 3.

[6] These concerns are longstanding: see, eg, *Davies v The King* (1937) 57 CLR 170; *Alexander v The Queen* (1981) 145 CLR 395; *Domican v The Queen* (1992) 173 CLR 555.

[7] See, eg, UEA ss 114–16, 165.

[8] On individualisation, see Michael J Saks and Jonathan J Koehler, 'The Individualization Fallacy in Forensic Science Evidence' (2008) 61 *Vanderbilt Law Review* 199; Simon A Cole, 'Forensics without Uniqueness, Conclusions without Individualization: The New Epistemology of Forensic Identification' (2009) 8 *Law, Probability & Risk* 233.

[9] *R v Tang* (2006) 65 NSWLR 681, 709 [120] (Spigelman CJ, Simpson J and Adams J agreeing); *Murdoch v The Queen* [2007] NTCCA 1 (10 January 2007) [300] (Angel ACJ, Riley J and Olsson AJ). However, because of a caveat in *Smith v The Queen* (2001) 206 CLR 650, 656–7

our intention to defend the current approach to visual identification evidence, especially the use of incriminating images for purposes of identification.[10] Our point is that, by contrast, the admission of voice evidence in Australia is hardly subjected to any regulation at all.

Turning to the discussion of voice evidence, we begin with a review of the dominant approaches to voice comparison (and identification), often derived from cases where lay strangers (ie those not familiar with a particular voice) positively identified an offender, usually on the basis of some kind of voice comparison exercise.[11] This review provides a useful background to our more detailed examination of the increasingly prominent role of the opinions of investigators, interpreters and other 'experts'. Most of the early cases are from New South Wales, though our analysis incorporates the common law and has implications for practice in both common law and UEA jurisdictions.

Judicial consideration of voice identification and comparison evidence, and particularly the use of voice recordings, is relatively recent.[12] Prior to the introduction of the UEA, courts in New South Wales began to consider voice identification evidence — usually where a sensory (or direct) witness positively identified a voice associated with a criminal act — by noting that risks associated with visual identification might apply to voice identification, but in a manner that highlighted some of their occasionally archaic and sometimes superficial concerns. While purporting to develop an admissibility jurisprudence, most courts stopped short of strictly imposing mandatory conditions for the admissibility of voice identification by sensory witnesses. The judges hearing the common law appeals in *R v Smith* ('*E J Smith*'),[13] *R v Brownlowe* ('*Brownlowe*'),[14] *R v Corke*[15] and *R v Brotherton* ('*Brotherton*')[16] — and even

---

[13]–[15] (Gleeson CJ, Gaudron, Gummow and Hayne JJ), Australian investigators are able to proffer positive identification evidence in circumstances where the reliability of such evidence is highly questionable. In the United Kingdom, the approach to images is largely unregulated and, in consequence, is similar to modern Australian approaches to voices: see *A-G's Reference (No 2 of 2002)* [2003] 1 Cr App R 21. In terms of warnings, there appears to be no substantial difference between visual, voice and other kinds of identification: *R v Lowe* (1997) 98 A Crim R 300, 317 (Hunt CJ at CL).

10 For a critical discussion of the forensic use of images, see Gary Edmond et al, 'Law's Looking Glass: Expert Identification Evidence Derived from Photographic and Video Images' (2009) 20 *Current Issues in Criminal Justice* 337; Gary Edmond et al, '*Atkins v The Emperor*: The "Cautious" Use of Unreliable "Expert" Evidence' (2010) 14 *International Journal of Evidence & Proof* 146; Glenn Porter, 'A New Theoretical Framework Regarding the Application and Reliability of Photographic Evidence' (2011) 15 *International Journal of Evidence & Proof* 26.

11 See generally Craig Carracher, 'Voice Identification Evidence' [1993] *Australian Bar Review* 75; David C Ormerod, 'Sounds Familiar? Voice Identification Evidence' [2001] *Criminal Law Review* 595; David Ormerod, 'Sounding Out Expert Voice Identification Evidence' [2002] *Criminal Law Review* 771.

12 Expansion in the use of voice recordings is a response to rapid advances in technological developments, the proliferation of communication technologies, and ever greater state-sponsored surveillance following terrorist attacks. See generally Kevin D Haggerty and Richard V Ericson (eds), *The New Politics of Surveillance and Visibility* (University of Toronto Press, 2006).

13 (1986) 7 NSWLR 444, on appeal from *R v Smith* [1984] 1 NSWLR 462.

14 (1986) 7 NSWLR 461.

15 (1989) 41 A Crim R 292.

16 (1992) 29 NSWLR 95.

appeals under the nascent UEA in *R v Colebrook*[17] and *R v Watson*[18] — focused attention on the quantity and quality of material available to the witness, the distinctiveness of the voice in question, the level of the listener's familiarity, and whether voices were compared under similar conditions (eg yelling in anger).[19] In practice, however, such considerations infrequently led to the exclusion of positive identifications by strangers. Rather, appellate judges required that limitations and problems with voice identification evidence should be brought to the attention of the jury through specific directions and warnings from the trial judge.[20] We can observe these tendencies in *E J Smith*, *Brownlowe* and *Brotherton*.

In *E J Smith*, the case that comes closest to imposing admissibility conditions on voice identification evidence, the trial judge (O'Brien CJ Cr D) insisted that a person purporting to identify the voice of the accused must either have recognised it because of previous familiarity or on some subsequent occasion because of its distinctiveness:

> Basically then for identification to be reliable of a voice with which one is not previously familiar, the law requires that the voice unlike the appearance of a person — must be found to have very distinctive characteristics, … firstly because of the intrinsic qualities of the voice and secondly because of the circumstances in which it was used so that the totality of the qualities of the voice, both its intrinsic qualities and those brought out by its use in those circumstances, make it readily *recognisable* to a witness who is not previously familiar with that voice.[21]

For an unfamiliar voice, it was for the jury to decide whether the voice in question demonstrated characteristics so distinctive and remarkable as to make it readily and reliably recognisable if heard again in similar circumstances. That is, where these conditions might be satisfied it was incumbent upon the trial judge to bring them to the jury's attention and for them to decide. According to O'Brien CJ Cr D, the jury would need to accept that there was a 'very distinc-

---

[17] [1999] NSWCCA 262 (27 August 1999).

[18] [1999] NSWCCA 417 (21 December 1999).

[19] In *R v Colebrook* [1999] NSWCCA 262 (27 August 1999), a woman sexually assaulted in her house at night subsequently recognised the voice of the attacker as a former boarder. This identification evidence, of a voice with which the witness was already reasonably familiar, was deemed admissible provided there were appropriate directions which referred to her gradual recollection and the notorious unreliability of voice identification evidence: at [31] (Simpson J, Mason P and Abadee J agreeing). See also *Watson*, ibid [36]–[39] (Newman J), where the UEA seems to have been effectively ignored; *R v Cassar [No 11]* [1999] NSWSC 321 (14 April 1999) [26]–[27], where Sperling J considered himself bound by the earlier appeal in *E J Smith*.

[20] In effect, this mimicked the concerns about visual and eyewitness identification (re-)emerging from cases such as *Alexander v The Queen* (1981) 145 CLR 395 and *Domican v The Queen* (1992) 173 CLR 555.

[21] *E J Smith* (1986) 7 NSWLR 444, 450 (Lee J) (emphasis added), quoting with approval the summing up of O'Brien CJ Cr D. See also the trial judgment of O'Brien CJ Cr D in *R v Smith* [1984] 1 NSWLR 462, 477, 482. The term 'recognisable' does not refer to instantaneous recognition.

tive' quality in the voice capable of leaving an 'indelible mental impression' in the witness's mind.[22]

In *E J Smith*, a teenager who overheard a home invasion, lasting about 10 minutes and resulting in the death of her father, gave positive voice identification testimony. She told investigating police that the intruder's voice was 'a distinctive voice … being rough, whiney at times, a whingey sound about it.'[23] Some nine months after the event, police officers took the daughter to observe proceedings in the Court of Petty Sessions — where their main suspect was representing himself in unrelated criminal proceedings — and asked her if she was able to recognise any of the voices.[24] In a session where only five persons — the judge, the prosecutor, two witnesses and the accused — spoke, the teenager indicated that the accused's was the voice she had overheard from her bedroom.[25]

On appeal, the New South Wales Court of Criminal Appeal ('NSWCCA') described the questions of whether the original voice had imprinted itself on the witness's memory, and whether the circumstances in which the voices were heard were sufficiently similar, as critical.[26] The NSWCCA stressed that the jury should be told that it must be satisfied with the honesty and reliability of the witness and satisfied beyond reasonable doubt that she was correct in her identification when the voice was subsequently heard in the Court of Petty Sessions.[27] Notwithstanding the trial judge's extensive directions, the NSWCCA was not satisfied that the daughter's description of the intruder's voice was sufficiently accurate or distinctive and concluded that the jury had not been adequately instructed in relation to the need to compare the witness's description of the voice of the offender with a recording of the earlier proceedings where she had purported to make a positive identification. The NSWCCA was concerned that the voice 'was not so singular that error might not occur [and that] [s]uch a state of affairs was never directly drawn to the jury's attention.'[28]

The main issue in the *Brownlowe* trial was the identity of armed robbers. Part of the largely circumstantial case against Brownlowe was voice evidence, based on a few sentences spoken during a bank robbery. Witnesses described one of the robbers as calm, quietly spoken and possessing an Australian accent. These witnesses, having been told that Brownlowe was charged with the robbery, were also taken to court where they heard him represent himself for about 10–15

---

[22] *R v Smith* [1984] 1 NSWLR 462, 482, 485. This is paraphrased in *Brownlowe* (1986) 7 NSWLR 461, 463 (Hunt J).

[23] *E J Smith* (1986) 7 NSWLR 444, 449 (Lee J). On appeal, Lee J described a recording of the accused's voice (from an earlier proceeding) in somewhat different terms: at 454.

[24] Ibid 448. This kind of procedure was subject to strong censure by King CJ in *R v Hallam* (1985) 42 SASR 126, 130. See also the discussion of United States jurisprudence on 'suggestion' in *State v Thibodeaux*, 750 So 2d 916, 932 (Traylor J) (La, 1999).

[25] *E J Smith* (1986) 7 NSWLR 444, 448 (Lee J).

[26] Ibid 458 (Lee J, Street CJ and Maxwell J agreeing).

[27] Ibid 458–9.

[28] Ibid 457–8. The Court was concerned that it was not made sufficiently clear that the jury were not to base their decision on the obvious similarities between the self-represented defendant's voice and the recording of the defendant in earlier proceedings (upon which the daughter had based her identification). See also *Brownlowe* (1986) 7 NSWLR 461, 465 (Hunt J).

minutes in relation to another matter.[29] At Brownlowe's trial, one witness 'said that she was fairly certain that it was the same voice because it was so similar.'[30] On appeal, the NSWCCA concluded that the evidence of witnesses to the robbery was wrongly admitted because it was only similarity evidence but was presented to the jury as evidence of identification or evidence capable of supporting identification: yet there was 'no way in which the jury could draw the necessary conclusion that the two voices were identical'.[31] Following *E J Smith*, the NSWCCA required that the witness identifying the voice must have prior familiarity or have recognised it subsequently because of distinctive features.[32] *Brownlowe* appears to have been amongst the most onerous responses to the reception of voice identification evidence given by direct, though non-familiar, witnesses.

In *Brotherton*, the NSWCCA reiterated the stipulation from *E J Smith* that an unfamiliar voice must be 'sufficiently distinctive as to have left an indelible mental impression in the witness's mind, thus permitting the conclusion safely to be drawn that the two voices were the same.'[33] However, in this case the victim of a sexual assault claimed that she 'recognised' the assailant's voice and hairstyle based on a brief (about 10 minute) exchange two days before the assault.[34] She described his voice as 'a really low husky voice' and told the police that 'it was "the *same* voice" that she had heard' previously.[35] Writing for the Court, Hunt CJ at CL rejected the need, in such circumstances, for the voice to be 'sufficiently distinctive as to make its characteristics memorable.'[36] He concluded that the complainant was sufficiently familiar with the accused and that any dangers would be addressed by the jury being 'warned (as in visual identification cases) that mistakes are sometimes made in the recognition of even close friends and relatives'.[37]

Overall, at common law, the courts in New South Wales were not particularly exclusionary in their orientation. In *E J Smith*, despite what might seem to have

---

[29] *Brownlowe* (1986) 7 NSWLR 461, 462–3 (Hunt J). As in *E J Smith*, this resembles the manner in which investigators exposed an eyewitness to the accused in the court precinct in *Festa v The Queen* (2001) 208 CLR 593. See also *Kelly v The Queen* (2002) 129 A Crim R 363, 371 [33], 373 [45] (McKechnie J).

[30] *Brownlowe* (1986) 7 NSWLR 461, 463 (Hunt J). The trial commenced two days after the first *E J Smith* decision was handed down and was conducted in ignorance of that decision.

[31] Ibid 466. See also discussion of similarity in *Craig v The King* (1933) 49 CLR 429, 446 (Evatt and McTiernan JJ).

[32] *Brownlowe* (1986) 7 NSWLR 461, 466 (Hunt J).

[33] *Brotherton* (1992) 29 NSWLR 95, 106 (Hunt CJ at CL).

[34] Ibid 97, 105 (Hunt CJ at CL). The evidence was that during the assault the complainant recognised the attacker, based on their brief discussion, and indicated as much. Whether this should be understood as 'recognition' or 'opinion' evidence is an issue to which we will return.

[35] Ibid 105 (emphasis in original).

[36] Ibid 106.

[37] Ibid, citing *R v Turnbull* [1977] 1 QB 224, 228 (Lord Widgery CJ for Lord Widgery CJ, Roskill and Lawton LJJ, Cusack and May JJ). The complainant's description of a tattoo on her attacker's thigh, 'not markedly different' from a tattoo on the accused, was used to support her voice identification evidence, in combination with other incriminating circumstantial evidence, such as the attacker's apparent familiarity with the residential complex where the attack took place and Brotherton had previously lived.

been a more restrictive approach, neither the trial judge nor the NSWCCA questioned the admissibility of the opinion (treated as 'recognition' or direct evidence) of a stranger obtained in highly suggestive circumstances. If voice 'distinctiveness' and the need for 'an indelible mental impression' were admissibility requirements for the impressions of non-familiars, then typically they were interpreted in a very accommodating fashion. With the exception of *Brownlowe*, positive voice identification evidence was either admitted or treated as admissible in all of the major appeals.[38] Even in *Brownlowe*, it seems that the characterisation of the testimony as identification (as opposed to similarity) evidence, rather than admissibility per se, was the main obstacle. In most of the early cases it was the adequacy of the directions to the jury that grounded the issue on appeal.

Nevertheless, courts of appeal in other Australian jurisdictions declined to follow the *E J Smith* line of authority, instead holding that familiarity and any 'distinctiveness as will have left an indelible mental impression goes to weight rather than admissibility'.[39] In *R v Hentschel*,[40] the Full Court of the Supreme Court of Victoria held that voice identification evidence was admissible even though the stipulations from *E J Smith*, reiterated in *Brownlowe* (and *R v Colebrook*), had not been satisfied.[41] Murphy J explained:

> The difficulty which I have with the decision in *R v Smith (E J)* … is that it purports to lay down as a rule of law apropos aural identification evidence, propositions which cannot, I believe, be supported as a matter of principle. Moreover, it lays down these propositions as conditions of the admissibility of such evidence, when I believe that at most they can only go to the weight of the evidence to be led.[42]

Notwithstanding these less onerous requirements, Murphy J recognised that it might be unsafe to convict on voice identification evidence standing alone.[43] Brooking J also referred to the earlier decision of *R v Harris [No 3]* ('*Harris*'), where Ormiston J considered the judicial discretion to exclude evidence of voice identification where it was insufficiently probative.[44]

The Victorian common law position was authoritatively summarised by Winneke P in *R v Callaghan*:

> there is no rule of law which obliges the trial judge to exclude such [lay voice comparison] evidence in the absence of evidence of prior familiarity or distinct-

---

[38] See also *R v Hampson* (Unreported, New South Wales Court of Criminal Appeal, Yeldham, Finlay and Brownie JJ, 23 July 1987).

[39] Noted in *Bulejcik v The Queen* (1996) 185 CLR 375, 394 (Toohey and Gaudron JJ) and endorsed in *Nguyen v The Queen* (2002) 26 WAR 59, 75 [62] (Malcolm CJ), 87 [124]–[125] (Anderson J, Steytler J agreeing) ('*Nguyen*').

[40] [1988] VR 362.

[41] We accept that in many cases, exemplified by the facts in *Brotherton* and *Callaghan*, the case against the particular accused may be compelling.

[42] *R v Hentschel* [1988] VR 362, 364. See also at 367–70 (Brooking J), explaining his reasons for rejecting *E J Smith*.

[43] Ibid 364.

[44] Ibid 369, citing *Harris* [1990] VR 310, 318–23.

iveness, although he may, in the exercise of his discretion, exclude it on grounds of prejudice or unfairness.[45]

This approach, perhaps in the absence of authoritative support for the line of cases following *E J Smith*, has been influential in other Australian jurisdictions. The Victorian response has been endorsed by the Supreme Court of Tasmania, and has found favour in South Australia and Queensland.[46] Courts in the Australian Capital Territory have ruled that 'voice identification will be admitted if it is relevant', subject to the court's discretion to exclude evidence.[47] Western Australia has an extensive jurisprudence that effectively mirrors the Victorian rejection of any special rules for voice identification evidence.[48] Consequently, the Victorian approach represents the orthodox position at common law (and, as we shall see, under the UEA).

Perhaps unexpectedly, notwithstanding a purportedly less onerous (or perhaps less prescriptive) approach to admissibility, judges in Victoria appear to have been more willing than judges in other jurisdictions to exclude otherwise admissible voice identification evidence on the basis of their exclusionary discretion. In *Harris* and *R v Rich [No 6]* ('*Rich*'), Ormiston J and Lasry J respectively each excluded positive identification evidence because they were concerned that its probative value was outweighed by the danger of unfair prejudice to the accused.[49] In *Rich*, the actual circumstances were similar to, though perhaps not quite as suggestive as, the manner in which the positive identification was obtained in *E J Smith*.

Considering voice comparison evidence in *Bulejcik v The Queen* ('*Bulejcik*')[50] — specifically, whether a recording of the accused's unsworn statement and an incriminating recording could be left to the jury to compare — the High Court did not express a final opinion on the status of *E J Smith* and the New South Wales approach to voice identification evidence. McHugh and Gummow JJ expressed doubts about the conditions imposed in *E J Smith*,[51] and Gaudron and Toohey JJ placed emphasis on whether the 'quality and quantity of the material is sufficient to enable a useful comparison to be made', noting that 'the greater the amount of material, the greater the similarity in the circumstances in which the voices were spoken or recorded and the greater the number of similar words used, the more useful the comparison.'[52] Brennan CJ

---

[45] (2001) 4 VR 79, 94 [27].
[46] *Greaves v Aikman* (1994) 4 Tas R 196, 208 (Cox J); *R v Bueti* (1997) 70 SASR 370, 379–80 (Doyle CJ); *R v Andrews* [2005] SASC 15 (21 January 2005) [41]–[43] (Debelle J); *Corke v The Queen* (1989) 41 A Crim R 292, 296 (Derrington J).
[47] *R v Miladinovic* (1992) 107 FLR 241, 245 (Miles CJ). See also *Tomicic v The Queen* (Unreported, Federal Court of Australia, Kelly, Jenkinson and von Doussa JJ, 23 August 1989) [29]–[30] (Kelly and von Doussa JJ); *R v Omar* [1991] 58 A Crim R 139, 146–7 (Miles CJ).
[48] See, eg, *Nguyen* (2002) 26 WAR 59; *Neville v The Queen* [2004] WASCA 62 (2 April 2004) ('*Neville*').
[49] *Harris* [1990] VR 310; *Rich* [2008] VSC 436 (23 October 2008). Cf *R v Mackay* [1985] VR 623.
[50] (1996) 185 CLR 375.
[51] Ibid 406–7.
[52] Ibid 395. In the circumstances, they considered the directions insufficient, particularly the failure to direct attention to the different contexts in which the recordings were obtained, the difficulty

doubted the existence of any particular rule (or the need for exhaustive jury instructions), and suggested it would not be relevant to comparisons by the jury anyway.[53]

More recently, after the introduction of the *Evidence Act 1995* (NSW), courts in New South Wales formally resiled from their increasingly idiosyncratic common law position by removing preconditions on the reception of voice identification evidence.[54] With the transition to the UEA regime, the trend has been to reject the imposition of specific conditions on admissibility and to instead characterise voice identification evidence as *recognition* (ie direct or fact) evidence governed solely by relevance (ss 55 and 56), the mandatory and discretionary exclusions (ss 135 and 137), and directions and warnings (ss 116 and 165). Voice identification evidence is treated as admissible if it is relevant: that is, it will be admissible where, *if accepted*, it could rationally affect the assessment of the probability of facts in issue. Directions and warnings, and to a lesser extent mandatory and discretionary exclusions, appear to be the preferred way to manage the problematic dimensions of evidence derived from voices and comparisons of voices. Where recorded evidence is available the tribunal of fact is frequently encouraged to undertake its own comparison.[55] Now, voice identification and comparison evidence is routinely admitted and questions about probative value and reliability are left for weight and the tribunal of fact. In consequence, all Australian jurisdictions have either abandoned or elected not to follow the restrictive approach associated with *E J Smith* and the courts of New South Wales pre-1995 (but which operated until 2000).[56]

Typically, voice evidence is characterised as recognition evidence: that is, it is treated as a kind of unconscious or non-reflective process of recognition leading to identification.[57] Classifying voice evidence in this way tends to confer the status of *fact* upon it, thereby avoiding any need to address interpretive issues and exclusionary rules associated with *opinion* evidence. In reality, the vast majority of voice comparison and recognition evidence from non-familiars is interpretive and therefore opinion. For practical reasons, most voice evidence — including positive identification evidence and even much of the evidence of close familiars (eg family members and longstanding friends) — is best concep-tualised as interpretative.[58] The alternative is for a messy inquiry into whether,

---

of comparing two unfamiliar voices, and the 'risk' that a jury 'might conclude too readily that a foreign accent on a tape is that of the accused where the accents are similar': at 397.

[53] Ibid 382.

[54] *R v Adler* (2000) 52 NSWLR 451; *Li v The Queen* (2003) (2003) 139 A Crim R 281.

[55] The appeal in *Bulejcik* was successful not because of the actual jury comparison exercise, but because of the inadequacy of warnings (and reliance on a tape recording that was not in evidence). For a more recent example of a jury comparison case, see the discussion of *R v Korgbara* (2007) 71 NSWLR 187 below in Part V.

[56] *R v Adler* (2000) 52 NSWLR 451.

[57] This process need not be instantaneous, and can encompass gradual recollection.

[58] The line between opinion and fact is notorious. See, eg, *R v Leung* (1999) 47 NSWLR 405, 414 [43] (Simpson J); *R v Smith* (1999) 47 NSWLR 419, 422–3 [16]–[22] (Sheller JA); *Neville* [2004] WASCA 62 (2 April 2004) [44]–[46] (Miller J). See also the discussion in Paul Roberts and Adrian Zuckerman, *Criminal Evidence* (Oxford University Press, 2004) 132–46 and Déirdre Dwyer, *The Judicial Assessment of Expert Opinion* (Cambridge University Press, 2008) 76–97.

when hearing a voice or comparing voices, the witness — stranger or familiar — made the positive identification instantaneously and without reflection, or consciously considered the identity of the speaker, or gradually recollected similarities or identity.[59] With the exception of non-reflective instantaneous recognition, all of this evidence would seem to be opinion evidence, regardless of how the witness, lawyer or judge classifies it.

In consequence, in most cases there is a need for lawyers and judges to consider whether voice identification evidence satisfies the rules governing the admission of opinion evidence, or to formally develop exceptions. Exceptions might be granted to those who are very familiar with a voice, and who may well recognise a voice instantaneously and unconsciously (though often these witnesses will be giving fact evidence). The voice identification and comparison evidence of those lacking familiarity should be treated as interpretive and, therefore, as opinion evidence: that is, as an opinion about whether two (or more) voices are derived from the same or similar source. There is also, as we explain below, an additional need to consider whether the limited probative value of much, though certainly not all, voice comparison and recognition evidence outweighs the very real danger of unfair prejudice,[60] particularly the prejudice caused by suggestion and extremely high levels of error, as in positive voice identifications subject to long delays.

Most of the cases discussed so far involved positive voice identification evidence — where a sensory witness attributes spoken words to a specific individual based on a comparison or limited familiarity — from those who had witnessed events relevant to criminal proceedings. In most of these cases, lawyers and judges simply assumed the evidence was admissible without explicitly adverting to the basis for admission. Common law receptivity is, however, mentioned in *Harris*. There, Ormiston J accepted that non-expert sensory witnesses should be allowed to express opinions derived from voice comparison, though without explaining the precise basis of admission. He stated: 'this is clearly a field in which non-expert opinion may be received, even if it were to involve opinion rather than observation in the widest sense.'[61]

In many cases, by classificatory fiat or elision, incriminating *opinions* about the identity of a speaker, based on the comparison of sounds, are treated as

---

[59] This approach avoids the need to determine, in every case, whether a particular mental process is unconscious recognition as opposed to conscious interpretation. It also focuses attention on whether the opinion about identity is 'specialised knowledge' based on sufficient exposure to the accused. Treating this as evidence of opinion avoids the anomalous position of allowing some interpretations (whether conscious or not) to be treated as evidence of fact. We could accept a 'factual' exception for the recognition evidence of family members, colleagues and those with considerable familiarity, provided this did not routinely extend to the evidence of investigators, translators and police acquired during the course of an investigation. See, eg, *R v Robinson* [2007] QCA 99 (30 March 2007) [20]–[25] (Keane AJ); *R v Trudgett* (2007) 70 NSWLR 696, 700–1 [19]–[33] (Spigelman CJ); *Neville* [2004] WASCA 62 (2 April 2004) [83], [90] (Heenan J); *Harris* [1990] VR 310, 318 (Ormiston J); *Bulejcik* (1996) 185 CLR 375, 381 (Brennan CJ). See also as an example of variable familiarity *Mills v Western Australia* (2008) 189 A Crim R 411. See also the discussion of UEA s 78 below in the text accompanying nn 85–88.

[60] See *R v Christie* [1914] AC 545; UEA ss 135, 137.

[61] [1990] VR 310, 318.

evidence of recognition. Consequently, the rules applicable to opinion evidence are rarely applied. Where they are considered, they are often circumvented through classification as fact or recourse to questionable and contorted common law categories such as 'ad hoc expertise'.[62]

In the remainder of this article, we are primarily interested in the evidence of those who were not direct witnesses and those whose only familiarity with voices emerges during the course of an investigation.[63] That is, we are most concerned with the evidence of investigators, interpreters and others classified (if only by courts) as voice comparison 'experts'. Much, and perhaps all, of their evidence is interpretive and, in consequence, should be treated as opinion evidence. These witnesses — frequently police officers, interpreters and a variety of formally qualified individuals (such as linguists) — are routinely allowed to express incriminating opinions based on their exposure to voices through surveillance or translation, and/or on the basis of analysis: usually repeated listening to a set of recordings. Whatever the common law might allow for direct or sensory witnesses (those we might characterise as 'earwitnesses'), there are rules governing the ability of displaced (or indirect) witnesses — such as investigators, translators and purported experts — to proffer their incriminating opinions, whether at common law or under the UEA.[64] Yet, notwithstanding these rules, many courts seem to have merely extended the common law receptivity to direct witnesses, and/or developed a superficial response to rules governing opinion, to enable displaced listeners to proffer their incriminating opinions.

At common law and under the UEA witnesses are obliged to give evidence of facts (ie description or unreflective recognition) and are prevented from expressing opinions unless those opinions are incidental or necessary to understand the testimony.[65] This seems to be the basis on which sensory witnesses are entitled to express opinions — recognised implicitly by Ormiston J in *Harris*, as discussed above — about identity derived from hearing (and seeing). Things, however, are different for those who are not direct (or sensory) witnesses. At common law (and *in practice* under the UEA), most witnesses can only express opinions if

---

[62] See, eg, *R v Colebrook* [1999] NSWCCA 262 (27 August 1999) [31] (Simpson J); *R v Watson* [1999] NSWCCA 417 (21 December 1999) [39] (Newman J); *Li v The Queen* (2003) 139 A Crim R 281, 286–7 [39]–[42] (Ipp JA).

[63] We are primarily interested in those who did not perceive the relevant sounds (as direct or sensory witnesses) as part of a crime, its preparation or its aftermath, whether as conversations, exchanges or commands. Our main focus attaches to displaced (or remote) listeners, and particularly those who are not familiar with the alleged speaker. We are, in consequence, primarily interested in those who compare unfamiliar voices remotely, although the issue of familiarity and related conceptions of recognition, identification and opinion will re-emerge throughout the article. In virtually all of the cases involving non-familiars and those who were not familiar with the suspects before the investigation, the witness is expressing an opinion about the identity of the speaker based on an interpretation (ie an incriminating opinion).

[64] Earwitnesses are the sound equivalent of eyewitnesses. That is, they witness an event and have a direct sensory experience.

[65] See UEA ss 76, 78; Andrew Ligertwood and Gary Edmond, *Australian Evidence: A Principled Approach to the Common Law and the Uniform Acts* (LexisNexis Butterworths, 5th ed, 2010) 603–11; Jeremy Gans and Andrew Palmer, *Uniform Evidence* (Oxford University Press, 2010) 134–8.

they have 'expertise' in a 'body of knowledge or experience' and the opinion will assist the tribunal of fact.[66] In theory, at least, the situation is more complicated under the UEA. First, the only bases for sensory witnesses to express opinions about identity based on voice comparison are provided by ss 78 and 79.[67] Of course, if the witness is giving factual (eg descriptive) evidence, then their evidence is admissible if relevant[68] and not caught by some exclusionary rule. The problem with most voice identification evidence and virtually all displaced listening is that where the witness is not already familiar with the voice, they will normally be expressing an opinion on the basis of some type of comparison, regardless of whether the evidence is characterised as recognition or direct evidence. Except where witnesses purport to identify features of a very familiar voice, any attempt at comparison or identification will generally be interpretive and, therefore, should be subject to the rules regulating the admission of opinion evidence.[69]

For us, the main problem is the admissibility pathway for the opinions of investigators, interpreters and qualified individuals about identity on the basis of displaced listening (and analysis) of sound recordings. Apart from the generally unsatisfactory decisions discussed below, there are relatively few decisions that attend to the question of 'expert' voice comparison evidence in Australia. The most prominent case, which predates the UEA and most of the modern Australian authority on voice comparison evidence, is, again, from New South Wales. Unlike the vast majority of the cases discussed below, it concerns the admissibility of 'expert' opinion evidence adduced by the *defence*.

In *R v Gilmore* ('*Gilmore*'),[70] the appellant challenged the exclusion of the opinion of a lecturer in English who specialised in phonetics.[71] Drawing on some authority from the United States,[72] the NSWCCA concluded that the opinion

---

[66] *Clark v Ryan* (1960) 103 CLR 486, 491 (Dixon CJ). See also *R v Bonython* (1984) 38 SASR 45, 46–7 (King CJ).

[67] See UEA s 76(1): 'Evidence of an opinion is not admissible to prove the existence of a fact about the existence of which the opinion was expressed.' Section 76 would appear to cover the field and eliminate any residual common law categories. There is no exception for ad hoc expertise, because 'specialised knowledge' seems to be a prerequisite. Arguably, the common law does not allow ad hoc experts to present opinion evidence pertaining to identification since the cases are concerned primarily with the use of transcripts: see *R v Menzies* [1982] 1 NZLR 41, 49 (Cooke J for Cooke, McMullin and Somers J and Sir Clifford Richmond) and *Butera v DPP (Vic)* (1987) 164 CLR 180; cf *Murdoch v The Queen* [2007] NTCCA 1 (10 January 2007).

[68] UEA ss 55–6.

[69] Where the witness is very familiar with the voice, as in the case of a family member or spouse, then the evidence is often characterised as 'recognition' and therefore evidence of fact. It might also satisfy an accommodating reading of the rules for expert opinion, especially under UEA s 79, which might allow an opinion about identity based on 'specialised knowledge' of a particular voice through long exposure (ie substantial experience across a wide range of situations and contexts) to be admitted. We discuss evidence supporting the general reliability, though certainly not infallibility, of voice identification by familiars in Part VI(B).

[70] [1977] 2 NSWLR 935.

[71] See also *R v McHardie* [1983] 2 NSWLR 733, 752–64 (Begg, Lee and Cantor JJ), where the admissibility of similar evidence was discussed.

[72] *Gilmore* [1977] 2 NSWLR 935, 939–41 (Street CJ, Lee and Ash JJ agreeing), citing *United States v Baller*, 519 F 2d 463 (4th Cir, 1975) and Henry F Greene, 'Voiceprint Identification: The Case in Favor of Admissibility' (1975) 13 *American Criminal Law Review* 171.

evidence was admissible. Subsequently, the particular technique (the use of spectrographs or voiceprints) relied upon by the defence in *Gilmore* was shown to be unreliable.[73] Since *Gilmore* there has been little sustained interest in the basis for the admissibility of opinion evidence, and most investigators, interpreters and 'experts' have been allowed to express their incriminating opinions on the basis of the rules governing ordinary earwitnesses (ie relevance) or through very accommodating readings of the rules governing opinion evidence. The latter approach finds expression in the English common law case of *R v Robb*:[74] a decision that is regularly followed and occasionally endorsed by Australian courts.[75] In *R v Robb*, the Court of Appeal upheld the admission of incriminating opinion evidence based solely on 'auditory techniques' (ie listening), even though the linguist purporting to identify Robb as the speaker on a ransom tape conceded that the 'great weight of informed opinion, including the world leaders in the field, was to the effect that auditory techniques unless supplemented and verified by acoustic analysis were an unreliable basis of speaker identification.'[76]

Perhaps because of the controversy associated with older voice comparison techniques, in conjunction with the sheer proliferation of voice recordings — obtained via methods ranging from telephone intercepts to covert listening devices — Australian investigators, prosecutors and judges facilitated new ways of admitting incriminating opinions. Unfortunately, these opinions were admitted before any credible research supporting the underlying techniques and assumptions was undertaken and notwithstanding a large body of scientific research reinforcing the difficulties of voice comparison. *Gilmore* demonstrates how the orthodox approaches to the admission of expert opinion evidence, where the primary interest is focused on qualifications and 'the field', circumvent the more fundamental inquiry into whether the technique is in fact valid and reliable.[77]

---

[73] See Committee on Evaluation of Sound Spectograms, Assembly of Behavioral and Social Sciences, National Research Council, *On the Theory and Practice of Voice Identification* (National Academy of Sciences, 1979). Interestingly, these problems were raised in *Gilmore* and expressed in *Harris* [1990] VR 310, 314 (Ormiston J) by scholars from Monash University.

[74] [1991] 93 Cr App R 161.

[75] See, eg, *R v Farquharson* (2009) 26 VR 410, 431–2 [90] (Warren CJ, Nettle and Redlich JJA). See also, in the United Kingdom context, *R v Chenia* [2004] 1 All ER 543, 573–4 [100]–[102] (Clarke LJ for Clarke LJ, Pitchford J and Judge Fabyan Evans); *R v Flynn* [2008] 2 Cr App R 20. *R v Robb* is analogous to the increasingly marginalised Australian tort case of *Commissioner for Government Transport v Adamcik* (1961) 106 CLR 292. Interestingly, as the influential *Makita (Australia) Pty Ltd v Sprowles* (2001) 52 NSWLR 705 decision implies, it is unlikely that this kind of evidence would be relied upon by a judge in modern Australian *civil* litigation. See also the discussion of *R v Robb* and *R v O'Doherty* [2003] 1 Cr App R 5 in *R v Korgbara* (2007) 71 NSWLR 187, 205–6 (McColl JA).

[76] [1991] 93 Cr App R 161, 165 (Bingham LJ for Bingham LJ, Hutchison and Buckley JJ). Recent writings by forensic linguists continue to emphasise the need for both auditory and acoustic techniques: Michael Jessen, 'The Forensic Phonetician: Forensic Speaker Identification by Experts' in Malcolm Coulthard and Alison Johnson (eds), *The Routledge Handbook of Forensic Linguistics* (Routledge, 2010) 378; John Olsson, *Forensic Linguistics* (Continuum, 2nd ed, 2008) 181; Malcolm Coulthard and Alison Johnson, *An Introduction to Forensic Linguistics: Language in Evidence* (Routledge, 2007) 149. On emerging approaches concerned with validation and reliability, see below Part VIII(C).

[77] In *Nguyen* (2002) 26 WAR 59, 74 [60] (Malcolm CJ), the issue of 'whether voice comparison is a recognised field of expertise' was raised too late — there had been no evidence regarding this point or the qualifications and experience of the interpreter at the trial.

Gilmore is also revealing because the appeal implies that prosecutors are likely to challenge, and judges more likely to scrutinise (and often exclude), 'expert' evidence adduced by defendants.[78]

Supplementary rules of admissibility, such as the basis rule — which requires the expert to explain the underlying technique used (and in some versions also the facts relied upon) to reach their opinion — and the ultimate issue rule — which, although no longer strictly applicable, should focus attention on evidence, especially opinions, that address an essential issue, such as the identity of an offender — tend to be trivialised.[79] What we can say is that there is a conspicuous lack of discussion of voice comparison evidence in terms of expert opinion evidence (or 'specialised knowledge'), and little interest in applying relevant rules strictly in the interests of ensuring the fairness of criminal proceedings.

Modern voice comparison cases exemplify a disconcerting willingness to recognise and admit incriminating opinions. That is, even in those cases where the admissibility of the incriminating opinions of investigators is considered, courts often excuse the inability to satisfy the terms of the exceptions to the statutory opinion rule (or its common law equivalents) by allowing those whose 'expertise' has been developed during the course of the investigation, mostly through repeated listening to voice recordings, to express their impressions as 'ad hoc experts', rather than as experts whose opinions are based on genuinely 'specialised knowledge' (under the UEA) or a 'body of knowledge or experience' (at common law) related to voice comparison.[80]

The idea of 'ad hoc expertise' is inconsistent with the explicit terms of UEA s 79(1) and represents a massive expansion of admissible opinion.[81] It enables the state to rely upon the incriminating opinions of investigators and those working closely with them. Recognition of 'ad hoc expertise' is convenient for investigators, prosecutors and courts, but it treats extant, if legally unknown,

---

[78] See also *R v Madigan* [2005] NSWCCA 170 (9 June 2005). This is certainly the experience in the United States: see, eg, D Michael Risinger, 'Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock?' (2000) 64 *Albany Law Review* 99; Jennifer L Groscup et al, 'The Effects of *Daubert* on the Admissibility of Expert Testimony in State and Federal Criminal Cases' (2002) 8 *Psychology, Public Policy and Law* 339.

[79] Compare the detailed attention paid to the basis of the opinion in civil cases such as *Makita (Australia) Pty Ltd v Sprowles* (2001) 52 NSWLR 705, 729–30 [59], 745–50 [87]–[102] (Heydon JA) and the recent High Court case of *Dasreef Pty Ltd v Hawchar* (2011) 85 ALJR 694, 704 [31] (French CJ, Gummow, Hayne, Crennan, Kiefel and Bell JJ). See also *R v GK* (2001) 53 NSWLR 317, 326–7 [40] (Mason P).

[80] There is an implicit, though never justified, confidence in the special abilities of police, interpreters and experts from cognate fields. See, eg, *Kelly v The Queen* [2002] WASCA 134 (17 May 2002) [20] (Anderson J) in relation to visual opinion evidence; *United States v Ladd*, 527 F 2d 1341, 1343 (Jones, Wisdom and Ainsworth JJ) (5th Cir, 1976).

[81] Gary Edmond and Mehera San Roque, 'Quasi-Justice: Ad Hoc Expertise and Identification Evidence' (2009) 33 *Criminal Law Journal* 8, 22–3. Cases where the concept of 'ad hoc expertise' was recognised include *Neville* [2004] WASCA 62 (2 April 2004) [45]–[46] (Miller J); *Li v The Queen* (2003) 139 A Crim R 281, 287 [42] (Ipp JA); *R v Drollett* [2005] NSWCCA 356 (4 November 2005) [63] (Simpson J); *R v Tang* (2006) 65 NSWLR 681, 709 [120] (Spigelman CJ); *Murdoch v The Queen* [2007] NTCCA 1 (10 January 2007) [296] (Angel ACJ, Riley J and Olsson AJ); *Irani v The Queen* (2008) 188 A Crim R 125, 128 [14] (Hoeben J). A legal fabrication, 'ad hoc expertise' is the ultimate in '*science* for litigation': see Gary Edmond, 'Supersizing *Daubert*: *Science for Litigation* and Its Implications for Legal Practice and Scientific Research' (2007) 52 *Villanova Law Review* 857.

scientific literature and research into voice comparison with disdain.[82] It allows investigators, translators and, occasionally, formally qualified individuals (such as linguists and those with an interest in phonetics) to express their incriminating opinions, on the basis of *whatever* familiarity or experience they have obtained during the course of an investigation or analysis, without having to satisfy the exception to the opinion rule for 'specialised knowledge'.

The investigators, interpreters and linguists routinely allowed to express incriminating opinions about identity frequently possess no relevant expertise. There is, as we shall see, considerable slippage and legal inattention to the considerable gap between translation (and interpretation) and identification.[83] Similarly, formal qualifications and experience (in linguistics or phonetics) tell us little about a person's ability to make reliable voice comparisons or understand methodological issues associated with voice comparison, particularly problems introduced by the suggestive way opinions are elicited.[84] Very few of the 'experts' featuring in the cases discussed below refer to relevant scientific research and none appear to have tested their actual ability.

As an alternative pathway for admission, several judges in UEA jurisdictions have suggested that s 78 might provide a basis to admit the opinions of displaced listeners.[85] This response is interesting. First, it explicitly recognises that these witnesses are expressing an *opinion*. Second, s 78 appears designed to allow the evidence of those whose opinion 'is based on what the person saw, heard or otherwise perceived' to be admitted where that 'opinion is necessary to obtain an adequate account or understanding of the person's perception of the matter or event'.[86] It seems curious that judges should read a statute in a manner that is inconsistent with its own terms in order to provide investigators and other displaced listeners with scope for expressing their incriminating opinions about

---

[82] See below Part VIII.

[83] On general problems with interpreters and translation in refugee and asylum courts, see Anthony Good, *Anthropology and Expertise in the Asylum Courts* (Routledge-Cavendish, 2007) ch 7; Livia Holden (ed), *Cultural Expertise and Litigation: Patterns, Conflicts, Narratives* (Routledge, 2011).

[84] It is not our intention to suggest that formal training as a linguist provides a basis for the admission of opinions based on voice comparison. In order to express an opinion that is relevant, there should be a demonstrably reliable technique. Without evidence of ability (or proficiency), the trappings of academic qualifications and university positions may be merely misleading.

[85] For example, the opinion evidence in *R v Leung* (1999) 47 NSWLR 405 was admitted at trial on the basis of s 78. Section 78 states that the opinion rule does not apply to evidence of an opinion expressed by a person if:
    (a) the opinion is based on what the person saw, heard or otherwise perceived about a matter or event; and
    (b) evidence of the opinion is necessary to obtain an adequate account or understanding of the person's perception of the matter or event.
    It embodies the common law 'sleight of hand', alluded to by Ormiston J in *Harris* [1990] VR 310, 314–15, that enables sensory witnesses to express opinions about identity rather than focusing attention upon the intractable fact/opinion distinction.

[86] This applies to all of the senses: see *AK v Western Australia* (2008) 232 CLR 438, 447 [21] (Gleeson CJ and Kiefel J), 454 [49] (Gummow and Hayne JJ), 461–4 [67]–[74] (Heydon J) for some discussion of taste, touch and smell.

the identity of speakers (and those in images).[87] This line of reasoning was formally considered and rejected by Kirby J in *Smith v The Queen* ('*Smith*').[88]

Smith is also instructive when considering investigative bias and relevance. *Smith* was an appeal concerned with police identification evidence based on security images from a bank. Kirby J's observations seem highly pertinent to the voice comparison evidence of investigators:

> The experience of the law, expressed with increasing conviction during the last two decades, is that very great risks of wrongful conviction and miscarriages of justice can attend identification (and recognition) evidence generally, and particularly where such evidence is based on photographs. In this sense, I see no difference in the dangers caused by evidence of identification from photographs of the offender in action, such as produced by bank surveillance, and identification from photographs of the accused and other suspects held by police. The risks, already large, may be enhanced by the natural desire of a person performing the act of identification to produce an affirmative outcome rather than to admit to incapacity and failure. The risks are still further increased where the person concerned has a relevant professional motivation (even if only subconsciously) to identify a person.[89]

The relevance of the voice identification evidence of displaced witnesses has been treated inconsistently in response to challenges to voice comparison evidence. In *Smith*, the witnesses were police officers, with limited exposure to Smith, purporting to identify him from CCTV images of a bank robbery. A majority of the High Court concluded that where the jury was in a similar position to the displaced witnesses, in respect to comparing incriminating images with the accused in the dock, then the witnesses' evidence was irrelevant. It is arguable that the majority conflate a degree of redundancy with relevance. The police officers' opinions about identity are relevant (even if they possess low probative value), but should not be admitted because they are opinions without an admissibility pathway (contra s 76).[90] By analogy, in voice comparison cases, the investigators do not hear or otherwise perceive 'the matter' (s 78) and generally do not possess 'specialised knowledge' relevant to voice comparisons (s 79).

---

87 Indeed, this approach was not followed in *R v Drollett* [2005] NSWCCA 356 (4 November 2005) [63] (Simpson J) and *R v Leung* (1999) 47 NSWLR 405, 410–12 [26]–[35] (Simpson J) (Spigelman CJ and Sperling J preferred not to express an opinion on the scope of s 78). In *R v Leung* the evidence was admitted as 'ad hoc expertise' via s 79. Simpson J maintained a stricter view in the non-expert case of *R v Whyte* [2006] NSWCCA 75 (24 March 2006) [56]–[57], contra Spigelman J at [35]–[36]. Applying s 78 to remote and displaced audiences seems inconsistent with the text of the provision and would appear to allow us all to become voice and visual 'ad hoc experts' to the extent that we could be bothered listening to, or watching, incriminating recordings.

88 (2001) 206 CLR 650.

89 Ibid 668 (citations omitted). See also *R v Crouch* (1850) 4 Cox CC 163, 164 (Maule J). The fact that these exposures and interpretations are obtained in conditions where the identity of the speaker was suggested, directly or indirectly, by investigators, or the speaker was identified by an unfamiliar investigator, tends to be trivialised: contra *R v Gaunt* [1964] NSWR 864, 866–7 (Herron CJ, Ferguson and Nagle JJ).

90 Here we agree with the analysis by Kirby J (and the overall outcome) in *Smith* (2001) 206 CLR 650. Cf, eg, *Neville* [2004] WASCA 62 (2 April 2004) [97]–[98] (Heenan J).

Where the defence has challenged the admissibility of incriminating opinions about the voices of non-familiars (such as the police with limited familiarity of Smith), most courts have distinguished the voice identification cases, often on the pragmatic basis that not admitting the evidence would require the jury to listen to voice recordings which are often of low quality, very long, and contain much content of little, if any, significance. Sometimes, in addition, the content and whether it is actually incriminating is contentious.[91] Nevertheless, because most judges approach the admissibility of voice evidence primarily on the basis of whether it is relevant, the key protections are, in effect, the discretionary (and mandatory) exclusions and warnings to the jury. Notwithstanding serious problems with much voice comparison evidence, few judges have excluded this evidence or prevented the jury from considering it except where the recordings were of very low quality.[92] On average, lawyers and judges, in common law and UEA jurisdictions, tend to be reluctant to fulfil their gatekeeping responsibilities when confronted with the incriminating opinions of displaced listeners.[93]

The low level of attention focused on the admissibility of evidence about the identity of voices places considerable weight on judicial directions and warnings.[94] Judges, as the cases discussed above indicate, have a tendency to admit voice comparison evidence and then attempt to address limitations, problems and dangers through directions and warnings. There is an expectation that judges will address specific issues.[95] In cases involving expert witnesses, the trial judge should also explain to the jury how they might respond to such evidence. We discuss the adequacy, and the scientific foundation, of such warnings and directions below in Part VIII(B). For the moment, we merely need to advert to the lack of attention to any scientific research, particularly research on the very high levels of error, the dangers created by suggestive voice identification procedures and, perhaps most disconcertingly, given the preference for admission and the reliance placed upon them, the apparently limited efficacy of judicial instructions, directions and warnings. There is a failure to treat voice comparison evidence as evidence of opinion and a reluctance to exclude incriminating opinions, even when they are likely to be unreliable, and therefore of

---

[91] See, eg, *Dodds v The Queen* (2009) 194 A Crim R 408, 414 [19]–[26] (McLellan CJ at CL); *Keller v The Queen* [2006] NSWCCA 204 (26 July 2006) [24] (Studdert J).

[92] See *Neville* [2004] WASCA 62 (2 April 2004) [88] (Heenan J) for an orthodox common law response to the discretionary exclusion. *R v Hall* [2001] NSWSC 827 (17 September 2001) was a case where the sound quality of purported 'admissions' was low. Ironically, sometimes the poor quality of voice recordings provides a basis for the admission of an incriminating transcript and 'expert' voice comparison evidence. See also *R v Murrell* (2001) 123 A Crim R 54, where fresh evidence suggested that an incriminating transcript prepared by investigating police officers contained significant and unfairly prejudicial mistakes; *Butera v DPP (Vic)* (1987) 164 CLR 180; *R v Solomon* (2005) 92 SASR 331, 350–1 [74]–[75] (Doyle CJ); *R v O'Neil* [2001] VSCA 227 (14 December 2001) [43]–[50] (O'Bryan AJA).

[93] See generally Gary Edmond, 'Specialised Knowledge, the Exclusionary Discretions and Reliability: Reassessing Incriminating Opinion Evidence' (2008) 31 *University of New South Wales Law Journal* 1; Tim Smith and Stephen Odgers, 'Determining "Probative Value" for the Purposes of Section 137 in the Uniform Evidence Law' (2010) 34 *Criminal Law Journal* 292.

[94] See UEA ss 116, 165.

[95] See below Part VIII(B).

limited probative value and likely to produce very real dangers of unfair preju-dice to the defendant.[96]

Among the witnesses appearing in the cases discussed in Part III, almost none had prior familiarity with the voices of suspects, and there was little, if any, prior experience or expertise in voice comparison. None were involved in the study of voices or voice comparison, and none had attempted to validate or assess the accuracy of their methods. Most of the opinions currently relied upon by investigators and prosecutors in Australia have never been subjected to any kind of validation or reliability study. We do not even know if those allowed to express incriminating opinions, as 'experts' or 'ad hoc experts' (or lay wit-nesses), can actually do what they contend. None of the current methods are demonstrably reliable.[97]

III   VOICE COMPARISON CASES: AN INTRODUCTORY SAMPLE

The cases discussed in this Part exemplify both the lack of judicial concern about the basis for the reception of 'expert' voice comparison evidence, and a failure to take sufficiently seriously the procedural or investigative biases that are often apparent. We have selected a sample of recent cases, primarily from the NSWCCA, to illustrate these limitations along with the exaggerated confidence invested in the trial and its ability to identify and adequately convey them. Let us begin with an appeal decided shortly after the approach from *E J Smith* and *Brotherton* was formally abandoned in *R v Adler*.[98]

In 2002, the NSWCCA heard the appeal in *R v Riscuta* ('*Riscuta*'), which concerned two co-accused, Riscuta and Niga.[99] This was an appeal from a conviction for the supply of heroin, with one ground focusing on the admission of incriminating voice identification evidence of an interpreter, Clarice Kandic. Kandic had initially been called as a witness in the 2001 trial, to prove some translations she had made of covert recordings from Romanian into English.[100] These translations had been completed in 1994. Eighteen months earlier, in 1993, she had been requested by the New South Wales Crime Commission to attend a short interview with Mariana Niga in case her interpretation skills were required.

---

[96] See, eg, *R v Miladinovic* (1992) 109 ACTR 11, affd *Miladinovic v The Queen* (1993) 47 FCR 190. See also the reference to the need for caution in *R v Makin* (1995) 120 FLR 9, 13–14 [20]–[21] (Crockett, Southwell and Vincent JJ), even though all parties agreed that no instruc-tions were required in this case.

[97] See Gary Edmond and Andrew Roberts, 'Procedural Fairness, the Criminal Trial and Forensic Science and Medicine' (2011) 33 *Sydney Law Review* (forthcoming).

[98] (2000) 52 NSWLR 457.

[99] [2003] NSWCCA 6 (6 February 2003).

[100] Ibid [7] (Heydon JA). Thus Kandic was a displaced listener and Kandic's opinion evidence was obtained in circumstances which bear many of the hallmarks of the 'ad hoc expert' cases, though in this case her *initial* exposure to the voice of the accused was in person. There is a suggestion that, while most of the tapes were translated days or months after they were made, at some point Kandic may also have been listening to the calls in question in 'real time'. In this respect it may be that the NSWCCA was treating her as an 'earwitness' to the events in question. Heydon JA, in pointing out that s 116 applies to voice identification evidence, and that in this case the warnings did not express the special need for caution mandated in s 116, did not engage directly with the difference between earwitnesses and displaced listeners: at [38], [61].

That interview, lasting approximately 30 minutes, during which Niga spoke for 15 to 20 minutes, proceeded in English. During her examination-in-chief, Kandic testified that based on her presence at the 1993 interview, she had 'recognised' one of the voices on the 1994 tapes as belonging to Mariana Niga. However, as the trial progressed, the defence requested that a voir dire be held in relation to that 'identification' and during the voir dire it became apparent that it was only in 2001, while talking to the Crown prosecutor just before Niga's trial was about to commence, that Kandic had identified the voice on the tapes as that of the woman she had observed being interviewed in English at the Crime Commission in 1993.[101] This was the first time Kandic disclosed to the prosecution that she believed the voice on the tape belonged to Niga. After a lengthy voir dire, in which the defence argued that her evidence ought to be excluded under s 137, the incriminating opinion evidence of Kandic, linking the voices on the tape to the person she had seen being interviewed in 1993, was admitted at trial.[102]

On appeal counsel for Niga advanced a range of reasons why the voice identification by Kandic ought to have been excluded. While Kandic claimed that the voice she heard both at the 1993 interview and on the tapes was 'a very specific voice', she testified that she recalled no unusual or distinctive features in the voice from the interview.[103] She had, however, been told by the investigating police that they believed the voice on the surveillance tapes was the woman (Niga) she had seen interviewed in English at the Crime Commission and that the recordings she transcribed in 1994 were from Niga's phone. The implication is that she had this information at the time she was asked to transcribe the tapes in 1994, and certainly before she disclosed the identification to the Crown prosecutor in 2001. At trial, Kandic also conceded that she had relied on the presence of the Christian name 'Mariana' on the tapes in coming to her conclusion about the identity of the speaker. Despite the long delay between hearing the voice and making the identification, and the fact that she could not recall any other specific details from the 1993 interview, she testified that her memory never failed her and was unwilling to acknowledge the possibility of error.[104] Finally, it was not until a week before the trial in 2001, in the circumstances described above, that Kandic disclosed that she 'recognised' the voice on the tape as that of Niga. It was in this context that Kandic was permitted to positively identify Niga as the voice of 'Mariana' on the covert recordings.

Remarkably, in a prosecution and appeal where the admissibility of the positive identification of Niga's voice was robustly contested, the NSWCCA (Heydon JA, Hulme J and Carruthers AJ agreeing) does not provide a clear explanation as to the basis for the admissibility of Kandic's evidence. There is no

---

[101] Ibid [18].
[102] Ibid [24]. In his ruling, over the objection of the defence, the trial judge not only envisaged that Kandic would give evidence, but also that the jury would compare tapes, where the speaker identifies herself as Mariana, with the other contested recordings: at [24].
[103] Ibid [27], [54].
[104] Ibid [18], [21], [42], [59]. On the voir dire, Kandic claimed that the memory came to her 'like a flash of light' as she was talking to the Crown Prosecutor: at [18]. However, she conceded that she had been told the name of the accused on a number of occasions: at [21].

discussion of the fact that Kandic was expressing opinions about identity that were not based on her 'specialised knowledge' as an interpreter. The relevance and, more problematically, the admissibility of her opinion evidence appear to have been taken for granted.

The trial judge and the NSWCCA thought that Kandic's voice identification evidence was properly admitted, the NSWCCA confirming that as long as the voice identification was relevant it was admissible unless excluded under ss 135, 137 or 138,[105] and rejecting the defence argument that that the significant problems in the way that the evidence was obtained triggered s 137.[106] For the NSWCCA, the main problem was that the trial judge had not adequately warned the jury about the particular dangers of the voice identification evidence according to s 165 of the *Evidence Act 1995* (NSW) — specifically the cross-lingual nature of the comparison — nor had the trial judge pointed to the *special* need for caution as required by s 116.[107] Despite some obvious dangers and inadequate warnings, in what was characterised as a compelling circumstantial case, the NSWCCA thought Kandic's identification evidence was properly admitted and, applying the proviso,[108] dismissed the appeal. The acknowledged inadequacy of the warnings was insufficient to overturn the conviction.

A similar approach was adopted in *R v El-Kheir*[109] where, once again, the NSWCCA did not concern itself with the admissibility of the translator's opinion evidence about the identity of speakers in a residence subject to covert surveillance, notwithstanding that:

- the sound recording was 'very poor' (rated at 2 on a scale from 0 to 10);

- the translator's level of confidence about who spoke the allegedly incriminating words was at the level of chance;

- there was considerable background noise;

- there were 'extended breaks where nothing could be heard';

- 'words could be heard but not understood';

- 'bits and pieces [were] missing'; and

- 'at times there was insufficient detail in the quality of the soundtrack to form a definite opinion as to who was speaking to whom'.[110]

---

[105] Ibid [34] (Heydon JA, Hulme J and Carruthers AJ agreeing).

[106] Ibid [60].

[107] Ibid [61]. Notwithstanding *Riscuta* and other cases such as *R v Camilleri* (2001) 127 A Crim R 290, s 116 of the UEA would not appear to apply to *displaced* (or indirect) voice identification evidence. See the definition of 'identification evidence' at above n 2.

[108] *Criminal Appeal Act 1912* (NSW) s 6(1).

[109] [2004] NSWCCA 461 (20 December 2004).

[110] Ibid [97], [103] (Tobias JA). The clearest parts of the recording (apparently) enabled the interpreter to distinguish between the respective abilities in Arabic of the two speakers; nevertheless, 'the quality of the utterances and terms of the recording were poor and … at times the language was such as to be either inaudible or indecipherable. At times there was corruption in the phonemic structure of the speech that made it difficult to understand': at [98].

In the aftermath of the surveillance operation, the translator, Dr Gamal, listened to the recordings 'again and again and again' in order to prepare a transcript and identify the speakers.[111] In relation to one of the allegedly incriminating statements, he testified that it could have only been one of two male voices. He 'accepted that there was a 50% chance that the statement he attributed to M2 [identified as El-Kheir] was attributable to M1', but was 'adamant that either M1 or M2 … made the statement.'[112]

Referring to *Li v The Queen* ('*Li*') (discussed below), the NSWCCA (Tobias JA, Hoeben J and Smart AJ agreeing) decreed that 'the admission of voice identification evidence was a matter for judicial discretion'.[113] Without troubling itself with the exclusionary opinion rule and the exception for 'specialised knowledge', the NSWCCA upheld the admission of the positive identification evidence from Dr Gamal where there were real doubts about its independence,[114] probative value and — in circumstances where only one of a few persons in the house could have uttered the allegedly incriminating words — necessity.[115]

The case of *R v Madigan* ('*Madigan*')[116] affirms this general trend while throwing the emerging contrast between the latitude afforded to the ('ad hoc expert') opinions of investigators and the restrictions placed on more conventional experts — particularly experts called by the defence (after *Gilmore*)[117] — into sharp relief. In *Madigan* the investigating police officers spent a total of 'maybe 50 hours, maybe more' listening to covert recordings and producing transcripts.[118] They 'replayed some tracks up to 20 times in an attempt to make out the words.'[119] One officer had interacted with Madigan several years earlier, and the other had had very limited exposure — some 2–3 minutes during fingerprinting and a police interview in which Madigan said very little.[120] On the basis of their repeated listening to the covert voice recordings they were allowed to give positive voice identification testimony.

Wood CJ at CL (Grove J and Hoeben J agreeing) concluded, on the basis that the accused and others had identified themselves — using nicknames and Christian names — in incriminating recordings from their phones, that there was

---

[111] Ibid [100].
[112] Ibid [103]. It seems Dr Gamal was told by investigating police that there were only two adult men in the house at the time of the recordings: at [103], [109].
[113] Ibid [96].
[114] See the discussion of contextual bias below in Part VI(C).
[115] We accept that these issues might not have been raised on appeal by the lawyers, but they are undoubtedly front and centre.
[116] [2005] NSWCCA 170 (9 June 2005).
[117] [1977] 2 NSWLR 935. See the discussion in the text accompanying above nn 70–78.
[118] [2005] NSWCCA 170 (9 June 2005) [21] (Wood CJ at CL).
[119] Ibid. Cf *R v Bain* [2010] 1 NZLR 1, where it was four different experts (three forensic consultants and a linguist), rather than the investigating police officers, who compiled the transcripts. In *Madigan*, the levels of exposure, apart from through listening to the tapes, seem to have been more limited than the interactions between the police officers and the accused in *Smith* (2001) 206 CLR 650, although we acknowledge that in *Madigan* the investigating police officers appear to have listened to a good deal of recorded material.
[120] *Madigan* [2005] NSWCCA 170 (9 June 2005) [22], [25] (Wood CJ at CL).

little risk that the jury might misuse or improperly value the positive identification evidence of the investigating police officers.[121] This merely raises the question of why these incriminating opinions were considered necessary or relevant (following the majority in *Smith*) in the first place.

Perhaps the most striking aspect of *Madigan*, however, was the exclusion of testimony from an expert witness called by the defence.[122] Madigan sought to adduce the testimony of a linguist (Ms Elliot) to describe alternative, and apparently more rigorous, approaches to voice comparison.[123] According to the NSWCCA:

> It does not however follow that the defence should have been permitted to call Ms Elliot to give her expert opinion on the 'methodology'. All that she was able to offer was to describe an approach to voice identification that differed from the method of identification by a person who had the opportunity of listening to the tapes and having some familiarity with the voices of the speakers, either as direct evidence or as ad hoc expert evidence, which has been accepted by the courts …
>
> She had not undertaken any acoustic analysis herself and was not in a position to offer an opinion as to whether the speakers were the Appellant, Woods and Ms Walker. …
>
> The defining point for the rejection of her evidence was that it did no more than identify an alternative method of voice identification that was dependent upon acoustic analysis, without placing in issue that which was led by the Crown.[124]

Challenging, directly or implicitly, the approach and 'expertise' of the investigating police officers was not enough. To the extent that the defence were able to point to the existence of qualified experts who could testify about scientific methods and, most importantly, about notorious problems, this response seems difficult to reconcile with principle, particularly the aim of doing justice in the pursuit of truth.[125]

---

[121] Ibid [98]. In *R v Jones* (1989) 41 A Crim R 1, the voice identification evidence of a builder who had carried out repairs for the accused was offered in conjunction with circumstantial evidence of the telephone intercept on the house occupied by the accused. See also *R v Watson* [1999] NSWCCA 417 (21 December 1999); *R v Ryan* (1984) 55 ALR 408, 412–13 (Street CJ).

[122] A more generous approach to evidence adduced by the accused, exemplified in *Gilmore*, seems to have been eroded in recent decades.

[123] *Madigan* [2005] NSWCCA 170 (9 June 2005) [102]–[103] (Wood CJ at CL). Somewhat ironically, given the basis for exclusion, the proposed rebuttal evidence may actually have been evidence of fact (or the basis for an opinion): the *description* of notorious difficulties with voice identification and standardised scientific techniques might be considered as evidence of fact(s) rather than opinion. Moreover, it would certainly appear to be relevant to the facts in issue and the only grounds for discretionary exclusion would seem to be that it would cause or result in undue waste of time: UEA s 135(c).

[124] *Madigan* [2005] NSWCCA 170 (9 June 2005) [107]–[109] (Wood CJ at CL, Grove J and Hoeben J agreeing). See also *Sook v Minister for Immigration and Multicultural Affairs* (1999) 86 FCR 584, 602 [43] (Moore J). The cases that support the admission of incriminating opinions by 'ad hoc experts' are discussed below.

[125] See generally H L Ho, *A Philosophy of Evidence Law: Justice in the Search for Truth* (Oxford University Press, 2008).

Other cases reinforce these trends. In *R v Camilleri*,[126] a police officer was allowed to positively identify the voice on covert recordings obtained via a listening device on the basis of a few words exchanged during the execution of a search warrant and a formal interview where the defendant refused to answer any questions. According to the NSWCCA:

> The fact that the police officer had such limited familiarity with the voice and the fact that he was told in advance that it was the accused'[s] voice on the tapes which he was asked to identify, did not mean that the evidence should not have been admitted.[127]

The appeal focused on the adequacy of the warning without any consideration of the admissibility or probative value of the incriminating *opinion*.

In *Irani v The Queen*,[128] a decision rejecting a s 137 challenge to the admissibility of a police voice identification, Hoeben J rehearsed all of the cases discussed in this Part in light of a defence concession that the police officer making the positive identification was qualified as an 'ad hoc expert'.[129] Consequently, the police officer's opinion about a voice recorded by a police informant in a nightclub was admitted even though the police officer had no familiarity with the accused's voice and was told who spoke the incriminating words by the police informant (who had indemnity from prosecution). In addition, the informant was with the police officer during the preparation of the transcripts and the positive 'identification'. The NSWCCA accepted that the opinion evidence was admissible and that any prejudicial effects (such as the appearance of independent corroboration) could be cured by clear directions to the jury and were outweighed by the probative value of the evidence.[130]

In *Dodds v The Queen*,[131] a police officer with limited exposure to the accused's voice was allowed to express an opinion about identity even though a co-accused with considerable familiarity identified Dodds as the speaker on a number of intercepted phone calls and some of the information on those calls fitted neatly with the peculiar life circumstances of the accused, dramatically reducing the need for speculative opinion evidence. The prosecution's failure to call an appropriate expert or undertake scientific comparisons was (apparently) rejected as a ground of appeal by the NSWCCA. Without addressing the issue in detail, McClellan CJ at CL seemed satisfied that the jury had been alerted to the

---

[126] (2001) 127 A Crim R 290.

[127] This extract is Hoeben J's description of *Camilleri* in *Irani v The Queen* (2008) 188 A Crim R 125, 130 [21]. It is unclear, in the absence of recordings, just how the jury is to fairly assess this evidence, especially if there are pervasive beliefs that police have special sensory prowess because of training and experience.

[128] (2008) 188 A Crim R 125.

[129] Ibid 129–130 [19]–[24]. Interestingly, Hoeben J at 132 [31] supported the trial judge's references to *R v Menzies* [1982] 1 NZLR 40, 49 (Cooke J for Cooke, McMullin and Somers JJ and Sir Clifford Richmond) and *Butera v DPP (Vic)* (1987) 164 CLR 180, even though these cases primarily involved the preparation of transcripts rather than voice comparison and identification.

[130] *Irani v The Queen* (2008) 188 A Crim R 125, 132 [32] (Hoeben J, McClellan CJ at CL and Harrison J agreeing).

[131] (2009) 194 A Crim R 408.

fact that the police officer had 'accepted that there was always room for error in voice comparison.'[132]

There is, evidently, confidence in the ability of police officers and interpreters to provide probative testimony on the issue of identity derived from exposure to voice recordings. In New South Wales, at least, there is an obvious preference for admission and a tendency to underestimate the risks and dangers associated with error and contamination. Overall, the cases discussed above demonstrate that neither concerns about process, nor uncertainty as to the principled basis for admission, are sufficient to temper the enthusiasm for incriminating voice evidence.

IV Cross-Racial and Cross-Lingual Comparisons by Displaced Listeners

A recurring feature in many of the voice identification cases (such as *Riscuta*) is the reliance on opinions based on cross-lingual comparisons and the reluctance of the courts to exercise any form of control, discretionary or otherwise, over the admission of this evidence.[133] This runs parallel to the general reluctance to consider, in a systematic way, the different methods that might be used to make the process of cross-cultural comparisons more reliable. In Part V, we consider how the disinclination to impose restrictions on the admission of opinions about identity is mirrored where the task of cross-lingual voice comparison and identification is left to the jury. Here we focus on the use of displaced witnesses purporting to assist the tribunal of fact to ascertain the identity of incriminating voices speaking foreign languages.

The evidence challenged on appeal in *R v Leung*[134] included the testimony of an accredited interpreter, Mr Fung, working with the Australian Federal Police. Fung was given a series of covert recordings of conversations in Cantonese, Mandarin and a third dialect, possibly Shanghainese.[135] These were described as 'the DAT tapes'. He translated the recorded conversations into English and in so doing isolated three different speakers, designated as 'M1', 'M2' and 'M3'. These transcripts were produced in November and December of 1997. In August of 1998, just before the trial, Fung was asked to listen to a number of brief recordings of different conversations between Leung and police officers and Wong and police officers ('the police tapes'). Fung was then asked to compare the voices recorded on the police tapes with the voices recorded on the DAT tapes and to give his opinion as to the identity of the speakers on the DAT tapes.[136] The majority of the conversations on the police tapes involving Leung

---

132 Ibid 432 [92].

133 A similar trend is apparent in visual identification cases, many of which allegedly involve cross-racial identifications: see the discussion in Edmond et al, 'Law's Looking Glass', above n 10.

134 (1999) 47 NSWLR 405.

135 The difficulty in even identifying the language (or dialect) indicates some of the underlying problems with translation and semantics (and sound quality), let alone identification: see Good, above n 83; Holden, above n 83.

136 *R v Leung* (1999) 47 NSWLR 405, 409–10 [18]–[19] (Simpson J). In other cases, trial judges have limited police investigators to characterising a voice as the same as another (usually un-

were conducted in Cantonese. The conversations on the police tapes with Wong were in English. Fung expressed the opinion, later repeated in evidence, that the speakers he had identified as M1 and M3 were, respectively, Leung and Wong.[137]

Significantly, there was some debate at trial as to the admissibility of this opinion evidence. It was conceded that the interpreter's opinion did not derive from 'specialised knowledge based on … training, study or experience'.[138] Fung 'volunteered', during cross-examination, 'that he was not a voice expert, but said that he had done his best to identify the voices.'[139] The trial judge referred to a number of common law cases concerned with voice identification, most promi-nently *Bulejcik*,[140] but concluded that s 78 of the UEA provided an admissibility pathway for Fung's opinion.[141] Notwithstanding the concession made at trial, on appeal the Crown resiled, arguing that Fung's incriminating identification evidence was admissible, despite his lack of formal qualifications and training in voice identification, because he 'fell into the category of "ad hoc expert"' as recognised and developed through the common law.[142]

The NSWCCA, in some detail, acknowledged the constraints under which Fung performed the task of voice comparison and identification. These included the brevity of the police tapes;[143] the very different circumstances in which the DAT and police tapes had been obtained; the fact that for all of the Wong tapes and at least one of the Leung tapes the comparison was made between different languages;[144] and Fung's concession that describing the characteristics of voices, as a layperson, is difficult and different to recognising a familiar voice.[145] For the Court, however, these limitations went to the weight of the evidence rather than the admissibility of Fung's ('ad hoc expert') opinion.

---

known) voice, without actually identifying the speaker. Identification, or perhaps more accu-rately differentiation, of speakers is often an implicit component of transcript preparation: see, eg, *R v Solomon* (2005) 92 SASR 331, 337 (Doyle CJ); *Dodds v The Queen* (2009) 194 A Crim R 408, 417–19 (McClellan CJ at CL).

[137] *R v Leung* (1999) 47 NSWLR 405, 410 [19] (Simpson J).

[138] UEA s 79. There was no challenge to Fung's ability, as a qualified *interpreter*, to prepare a transcript from the DAT tapes.

[139] *R v Leung* (1999) 47 NSWLR 405, 410 [21] (Simpson J).

[140] (1995) 185 CLR 375.

[141] *R v Leung* (1999) 47 NSWLR 405, 410 [23] (Simpson J). Recourse to s 78 is, in this context, somewhat anomalous, and on appeal it was decided by Simpson J (Spigelman CJ and Sperling J reserving their opinions) that s 78 was not an appropriate basis for admission: at 412 [34]–[35].

[142] Ibid 412 [31].

[143] See ibid 408 [8], 413 [42].

[144] Ibid 413 [42].

[145] Ibid 410 [21]. Simpson J also points out that when Fung was asked to make the comparison he would have 'approached his task on the assumption that the two voices on the police tapes were in fact the same as two of the voices on the DAT tapes' and that in situations where the identity of the speakers on the tapes remained open there might be 'real questions of propriety' in relation to identifications made under such circumstances: at 414 [45]. This argument is taken up in *Li* (2003) 139 A Crim R 281, where the appellant argued that the translator's identification was tainted because he knew, when handed the police interview tape, that Li was already a suspect. However, the NSWCCA Court rejected this argument, in part because of what was perceived to be the practical difficulty of setting up a voice 'line-up' (or parade), but primarily because analo-gising between visual and voice identification was considered inapposite: at 289 [60] (Ipp JA).

In *Li*,[146] cross-lingual voice comparison and identification evidence was prof-fered by an interpreter (Stephen Chan), a police officer (Sergeant Lee) and a senior lecturer in linguistics from the University of Sydney (Dr Gibbons). Each had been asked to express an opinion as to whether a person speaking Cantonese on a surveillance tape (referred to as 'tape 6') was the voice of the appellant. Tape 6 recorded one side of an incriminating telephone conversation. The defence argued that the opinions of Chan, Lee and Gibbons purporting to identify the voice on the tape as that of the appellant should not have been admitted and, further, that the trial judge had not given an adequate warning about the dangers of voice identification and voice similarity evidence.[147]

In 1998 Chan was provided with a number of surveillance tapes which in-cluded tape 6. He was asked to transcribe and translate the contents of these tapes, which included more than one voice and were primarily in Cantonese.[148] He designated one of the voices on tape 6 as 'M1' and gave his opinion that the voice of M1 appeared on all five of the tapes supplied to him.[149] About a year later Chan was asked to listen to part of the audio recording of the appellant's police interview, apparently conducted in English, and to give his opinion as to whether the voice he had identified as M1 was that of the appellant. He listened to the original tapes but 'conceded that it might have only been once.'[150] Chan then identified M1 as Li. The trial judge concluded that Chan's opinion about the identity of the speakers was relevant and admissible.[151]

The appellant identified 10 problems with Chan's evidence. They included that Chan 'was not a voice recognition expert'[152] and gave 'an ordinary man's opinion' as to the similarity between the voices on the tapes.[153] The combined effect of these (and other) weaknesses, the defence argued, meant that the identification evidence ought to have been excluded via s 137 of the *Evidence Act 1995* (NSW) because its probative value was outweighed by the danger of unfair prejudice to the accused. The appellant also argued, following *Smith*,[154]

---

[146] (2003) 139 A Crim R 281.

[147] There is some slippage in the language used to describe the type of evidence given by these different witnesses and the judgment seems to refer to 'voice identification' and 'voice similarity' evidence interchangeably. The voice evidence is initially referred to as 'voice similarity opinion evidence', though it is clear that the evidence goes beyond evidence of similarity and in fact purports to make a positive identification of the appellant's voice: see, eg, ibid 284 [18] (Ipp JA).

[148] Chan listened to the tapes numerous times and isolated a number of different speakers. Here the issue of identification or, perhaps more accurately, differentiation raises its head.

[149] *Li* (2003) 139 A Crim R 281, 285 [32] (Ipp JA).

[150] Ibid 286 [36].

[151] Ibid 286 [37].

[152] Ibid 287 [45].

[153] Ibid 288 [45]. Other problems with Chan's evidence raised by the appellant were: that he 'would not say there were any special features of the voice'; that he agreed that 'people speaking on a telephone have a different type of speech from people speaking face to face'; and that he had 'no training, knowledge or experience in comparing voices speaking in English and those speaking in Cantonese'.

[154] (2001) 206 CLR 650.

that the comparison was one that could have been conducted by the jury and was thus irrelevant.[155]

Ipp JA (Whealy J and Howie J agreeing), however, held that the evidence was relevant. He did not accept that the combined effect of these weaknesses meant that the evidence ought to have been excluded. Weaknesses in Chan's incriminating opinion evidence were characterised as issues for the jury. In particular, Ipp JA was not persuaded that there were fundamental problems with Chan comparing voices speaking Cantonese with a voice speaking English. He saw 'no reason why the cross-lingual element in the comparison that Mr Chan was required to undertake detracted significantly from his ability to express a *reliable* opinion.'[156]

The arguments rehearsed in relation to Chan were extended to cover the opinions of the two other witnesses who also — though perhaps not independently — identified the voice on tape 6 as that of Li. Sergeant Lee, a police officer fluent in Cantonese and English, and familiar with Mandarin, with some experience in Cantonese to English and English to Cantonese translation, first heard the incriminating speech via audio surveillance. Lee transcribed and translated a tape of what had been spoken. He subsequently listened to two other tapes which contained short passages of the appellant speaking in both Mandarin and Cantonese, had access to the incriminating conversation from tape 6, and reached the conclusion that the voice on tape 6 was that of Li.[157] The defence raised concerns about Lee's evidence, identifying limitations with the samples, the possibility of bias, and the lack of specific training or experience in voice identification and cross-lingual comparisons.[158] Once again the Court considered that these issues went to weight and as such were matters for the jury.[159]

The third prosecution witness, Dr Gibbons, listened to the audio recording of the police interview with the accused (this became his 'base' tape). Dr Gibbons identified a number of specific characteristics of the accused's voice on the base tape, and then compared the base tape (where the voices were speaking in English) with the surveillance tapes, including tape 6 (where the voices were speaking both Mandarin and Cantonese). He identified the voice on tape 6 as that of Li, based on 'general voice properties' as well as the presence of several apparently distinctive characteristics.[160] In cross-examination, Dr Gibbons conceded that he had no specific expertise in either Cantonese or Mandarin, and

---

[155] *Smith* is discussed above in Part II. Here, the invocation of *Smith* appears to be tactical, drawing on tensions in appellate authority rather than on principle or scientific research.

[156] *Li* (2003) 139 A Crim R 281, 289 [56] (emphasis added).

[157] Ibid 290 [65]–[69]. There is no indication of the number of times that Lee had listened to any of these tapes, nor how long he had spent transcribing and translating the original conversation.

[158] Ibid 290–1 [70]. See also *R v Gao* [2003] NSWCCA 390 (16 December 2003) [20]–[24] (Greg James J, Sully and Adams JJ agreeing), where the NSWCCA upheld the admissibility of an opinion from an interpreter that the voice he heard during a very brief police interview — where the accused indicated (in English) that he would not answer any questions — was the same voice he had heard during telephone interceptions of Cantonese speakers.

[159] Ibid 291 [71]. Drawing upon civil justice authority, Ipp JA explained that the 'risk of bias (unconscious or otherwise) is no reason not to admit evidence of an expert'. See also *R v Galea* (2004) 148 A Crim R 220, 241–2 [135]–[144] (Ipp JA).

[160] *Li* (2003) 139 A Crim R 281, 291 [74]–[75] (Ipp JA).

that he was not an expert in cross-lingual comparisons between English and those languages. He also conceded that he had no statistical information about the frequency and distribution, amongst Cantonese speakers, of the 'distinctive' features that he had identified.[161] Indicating that the opinion evidence of Dr Gibbons was properly admitted, once again Ipp JA explained that such problems went merely to the weight of the evidence and that Dr Gibbons was properly qualified to give expert opinion evidence positively identifying the voice of the accused on the relevant tapes. Overall, Ipp JA doubted that weaknesses in the voice identification evidence gave rise to any unfair prejudice to the appellant.[162]

V Cross-Lingual Jury Comparisons

While our primary concern is with the admission of incriminating voice comparison evidence, we want to briefly consider cases where the jury is asked to make voice comparisons instead of, or in addition to, an investigator or other (ad hoc) 'expert'.[163] Cases where the displaced listeners are members of the jury reflect the permissive trends discussed above, and raise their own set of analogous concerns. The appeal in *R v Korgbara* ('*Korgbara*') offers a particularly striking example.[164] This case provides a stark indication of the judicial unwillingness to consider the various methods by which voice comparison could (at least arguably) be conducted more reliably, and the refusal to impose restraints on the admissibility of voice comparison evidence for the purpose of identification.

In *Korgbara*, the Crown relied upon recordings of a number of intercepted telephone calls made to and from a mobile phone that was alleged to belong to the appellant. Apart from one call, in which it was conceded that Korgbara had called the NRMA and spoken in English, all of the recorded conversations were in a Nigerian language called Igbo. Translators were called to give evidence of the content of the intercepted conversations, and the Crown alleged that the appellant was the intended recipient and a party to most of the Igbo calls. It was the Crown's contention that as the receiver of those calls the appellant was revealed to be knowingly concerned in the importation of cocaine. The appellant gave evidence in English and denied speaking in any of the Igbo recordings. There was no verified sample of the appellant speaking Igbo, though the appellant was from Nigeria and did in fact speak Igbo.[165] In the end, the jury were invited to make their own comparison between the defendant's voice on the tape in the NRMA call and the other Igbo calls, and between the defendant speaking in court and the recorded voice of the receiver of the relevant Igbo

---

[161] Ibid 292 [77].

[162] Ibid 292 [78].

[163] This approach is endorsed by both John Henry Wigmore and Rupert Cross: see Twining, *Rethinking Evidence*, above n 5, ch 5. See also the earlier English authority *R v Bentum* (1989) 153 JP 538 and the implicit endorsement of the procedure, by the High Court, in *Bulejcik* (1996) 185 CLR 375.

[164] (2007) 71 NSWLR 187. See also Transcript of Proceedings, *Korgbara v The Queen* [2007] HCATrans 485 (31 August 2007).

[165] *Korgbara* (2007) 71 NSWLR 187, 190 [8] (McColl JA).

calls, with a view to determining whether the recorded voice was the appellant.[166]

On appeal, it was argued that in the absence of expert analysis of the recorded telephone calls, it should not have been left to the jury to make a comparison between a voice speaking English and a voice speaking a foreign language.[167] The appellant's counsel argued that the courts should adopt a cautionary approach and require expert analysis as a prerequisite if a jury is asked to perform this kind of voice comparison task.[168]

McColl JA (James J agreeing) reviewed the Australian and overseas authorities relied upon by the appellant and concluded that it was not possible for the Court to 'establish a prescriptive rule that voice comparison evidence should only be admitted where supported by expert testimony.'[169] For the majority, the absence of controls regulating voice identification evidence in the UEA, in contrast to those regulating the admissibility of visual identification evidence in pt 3.9, meant that there was no intention to place restrictions on voice evidence, even where that evidence involved a cross-lingual comparison.[170] The majority

---

[166] Ibid 191–4 [20]. It was the comparison between the voice recordings that had been initially anticipated by the Crown when seeking to have the calls admitted.

[167] Ibid 194 [21].

[168] Ibid 194–5 [23], [27].

[169] Ibid 207 [74]. See also *R v Smith* (1990) 50 A Crim R 434, 453–4 (Young CJ, Crockett and Southwell JJ); *Nguyen* (2002) 26 WAR 59, 74 [57], 76 [67] (Malcolm CJ), 89 [134], 90 [138] (Anderson J). In *Nguyen*, Malcom CJ and Anderson J agreed that jurors should be allowed to make cross-lingual comparisons, relying on Brennan CJ's assertion in *Bulejcik* (1996) 185 CLR 375, 381 that recognition of a speaker's voice is 'a commonplace of human experience'. Discussing the jury's comparison of telephone recordings of spoken Vietnamese and the accused speaking in English in the context of jury warnings, Anderson J wrote (at 89 [134]): 'I think it would have been inappropriate for the jury to be warned of the dangers which arise from weaknesses in "human perception and recollection"', and (at 90 [138]):

> I cannot accept the submission that the jury should have been warned not to embark upon a process of comparison themselves. I see no reason why the jury are not entitled to compare voice recordings in order to come to their own conclusions. Voice recognition is not, of itself, an expert process.

In *Nguyen* there was also incriminating opinion evidence from an interpreter who had translated intercepts from the accused's mobile phone every day for two months. *Nguyen* was endorsed in *Neville* [2004] WASCA 62 (2 April 2004) [41], [66]–[68] (Miller J), [101]–[102] (Heenan J), where the jury's entitlement to make voice comparisons was explicitly recognised, and in *Asfoor v The Queen* [2005] WASCA 126 (15 December 2004) [88]–[90] (Templeman J), where a witness identified a familiar person speaking in a foreign language that the witness did not understand. Cf *R v Morgillo* (Unreported, New South Wales Supreme Court, Campbell J, 28 July 1992), where the judge declined to allow a jury to compare voices where there was only 36 minutes of voice recording available. The correctness of *R v Morgillo* was doubted in *R v Bulejcik* (Unreported, New South Wales Court of Criminal Appeal, Hunt CJ at CL, Carruthers and Bruce JJ, 21 July 1994), as noted by the High Court in *Bulejcik* (1996) 185 CLR 375, 396 (Toohey and Gaudron JJ). In *Evans v The Queen* (2006) 164 A Crim R 489 and *Evans v The Queen* (2007) 235 CLR 521, 530 [27] (Gummow and Hayne JJ), 568–9 [178]–[182] (Heydon J), voice 'comparison' seems to have been taken to extremes, with the accused being required to undertake an in-court re-enactment (rather than a demonstration within the meaning of s 53 of the UEA) so that the jury could compare his voice with a sensory witness's description of a voice from an armed robbery.

[170] *Korgbara* (2007) 71 NSWLR 187, 203 [59] (McColl JA, James J agreeing). McColl JA thus endorsed Ipp JA's contentions in *Li* (2003) 139 A Crim R 281, 289–90 [56], [61] that 'the admission of voice identification evidence turns on judicial discretion' and that cross-lingual comparisons can be considered in the same way as comparisons between voices speaking the same lan-

emphasised the discretionary nature of the decision to admit voice comparison evidence, in a manner consistent with the Victorian common law approach to direct witnesses and the UEA cases discussed in the previous Parts. In explaining its decision, the majority used the likelihood of differences of opinions about the best method(s) for conducting voice identifications as a reason for not requiring them.[171] Perversely, judicial suspicion about the absence of standardised methods among professionals is used to require the jury to undertake this formidable (and error-prone) task without assistance. McColl JA concluded that the relevant test, described in the common law decision of *Bulejcik*, is simply 'whether the quality and quantity of the material is sufficient to enable a useful comparison to be made.'[172] The implication is that any restrictions on allowing the jury to engage in such a comparison will, relying on *Bulejcik*, be minimal.[173]

In dissent, Grove J accepted that where the jury is comparing voices speaking in English, the authorities do not support the imposition of a prescriptive rule (for example, a mandatory requirement that the identification must proceed by way of a specific form of acoustic analysis).[174] However, he did not consider imposing restrictions on cross-lingual comparisons as incompatible with the statutory framework of the UEA:

> In my view, permitting the comparison of one language with a different language without suitable material which I would contemplate as evidence of someone either possessing relevant expertise or familiar with the voice of the accused in the language used where identity is challenged (an 'ad hoc' expert) is not to establish a prescriptive rule but, to the contrary, to extend the scope of what is permissible beyond recognised boundaries.

> The general incantation of the admissibility of matters of relevance in s 55 of the *Evidence Act 1995* and the inclusion of 'aurally' as a species of identification evidence defined in the dictionary to that Act does not, in my opinion, establish a statutory scheme governing the admissibility of voice identification evidence without restriction. It is noteworthy that the statute expressly preserves the common law where it is itself relevantly silent: see s 9.[175]

While we do not want to endorse Grove J's recourse to the 'ad hoc expert' as an appropriate mechanism to regulate expert assistance with voice comparison evidence or his implicit support for leaving voice comparison to the jury, his concerns about the difficulties of cross-lingual comparisons are salutary:

> It is self evidently not a commonplace human experience to recognise a speaker's voice in a language other than that which one is otherwise familiar, and familiar in the language in which the person is articulating.

---

guage, thereby further extending the latitude established in *R v Adler* (2000) 52 NSWLR 451, 455 [18] (Smart AJA).
171 *Korgbara* (2007) 71 NSWLR 187, 208 [78] (McColl JA, James J agreeing).
172 Ibid 196 [35], 208 [79], quoting *Bulejcik* (1996) 185 CLR 375, 395 (Toohey and Gaudron JJ).
173 In *Bulejcik* (1996) 185 CLR 375, 395, Toohey and Gaudron JJ noted that '[t]he defence may wish to call expert evidence where the jury may have difficulty in drawing a distinction between two voices of a particular nationality or dialect.'
174 *Korgbara* (2007) 71 NSWLR 187, 209–10 [113].
175 Ibid 210 [113]–[114]. Here Grove J appears to be invoking the tradition associated with *E J Smith* (1986) 7 NSWLR 444.

> In the present case, there was no evidence to describe the nature of communication which is constructed to comprise the Igbo tongue. For all that is known, the language may be constructed, for example, upon variations in tone. It may use sound production techniques which are entirely divorced from those which constitute the English language. It would be mere guesswork, unless relevantly informed, to assume that human vocal faculties are utilised so as to produce comparable sounds when articulating in English and in Igbo.[176]

Grove J's cautionary response is unusual. Most Australian courts deal with cross-lingual comparisons, including identifications where the witness does not speak the foreign language but claims to be familiar with the person allegedly speaking it, through admission and warnings.[177] Thus, Toohey and Gaudron JJ stated in *Bulejcik*:

> Where the jury is itself asked to make a comparison of voices … very careful directions are called for. It is not irrelevant that in the case of handwriting comparisons, it has been said to be unsafe to leave the matter to the jury without the guidance of an expert. It is unnecessary to go that far in the case of a voice comparison but, in our view, 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.[178]

Cross-lingual comparisons are routinely facilitated and judges purport to recognise the dangers inherent in leaving voice comparison to the jury.

Regardless of whether comparisons are undertaken by lay witnesses, purported experts or even juries, trial and appellate judges have been resistant to the exclusion of this evidence on the basis of the mandatory and discretionary exclusions — that is, on the basis that the unknown but often questionable probative value of the evidence is outweighed by the very real danger that the jury will overvalue the evidence or make a mistake, especially where the accused speaks the impugned language.[179] Judges seem to be remarkably confident in the adversarial trial, its safeguards, and the ability of lay fact-finders to appreciate the significance of the dangers even though they are rarely mentioned, and almost never explained in any detail, during the course of trials and appeals.

Cross-lingual comparisons seem to be symptomatic of an unprincipled and empirically indifferent approach to admissibility, reliability, and decision-making by investigators, prosecutors, judges and, in consequence, juries. In the following

---

[176] *Korgbara* (2007) 71 NSWLR 187, 210 [118]–[119]. The phrase 'commonplace of human experience' refers to a statement by Brennan CJ in *Bulejcik* (1996) 185 CLR 375, 381, where the recorded voices were not cross-lingual but accented.

[177] See, eg, *Asfoor v The Queen* [2005] WASCA 126 (15 December 2004) [84] (Templeman J).

[178] (1996) 185 CLR 375, 398–9 (citations omitted). See also *R v Solomon* (2005) 92 SASR 331, 349 [66] (Doyle CJ); *R v Mouhalos* (1998) 197 LSJS 483, 489 (Doyle CJ). It is worth noting that in early fingerprint cases, photographs of latent prints and reference fingerprints were provided to the jury, although more recent cases insist that it is latent fingerprint examiners who should undertake the comparisons: *R v Lawless* [1974] VR 398, 423 (Winneke CJ, Gowans and Kaye JJ); see also *Bennett v Police* [2005] SASC 167 (4 May 2005) [52]–[56] (Doyle CJ).

[179] See UEA s 137. See also s 135, which gives the court discretion to refuse to admit evidence where its probative value is substantially outweighed by the danger that the evidence might be unfairly prejudicial, misleading or confusing, or an undue waste of time.

Parts we consider scientific research on voice comparison as well as the effectiveness of the adversarial trial and its safeguards in dealing with identification evidence.

VI Scientific Research: Human Voice 'Identification' beyond the Courts

In this Part, we provide an overview of research relevant to the reception and assessment of voice comparison and identification evidence that, we argue, should inform the decisions made by courts and prosecutors about voice identification evidence more broadly, and the decisions about opinion evidence proffered by 'experts' more specifically. The failure to take seriously the problem of investigative bias, the courts' over-reliance on the use of directions, and the inadequacy of traditional adversarial safeguards such as the use of defence experts or cross-examination, mean that the courts should be looking to alternative mechanisms to control the admission of this evidence. One alternative is to include the use of validated forensic voice comparison methods and associated probabilistic evidence; another is to use voice identification parades combined with a more rigorous approach to assessing the reliability and thus the admissibility of voice identification evidence generally.

A *Introduction and Some Conceptual Clarification*

Initially, we should address some of the conceptual confusion that attends the reception of this evidence in criminal trials. 'Voice comparison' and 'voice identification' may be practically and conceptually distinct tasks. Some voice identifications are based on comparisons while others are based on *recognition* or *recollection*. Comparison is a deliberative process, while recognition often refers to identifications that are instantaneous. Recollection would seem to comprise a subgroup of recognition (usually, though not invariably, at the deliberative end). Voice recognition may be distinct from voice comparison where it does not involve conscious deliberation or interpretation. Unfortunately, Australian courts have used these and other terms loosely and sometimes interchangeably.[180] It is probably too late in the day, and analytically too cumbersome, to try to clearly and definitively define these terms for forensic purposes. Rather than focusing on pedantic definitions, the more important point is to appreciate how extant research illuminates the frailties of investigative and legal responses to voice evidence, *however characterised*.

It is, nevertheless, useful to distinguish 'scientific voice comparison' (or technical speaker identification) from 'naive speaker identification' (whether based on comparison *or* recognition). Scientific voice comparison, as the name implies, involves comparison and technical analysis, almost always by those unfamiliar with the voices and possible speakers. Features and characteristics of two or more voices are compared in order to determine whether there is sufficient

---

[180] See, eg, *R v Leung* (1999) 47 NSWLR 405; *Li* (2003) 139 A Crim R 281.

similarity or dissimilarity to determine the likelihood that a source (eg perpetrator) and a target (eg suspect) utterance shared the same origin.[181] The plasticity of the speech organs and language[182] means that no two utterances by the same person will ever be identical, or necessarily distinct from the utterances made by another individual.[183] Thus, any comparison between two speech samples can only be probabilistic, rather than categorical; that is, it can indicate that the source of the utterances is *likely* the same or *likely* different, but not that the source *is* the same or *is* different.[184] In order for a valid and reliable voice comparison of two utterances to be made, it is first necessary to identify and measure the features present in the sample that are likely to be useful for discriminating between the origins of the utterances. Secondly, it is necessary to calculate the likelihood that two voices will share a certain proportion of these characteristics, distinctive or otherwise, by chance alone. Ignorance about the frequency of features and their interrelationships among the relevant populations may result in mistaking reasonably common voice characteristics or speech habits for powerful discriminating evidence.[185] Conversely, information about the frequency of voice characteristics and features may produce highly probative, if necessarily probabilistic, evidence.[186] The issues and challenges associated with scientific voice comparison are considered briefly below in Part VIII(C). Because most of the testimony of displaced listeners involves naive speaker identification, the remainder of this Part is oriented in that direction.

Naive speaker identification, which is simply lay voice identification that incorporates both comparison and recognition evidence, relies on no such informed decision-making or analytical process. It is based entirely on human perceptual capacities and limitations (such as encoding, storage and retrieval) and contextual factors (such as familiarity and levels of exposure).[187]

---

[181] This is discussed briefly below in Part VIII(D).

[182] See generally Francis Nolan, *The Phonetic Bases of Speaker Recognition* (Cambridge University Press, 1983).

[183] Richard Hammersley and J Don Read, 'Voice Identification by Humans and Computers' in Siegried Ludwig Sporer, Roy S Malpass and Guenter Koehnken (eds), *Psychological Issues in Eyewitness Identification* (Lawrence Erlbaum Associates, 1996) 117; Francis Nolan, 'Speaker Identification Evidence: Its Forms, Limitations, and Roles' (Paper presented at the Conference on Law and Language: Prospect and Retrospect, University of Lapland, Finland, 12–15 December 2001).

[184] However, where the likelihood is high some analysts may be willing to make categorical calls. In contrast, naive speaker identification (and comparison) routinely involves categorical calls about individualisation.

[185] See the general comments by Commissioner Shannon in South Australia, Royal Commission of Inquiry in Respect to the Case of Edward Charles Splatt, *Report* (1984) 39.

[186] Such evidence will be produced to the extent that features can be stabilised to result in a DNA-like analysis and probabilistic expression. See generally Philip Rose, *Forensic Speaker Identification* (Taylor & Francis, 2002); Joaquin Gonzalez-Rodriguez et al, 'Emulating DNA: Rigorous Quantification of Evidential Weight in Transparent and Testable Forensic Speaker Recognition' (2007) 15 *IEEE Transactions on Audio, Speech, and Language Processing* 2104.

[187] Nolan, 'Speaker Identification Evidence', above n 183. Consider the facts in *R v Morris* (1996) 88 A Crim R 297, where an inaccurate newspaper report had a displacement effect on the recollection of the instructing solicitor and others present of what was said during the summing up.

B *Familiarity*

Just as there is slippage in the use of terminology in relation to voice comparison, identification and recognition, so too is there conceptual confusion regarding the use and interpretation of the words 'familiar' and 'familiarity' in relation to speaker identification.[188] Specifically, there does not appear to be a consistent application of these terms, despite the fact that they are integral to both general earwitness performance and to admissibility determinations in the case of 'experts'. Further, the way in which the terms are used in legal decisions is sometimes at odds with their use in the experimental work on voice comparison.

While 'familiarity' can reasonably be used to describe any point on a continuum of exposure ranging from incidental to in-depth — as demonstrated by the Court in *R v Leung*[189] — in much empirical voice identification literature the term 'familiar' is used to denote a threshold of perception whereby something or someone becomes recognisable or identifiable.[190] A person's voice is considered familiar to an individual when that individual can put a name to that voice, or link that voice to a prior exposure, with a particular level of accuracy. These familiarity-based decisions occur more rapidly than purposeful comparison-based decisions and are best construed categorically — eg 'that voice does, or does not, belong to my mother'.[191] These are the types of displaced voice identification that might more readily fit within the exceptions to exclusionary opinion evidence rules.[192] However, having simply heard a voice before does not necessarily make it familiar within this more precise usage of the term. Indeed many people will not achieve this threshold of familiarity with a voice until they have been exposed to it many times, on many different occasions.[193] Moreover, in the general population, individual differences in ability mean that some people

---

[188] The identification evidence of familiars is conventionally considered to be more reliable than the evidence of strangers: see, eg, the eyewitness case *Ilioski v The Queen* [2006] NSWCCA 164 (10 July 2006) [68]–[70] (Hunt AJA).

[189] (1999) 47 NSWLR 405. See the discussion above in Part IV.

[190] See, eg, Anthony P Weiss et al, 'Distinguishing Familiarity-Based from Source-Based Memory Performance in Patients with Schizophrenia' (2008) 99 *Schizophrenia Research* 208; Kanae Amino and Takayuki Arai, 'Effects of Linguistic Contents on Perceptual Speaker Identification: Comparison of Familiar and Unknown Speaker Identifications' (2009) 30 *Acoustical Science and Technology* 89.

[191] Andrew P Yonelinas and Larry L Jacoby, 'Dissociations of Processes in Recognition Memory: Effects of Interference and of Response Speed' (1994) 48 *Canadian Journal of Experimental Psychology* 516; Douglas L Hintzman, David A Caulton and Daniel J Levitin, 'Retrieval Dynamics in Recognition and List Discrimination: Further Evidence of Separate Processes of Familiarity and Recall' (1998) 26 *Memory & Cognition* 449.

[192] Recent cases involving voice identification evidence of familiars include *Re Dickson* [2008] VSC 516 (26 November 2008) [28]–[29] (Lasry J); *Savic v The Queen* [2008] NSWCCA 312 (16 December 2008) [46] (Allsop P). See also the evidence of familiars in response to images in *R v Murdoch [No 4]* (2005) 195 FLR 421, 431–5 [56]–[81] (Martin (BR) CJ); *Murdoch v The Queen* [2007] NTCCA 1 (10 January 2007) [203]–[245] (Angel ACJ, Riley J and Olsson AJ).

[193] For example, people with phonagnosia, normally acquired through damage to the right cerebral hemisphere, are incapable of recognising or experiencing 'familiarity' with even the voices of their family, despite the fact that these voices are not in any way novel to them: Diana Roupas Van Lancker et al, 'Phonagnosia: A Dissociation between Familiar and Unfamiliar Voices' (1988) 24 *Cortex* 195.

are able to recognise voices (or faces) more quickly and more reliably than others.[194]

The precise threshold for 'familiarity' is difficult to isolate, though a great deal of research has been conducted on human ability to identify the voices of people known to listeners as well as their ability to identify the voices of strangers. The evidence suggests that the identification of voices of family, colleagues, famous people and some acquaintances can be reasonably accurate, even in demanding circumstances.[195] In one influential study an individual was exposed to 29 voice recordings of family members and acquaintances. Identification (ie naming) accuracy of friends and acquaintances was 31 per cent on the basis of the utterance 'hello', 66 per cent based on a single sentence and 83 per cent after a 30 second recording.[196] These findings were broadly replicated for famous voices.[197] Overall, while there is substantial variability in the literature, and for individual listeners, accuracy rates for the recognition of well-known voices are not uncommonly higher than 80 per cent.[198] Experimental evidence also suggests that individuals are able to identify their own voice with around 84 per cent accuracy.[199]

Such high levels of accuracy do not extend to listeners who are attempting to identify (ie compare or recollect) the voices of strangers.[200] In an experiment

---

[194] Richard Russell, Brad Duchaine and Ken Nakayama, 'Super-Recognizers: People with Extraordinary Face Recognition Ability' (2009) 16 *Psychonomic Bulletin & Review* 252; A Schmidt-Nielsen and Karen R Stern, 'Identification of Known Voices as a Function of Familiarity and Narrow-Band Coding' (1985) 77 *Journal of the Acoustical Society of America* 658. It is possible to test abilities, though it may be unethical to test (at least with ecological validity) in some very stressful situations, such as an armed robbery or sexual assault.

[195] Such as where participants are given a very large set of voices from which to make their identification (ie with no priming with regard to who they might hear) and restricted or distorted speech samples (eg single words/sounds, filtered/altered utterances, backward samples and rate-altered voices): Diana Van Lancker, Jody Kreiman and Karen Emmorey, 'Familiar Voice Recognition: Patterns and Parameters — Part I: Recognition of Backward Voices' (1985) 13 *Journal of Phonetics* 19; Diana Van Lancker, Jody Kreiman and Thomas D Wickens, 'Familiar Voice Recognition: Patterns and Parameters — Part II: Recognition of Rate-Altered Voices' (1985) 13 *Journal of Phonetics* 39.

[196] Peter Ladefoged and Jenny Ladefoged, 'The Ability of Listeners to Identify Voices' in *UCLA Working Papers in Phonetics 49* (UCLA Phonetics Laboratory Group, 1980) 43, 48–9 <http://escholarship.org/uc/item/5w14p7x2>.

[197] Van Lancker, Kreiman and Wickens, above n 195.

[198] Daniel Read and Fergus I M Craik, 'Earwitness Identification: Some Influences on Voice Recognition' (1995) 1 *Journal of Experimental Psychology: Applied* 6; A Daniel Yarmey et al, 'Commonsense Beliefs and the Identification of Familiar Voices' (2001) 15 *Applied Cognitive Psychology* 283; Amino and Arai, above n 190.

[199] Schmidt-Nielsen and Stern, above n 194, 662.

[200] The distinction between recognition and discrimination is an important one. A recognition task does not limit listeners to a set of speakers from which they may or may not select a voice. The task is more akin to picking up the telephone and hearing any one of all possible people you know speaking. By contrast, a discrimination task is one where boundaries are enforced for the response set. For example, you may be told that you will hear the voices of your colleagues, or be presented with a fixed number of 'foils' or alternatives from which to select. Importantly, a discrimination task is relatively simpler, as the response options are limited and cognitively less demanding selection processes can be used (eg by comparing your memory of the voice to the others in the set rather than comparing your memory to all other voices you have ever been exposed to, or to all other familiar voices). On the other hand, as is evident from the cases, when

where participants were exposed to either 30 or 70 seconds of a previously unknown voice, listeners were able to correctly identify the voice of a target in 42 per cent of the instances in which it was presented (also known as a 'hit').[201] However, when that voice was not present, listeners identified another previously unheard (or 'innocent') voice as the target voice 51 per cent of the time (a 'false alarm' or false positive). While this disconcerting rate of false alarms has been replicated,[202] substantial variability has also been noted for both false alarms and hit rates where unfamiliar speaker identification has been tested.[203] Overall, the experimental research indicates that familiars tend to be much more accurate than non-familiars, but that even familiars experience a significant rate of error and inaccuracy in the identification of known voices, and results can vary markedly as a result of factors such as health, fatigue, intoxication or emotional state.[204] Those not familiar with a voice tend to have relatively high levels of error when trying to identify that voice, and the accuracy for all listeners is affected by the circumstances and conditions in which any comparison or recollection exercise is undertaken.

C *Factors Affecting Voice Comparison and Recognition*

In the absence of the type of familiarity that is gained through repeated and variable exposure to a particular voice (as in the case of family members, friends and colleagues), many other factors have been shown to affect the accuracy of voice identifications.[205] Recognition of previously heard voices is less accurate if the quality of the speech is poor (eg if the speech is heard through a telephone, whispered, or part of a low quality recording),[206] if the tone or pitch of the voice has been altered,[207] if the exposure time[208] or speech duration is short,[209] or if

---

conducted by those engaged in the investigation, such a discrimination task is perhaps more prone to bias. See the discussion below in Part VI(C).

[201] José H Kerstholt et al, 'Earwitnesses: Effects of Speech Duration, Retention Interval and Acoustic Environment' (2004) 18 *Applied Cognitive Psychology* 327.

[202] José H Kerstholt et al, 'Earwitnesses: Effects of Accent, Retention and Telephone' (2006) 20 *Applied Cognitive Psychology* 187.

[203] Brian R Clifford, 'Voice Identification by Human Listeners: On Earwitness Reliability' (1980) 4 *Law and Human Behavior* 373; Yarmey et al, above n 198.

[204] Dominic Watt, 'The Identification of the Individual through Speech' in Carmen Llamas and Dominic Watt (eds), *Language and Identities* (Edinburgh University Press, 2010) 76, 79, citing Francis Nolan, 'Forensic Speaker Identification and the Phonetic Description of Voice Quality' in W J Hardcastle and J Mackenzie Beck (eds), *A Figure of Speech: A Festschrift for John Laver* (Lawrence Erlbaum Associates, 2005) 385.

[205] A Daniel Yarmey, 'Earwitness Speaker Identification' (1995) 1 *Psychology, Public Policy, and Law* 792.

[206] Yarmey et al, above n 198; Tara L Orchard and A Daniel Yarmey, 'The Effects of Whispers, Voice-Sample Duration, and Voice Distinctiveness on Criminal Speaker Identification' (1995) 9 *Applied Cognitive Psychology* 249.

[207] Orchard and Yarmey, above n 206; Howard Saslove and A Daniel Yarmey, 'Long-Term Auditory Memory: Speaker Identification' (1980) 65 *Journal of Applied Psychology* 111.

[208] Susan Cook and John Wilding, 'Earwitness Testimony: Never Mind the Variety, Hear the Length' (1997) 11 *Applied Cognitive Psychology* 95; Yarmey, above n 205, 804–5.

[209] Susan Cook and John Wilding, 'Earwitness Testimony: Effects of Exposure and Attention on the Face Overshadowing Effect' (2001) 92 *British Journal of Psychology* 617.

there is a delay between original exposure and subsequent identification.[210] Accuracy rates of identifying incidentally heard voices have at times been shown to peak at 49 per cent after a delay of one week, only to decline to approximately 8 per cent after three weeks.[211] Conversely, additional speech utterance variety,[212] contextual consistency and distinctiveness have been associated with improved voice identification accuracy.[213]

With regard to the types of voice identification arising from the Australian case law, at least two further considerations emerge. The first relates to human decision-making biases where an interpreter or investigator (and sensory witnesses, such as in *E J Smith* and *Brownlowe*) identifies a voice that is heard in the context of an investigation. The second results from an identification process occurring across languages (a process that also applies to some jury comparisons).

First, the term 'confirmation bias' describes a situation where people are inclined to interpret evidence in a manner consistent with their expectations, rather than at face value.[214] In the voice identification context, where interpreters and investigators are provided with clear cues that others believe the source and target voices came from the same person, this tendency is liable to translate into an elevated likelihood that the interpreter or investigator will declare a match between the two voices, even where they originate from different speakers. Evidence of this tendency has been demonstrated in experiments where forensic scientists (fingerprint examiners) have been given inaccurate impressions (ie misleading or extraneous information about the case) and produced mistakes (and indeed reversals of previously expressed opinions).[215] Confirmation bias affects highly skilled experts, including those using widely accepted protocols.[216] Extrapolating from studies of latent fingerprint examiners, which have suggested that contextual cues may be subtle and may even operate unconsciously, formal training and experience are unlikely to protect the listener (or analyst) from error in voice comparison.[217]

Even in cases where the expectations of a match between the perpetrator and the suspect are less obvious, the comparison or recollection process itself can

---

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

[211] Clifford, above n 203, 383.

[212] Read and Craik, above n 198.

[213] Ibid; Saslove and Yarmey, above n 207.

[214] Gretchen B Chapman and Eric J Johnson, 'Incorporating the Irrelevant: Anchors in Judgments of Belief and Value' in Thomas Gilovich, Dale Griffin and Daniel Kahneman (eds), *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge University Press, 2002) 120, 133.

[215] Itiel E Dror et al, 'When Emotions Get the Better of Us: The Effect of Contextual Top-Down Processing on Matching Fingerprints' (2005) 19 *Applied Cognitive Psychology* 799; Itiel E Dror, David Charlton and Ailsa E Péron, 'Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications' (2006) 156 *Forensic Science International* 74.

[216] This is why most drug trials are double blind. See, eg, the discussion in R Barker Bausell, *Snake Oil Science: The Truth about Complementary and Alternative Medicine* (Oxford University Press, 2007).

[217] In addition, it is very difficult to meaningfully cross-examine upon such issues: see, eg, *Nguyen* (2002) 26 WAR 59, 87 [124] (Anderson J).

play a substantial role in the likelihood that an identification will be made. Where a listener is asked to identify a previously heard voice from a set of voices, the likelihood that the listener will choose the suspect by chance alone is influenced by many factors, including the size of the parade,[218] the instructions accompanying the procedure,[219] the presence of feedback (not necessarily deliberate or even conscious) from the parade administrator,[220] the circumstances in which the comparison is undertaken, and discussion with other witnesses.[221] For voice identification, unlike for eyewitness identification, there are relatively few 'voice parades', very few constraints on how voice identification evidence is obtained and limited application of exclusionary rules. Nonetheless, there is no compelling argument as to why such factors should not be taken into consideration when assessing the relevance, admissibility and probative value of all voice identification evidence — particularly given the impression among psychologists that voice identification is substantially less reliable than eyewitness identification.[222] This makes the tolerance for the opinions of investigators, and the reluctance of judges to impose some kind of regulation on voice comparison and identification, all the more remarkable.

Secondly, cross-lingual voice identifications played a role in several of the cases previously discussed.[223] In each of these cases the source speech was produced in a foreign language (eg Romanian, Cantonese, Mandarin and Igbo), while the target speech provided by the suspect, usually in a police interview, was in English. In these cases the interpreters or investigators were asked if the source speech was produced by the same person as the English target speech. From a practical standpoint, cross-lingual identifications are only possible if language-independent cues exist and remain consistent across different languages. These cues may include age, sex, and the size and shape of the speaker's vocal tract, nasal cavities and vocal folds.[224] The evidence supporting the utility of these language-independent cues also suggests that cross-lingual speaker identification can be influenced by many factors, for example: the types of languages being compared,[225] the origin and experience of the speaker,[226] the

---

[218] A Daniel Yarmey, A Linda Yarmey and Meagan J Yarmey, 'Face and Voice Identifications in Showups and Lineups' (1994) 8 *Applied Cognitive Psychology* 453.

[219] See Nancy Mehrkens Steblay, 'Social Influence in Eyewitness Recall: A Meta-Analytic Review of Lineup Instruction Effects' (1997) 21 *Law and Human Behavior* 283.

[220] See Sarah M Greathouse and Margaret Bull Kovera, 'Instruction Bias and Lineup Presentation Moderate the Effects of Administrator Knowledge on Eyewitness Identification' (2009) 33 *Law and Human Behavior* 70.

[221] See Helen M Paterson, Richard I Kemp and Jodie R Ng, 'Combating Co-Witness Contamination: Attempting to Decrease the Negative Effects of Discussion on Eyewitness Memory' (2011) 25 *Applied Cognitive Psychology* 43.

[222] See, eg, Clifford, above n 203, 391; Lawrence M Solan and Peter M Tiersma, 'Hearing Voices: Speaker Identification in Court' (2003) 54 *Hastings Law Journal* 373, 432.

[223] See above Parts IV–V.

[224] Steven J Winters, Susannah V Levi and David B Pisoni, 'Identification and Discrimination of Bilingual Talkers across Languages' (2008) 123 *Journal of the Acoustical Society of America* 4524, 4525–6.

[225] See Olaf Köster and Niels O Schiller, 'Different Influences of the Native Language of a Listener on Speaker Recognition' (1997) 4 *Forensic Linguistics* 18.

language(s) spoken by the listener,[227] the listener's proficiency in the speaker's language,[228] and whether the listener is familiar with the voice.[229]

Taking into account this complex array of factors it may come as a surprise that a few researchers have, at least in the context of their studies, characterised some cross-lingual identifications as reliable.[230] Closer consideration, however, reveals the importance of context when drawing conclusions from this work. Specifically, identification accuracy rates described as reliable in one study ranged from 45 to 60 per cent.[231] Such figures are not generally synonymous with reliability, particularly as accuracy rates in this particular study were inflated by the removal of participants who did not satisfy the minimum performance criterion in its training phase.[232] In another study, Goldstein and colleagues concluded that their data demonstrated that accented voices speaking an unfamiliar language are as well-remembered as are voices speaking incomprehensible words in a foreign language; however, the accuracy rates were 58 per cent and 57 per cent respectively.[233] More generally, Goggin and colleagues reported accurate identification rates of between 12 per cent and 35 per cent for listeners making identifications across languages,[234] while others present accuracy rates between 47 per cent and 70 per cent with the false alarm rate above 67 per cent even when the second language was familiar.[235] Thus, the 'reliability' of cross-lingual identifications must be evaluated against an appropriate threshold of performance given the particular context. While a 57 per cent voice identification accuracy rate might be considered good enough in most day-to-day settings (eg when answering the telephone), it is not appropriate in a forensic context, given the serious consequences associated with an error and the difficulty of conveying limitations to a lay jury in the context of an accusatorial trial. Where jurors are asked to undertake voice comparison themselves they may, even *with* such information, have an exaggerated confidence in their ability to make reliable comparisons, or use — whether they know it or not — other incriminating evidence to supplement their analysis.[236]

---

[226] On origin, see Nathan Daniel Doty, 'The Influence of Nationality on the Accuracy of Face and Voice Recognition' (1998) 111 *American Journal of Psychology* 191. On experience, see ibid.

[227] Charles P Thompson, 'A Language Effect in Voice Identification' (1987) 1 *Applied Cognitive Psychology* 121; Köster and Schiller, above n 225.

[228] Judith P Goggin et al, 'The Role of Language Familiarity in Voice Identification' (1991) 19 *Memory & Cognition* 448.

[229] Kirk P H Sullivan and Frank Schlichting, 'Speaker Discrimination in a Foreign Language: First Language Environment, Second Language Learners' (2007) 7 *Forensic Linguistics* 95.

[230] See, eg, Winters, Levi and Pisoni, above n 224.

[231] Ibid 4529 (figure 1).

[232] Ibid 4527. The training phase is where listeners are exposed to the target voice in order that they might be able to identify it given the experimental conditions in subsequent recognition phases.

[233] Alvin G Goldstein et al, 'Recognition Memory for Accented and Unaccented Voices' (1981) 17 *Bulletin of the Psychonomic Society* 217, 219.

[234] Goggin et al, above n 228, 451.

[235] Axelle C Philippon et al, 'Earwitness Identification Performance: The Effect of Language, Target, Deliberate Strategies and Indirect Measures' (2007) 21 *Applied Cognitive Psychology* 539, 544–5. See also Köster and Schiller, above n 225.

[236] Some of these issues are raised (though not necessarily in the voice recognition context) in *R v Lam* [2005] VSC 299 (10 June 2005) [20]–[28] (Redlich J); *R v Bennett* (2004) 88 SASR 6,

VII Reconsidering *Riscuta* and *Korgbara*

For the purpose of clarity, it is useful to attempt to apply the results of experimental research to the facts of *Riscuta* and *Korgbara*.[237] In the case of *Riscuta* it is unlikely that the interpreter, Kandic, was sufficiently exposed to the voice of Niga during the 30 minute interview at the Crime Commission in 1993 to consider the voice familiar or 'known' — that is, recognisable to the extent that Kandic could have named Niga were she to, say, answer a telephone call from her.[238] There are several factors which threaten the accuracy of Kandic's positive identification evidence. Kandic spent only 30 minutes with Niga in 1993, during an interview that was conducted in English. In 1994 she translated a number of surveillance tapes which allegedly had Niga's voice on them. However, there was no indication that Kandic had independently recognised or identified Niga's voice in 1994. Nor was there any indication of such recognition for another seven years. Further, there was evidence to suggest that the police had disclosed to Kandic their belief that the voices from 1993 and 1994 were the same, and Kandic also conceded that she was relying, in part, on contextual information to come to her conclusion that the voice on the tapes was that of Niga.[239]

So in this case we are considering a situation where a person is thinking back eight years (from 2001 to 1993) to match a voice they heard seven years ago (in 1994) and not since. The experimental evidence indicates that our ability to correctly identify voices degrades over time. More specifically, incidentally heard voices were identified at best with 49 per cent accuracy one week after exposure, declining to 8 per cent accuracy after three weeks.[240] And although the accuracy for familiar voice identification is likely to start much higher than this — at around 80 per cent[241] — the decline anticipated in *Riscuta* over the 18 months between the interview and the covert recordings, or indeed the further seven years until the identification, can reasonably be assumed to be considerable.

In *Riscuta* we also confront a situation where the likelihood that confirmation bias (or suggestion) has influenced the identification is high. So, in this case, where the expectation of a match between the person from 1993 and the person from 1994 had clearly been conveyed to Kandic by the police, her identification, whenever made, was contaminated by that expectation rather than being based

---

19–20 [80] (Doyle CJ); *R v Coxon* (2002) 82 SASR 412, 419 [32]–[34] (Prior J); *Festa v The Queen* (2001) 208 CLR 593, 643 [166] (Kirby J). See also *R v Burchielli* [1981] VR 611, 616, 620–1 (Young CJ and McInerney J); *R v Haidley* [1984] VR 229, 231–2 (Young CJ); *E J Smith* (1986) 7 NSWLR 444, 458 (Lee J).

237 The facts of *Riscuta* [2003] NSWCCA 6 (6 February 2003) are set out above in Part III. The facts of *Korgbara* (2007) 71 NSWLR 187 are set out above in Part V.

238 Contrast *Nguyen* (2002) 26 WAR 59, 67 [28] (Malcolm CJ), where the interpreter listened to more than 600 telephone calls involving the accused, and *Neville* [2004] WASCA 62 (2 April 2004) [10] (Miller J), where the police officer had listened to at least 78 telephone calls, 21 of which ran for a total of over three hours when played in court.

239 *Riscuta* [2003] NSWCCA 6 (6 February 2003) [23] (Heydon JA).

240 Clifford, above n 203, 383.

241 See above Part VI(B).

solely on her own perceptual experience — that is, on the presence or absence of any recollection of the voice from 1993 to 1994.

Kandic also indicated that the voice from 1993 did not have any unusual features.[242] Evidence suggests that with lower levels of exposure to a particular voice, factors such as distinctiveness become increasingly informative regarding the likely accuracy of an identification. For instance, where the quality of the speech is poor (as in the case of some recordings or whispered conversations), the tone or pitch has been altered by way of disguise, the exposure time is short, or the speech offers limited variability, the likelihood of an accurate identification is reduced. Further, this is pronounced where identifications are made across languages, as in both *Riscuta* and *Korgbara*.

It *is* possible for identifications to be made across languages with relatively high levels of reliability. However, for this to occur there need to be sufficient language-independent cues. Ideally, there would also be a pre-existing familiarity with the voice (eg repeated exposure) in both languages. This would allow prior experience of language-independent cues to inform any subsequent identification. In the cases at hand, however, cross-lingual identification is unlikely to be highly reliable. In *Riscuta* the comparison was made between an unfamiliar voice speaking in Romanian and an unfamiliar voice speaking in English. In *Korgbara*, where the comparison was made between English, a familiar non-tonal language (and one spoken by the listener), and Igbo, a previously unheard tonal language, it is uncertain that relevant language-independent cues were even available, let alone sufficient, to facilitate an identification with much probative value. Indeed, the available empirical evidence suggests that accurate identification is unlikely, with rates of cross-lingual identification accuracy ranging from 12 per cent at worst to 70 per cent (ie a 30 per cent rate of error) *at best*.[243] This is clearly a far cry from the levels of performance necessary to generate confidence that the correct individual has been identified in a forensic context, and is certainly not a credible basis for leaving cross-lingual comparison to a jury as occurred in *Korgbara*.

One response in *Riscuta* would have been to ensure that the many limitations with Kandic's opinion were canvassed in the trial and then reiterated through a clear set of directions and warnings. It does not follow, however, that adequate explanation of the limitations with such evidence will always occur and that, even where it does, the extent of human frailties — including the frailties of interpreters and investigators — will be appreciated.[244] Moreover, where

---

[242] *Riscuta* [2003] NSWCCA 6 (6 February 2003) [54] (Heydon JA).
[243] See above Part VI(C).
[244] See the discussion below in Part VIII(B). Both *Riscuta* [2003] NSWCCA 6 (6 February 2003) and *Korgbara* (2007) 71 NSWLR 187 appear prominently in Judicial Commission of New South Wales, *Criminal Trial Courts Bench Book* (2011) [3-100]–[3-110] ('Identification Evidence — Voice') <http://www.judcom.nsw.gov.au/publications/benchbks/criminal/identification_evidence-voice.html>. *Korgbara* is cited as confirming that there are no preconditions for the admission of voice identification evidence other than relevance, and as establishing the principle that 'there is no prescriptive rule that voice comparison evidence in relation to foreign languages should only be admitted where it is supported by expert testimony': at [3-100]. In referring to Heydon JA's judgment in *Riscuta*, the *Bench Book* indicates that the directions given by the trial judge were

interpreters and police express opinions that were formed in ways that ignored corrosive contamination and bias and were presented as part of a more extensive prosecution case, then the weakness of the voice comparison and identification evidence may not be recognised, conveyed or accepted. It may be that other incriminating evidence will act as a makeweight, or that the very strong corrosive potential of suggestion will be underestimated by jurors who prefer to interpret contaminated opinions, inappropriately, as (independent) corroboration. This is certainly how judges have explained their own responses when upholding convictions.[245]

Cross-lingual comparisons accentuate the ordinary problems with identification experienced by laypersons and 'experts' not familiar with the person of interest, and the methodological problems.[246] These concerns are compounded in cases where sound recordings are of poor quality, of brief duration, have been obtained in different circumstances, or have been presented to the witness in conditions where there is a risk of suggestion. Positive identifications obtained in such circumstances are likely to carry a non-trivial risk of error unless there is some persuasive reason to believe otherwise. Unless comparisons are undertaken by familiars — free from bias or focused expectations — or by those with demonstrably reliable techniques in circumstances where analysis is undertaken without any suggestion about the identity of the relevant voice(s), comparisons and identifications are likely to compound, rather than expose, investigative mistakes. Where the accused is one of a small minority who actually speaks the relevant language, as in *Korgbara*, allowing the tribunal of fact to undertake its own comparison, in circumstances where there is *other* evidence, may make it difficult and perhaps impossible for the trial to be fair. In the context of an

---

inadequate within the terms required by s 116 of the *Evidence Act 1995* (NSW), and draws on *Riscuta* both in relation to the need to inform the jury of the *special* need for caution and in identifying the factors that need to be brought to the attention of the jury: at [3-110]. Note, however, the discussion above in Part II: displaced listeners are not caught by the definition of identification evidence in the UEA and thus whatever protection is offered by s 116 is inapplicable to evidence given by such listeners. The New South Wales *Bench Book* thus compounds the conceptual confusion surrounding this area in so far as it draws, primarily, on cases involving 'ad hoc experts' or other displaced listeners, but does not explicitly address the distinction between direct and displaced listeners. By contrast, the Victorian equivalent does contain a separate section for comparison evidence: Judicial College of Victoria, *Victorian Criminal Charge Book* (2011) [4.12.5] ('Charge: Comparison Evidence') <http://www.justice.vic.gov.au/emanuals/CrimChargeBook/default.htm>. This section covers jury comparisons and comparisons made by 'witnesses comparing people or items about which they have greater knowledge than the jury', but a warning is not required for comparisons undertaken by those with expertise: see para 99 in [4.12.1D] ('When to Give an Identification Evidence Warning').

245 This is apparent in the majority of cases discussed above in Parts III–V. There is no advantage in treating Kandic's claims in *Riscuta* as 'recognition' (or fact) rather than opinion evidence. Such an approach merely displaces the main epistemological issues. It purports to circumvent the exclusionary opinion rule without addressing the questions of what level of familiarity is required before such comparison might be minimally reliable, and how the conditions in which the identification was obtained affect reliability.

246 See generally the discussion of analogous problems with eyewitness identification in Jim Dwyer, Peter Neufeld and Barry Scheck, *Actual Innocence: Five Days to Execution and Other Dispatches from the Wrongly Convicted* (Doubleday, 2000); Sheri Lynn Johnson, 'Cross-Racial Identification Errors in Criminal Cases' (1984) 69 *Cornell Law Review* 934; Brian L Cutler and Margaret Bull Kovera, *Evaluating Eyewitness Identification* (Oxford University Press, 2010) 37–40.

accusatorial trial, hearing the voice of a black African sitting in the dock who speaks the impugned language, combined with voice evidence or suggestive comparisons, may be a form of unfair prejudice.[247]

In a case like *Korgbara*, it is likely that jurors will make errors evaluating the probative value of the fact that both the perpetrator and the suspect speak a rare Nigerian dialect. There is a real risk that jurors will misattribute the rarity of Igbo in Australia as evidence that increases the likelihood that the perpetrator and the suspect are the same person. The reasoning runs as follows: very few people in Australia speak Igbo, therefore it is very unlikely that both the perpetrator and the suspect would speak Igbo by chance alone — ergo, because both these people speak Igbo, the suspect must be the perpetrator. This reasoning and attribution is mistaken. In reality, the fact that both the perpetrator and the suspect in the case speak Igbo is far from coincidental, as it would need to be to sustain the attribution just described. Rather, every suspect must speak Igbo in order to be considered a suspect. Therefore, the fact that the suspect speaks Igbo does not add anything to the likelihood that this particular suspect is also the perpetrator. The probability that a defendant in this trial speaks Igbo is a prerequisite; it cannot be used to discriminate between innocent and guilty suspects. The fact that the suspect speaks Igbo is therefore not relevant to calculating the likelihood that the suspect is the perpetrator and should not be confused with the very rare event that a randomly selected person in Australia would speak Igbo.[248]

Finally, it may be that in many cases, including the circumstances in *Korgbara*, if there is no demonstrably reliable means of comparing the voices then recordings should not be presented to juries for purposes of comparison and identification. The existence of *other* incriminating evidence does not overcome this deficiency, but instead is likely to compound it, making even more critical the admissibility decisions on evidence that involves identification (or similarities) whether by lay or 'expert' witnesses or juries. Although unpalatable to those reared in the tradition of Bentham, Wigmore and Cross, it seems that we cannot be confident that the trial and the tribunal of fact are capable of consistently and adequately dealing with some forms of voice evidence, especially when compounded by other suggestive evidence in an accusatorial proceeding.[249]

---

[247] One aspect of this is that there is a danger in a case like *Korgbara* (as noted in *Bulejcik* (1996) 185 CLR 375, 397 (Toohey and Gaudron JJ)) that the jury 'might conclude too readily that a foreign accent on a tape is that of the accused where the accents are similar.'

[248] Some of these forms of reasoning resemble fallacies associated with misinterpretations of DNA evidence, discussed in *Aytugrul v The Queen* [2010] NSWCCA 272 (3 December 2010) [78]–[95] (McLellan CJ at CL). Special leave to appeal from that decision has been granted: Transcript of Proceedings, *Aytugrul v The Queen* [2011] HCATrans 238 (2 September 2011).

[249] See Edmond and Roberts, above n 97. Cf Larry Laudan, *Truth, Error, and Criminal Law: An Essay in Legal Epistemology* (Cambridge University Press, 2006) ch 1. This is not to suggest that recordings should not be admissible, but rather to focus on the way the evidence is used.

VIII   DEAF AND DUMB JUSTICE: SCIENTIFIC RESEARCH AND LEGAL PRACTICE

Why have prosecutors, defence lawyers and judges not engaged with mainstream, credible and cautious scientific research?

The way rules of evidence have been interpreted seems to have given prosecutors and investigators an easy ride at the expense of the accused and, in many cases, prevented courts and jurors from finding out about the extent of weaknesses in many types of incriminating opinion evidence or about unacceptable investigative procedures. While we appreciate that judges tend to be dependent on the parties, if the parties — and here we are talking about the state in most cases — are unable or unwilling to provide appropriate expertise or evidence about serious problems and limitations, then we must wonder about the value of the rules and practices that have been developed around voice evidence. In the following sections we review some possible 'solutions' to the difficulties posed by incriminating voice identification evidence. These include the use of additional experts to inform the jury, judicial responses to incriminating opinions about voices, emerging techniques of voice comparison that are endeavouring to overcome some of the limitations associated with unaided listening by non-familiars, and finally, the use of voice identification parades.

A   *Remedial Psychologists?*

Before turning to the more conventional remedy of *judicial* warnings or directions, we want to consider whether current practice might be redeemed through recourse to expert witnesses (eg experimental psychologists) informing the tribunal of fact about the results of experimental scientific research.[250]

We should first note that such recourse to psychologists is at odds with judicial protection of jurors from overexposure to expert evidence, especially in areas where they believe laypeople are competent based on life experiences.[251] Historically, Australian judges have jealously guarded their control over what jurors should be told about ordinary human abilities, experiences and tendencies. In general, they have been indifferent to experimental research by psychologists and other non-medical scientists, particularly in relation to informing admissibility jurisprudence. This is, we suggest, an unfortunate state of affairs, and has led legal practice in directions that are difficult to reconcile with the rational tradition of evidence and proof as well as what is known beyond the courts.

However, it is our contention that allowing the defence to call psychologists (or others with relevant research interests and competence in experimental methodologies) to explain the limitations of voice comparison and identification

---

[250] Some of the many complexities associated with informing the tribunal of fact of such research are discussed by David L Faigman, *Constitutional Fictions: A Unified Theory of Constitutional Facts* (Oxford University Press, 2008) ch 8. See also John Monahan and Laurens Walker, *Social Science in Law: Cases and Materials* (Foundation Press, 7th ed, 2009).

[251] There is a general reluctance to admit psychological evidence: see, eg, *Smith v The Queen* (1990) 64 ALJR 588; *R v Smith* (2000) 116 A Crim R 1, 8–13.

evidence is not a viable solution to the difficulties besetting current practice.[252] The adversarial nature of proceedings and the almost certain presence of additional incriminating evidence mean that the trial is not conducive to a neutral tutorial. Allowing the defence to call experts to offer (sometimes abstract) information, qualifications and criticisms, which will not always match the precise conditions of the instant case, is unlikely to render the opinions of displaced listeners probative or reduce the danger of unfair prejudice.[253] It may in fact have the perverse effect of strengthening the prosecution's case, by casting the problem for the jury as merely a conflict of interpretation rather than as a fundamental question of reliability. Further, since defence witnesses are almost always able to be portrayed as more partisan than state-employed investigators and consultants, they are unlikely to exert the same sort of influence as the incriminating opinions of 'experts' appearing for the prosecution. Similarly, explaining methodological limitations — eg that suggestions and cues are likely to substantially impact interpretations — might not influence the thinking of judges or juries, especially in the context of the overall case. Moreover, most of the experimental studies have not exposed participants to additional information when asking them to make their comparisons.[254] It is highly likely that supplementary information, such as the opinions of prosecution 'experts', will dramatically influence lay responses — and it is highly likely that these opinions will be influential, regardless of whether they are correct.[255]

We would contend that critical insights should lead to the exclusion rather than admission — however qualified — of a great deal of voice evidence from displaced listeners who do not have demonstrably reliable methods. Moreover, requiring psychologists to rehearse a range of relevant and quasi-relevant studies in ways that might inform juries in order to convince them to approach 'expert' opinion carefully is a very cumbersome, expensive and risky way to proceed. Rather than the state being required to develop more reliable procedures and techniques for collecting, analysing and reporting voice evidence, jury after jury is to be taught about problems with unreliable forms of incriminating opinion evidence, in circumstances where the fairness of proceedings may depend upon the success of this one-sided tutorial. In addition, the accused is tasked with identifying a suitable alternative expert witness to discredit evidence that is of a type that is known to be inaccurate, and bears the risk of the reliance on traditional safeguards — such as exclusionary discretions, directions and warnings — that seem to have, at best, inconsistent application and mixed efficacy. It is the obligation of the state to prove guilt beyond reasonable doubt and this should not

---

[252] See the analysis of analogous problems with eyewitness evidence in Kristy A Martire and Richard I Kemp, 'Can Experts Help Jurors to Evaluate Eyewitness Evidence? A Review of Eyewitness Expert Effects' (2011) 16 *Legal and Criminological Psychology* 24.

[253] Note also the judicial reluctance to consider methodological limitations in *Madigan* [2005] NSWCCA 170 (9 June 2005) and *Korgbara* (2007) 61 NSWLR 187.

[254] Interestingly, these are the very conditions in which the tribunal of fact is expected to undertake its assessment of the evidence once it is admitted.

[255] Richard Kemp, Stephanie Heidecker and Nicola Johnston, 'Identification of Suspects from Video: Facial Mapping Experts and the Impact of Their Evidence' (Paper presented at the 18th Conference of the European Association of Psychology and Law, Maastricht, 2–5 July 2008).

be subtly eroded or shifted by the admission of unfairly prejudicial evidence, especially the subjective and contaminated opinions of non-expert investigators, and by cross-lingual comparisons by juries. The state, after all, has greater epistemic and ethical obligations than other parties, considerable resources at its disposal, and a high standard of proof designed to protect the innocent.

B *Judicial Directions and Other 'Solutions'*

Undoubtedly, the preference of Australian judges for managing the potential dangers of incriminating voice evidence is to issue 'very careful instructions' to the jury, as expressed by the High Court in *Bulejcik*:[256]

> Where a witness identifies a voice on the basis of having heard it before, the witness needs to have heard a sufficient amount of the accused's speech to be familiar with it because, in saying that the voice at the crime scene is that of the accused, the witness is relying on his or her memory of the accused's voice. Where a witness identifies a voice on the basis of having heard it subsequently, there should be something about the voice at the crime scene to sufficiently embed it in the witness's memory so as to enable him or her to say that it is the same as a voice which he or she heard subsequently. The greater the distance in time between when the two voices compared were heard, the greater the desirable degree of familiarity or distinctiveness. …

> This Court would be slow to depart from a trial judge's assessment that material was of sufficient quality and quantity for the jury to be permitted to make the necessary comparison. The question rather is whether the jury were given sufficient warning of the difficulties involved.[257]

Without reference to empirical studies or relevant scientific literature, the trial judge is required to provide 'very careful directions as to those considerations which would make a comparison difficult and … a strong warning as to the dangers involved in making a comparison'[258] — though even here Brennan CJ resisted, noting that the sufficiency of any warning is 'not assessed by reference to a formula nor by postulating a hypothetical warning against risks of which a

---

[256] Against the background of this preference for jury directions, it is worth noting that directions represent a significant area of concern, with the New South Wales Law Reform Commission currently conducting an inquiry into them: New South Wales Law Reform Commission, *Jury Directions*, Consultation Paper No 4 (2008) ('*NSW Consultation Paper*'). The Queensland Law Reform Commission published its report on jury directions in December 2009: Queensland Law Reform Commission, *A Review of Jury Directions*, Report No 66 (2009) ('*Queensland Report*'). The Victorian Law Reform Commission also published its final report in 2009: Victorian Law Reform Commission, *Jury Directions: Final Report*, Report No 17 (2009) ('*Victorian Report*'). Each of these publications features a section on the directions required in relation to identification evidence: *Queensland Report*, 524–36; *Victorian Report*, 50–2. The *NSW Consultation Paper* does not discuss voice identification directly, though it does raise the question of directions in relation to juries making their own assessment of CCTV and photographic evidence: at 136–8. Neither the *Queensland Report* nor the *Victorian Report* addresses voice identification directly.

[257] (1996) 185 CLR 375, 394–5, 397 (Toohey and Gaudron JJ).

[258] Ibid 398–9. Judges are as likely to refer to notorious or classical misidentifications, as in Genesis 27:1–22, as they are to empirical literature: see, eg, *Neville* [2004] WASCA 62 (2 April 2004) [85] (Heenan J).

reasonable jury would be as well aware as the trial judge.'[259] The Chief Justice expressed a reluctance to 'impose … an artificial restraint on the jury's employment of their common sense.'[260]

Without wanting to adopt a totally deprecatory attitude to judicial experience (or the wisdom of 'the Law'), or even to the ability of many instructions to touch upon salient issues and problems, it would be a mistake to equate legally recognised limitations of voice comparison and identification evidence and espoused faith in the value of directions and warnings with the rather more extensive, detailed and critical scientific research. Apparently unwittingly, lawyers and trial and appellate judges routinely overlook relevant research and/or embrace popular misconceptions, such as the appeal to 'indelible impression' by the trial judge in *E J Smith*.[261] In addition, prosecutors and judges have tended to trivialise the way in which voice identification evidence is obtained, even though suggestive procedures have a demonstrated tendency to contaminate interpretations.[262]

We can obtain some sense of the limits of judicial warnings by reviewing Winneke P's judgment in *R v Callaghan*.[263] This case involved a bank robbery and was one where, unusually, the Victorian police organised a voice parade. In response to the impugned voice identification evidence of bank staff — ie direct unfamiliar witnesses — in the aftermath of the robbery, Winneke P complimented the 'full instructions' of the trial judge. By way of summary we are told:

> In the course of his directions to the jury, the [trial] judge gave what *appear to me to be full instructions as to the caution with which they should treat the evidence of identification*. It is, I think, unnecessary to set them out in full. Amongst other things, he directed them, with the full authority of his office, that:
>
> - The caution which courts are required to give in relation to visual identification 'must apply even more so to witnesses giving evidence of voice identification'.
>
> - They must take into account factors which, of necessity, reduce the weight of the evidence; for example that the witnesses had never before heard the voice of the offender behind the tellers' counter; that it is much easier to identify a voice which is familiar; that mistakes can occur even when a voice is familiar; that the tone of the voice of the offender was 'much more demanding and insisting than the tone of the recorded voices

---

[259]   *Bulejcik* (1996) 185 CLR 375, 384. In practice, it may be impossible to prevent the jury making the comparison where such evidence is admitted: see, eg, *R v O'Sullivan* [1969] 1 WLR 497, 503 (Winn LJ for Winn and Widgery LJJ and Lawton J).

[260]   *Bulejcik* (1996) 185 CLR 375, 383.

[261]   *R v Smith* [1984] 1 NSWLR 462, 482, 485 (O'Brien CJ Cr D). Interestingly, Smith was unrepresented, so the literature and research on which the trial judge relied, which was primarily legal, was probably the result of his own endeavours.

[262]   See generally Dror, Charlton and Péron, above n 215; D M Risinger et al, 'The *Daubert*/*Kumho* Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion' (2002) 90 *California Law Review* 1.

[263]   (2001) 4 VR 79.

including the accused'; that the event in the bank was short, and the words spoken were 'short and sharp'.

- There were very limited opportunities for the voice to become recognisable to the witnesses, and there 'were no really distinguishing features about the voice they described'; the voice was 'Australian' rather than foreign; nothing to suggest they were particularly distinctive.

- The jury must take account of the fact that the experience must have been frightening and that, whilst some people might be capable of making accurate observations under situations of strain, others might have their powers of observation and hearing quite diminished by the terror of it all.

- The lapse of time between the event and the later 'identification' is important in that 'the greater the time, the more opportunity for the natural fallibility of human memory to be increased'.

- The jury should consider how positive the witness was, without forgetting the personality. Some witnesses can be positive but mistaken; others cautious but correct, albeit not confident.

- That some witnesses may have 'better ear for sound than others'.

- That the jury 'should consider the evidence of personal identification' most carefully before acting upon it. Where possible 'you should look for some feature or features of the evidence which tend to make it reliable'.[264]

Disregarding the manner in which the comparison was undertaken and the opinion evidence was collected enables us to focus on how a tribunal of fact should approach and apply instructions about voice identification evidence.[265] Notwithstanding the potential value of these instructions, it is not obvious how they could be understood and applied by a jury in the absence of empirical information about actual capabilities and limitations. Although legally orthodox, these directions do not provide any indication of:

- the actual effects of contextual factors;

- just how corrosive delayed comparisons and recollections can be;

- how limited exposure dramatically reduces accuracy;

- how tone and type of speech and recording type influence accuracy;

- the very high risk of error;

---

[264] Ibid 96 [29] (emphasis added). See also, quoting this passage, *R v Lam* [2005] VSC 299 (10 June 2005) [14] (Redlich J). In *Bulejcik* (1996) 185 CLR 375, 395 (Toohey and Gaudron JJ), it was noted that the jury had been (properly) 'warned to consider the different acoustics and not to bear it against the appellant that English was not his mother tongue.' Cf the discussion below in Part VIII(D).

[265] In our discussion of *R v Callaghan*, we will 'bracket' the manner in which the comparison was obtained. Without wanting to condone the method used in the case, there is insufficient information about the actual process followed for a full discussion to be undertaken. Nevertheless, the approach adopted — a voice parade — seems to have been far less problematic than the very suggestive processes routinely used by investigators, translators and 'experts'.

- the way witness confidence is often misleading;
- how witness variability might apply in the specific circumstances;
- how witness interactions and investigator confirmation may produce (mistaken) consensus and inflate levels of confidence; and
- how even the most subtle clues from honest investigators can contaminate virtually any identification.

Things would seem to become more complicated, and more error-prone, when such factors are combined. Nevertheless, in the absence of detail drawn from relevant and publicly available scientific research, jury instructions may be worthless. They might appear to render a trial formally fair by drawing attention to legally notorious dangers, but there must be genuine doubt about whether they practically assist juries to rationally assess incriminating voice evidence.[266]

As things stand, jurors are somehow expected to 'take into account' or 'consider … most carefully'[267] a range of contextual factors without information on how such factors might influence accuracy whether individually or collectively. There is an assumption that mere advertence is enough to discharge the obligation of dealing with a type of evidence which is demonstrably prone to error, and far less accurate than most jurors and judges are likely to assume, even after conventional warnings. There is also evidence that laypersons and 'experts' tend to dramatically underestimate how suggestion, or even prior information, shapes interpretations and analyses. This is important, particularly for jury comparisons undertaken in conjunction with exposure to other incriminating information or evidence that the accused speaks the impugned language. Furthermore, how should the jury 'take into account' the impact of fear? And can they ignore this (somewhat contradictory) warning by simply accepting (without any evidence) that the witness is not the kind of person likely to be affected, because of imputed accuracy on the basis of training as a bank teller or experience as a police officer?

In addition, where witnesses are qualified by the courts as 'experts', whether through formal qualifications or experience or as 'ad hoc experts', the warnings about problems with identification might not be given in relation to their 'expert' opinion evidence, even though the same problems will almost always arise. In the absence of validated methods, the problem is that the 'expert' does not have a demonstrably reliable method of overcoming these kinds of problems or ascertaining their level of accuracy. Rather, juries are likely to be told in general terms that there are dangers with expert evidence and that the decision is ultimately for them. They are not always told that the individuals expressing opinions may have been exposed to other contextual information, do not have validated methods, or do not necessarily appreciate the significance of this failure; nor are they always told that lay and 'expert' witnesses may not be able to do what they

---

266 See Kristy A Martire and Richard I Kemp, 'The Impact of Eyewitness Expert Evidence and Judicial Instruction on Juror Ability to Evaluate Eyewitness Testimony' (2009) 33 *Law and Human Behavior* 225.

267 *R v Callaghan* (2001) 4 VR 79, 96 [29] (Winneke P).

claim, and that some of the witnesses have no relevant expertise and are no more likely to be accurate than a person selected randomly from the street.[268]

There is, in addition, little evidence that police, translators and interpreters, and even linguists perform much better than average or are particularly accurate at comparisons across the many different conditions confronting earwitnesses and listeners. Moreover, even if interpreters, investigating police and linguists were slightly or even significantly better than unfamiliar laypersons, there would still be the issue of how much better and how reliable their incriminating opinion testimony ought to be before it is admitted as an exception to the opinion rule based on 'specialised knowledge' or 'experience'.[269] There are, after all, few means of credibly challenging this evidence without extensively canvassing the specialist literature. We also recognise that repeatedly listening to a voice *may* improve an ability, but this raises the question of whether jurisprudence should expediently *construct* 'experts', especially where these are investigators or persons involved in an investigation (eg translators) and not part of the specialist communities actually involved in scientific voice comparison research.

Returning to the content of instructions, there is no expectation that judges will explain every relevant aspect of contested identification evidence in every case. Provided the trial judge broadly canvasses the issue in a way that draws attention to what the lawyers and judge consider are the major issues or potential defects, based on judicial experience rather than scientific study, that will suffice.[270] There are, for example, few judicial references to suggestion and contamination, despite the fact that the empirical research suggests that these can have incredibly powerful effects even where the suggestion is extremely subtle or unconscious. This means that investigators and witnesses of undoubted integrity can be sincerely mistaken if the evidence is not collected and analysed with sensitivity to risk of contamination. Where witnesses are allowed to speak to each other about the sound of a voice (or the appearance of a person) before making formal statements, they are very likely to influence (and reinforce) each other's assessments.[271] Yet judicial statements rarely warn in these terms and almost never recognise the corrosive potential of such apparently innocuous interactions.

It is important to recognise that the vast majority of available empirical studies suggest that jury directions, instructions and warnings seem to be ineffective.[272] Even if judges could provide detailed and scientifically predicated directions, the empirical research suggests that it would be difficult to understand and apply them to the particular evidence, especially in the overall context of the trial. In

---

[268] Because some of these witnesses may have acquired abilities and possess opinions that are probative, we suggest a procedure, outlined below, that might help to remove *some* of the most egregious aspects of the unfair prejudice associated with such 'ad hoc expert' opinions.

[269] There is also the question about the relevance of such opinions, which was advanced in the context of eyewitness identification by the majority in *Smith* (2001) 206 CLR 650, 654–6 [9]–[12] (Gleeson CJ, Gaudron, Gummow and Hayne JJ).

[270] *R v Haidley* [1984] VR 229, 230 (Young CJ), approved in *R v Callaghan* (2001) 4 VR 79, 98 [35] (Winneke P, Brooking JA and O'Bryan AJA agreeing).

[271] Paterson, Kemp and Ng, above n 221.

[272] These concerns are borne out in the various consultation papers and reports referred to at above n 256.

consequence, jury directions are doubly weak. First, legally orthodox warnings tend to present jurors with highly abstract information. Secondly, decades of research suggests that even technically and epistemologically sound directions are less efficacious than any safeguard could credibly claim to be.[273]

Interestingly, in response to analogous difficulties with the interpretation of incriminating images — such as CCTV recordings of robberies — judges have endeavoured to address evidentiary infirmities, not by excluding incriminating opinions of unknown probative value or developing scientifically predicated warnings, but rather by limiting the opinions of 'ad hoc experts' to descriptions of similarities (and in theory, differences). This, however, is a cosmetic response to a deeper set of epistemic and procedural problems. What is more, there is no evidence that this 'solution' makes any difference or alters the way the tribunal of fact approaches incriminating opinions.[274] What, after all, is the difference *in effect* between an 'expert' who testifies that X is Y (or appears to be Y) and an 'expert' who testifies, on the basis of an examination of the same images, that he or she could see no differences, only a high level of anatomical similarity?[275] Our limited vocabulary with respect to describing sounds and the features of voices makes this 'solution' impractical as a sufficient response to the admission of voice comparison and identification evidence.[276] In the absence of information about the frequency of alleged similarities among relevant populations,

---

[273] See Roselle L Wissler and Michael J Saks, 'On the Inefficacy of Limiting Instructions: When Jurors Use Prior Conviction Evidence to Decide on Guilt' (1985) 9 *Law and Human Behavior* 37; Joel D Lieberman and Bruce B Sales, 'The Effectiveness of Jury Instructions' in Walter F Abbott and John Batt (eds), *A Handbook of Jury Research* (American Law Institute–American Bar Association, 1999) 18-1; 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* (Guilford Press, 2005) 407.

[274] See Dawn McQuiston-Surrett and Michael J Saks, 'The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Factfinders Hear' (2009) 33 *Law and Human Behavior* 436.

[275] It is thus disappointing that recommendations 29–31 of the *Victorian Report*, above n 256, 16, in relation to jury directions, perpetuate the idea that these differences in expression are meaningful and state that '"identification evidence", "recognition evidence" and "similarity evidence" should be given distinct definitions' and that warnings should only be mandatory in cases involving 'identification evidence'.

[276] *R v Smith* [1984] 1 NSWLR 462, 478–9 (O'Brien CJ Cr D). O'Brien CJ Cr D observed at 478 that

> whilst many features of a person which are visually noticeable, such as age, height, size, colour of hair and eyes and the numerous other physical characteristics of a particular human being are fairly readily capable of description so as to give a reasonable reproduction in everyday vocabulary, the features of a voice are not by any means as readily capable of verbal description.

Moreover, he recognised the considerable variation in voices depending on 'the circumstances in which they are used and the purposes for which they are used. The voice of a man speaking affectionately to a child necessarily varies markedly from his voice if abusing a fellow motorist in an argument between drivers on the road': at 479. See also *Festa v The Queen* (2001) 208 CLR 593, 619–20 [84], where McHugh J stated (citations omitted):

> The risk of mistake in identifying a voice is at least as great as in identifying a person. The reliability of voice identification varies with such factors as the length and volume of speech heard, the witness's familiarity with the accused's voice and the time elapsing between the occasions when the witness heard the voice of the perpetrator and the voice of the accused.

See also *R v Golledge* [2007] QCA 54 (2 March 2007) [59] (Keane JA).

'experts' are as likely to mislead as to provide independent corroboration or reliable inculpatory information.

Finally, there is the issue of how voice comparison and identification evidence should be combined with other evidence. Leaving aside the testimony of lay earwitnesses, the admissibility of opinion evidence based on a 'body of knowledge or experience', 'specialised knowledge' or 'ad hoc expertise' should be considered independently of any other evidence.[277] Furthermore, the practical inadequacy of directions, the inability to effectively cross-examine, and the potentially misleading confidence and sincerity of the witnesses should be taken into consideration in any decision to admit or exclude. Incriminating opinion evidence of unknown probative value should not be admitted merely because the jury might accept it or because, notwithstanding weakness, it is more convenient than other alternatives, particularly further research or exclusion.

C *Scientific Voice Comparison and Probabilistic Evidence*

It is worth noting that there are emerging probabilistically oriented approaches to voice comparison. These approaches, which do not depend primarily upon memory or subjective human comparison, aim to eliminate, through a range of scientific methods, many of the problems associated with auditory voice comparison. Proponents tend to be reasonably conversant with psychological research and a range of complex technical and statistical issues. It is not our intention to formally endorse such approaches, which are by no means infallible, nor to indicate that they are sufficiently reliable for legal practice — although we note that they have been admitted in Australia and New Zealand.[278] Rather, we merely want to indicate that there are highly qualified technical experts endeavouring to develop and validate more rigorous approaches to the analysis of sounds and particularly the comparison of voices — and that this research is ongoing because of the limitations of human listeners and expanding forensic and security needs.[279]

Rather than transforming interpreters and police officers into voice comparison experts by contorting rules, subverting principle, or propagating 'familiarity', we should instead be encouraging and assessing these scientifically predicated techniques to determine if they are sufficiently robust to be incorporated into criminal investigations and proceedings. New forms of voice comparison *may*

---

[277] Once such opinion evidence is admitted, the jury should be allowed to combine various strands of direct and indirect evidence. Here, supplementary evidence may be used as a makeweight. This merely reinforces the importance of admissibility decision-making.

[278] Dr Philip Rose, for example, provided reports in *R v Hufnagl [No 1]* [2008] NSWDC 134 (24 June 2008) and *R v Bain* [2010] 1 NZLR 1.

[279] See generally Gonzalez-Rodriguez et al, above n 186; Philip Rose, *Forensic Speaker Identification* (Taylor & Francis, 2002); Geoffrey Stewart Morrison, 'Forensic Voice Comparison' in Ian Freckelton and Hugh Selby (eds), Thomson Reuters, *Expert Evidence* (at August 2011) ch 99; Geoffrey Stewart Morrison, 'Forensic Voice Comparison and the Paradigm Shift' (2009) 49 *Science and Justice* 298. Gary Edmond is currently engaged, with Geoffrey Stewart Morrison and others, in a research project sponsored by the Australian Research Council entitled 'Making Demonstrably Valid and Reliable Forensic Voice Comparison a Practical Everyday Reality in Australia'.

reduce some of the pre-modern commitments that continue to haunt contemporary legal experience and practice. Incriminating voice comparison evidence should be supported by empirical research that indicates that particular types of analytical practice, and the opinions derived from them, are *demonstrably reliable*.[280]

D *Voice Identification Parades for Those Who Become Familiar after the Fact*

Even without demonstrably reliable techniques, we could enact procedures that would reduce some of the most egregious aspects of voice comparison by those involved in investigations and translations. The value of voice identification evidence would be dramatically improved by the introduction of voice parades.

There is a long history of eyewitness identification parades or line-ups around the world and in Australia (under both the common law and the UEA), and they are the preferred method in relation to visual identification evidence under the UEA.[281] The use of identification parades has been informed by an extensive empirical literature investigating the strengths and weaknesses of procedures.[282] A similar, if smaller, research base exists (and could be extended) to inform voice identification parades.[283] However, concerns about preserving the accuracy and improving the assessment of voice identification evidence do not appear to have reached the same level as those exhibited in relation to visual identification and identifications derived from images. This is unfortunate, given the benefits that properly constructed voice identification parades might offer, particularly with regard to the challenges and dangers arising from 'ad hoc expert' testimony.[284]

---

[280] See Edmond and Roach, above n 4; Edmond, 'Specialised Knowledge', above n 93.

[281] See UEA ss 114, 115(5)–(6), though again it is important to note that such procedures do not apply to displaced viewers.

[282] See the discussion of this literature in Gary L Wells and Deah S Quinlivan, 'Suggestive Eyewitness Identification Procedures and the Supreme Court's Reliability Test in Light of Eyewitness Science: 30 Years Later' (2009) 33 *Law and Human Behavior* 1 and Gary L Wells et al, 'Eyewitness Identification Procedures: Recommendations for Lineups and Photospreads' (1998) 22 *Law and Human Behavior* 603.

[283] Such parades have been used (or recommended) in several cases, though primarily in relation to direct earwitnesses: *R v Callaghan* (2001) 4 VR 79, 84 [9] (Winneke P); *R v Daley* [2002] NSWSC 279 (14 September 2001) [165]–[174] (Simpson J); *R v Golledge* [2007] QCA 54 (2 March 2007) [33] (Keane JA); *Harris* [1990] VR 310, 314 (Ormiston J); *Burrell v The Queen* (2009) 196 A Crim R 199, 211 [62] (Beazley JA, Grove and Howie JJ). However, some judges have expressly dismissed the need for voice parades for earwitnesses (and, implicitly, for 'ad hoc experts'), even though identification parades are routinely used for eyewitnesses: *R v Jones* (1989) 41 A Crim R 1, 7 (Young CJ, Gobbo and Nathan JJ). See also *R v Smith* [1984] 1 NSWLR 462, 479 (O'Brien CJ Cr D); *R v Miladinovic* (1992) 109 ACTR 11, 16 (Miles CJ); *Li* (2003) 139 A Crim R 281, 289 [60] (Ipp JA); *Irani v The Queen* (2008) 188 A Crim R 125, 129 [16]–[19] (Hoeben J). Interestingly, in *Neville* [2004] WASCA 62 (2 April 2004) [35]–[36] (Miller J), [88] (Heenan J) and *Hirst v The Police* [2005] SASC 201 (2 June 2005), identification parades are discussed for eyewitnesses but ignored in relation to the voice identification evidence. Here, our discussion is restricted to 'ad hoc experts' and formally qualified 'experts' (such as linguists) who are not actually specialists in voice comparison.

[284] While we do not endorse their recommendations wholesale, see A P A Broeders and A G van Amelsvoort, 'A Practical Approach to Forensic Earwitness Identification: Constructing a Voice Line-Up' (2001) 47 *Problems of Forensic Sciences* 237 for a detailed consideration of the appli-

It is both theoretically and practically desirable to subject displaced (or indirect) listeners such as police officers and interpreters (hereafter 'investigative familiars') to voice parades,[285] just as it is possible to use such identification procedures with traditional eyewitnesses.[286] By doing so it is possible, if the parade is adequately constructed, to remove some of the previously discussed threats to the value of the comparison. First, having an investigative familiar listen to an assortment of different voices[287] and attempt to identify the voice which produced the incriminating utterance provides an indication of the likely accuracy of that identification and the strength of the suspicion. If the investigative familiar selects the voice of the suspect rather than a parade 'filler' (ie known innocent), their identification of the suspect as the speaker of the incriminating speech has substantially higher probative value than the 'identifications' currently being proffered in trials. Such selections also provide independent support for ongoing investigations.

Moreover, if the identification parade is presented to the investigative familiar in a fashion such that neither the witness nor the parade administrator knows which voice belongs to the suspect (ie a double blind procedure), it is possible to sanitise the identification of any corrosive contamination or confirmation bias, irrespective of the context in which the original 'witnessing' occurred, thereby making the identification *independent*. This is because while the witness may know that the police think person X committed crime Y, such knowledge cannot affect the witness's ability to recognise or 'know' a previously heard voice when presented with it. The voice of the suspect either is or is not the voice the witness heard, and the witness either is or is not able to recognise it from the voices they are presented with. The beliefs held by the police regarding the guilt or innocence of the suspect are of no consequence in a double-blind identification procedure. It is, however, important to be aware that the perpetrator of the crime in these instances of investigative familiarity is likely to be one of very few potential suspects (ie speakers of a certain language, visitors to a specific (monitored) location, recipients of calls from impugned numbers). In such circumstances, as with parades more generally, it is vital to construct the procedure in such a way that the fillers share sufficient characteristics with *descriptions* of the suspect, so that any voice could potentially be the voice of the perpetrator (eg they all speak the same dialect of Cantonese); however, the fillers should not be chosen based on their *similarity* to the voice of the suspect, as this

cability of the eyewitness identification procedures to earwitness evidence. See also Francis Nolan, 'A Recent Voice Parade' (2003) 10 *Forensic Linguistics* 277.

[285] Investigative familiars are not necessarily familiar in the sense of being able to make accurate categorical ascriptions, but rather they are those who are not complete strangers because they have satisfied some threshold of exposure — however limited — during the course of an investigation.

[286] Our one caveat is that individuals associated with an investigation should not be gratuitously exposed to recordings of incriminating voices merely to increase the chances of obtaining a positive identification. All voice identification parades should be disclosed to the defence.

[287] One of the voices is usually the voice of the person thought to have created the incriminating speech. The remainder of the voices would be known innocent foils.

would produce a parade of 'clones' and would make the comparison task unrealistically difficult.[288]

Voice parades might even help to resolve questions regarding the accuracy and validity of cross-lingual identifications. If, for example, the witness hears incriminating speech in Cantonese, and the police interview the suspect in English, English speech samples provided by a number of native Cantonese speakers could be used in the voice parade. Thus the analyst (here, most often an interpreter) could demonstrate that there are sufficient language-independent cues for them to recollect (or recognise) a speaker in the absence of any explicit knowledge of the speaker's status in the investigation. If the witness is able to do this, the issue of never having heard verified samples of the perpetrator's speech across languages is irrelevant because the witness has demonstrated that elements of the speech are consistent enough for the benefits derived from familiarity to be preserved.

Like analogous developments with eyewitness evidence, voice parades might substantially improve our understanding of the value of identification evidence. Requiring investigative familiars purporting to give positive identification evidence (or describe similarities) to successfully complete a voice parade before being entitled to express their opinions would reduce some of the most undesirable dimensions of current practice.[289] Parades might not, however, guarantee ability, and where the number of participants is small there remains a real risk of chance selection or selection based on the voice that is most similar to that remembered. Notwithstanding the potential for voice parades to improve the quality of voice-related evidence, *the strong preference must be for validated and reliable scientific voice comparison techniques*.

E *Discussion*

Generally, if voice identification evidence is not derived via direct (ie sensory) witnesses, familiars or experts with demonstrably reliable techniques (and without suggestion), in the vast majority of circumstances it should not be admitted. At the very least, investigators, interpreters and linguists should not be allowed to express their opinions about identity or similarities at trial unless they have been exposed to a considerable amount (ie many hours) of the voice in the conditions in which the comparison will be undertaken and as part of their routine duties,[290] and only where the identity was not suggested or disclosed. Even so, there should always be a very strong preference for lay witnesses with a high level of familiarity, for methods that do not depend upon the interpretations of investigators, and for investigators to demonstrate their ability in a voice

---

[288] C A Elizabeth Luus and Gary L Wells, 'Eyewitness Identification and the Selection of Distracters for Lineups' (1991) 15 *Law and Human Behavior* 43.

[289] The use of parades might help to sanitise otherwise odious 'expert' opinions, although the admissibility pathway for the opinions would remain problematic.

[290] As opposed to merely repeatedly listening in order to make a comparison or being asked about the identity of a voice with which they may have some limited familiarity.

parade.[291] The preparation of transcripts — whether in English or some other language — should not generally qualify a person to express an opinion about identity. The risks are so great and the difficulty of effectively exploring and challenging such ipse dixit is so pronounced that such practices should not be accommodated by legal institutions purporting to dispense justice. Opinion evidence from these sources, or derived in these ways, should not be admitted. While the ipse dixit of experts is unacceptable, the ipse dixit of investigators (as 'ad hoc experts') verges on scandalous.

We accept that in some circumstances, especially where, as in *R v El-Kheir*,[292] the voice could only have been that of one of a limited number of individuals, the exercise is different to that where the range of speakers is large or unconstrained.[293] Nevertheless, dangers and risks persist. Correctly identifying a speaker will not always equate to proof of guilt. In *R v El-Kheir*, for example, it is possible that a person visiting the house when a covert surveillance operation and police drug raid occurred, who was recorded speaking to the owner of the house on a hidden microphone, may not have been implicated in the importation. Sometimes there will be controversy not only about the identity of the speaker but also about the precise meaning of allegedly incriminating words.[294] Where the recording is poor and the meaning of words is credibly contested there is a danger that mere association may be equated with guilt.

Voice comparison by strangers tends to be error-prone, with error rates likely to increase significantly over time. Desirable as it may seem to allow direct witnesses to testify, ideally only factual descriptions and opinions about identity or features of a voice expressed roughly contemporaneously should be admissible. Descriptions and comparisons should be obtained in a neutral manner and as close in time to the actual event(s) as possible, otherwise the value of the description or opinion, regardless of the apparent credibility of the witness, is likely to be limited, and far more limited than the tribunal of fact is likely to appreciate. Allowing earwitnesses and investigators to express opinions in circumstances that do not take account of scientifically notorious frailties subverts the accuracy of legal processes and substantially increases the risk of convicting an innocent person.

Most of these problems are not as applicable to the identification evidence of those who are very familiar with the accused.[295] In general, 'true' familiars should be allowed to express opinions, including positive opinions about identity, as well as to give direct evidence of non-deliberative recognition. Both forms of evidence should, in the normal course of affairs, be admissible. While

---

[291] On police familiarity, see *Miladinovic v The Queen* (1993) 47 FCR 190; *R v Leaney* [1989] 2 SCR 393, 403–5 (Lamer J), 408–12 (Wilson J), 413 (McLachlin J).

[292] [2004] NSWCCA 461 (20 December 2004).

[293] See also the concerns raised by Simpson J in *R v Leung* (1999) 47 NSWLR 405, 414 [45] about potential contamination in relation to comparisons made where the identity of the suspect remains open, and the related discussion in *Li* (2003) 139 A Crim R 281, 289 [58]–[60] (Ipp JA). See the discussion in above n 145.

[294] See, eg, *Dodds v The Queen* (2009) 194 A Crim R 408; *Nguyen v The Queen* (2007) 173 A Crim R 557.

[295] Of course, factors such as size of sample and quality of recording may still be important.

obviously not infallible, the value of such evidence is generally warranted by experience as well as by replicated scientific research.[296]

IX SILENCE IN COURT?

Recently, after a long inquiry, an eminent group of scientists, mathematicians and engineers, joined by a few senior lawyers and judges, reported to Congress on the condition of the forensic sciences in the United States. Their findings were both surprising and disconcerting. They concluded that

> [w]ith the exception of nuclear DNA analysis … no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source. …
>
> The law's greatest dilemma in its heavy reliance on forensic evidence … concerns the question of whether — and to what extent — there is *science* in any given forensic science discipline.[297]

These concerns are generally applicable to the forensic sciences in Australia and to most of the methods of voice comparison and voice identification currently used by displaced listeners and investigative familiars accepted by Australian courts. We must have very serious misgivings about the foundations and reliability of purportedly expert voice identification evidence, particularly its non-institutionalised and ad hoc varieties.[298]

Notwithstanding, or perhaps because of, the lack of specialised knowledge in most areas of forensic voice comparison, our judges have, quite perversely, developed jurisprudence and practices that enable those without relevant training, study, experience or demonstrated ability and who have not given attention to relevant scientific research to, nevertheless, express their incriminating opinions in circumstances where the identity of the speaker is quite often the ultimate issue. Those without demonstrated proficiency are magically transformed into experts for the purpose of litigation. Moreover, lay jurors unfamiliar with the accused, their voice and even their language may be asked to compare voices speaking in different languages and under different conditions. These practices are not conducive to a fair trial or an accurate verdict.

---

[296] Such evidence will be relevant and admissible as fact if it is non-reflective recognition, and as opinion if it is 'specialised knowledge' based on considerable 'experience' (ie familiarity). The same cannot be said for the evidence of investigators and interpreters whose expertise and experience is not in voice comparison or whose familiarity is derived solely from participation in the investigation at hand.

[297] Committee on Identifying the Needs of the Forensic Science Community, Committee on Science, Technology, and Law Policy and Global Affairs and Committee on Applied and Theoretical Statistics (Division on Engineering and Physical Sciences), National Research Council, *Strengthening Forensic Science in the United States: A Path Forward* (National Academies Press, 2009) 7, 9 (emphasis in original).

[298] See Gary Edmond, 'Impartiality, Efficiency or Reliability? A Critical Response to Expert Evidence Law and Procedure in Australia' (2010) 42 *Australian Journal of Forensic Sciences* 83; Gary Edmond, 'Actual Innocents? Legal Limitations and Their Implications for Forensic Science and Medicine' (2011) 43 *Australian Journal of Forensic Sciences* 177.

Our lawyers (particularly prosecutors) and judges have been remarkably inattentive (or resistant) to the results of empirical research.[299] Even though comparison of sounds and identification from sounds is, in many situations, even less reliable than comparison or identification in relation to vision and images, judges have tended to adopt a less interventionist approach to voice evidence. Our current laws seem to admit much incriminating opinion evidence in circumstances where it is not clear that the frailties of the evidence are adequately recognised, let alone conveyed. Lawyers and judges do not cite, and very rarely refer to, relevant empirical and experimental literature. Rather, they tend to rely upon unsystematic impressions and experiences and the rather random way in which weaknesses and limitations may or may not be exposed and considered during trials and appeals.

Without wanting to suggest that the empirical literature will provide a straightforward or unambiguous basis for legal practice, it would seem that relevant expert literature could help to guide and improve practice and correct a range of strange anomalies and beliefs about both human perceptions and the ability of the adversarial trial and its safeguards to substantially address problems with sounds, voices and comparisons.

Interestingly, earlier concerns about dangers with voice comparison, the potential for prejudicial effects associated with investigators (including apparently well-intentioned investigators), and the manner in which voice identification evidence was obtained seem to have been largely abandoned. Here, the Victorian common law seems to offer something of a limited exception and example. Notably, in *Harris*, while Ormiston J effectively rejected the more demanding New South Wales requirements for voice identification evidence, he nevertheless excluded the evidence of a police officer, who had listened to hundreds of tape recordings, because of her limited familiarity with the accused and the suggestive manner in which she was initially introduced to the recordings. Detective Sergeant Corrie had had some exposure to the various accused, and much more exposure than encountered in many recent cases from New South Wales. Nevertheless, Ormiston J concluded that

> there was so much suggestive, direct and indirect, material involved in Miss Corrie's doubtless honest attempt at identification, that it should be excluded from evidence in the exercise of my discretion, for this is a kind of prejudice which cannot be removed at the trial merely by cross-examination or by other evidence. Merely because she is a police, and not a lay, witness can make no difference, nor the fact that she has heard the voices and the tapes many times thereafter. …
>
> In the end … the probative value of Miss Corrie's identification is too speculative and too overlaid with other material to allow it to be led before the jury, who may be irrationally impressed by it. The existence of other materials may indeed obscure the inherent weakness of her evidence, but it may be hard to

---

[299] However, the use and gradual improvement of identification parades and the considered response to empirical research in *Winmar v Western Australia* (2007) 35 WAR 159 suggest that (mediated) engagement is at least possible. See also the detailed discussion of empirical research in *R v Henderson* [2010] 2 Cr App R 24.

> persuade the jury that they should put out of mind what may appear to be a straightforward identification …[300]

We might note that instructions and warnings were apparently insufficient to overcome the defects and 'the often-praised commonsense of juries' to which Ormiston J had earlier alluded.[301] Ormiston J thought that the danger was of

> a jury being 'irrationally impressed' by certain identification evidence which is a proper discretionary basis for excluding some of that evidence where the means adopted are conducive to drawing false or unreliable and thus misleading conclusions.[302]

Without reference to relevant scientific research, Ormiston J adopted a cautious and exclusionary approach to voice identification and its potentially prejudicial effects.[303] This protective attitude, concerned with accuracy and fairness, seems to have lapsed in recent years (especially in New South Wales). It has lapsed in ways that appear inconsistent with substantial concerns about accuracy and fairness as well as with the results of ongoing scientific research programmes. Few judges now exclude voice comparison or identification evidence using admissibility rules or discretionary (or mandatory) exclusions.[304]

We can only wonder why legal practice is inconsistent with what is known. We can only speculate about why visual evidence is more regulated than forms of voice evidence. Evidently, both are error-prone. Our anxieties are accentuated by inconsistencies which systematically assist the state and subvert espoused principles of evidence law and criminal justice.

What we should do is yet another problem. It appears to us that we need to continuously refine practice in ways that accommodate and recognise the knowledge developed in other fields. Centuries ago, Saunders J declared that

> if matters arise in our law which concern other sciences or faculties, we commonly apply for the aid of that science or faculty which it concerns. Which is an honourable and commendable thing in our law. For thereby it appears that we don't despise all other sciences but our own, but we approve of them and encourage them as things worthy of commendation.[305]

Where long traditions and practices, such as placing confidence in lay abilities or juries, are threatened, we need to have multidisciplinary conversations about how the goals of criminal justice can be facilitated through revised practices and procedures. The social legitimacy of the courts can only be maintained through

---

[300] *Harris* [1990] VR 310, 322–3. Cf the more accommodating response in *Irani v The Queen* (2008) 188 A Crim R 125.

[301] *Harris* [1990] VR 310, 316.

[302] Ibid 319. Consider the concern expressed by the High Court in the early visual identification case of *Davies v The King* (1937) 57 CLR 170, 181 (Latham CJ, Rich, Dixon, Evatt and McTiernan JJ).

[303] See also *Rich* [2008] VSC 436 (23 October 2008) [38]–[41] (Lasry J); *Chedzey v The Queen* (1987) 30 A Crim R 451, 464 (Olney J).

[304] See Smith and Odgers, above n 93; Edmond, 'Specialised Knowledge', above n 93.

[305] *Buckley v Thomas* (1554) 1 Plowd 118, 125; 75 ER 182, 192.

the incorporation of exogenous knowledge, however disruptive or unsettling that may be.

In the interim, in the absence of evidence of ability and reliability, prosecutors and judges should be far more reticent about adducing and admitting the opinions of non-familiar witnesses. Until we have empirically-informed responses to our epistemic and legal infirmities, Australian courts should be a little quieter, though substantially more sound.