'Expert' opinion evidence, 'junk science' and the law

Professor Gary Edmond FRSN FAAL
Faculty of Law and Justice, UNSW, Sydney
Director, Program in Evidence, Expertise and Law
Chair, Evidence-based Forensics Initiative

g.edmond@unsw.edu.au

Never Stand Still

1. Introduction

2. Evaluating forensic science evidence

3. *Lang v R*: A recent Australian example

4. Critical observations

Reasons for receiving expert evidence – epistemic

'… if matters arise in our laws which concern other sciences and faculties we commonly call for the aid of that science or faculty which it concerns, which is an honourable and commendable thing for thereby it appears that we do not despise all other sciences but our own, but we approve of them and encourage them.'

Buckley v Rice Thomas (1554) 1 Plowden 118

'On certain matters, such as those of science or art, upon which the court itself cannot form an opinion, special study, skill or experience being required for the purpose, 'expert' witnesses may give evidence of their opinion.'

Folkes v Chadd (1782) 3 Doug KB 157

'An expert's opinion is admissible to furnish the court with scientific information which is likely to be outside the experience and knowledge of a judge or jury. If on the proven facts a judge or jury can form their own conclusions without help, then the opinion of an expert is unnecessary.'

R v Turner [1975] QB 834, 841
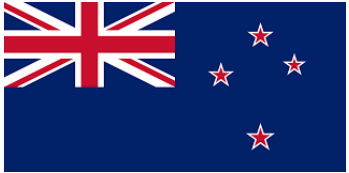
Adversarial basics: admissibility of expert evidence

Evidence Act 2006 (NZ)

25 Admissibility of expert opinion evidence

(1) An opinion by an expert that is part of expert evidence offered in a proceeding is admissible if the **fact-finder is likely to obtain substantial help** from the opinion in understanding other evidence in the proceeding or in ascertaining any fact that is of consequence to the determination of the proceeding.

Australian common law (*Bonython v R*)

'[The] judge must consider and decide two questions. … [first]: (a) whether the subject matter of the opinion is such that a person without instruction or experience in the area of knowledge or human experience would be able to form a sound judgment on the matter **without the assistance** of witnesses possessing special knowledge or experience in the area, and (b) whether the subject matter of the opinion **forms part of a body of knowledge or experience which is sufficiently organized or recognized to be accepted as a reliable body of knowledge or experience** … The second question is whether the witness has acquired by study or experience sufficient knowledge of the subject **to render his opinion of value** in resolving the issues before the court.'

Adversarial basics: admissibility of expert evidence

Uniform evidence law (UEL)

79 Exception to the opinion rule

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

'The focus must be on the words "specialised knowledge", not on the introduction of an extraneous idea such as "reliability".'
(*R v Tang*, 2006, Spigelman CJ; *Tuite v The Queen*, 2015)

'Section 79 is not concerned with reliability of the expert's opinions'.
(*Chen v R,* 2018, [62])

Some shared expectations in common law systems

Expert opinion evidence in criminal proceedings, should:

- assist trier of fact.
- be based on training, study or experience.
- be proffered by a witness with specialised skill or knowledge.
- disclose information about procedures, assumptions, limitations and quality assurance.
- provide reasons or basis for opinion.
- be impartial.
- be 'reliable' or based on reliable principles – though not necessarily in Australia.
- be capable of evaluation by an independent lay tribunal of fact – whether jury, trial judge or appellate court.

Adversarial basics: expert evidence

'It is a primary duty imposed on experts in giving opinion evidence to furnish the trier of fact with the criteria to enable the evaluation of the expert conclusion: *Makita (Australia) Pty Ltd v Sprowles* (2001). … The 'bare *ipse dixit*' of a scientist upon an issue in controversy should carry little weight: *Davie v Magistrates of Edinburgh* (1953).'

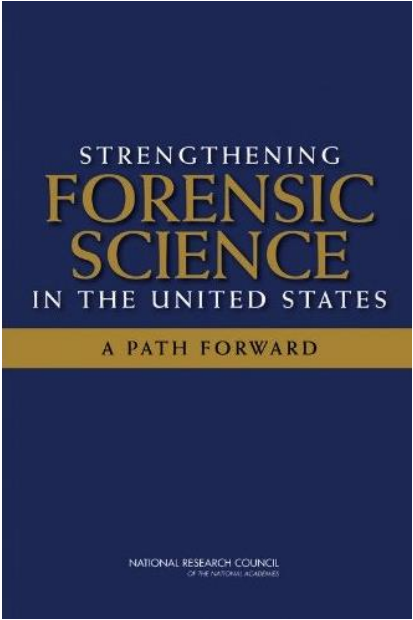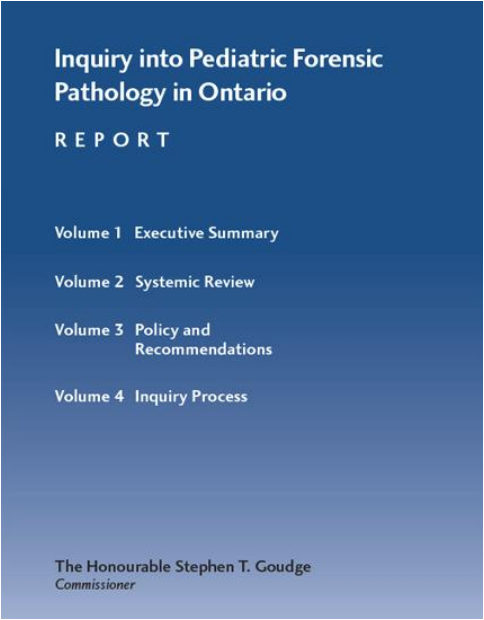*Hillstead v The Queen* [2005] WASCA 116, [48].

See also 'Forensic science evidence and the conditions for rational (jury) evaluation' (2015) 39 *Melbourne University Law Review* 75-123.

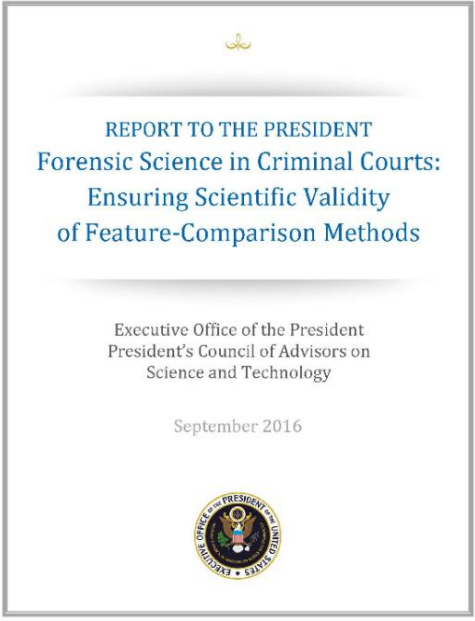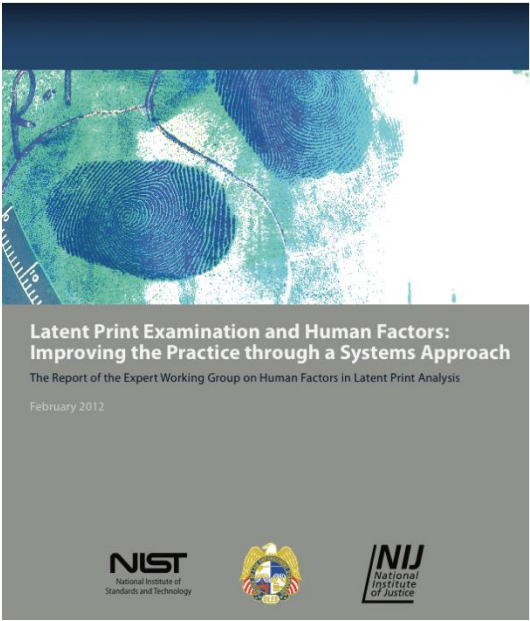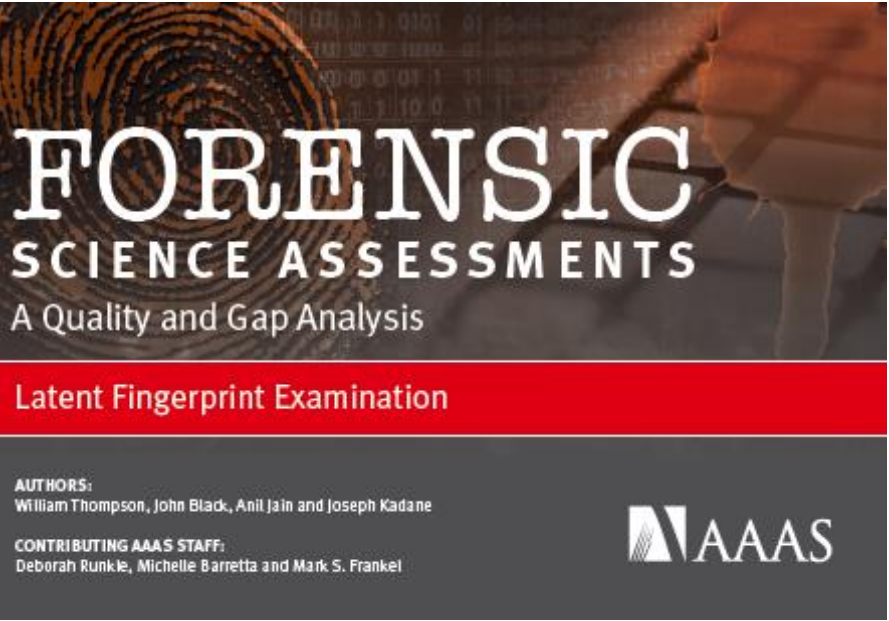2. **Evaluating forensic science evidence (according to mainstream scientists and peak scientific organisations)**

Scientific 'benchmarks'

Inquiry into Pediatric Forensic Pathology in Ontario
REPORT
Volume 1 Executive Summary
Volume 2 Systemic Review
Volume 3 Policy and Recommendations
Volume 4 Inquiry Process
The Honourable Stephen T. Goudge
Commissioner

STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES
A PATH FORWARD
NATIONAL RESEARCH COUNCIL OF THE NATIONAL ACADEMIES

Forensic Science Regulator
Code of Practice
March 2023
Version 1

NIST
FORENSIC SCIENCE

FORENSIC SCIENCE ASSESSMENTS
A Quality and Gap Analysis
Latent Fingerprint Examination
AUTHORS:
William Thompson, John Black, Anil Jain and Joseph Kadane
CONTRIBUTING AAAS STAFF:
Deborah Runkle, Michelle Barretta and Mark S. Frankel
AAAS

Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach
The Report of the Expert Working Group on Human Factors in Latent Print Analysis
February 2012
NIST National Institute of Standards and Technology
NIJ National Institute of Justice

The Fingerprint Inquiry | Scotland
The Fingerprint Inquiry Report
Volume 1

REPORT TO THE PRESIDENT
Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods
Executive Office of the President
President's Council of Advisors on Science and Technology
September 2016

Focusing on feature comparison methods

'A forensic feature-comparison method is a procedure by which an examiner seeks to determine whether an evidentiary sample (e.g., from a crime scene) is or is not associated with a source sample (e.g., from a suspect [or a weapon]) based on similar features. The evidentiary sample might be DNA, hair, fingerprints, bitemarks, toolmarks, bullets, tire tracks, voiceprints, visual images, and so on.' (PCAST, 46)
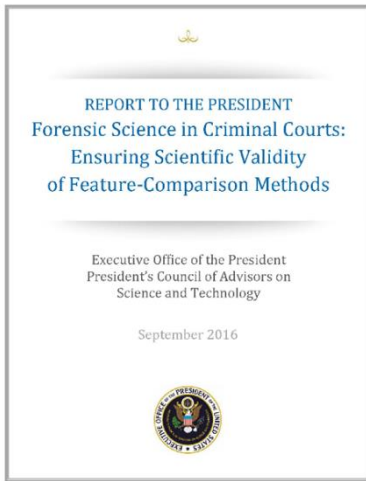
Forensic science: Guiding principles

'Two very important questions should underlie the law's admission of and reliance upon forensic evidence in criminal trials:

(1) the extent to which a particular forensic discipline is founded on a **reliable scientific methodology** that gives it the capacity to accurately analyze evidence and report findings [ie **validation**], and

(2) the extent to which practitioners in a particular forensic discipline rely on human interpretation that could be tainted by **error**, the threat of **bias**, or the absence of sound operational procedures and robust performance **standards**.' [ie **error, cognitive bias and standards**]

National Academy of Sciences, *Strengthening Forensic Science in the United States* (2009) ('NAS' or 'NRC' report)
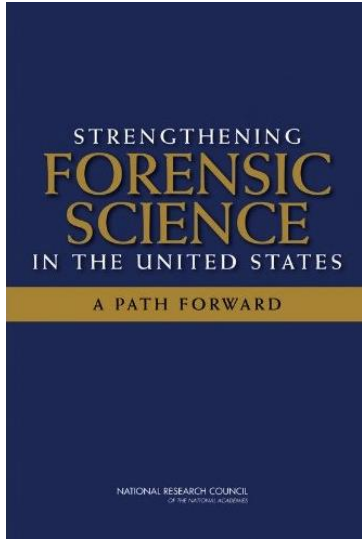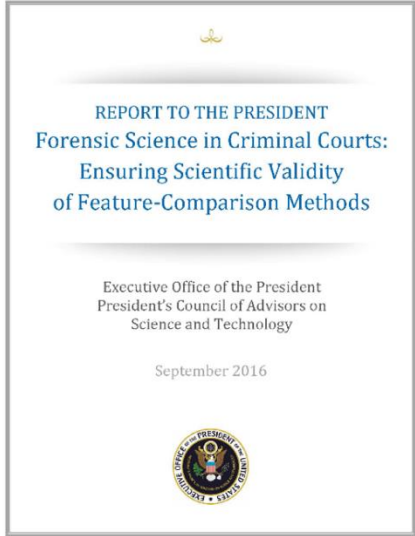
Validation

'… valid scientific knowledge can *only* be gained through *empirical* testing of specific propositions.'

'… methods must be presumed to be unreliable until their foundational validity has been established based on empirical evidence.'

Does the method work? In what conditions? How well? How do we know? Was the method applied appropriately in this case?

REPORT TO THE PRESIDENT
Forensic Science in Criminal Courts:
Ensuring Scientific Validity
of Feature-Comparison Methods

Executive Office of the President
President's Council of Advisers on
Science and Technology

September 2016

Are forensic feature comparison methods* valid?
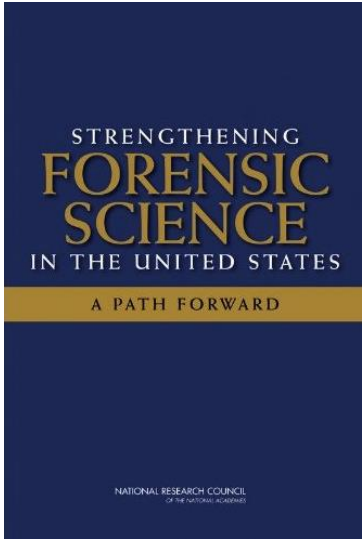
'With the exception of nuclear DNA analysis, however, no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source.'
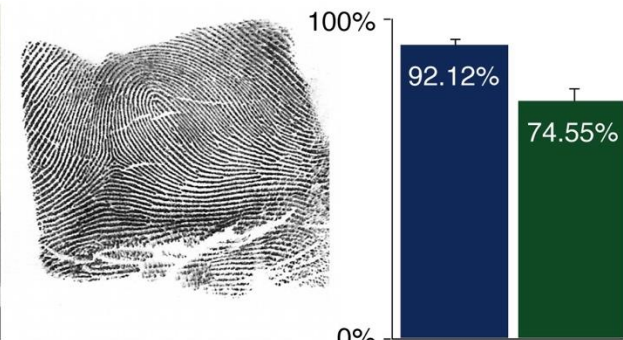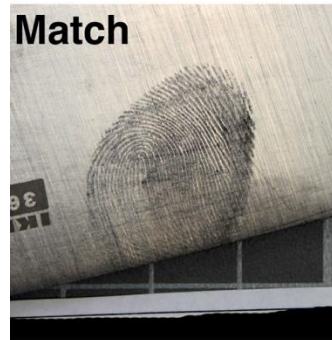
National Academy of Sciences, *Strengthening Forensic Science in the United States* (2009)

* e.g. latent fingerprint, shoe, foot, ear, tyre, bite mark, hair, fibre, soil, ballistic, toolmark, handwriting, bloodstain, voice and image comparisons, manner of death (for forensic pathology) and so on.
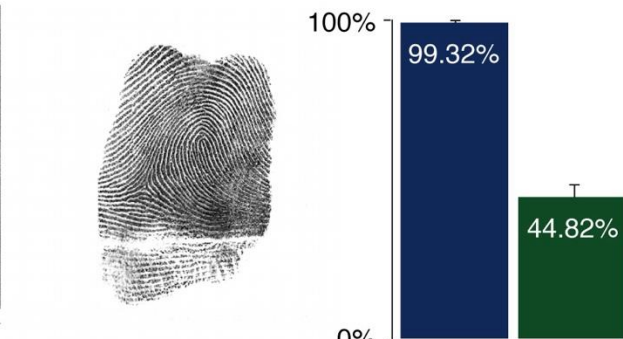
An early validation study for latent fingerprint comparison (2011)
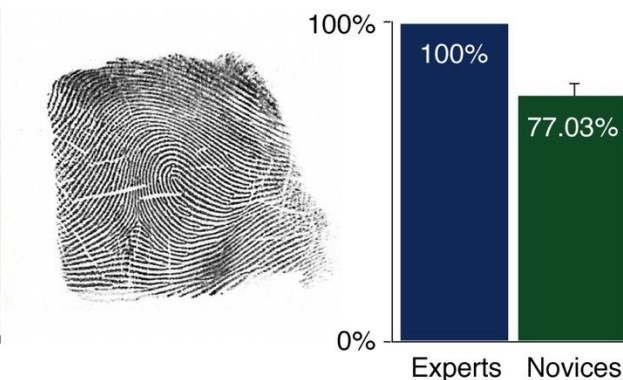

Matching prints

Similar non-matching prints (risk of false positives, especially for 'green' judges and jurors)

Non-similar non-matching prints

Navy – examiners
Green – students

Match

Similar

Not Similar

92.12%   74.55%
99.32%   44.82%
100%   77.03%

Experts   Novices

Jason Tangen, Matthew Thompson, and Duncan McCarthy, 'Identifying fingerprint expertise' (2011) 22 *Psychological science* 995.

Other validation studies for fingerprints (post NAS inquiry and report)

REPORT TO THE PRESIDENT
Forensic Science in Criminal Courts:
Ensuring Scientific Validity
of Feature-Comparison Methods

Executive Office of the President
President's Council of Advisors on
Science and Technology

September 2016
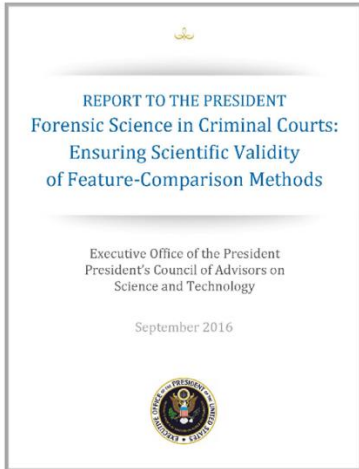
Study	False Positives			
	Raw Data	Freq. (Confidence bound)	Estimated Rate	Bound on Rate
Early studies				
Langenburg (2009a)	0/14	0% (19%)	1 in ∞	1 in 5
Langenburg (2009b)	1/43	2.3% (11%)	1 in 43	1 in 9
Langenburg et al. (2012)	17/711	2.4% (3.5%)	1 in 42	1 in 28
Tangen et al. (2011) ("similar pairs")	3/444	0.68% (1.7%)	1 in 148	1 in 58
Tangen et al. (2011) ("dissimilar pairs")	0/444	0% (0.67%)	1 in ∞	1 in 148
Black-box studies				
Ulery et al. 2011 (FBI)**	6/3628	0.17% (0.33%)	1 in 604	1 in 306
Pacheco et al. 2014 (Miami-Dade)	42/995	4.2% (5.4%)	1 in 24	1 in 18
Pacheco et al. 2014 (Miami-Dade) (excluding clerical errors)	7/960	0.7% (1.4%)	1 in 137	1 in 73

PCAST, 98

Using image (or facial) comparison for purposes of identification as an example

What do we need?
(yes, it's a question for you to think about)

Images from *R v Tang* (2006)

Validity of image comparisons (using passport quality photos)

Try the Glasgow Face Match Test (GFMT) online. Ordinary people make errors, when asked if these are the same person, about 20-25% of the time.

- Ask examiners (participants) to determine whether images are the same person or different persons over and over. (It is imperative that those managing the testing know the correct answers.)
- What tools, equipment or resources should examiners have during the testing? (Information relevant to task and, ideally, their normal tools.)
- What should the examiners know about the case?
- What do the tests reveal and what are their benefits?
- What are the limits of such validation studies?

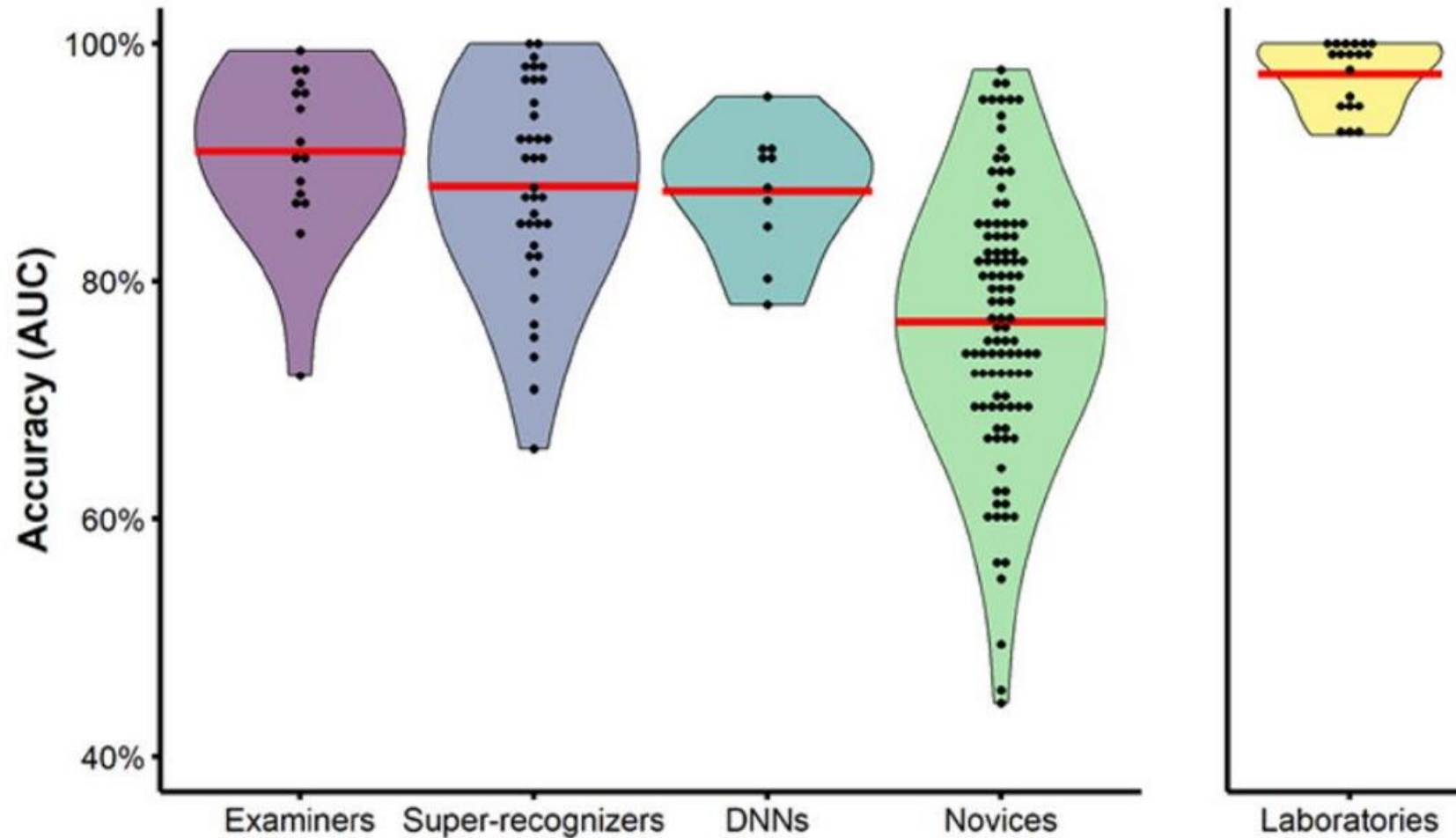# A recent validation study of facial comparisons based on moderate quality images from the 'wild'

Change of appearance?

Images from social media

A. Towler et al, 'Diverse types of expertise in facial recognition' (2023) 23 *Sci. Rep.* 11396.

Results of (a) validation study (moderate quality images): Who might testify?

Examiners – those trained and working in specialist units (and routinely tested).
DNN – deep neural networks (algorithms).
Novices – ordinary persons.
Super-recognisers – persons with demonstrably superior innate abilities.

A. Towler et al, 'Diverse types of expertise in facial recognition' (2023) 23 *Sci. Rep.* 11396.

Validity as applied (CCTV security images)

Does the foundational validity of the study of photo comparison with medium to good full-frontal images apply to comparison and analysis of these images (from *Tang v R*)?
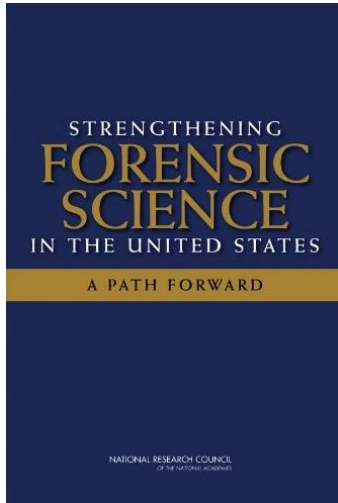
Standards

'Often there are no standard protocols governing forensic practice in a given discipline. And, even when protocols are in place, they often are vague and not enforced in any meaningful way. … These shortcomings obviously pose a continuing and serious threat to the quality and credibility of forensic science practice.'

Standards should be based on validation testing. They provide a clear indication of what information is required, and what is not required (and should be avoided) during any analysis.

**National Academy of Sciences, *Strengthening Forensic Science in the United States* (2009)**

Standards

- Is there a specific procedure (ie method or protocol) based on validation?
- What is it, and was it scrupulously followed? (For example, should features in the 'unknown' image – i.e. face – be documented before the reference image is examined and compared?)
- Is there an identifiable image quality or set of conditions or features that is required for the comparison? (e.g. does the image need to incorporate the full face, and can the examiner rely on gait or clothing?)
- Does image 'enhancement' work or matter?
- Can examiner access case information? (Is it documented?)

Accuracy, uncertainty and errors
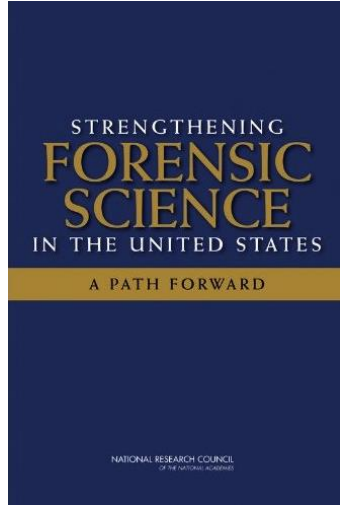
'Few forensic science methods have developed adequate measures of the accuracy of inferences made by forensic scientists. All results for every forensic science method should indicate the uncertainty in the measurements that are made, and studies must be conducted that enable the estimation of those values.'

National Academy of Sciences, *Strengthening Forensic Science in the United States* (2009).

Expressing errors and limitations (fingerprints)

'PCAST finds that latent fingerprint analysis [has] *a false positive rate that is substantial* and **is likely to be higher than expected by many jurors** based on longstanding claims about the infallibility of fingerprint analysis. The false-positive rate could be as high as 1 error in 306 cases based on the FBI study [Ulery et al] and 1 error in 18 cases based on a study by another crime laboratory [Miami-Dade]. **In reporting results of latent-fingerprint examination, it is important to state the false-positive rates based on properly designed validation studies.**'
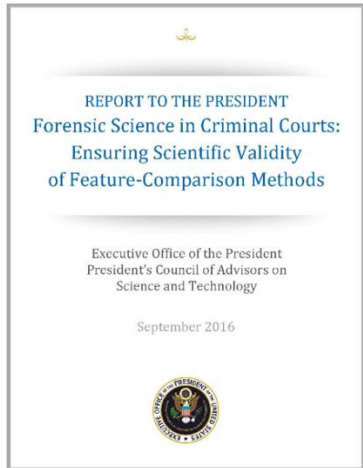
PCAST, 9-10

Terminology and expression

'many terms are used by forensic examiners in reports and in court testimony … Such terms include … "match," "consistent with," "identical," "similar in all respects tested," and "cannot be excluded as the source of." … the forensic science disciplines have not reached agreement or consensus on the precise meaning of any of these terms. This imprecision in vocabulary stems in part from the paucity of research'.
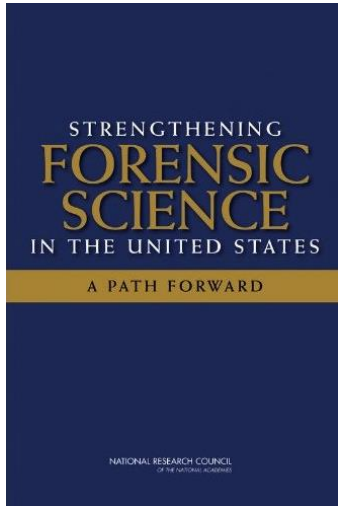
National Academy of Sciences, *Strengthening Forensic Science in the United States* (2009).

# Expression of comparison opinions - cacophony

- **DNA profiles**: statistical (frequentist or likelihood ratios). (Validated – derived from mainstream scientific research)
- **Latent fingerprints**: positive ID (i.e. individualisation). (PV)
- **Ballistics**: individualisation to a specific weapon. (PV)
- **Incriminating images** (e.g. CCTV): Individualisation and similarities. (PV)
- **Shoe prints**: varies, often now framed in Bayesian terms – e.g. strong support. (NV/PV)
- **Blood spatter and stain interpretation**: varies from positive assertions to qualified, but high error rates (NV/PV)
- **Incriminating voice recordings**: individualisation, even across languages. (NV)
- **Microscopic hair comparison**: historically, very similar or the same and cannot exclude (Invalid)
- **Bite marks**: historically individualisation. (Invalid)

V – validated; PV – partially validated; NV – not validated

Expression of results and accuracy (image comparison)

- How should the analyst's opinion be expressed? (Not on the basis of what they might be willing to opine.)
- Expression should be guided by the results of foundational validation studies. These are **indicative**, but provide an empirically-based framework for expression and evaluation.
- The foundational validation studies we have are based on moderate to good quality images, so we should not simply extrapolate to low quality images (as in *Tang*), for example.
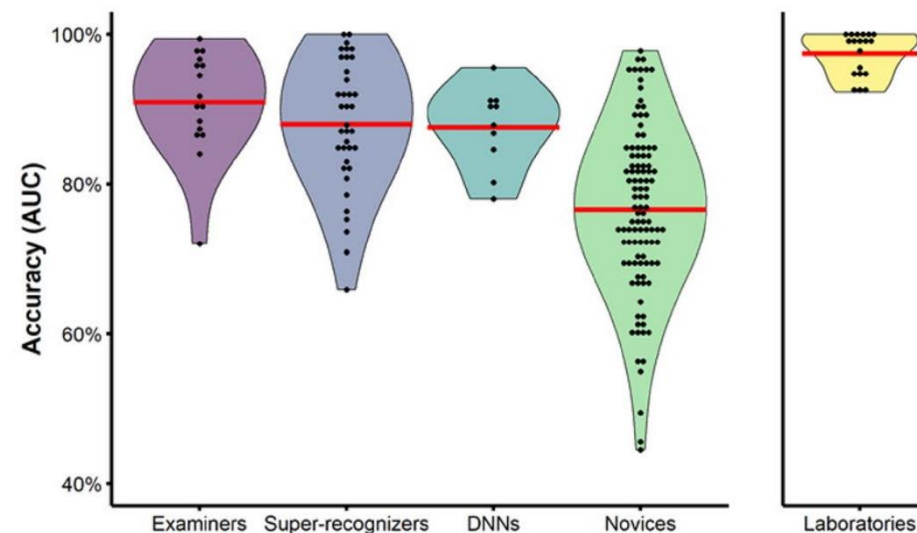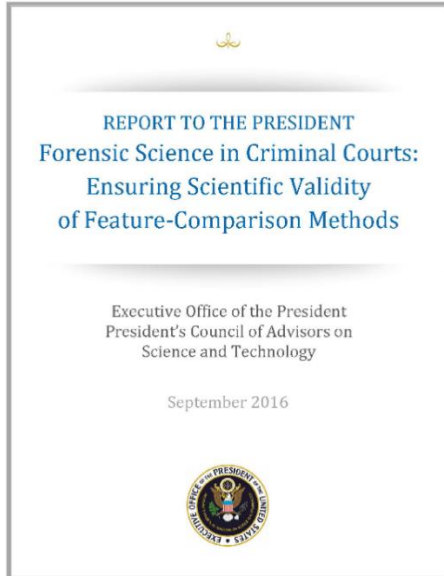
Don't rely upon ...

'We note, finally, that neither experience, nor judgment, nor good professional practices (such as certification programs and accreditation programs, standardized protocols, proficiency testing, and codes of ethics) can substitute for actual evidence of foundational validity and reliability.

…

Similarly, an expert's expression of *confidence* based on personal professional experience or expressions of *consensus* among practitioners about the accuracy of their field is no substitute for error rates estimated from relevant studies. **For forensic feature-comparison methods, establishing foundational validity based on empirical evidence is thus a *sine qua non.* Nothing can substitute for it.**'

President's Council of Advisers on Science and Technology (2016)

REPORT TO THE PRESIDENT
Forensic Science in Criminal Courts:
Ensuring Scientific Validity
of Feature-Comparison Methods

Executive Office of the President
President's Council of Advisers on
Science and Technology

September 2016

'Experience' example: Australian passport officers

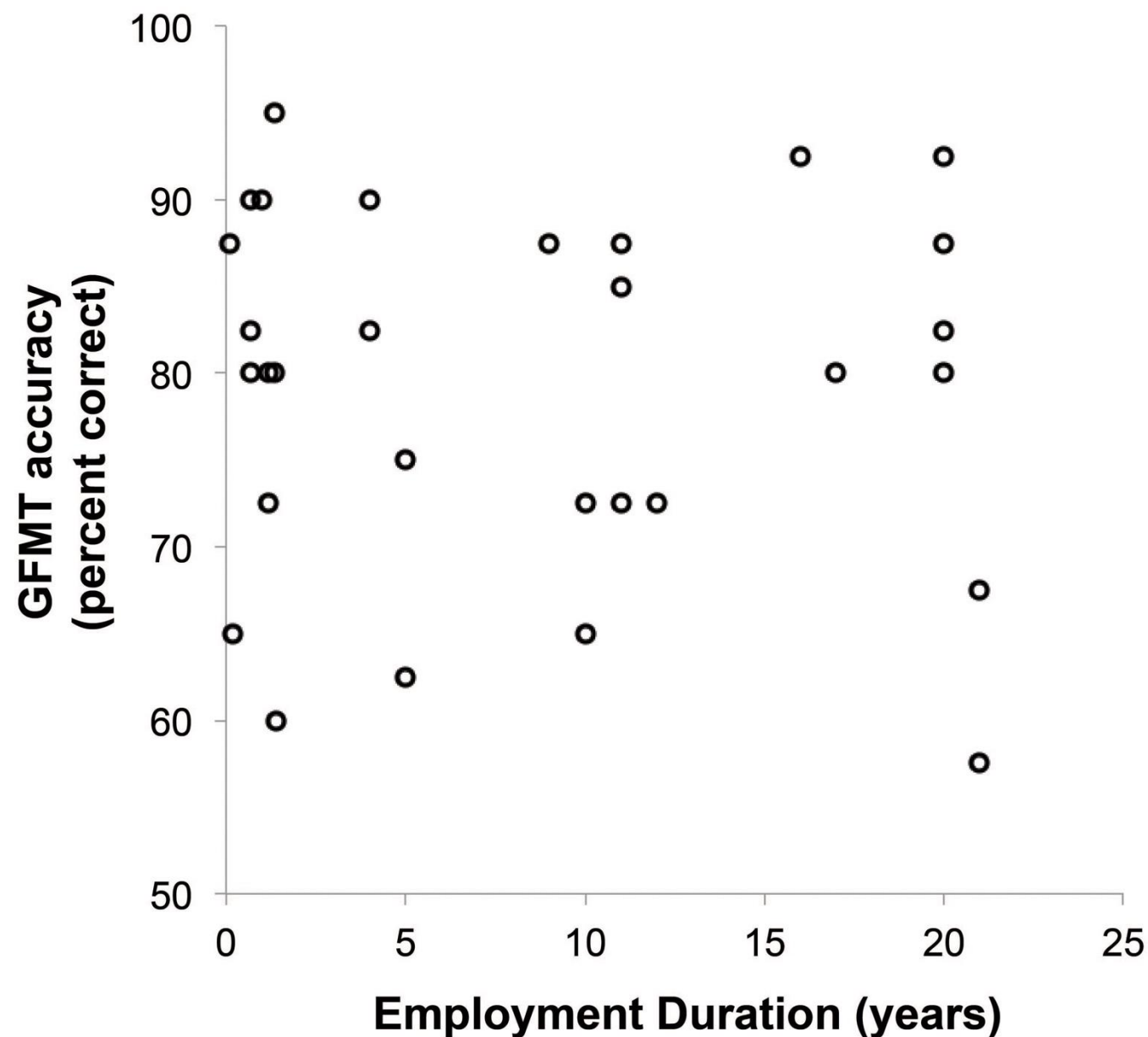# Australian passport officers' actual abilities

Normative (students):
M = 81.3%; SD = 9.7
Passport Officers:
M = 79.2%; SD = 10.4

White et al. (2014).
Passport Officers Errors
in face Matching. PLOS
ONE 9(8)

Human factors (especially cognitive biases)

'Some initial and striking research has uncovered the effects of some biases in forensic science procedures ... The forensic science disciplines are just beginning to become aware of contextual bias and the dangers it poses. The traps created by such biases can be very subtle, and typically one is not aware that his or her judgment is being affected.'
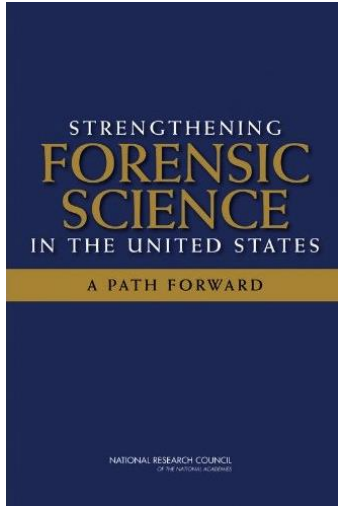
National Academy of Sciences, *Strengthening Forensic Science in the United States* (2009)

See also
Expert Working Group, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach* (US National Institute of Standards and Technology & National Institute of Justice, 2012);

EBFI, 'Thinking forensics: Cognitive science for forensic practitioners' (2017) 57 *Science & Justice* 144 and 'Just cognition' (2019) 82 *Modern Law Review* 633.

Human factors: a notorious example

Itiel Dror, David Charlton and Ailsa Péron, "Contextual information renders experts vulnerable to making erroneous identifications" (2006) *Forensic Science International* 174.
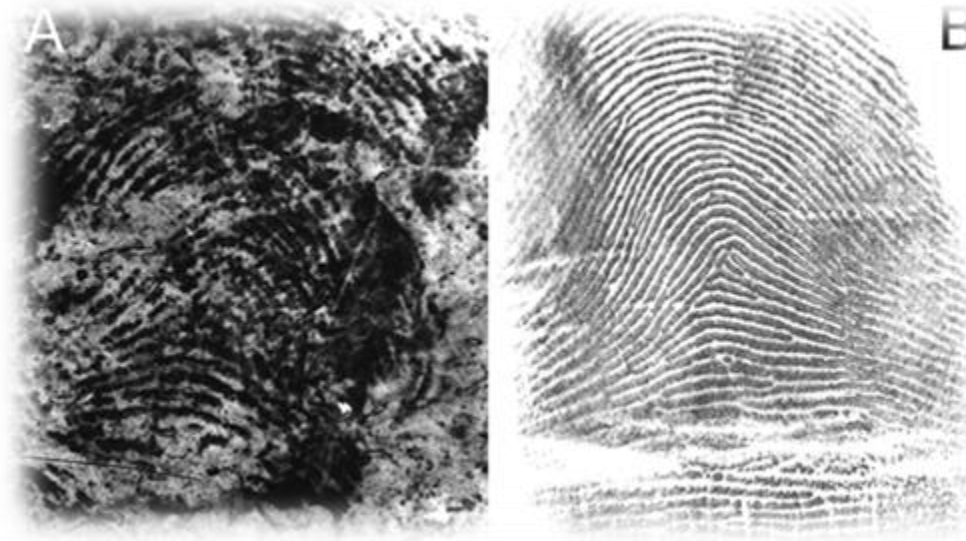
For a meta-analysis, see G. Cooper and V. Meterko, 'Cognitive bias research in forensic science: A systematic review' (2019) 297 *Forensic Science International* 35.

Validation enables the rational evaluation of opinions based on feature comparisons

Do they have pertinent expertise? Can they do the specific thing? (Rather than do they have general, or apparently relevant, expertise or have they done it before.)

How well can they do it and how do we know? (What studies support the *foundational* validity of the technique? Were appropriate standards or methods *applied* in the (case) specific application? What evidence supports this individual being demonstrably proficient? Did other ('blinded') proficient analysts agree?)

What form of expression is used and what is it based on? (Is it based on empirical studies with ground truth, or convention, beliefs and best guesses?)

How were cognitive biases managed? (How were the analyst or reviewer shielded from task-irrelevant information?)

Does the **expert report** (or statement or testimony) enable these questions to be answered?

3. An Australian case study

HIGH COURT OF AUSTRALIA

KIEFEL CJ,
GAGELER, GORDON, EDELMAN AND JAGOT JJ

THOMAS CHRIS LANG   APPELLANT

AND

THE QUEEN   RESPONDENT

Lang v The Queen
[2023] HCA 29
Date of Hearing: 12 May 2023
Date of Judgment: 11 October 2023
B57/2022

ORDER

Appeal dismissed.

A 68 year old woman was found dead in her bed, with a kitchen knife in her abdomen, in a penthouse at Kangaroo Point, Brisbane. The scene was tidy and there was no sign of a struggle.

Should a qualified and very experienced forensic pathologist be allowed to express an opinion as to whether the fatal abdominal stab wound was self-inflicted (and so suicide) or inflicted by another person (and so homicide)?

'It is more likely the wound was inflicted by another person than self-inflicted.'

High Court majority in *Lang v R* – opinion admissible

Jagot J

'… it cannot merely be assumed that Dr X, in giving any part of his evidence, was not basing his opinions on his specialist qualifications and work of 25 years as a forensic pathologist. … it must be recalled that that Dr X has worked as a forensic pathologist for some 25 years. He had performed anywhere between about 4,000 to 5,000 autopsies. … Dr X's opinion as to the likelihood of the fatal wounds being inflicted by another person rather than self-inflicted was not cloaked "with a spurious appearance of authority", and thereby did not involve any risk that "legitimate processes of fact-finding may be subverted".' (*Lang,* [461]-[490])

But, what do you need to *know* to determine admissibility, probative value, or to evaluate the opinion?

With respect to the forensic pathologist's opinion that it was 'more likely' homicide than suicide, what can we say about?

- Validation of method or skill: *foundational* or *as applied*

- Standards

- Uncertainty, limitations and error

- Expression – 'more likely' homicide

- Reasons (or reasoning)

- Scientific publications – e.g. systematic reviews

- Cognitive bias and managing exposure to task (or domain) irrelevant information

- (in)Consistency with earlier cases – e.g. *Velevski*.

See 'Legal pathology: Evaluating the High Court on expert evidence' (2025) 44(1) *University of Queensland Law Journal* (advance access).

4. Junk law?

Legal v scientifically-informed approaches

Courts tend to focus on: 'field'; folk impressions of validity and reliability; formal qualifications; experience; novelty; previous admission; source of the expert (privileging state employees); consensus statements; demarcation between scientific, technical and non-scientific; partiality and conflicts; prima facie plausibility; and whether opinion might assist the jury.

According to scientists, what matters are: validation; error rates and uncertainties; standards; demonstrable personal proficiency; managing cognitive bias; expressions based on data (ideally statistical); reference to publications (e.g. systematic reviews); and independence from law enforcement.

Scientific, technical and biomedical experts should be primarily driven by scientific norms and values, not what courts might happen to allow.

Studies of trial safeguards

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

A WARNING ABOUT JUDICIAL DIRECTIONS AND WARNINGS

ABSTRACT

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

FORENSIC SCIENCE EVIDENCE AND THE LIMITS OF CROSS-EXAMINATION

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

The ability to confront witnesses through cross-examination is conventionally understood as the most powerful means of testing evidence, and one of the most important features of the adversarial trial. Popularly feted, cross-examination was immortalised in John Henry Wigmore's (1863–1943) famous dictum that it is 'the greatest legal engine ever invented for the discovery of truth'. Through a detailed review of the cross-examination of a forensic scientist, in the first scientifically-informed challenge to latent fingerprint evidence in Australia, this article offers a more modest assessment of its value. Drawing upon mainstream scientific research and advice, and contrasting scientific knowledge with answers obtained through cross-examination of a latent fingerprint examiner, it illuminates a range of serious and apparently unrecognised limitations with our current procedural arrangements. The article explains the limits of cross-examination and the difficulties trial and appellate judges — and by extension juries — experience when engaging with forensic science evidence.

Forensic science and the myth of adversarial testing

Gary Edmond

EXPERT REPORTS AND THE FORENSIC SCIENCES

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

The Cool Crucible: Forensic Science and the Frailty of the Criminal Trial

Gary Edmond and Mehera San Roque*

Abstract

Recent developments overseas and in Australia have thrown into question the assumption that the incriminating expert opinion evidence (ie forensic science and forensic medicine) relied on, routinely, in criminal trials is epistemologically robust. At the same time, scepticism about the efficacy of traditional safeguards has been rendered more acute when considering the capacity of the criminal trial to effectively manage incriminating expert evidence in a manner that genuinely reflects commitments to a fair trial. Against this background, this article provides a succinct overview of some of the current concerns and limitations both of traditional adversarial safeguards and more contemporary tailored attempts to manage incriminating forensic science. Taking into account that much incriminating expert evidence is either unreliable or of unknown reliability, this article suggests that courts need to be willing to adopt a more exclusionary orientation towards incriminating forensic science and medicine evidence, and develop new mechanisms, responsive to empirical research, to manage such evidence during criminal proceedings.

Trial by Cognitive Ordeal: Irrational Approaches to the Opinions of Investigators, Trial Integrity and Proof

Gary Edmond*

This article suggests that lawyers and judges may not understand the effects of their rules and procedures upon the production of evidence and its evaluation in criminal trials and appeals. Focusing on case studies involving the opinions of police officers and other investigators, as well as experts, it explains how applicable rules, procedures and safeguards did not produce, and appear incapable of producing, the effects claimed by courts. Drawing on scientific research — on cognitive bias and voice and face comparison — the article demonstrates how judges have enabled investigators to express their biased and speculative opinions, treated investigators' opinions as expert (or special) without evidence of validity or ability, and encouraged jurors to make their own voice and face comparisons in the highly suggestive context of the accusatorial criminal trial. Courts have placed great reliance on trial safeguards, such as cross-examination and judicial directions, trivialised the difficulty of voice and image comparisons and overlooked the likelihood that juror interpretations will be incurably biased, and that the same evidence will be unwittingly counted more than once.