Australian Journal of Forensic Sciences, 2013
Vol. 45, No. 3, 315–322, http://dx.doi.org/10.1080/00450618.2013.782339

Taylor & Francis
Taylor & Francis Group

Identification personified

Jason M. Tangen*

School of Psychology, The University of Queensland, St Lucia, Queensland, Australia

(*Received 20 February 2013; final version received 25 February 2013*)

Forensic practitioners are beginning to acknowledge that there is more to identifying trace evidence than the methodology on which they rely. Instead, identification is based on *human* decision making, which is prone to error. The problem is that we know very little about how these decisions are made, how often errors arise, what to do about them if they occur, and how to acknowledge the risk of error in courts. Dealing with errors in forensic science is not simply a matter of 'getting rid of the bad apples'. Instead, we must develop resilient systems that identify and enhance the positive capacities of people and organisations that allow them to adapt effectively under pressure. The recognition of error in forensic practice is an excellent first step in strengthening the field, but 'more and better research' still needs to be conducted to understand better the nature of expertise in identification to improve training and the value of expert testimony.

Keywords: error; incident reporting; identification; expertise; evidence; judgement

Introduction

In May 2004, Brandon Mayfield, a lawyer from Oregon, USA, was arrested by the Federal Bureau of Investigation (FBI) in an investigation into a terrorist train bombing in Madrid, Spain. Three examiners from the FBI fingerprint lab, and one independent examiner, attributed a print lifted from a bag of detonators to Mayfield. Two weeks after Mayfield was arrested, the Spanish National Police identified Ouhnane Daoud – a known Algerian terrorist – as the source of the print. After the FBI realised its mistake, Mayfield was released from custody.

In an interview with the Chicago Tribune, John Massey, one of the three FBI examiners who attributed Daoud's print to Mayfield, acknowledged the error: 'We just did our job and made a mistake,' he said. 'That's how I like to think of it – an honest mistake.' He went on to say: 'I'll preach fingerprints till I die. They're infallible.'

The above quote nicely illustrates a common assumption that has been traditionally held by fingerprint examiners. Massey starts by acknowledging that he made a mistake, but in the very next breath, goes on to claim that fingerprint identification is infallible. Clearly, he recognised that he and his colleagues attributed the print to the wrong person. How then could he claim that '[fingerprints] are infallible?' The answer is that Massey was citing human error as the cause of the mistake, and that the methodology of fingerprint identification was not to blame – and remains infallible. But of course

*Email: jtangen@psy.uq.edu.au

there cannot be an identification without a human examiner, so there is no distinction between a human error – a 'mistake' in Massey's words – and an error due to the method that the examiner used to arrive at a decision. Prints don't make errors. People make errors.

In light of such claims of infallibility by Massey and others, and a general willingness to infer scientific validity from longstanding use, forensic science has recently faced several attacks on the competence of its examiners and integrity of its methods. These criticisms have taken the form of disputes over the legal admissibility of forensic evidence in the United States (e.g. Daubert v. Merrell Dow Pharmaceuticals, Inc.) and challenges posed by commentators from the peer reviewed legal and scientific circles who highlight the absence of solid scientific studies on the competence of forensic examiners and validity of their methodologies [1–7]. Several authoritative independent inquiries and reports have also placed considerable emphasis on the need for research on the capabilities and competence of forensic examiners including: a 2009 report by the US National Academy of Sciences (NAS)[8], an inquiry into the controversial case of Shirley McKie[9], and the Expert Working Group on Human Factors in Latent Print Analysis sponsored by the United States National Institute of Standards and Technology and National Institute of Justice[10]. As a result of these reports and criticisms, academics, lawyers, policy makers, forensic examiners, and managers in Australia have been scrutinising and modifying their practices and standards[11], reflecting on the nature of forensic science as a profession[12], and questioning the role and relevance of legal mandates in forensic science[13].

When the NAS recommended that: 'Research is needed to address issues of accuracy, reliability, and validity in the forensic science disciplines,' given that the human factor is intrinsically tied to the methodology of fingerprint identification, their demand for evidence is essentially a call for demonstrations of the competency of examiners themselves. That is, experiments designed to measure expertise and examine the factors that influence human decision making. Many of the criticisms raised in these reports and by critics are therefore not simply aimed at a particular process, methodology, or standard, but are directed at forensic examiners to demonstrate that their testimony is within the limits of their expertise – that they can in fact do what they claim.

It is clear that practitioners are doing their best to capture criminals and uphold civil liberties. They have a very high workload and are relentlessly coding evidence, making identifications, supporting detectives, searching and maintaining databases. Few examiners, however, have the methodological skills, funding, time, infrastructure, or experience with research to mount scientific studies of human performance to ensure that their discipline meets high scientific standards.

Applied cognitive scientists, on the other hand, have the reward structures already in place for conducting and publishing high quality research on human expertise. We spend our days designing, programming, and administering computer-based experiments, performing statistical analyses, communicating our findings in peer reviewed outlets and at academic conferences, and we conduct our research across a variety of professions. We are in an excellent position to help forensic examiners conduct 'more and better research' as recommended by the NAS, while ensuring that their practice satisfies legal admissibility criteria. One such collaboration is described below, followed by some lessons learned from diagnostic medicine, and a brief description of some future research directions.

Identifying fingerprint expertise

Between September 2010 to February 2011, Tangen, Thompson, and McCarthy[14] conducted an experiment in collaboration with the Forensic Services Branch at the Queensland Police Service, and the participation of 37 qualified practising fingerprint experts from the Australian Federal, New South Wales, Queensland, South Australia, and Victoria Police. These are the kind of data that are being called for by the NAS and legal commentators, and this collaboration provides an excellent example of how cognitive scientists can work with forensic examiners to engage in research to improve their field.

The aim of this experiment was to understand the extent to which qualified fingerprint experts could distinguish between matching and non-matching prints compared with novices. Aside from the original short report that we published, we discuss our design choices, results, and implications of this experiment at length elsewhere[15]. In short, the results from Tangen et al.[14] indicated that qualified court-practising fingerprint experts were exceedingly accurate compared with novices – but they did make mistakes – and tended to err on the side of caution by making misses (i.e. incorrectly declaring that two matching prints do not match) rather than false alarms (i.e. incorrectly declaring that two non-matching prints match)[16,17].

The expert fingerprint examiners in our experiment performed exceedingly well by correctly identifying 92.12% of the matching prints as 'matches' (hits), and correctly declaring 99.32% of the non-matching prints as 'non-matches' (correct rejections). Only three non-matching fingerprint pairs out of the 444 examined were incorrectly declared 'matches' (false alarms, 0.68%) by three different examiners. We gave the same test to 37 novice participants from The University of Queensland, and they made a large number of errors. In particular, novices mistakenly declared 55.18% (compared with 0.68% for experts) of the similar, non-matching distractor prints as 'matches'.

Note, however, that our experiment was not designed to measure the pervasiveness of errors in practice or even to assess the accuracy of individual examiners or forensic departments. We were measuring matching expertise rather than absolute performance. That is, how well fingerprint examiners perform relative to novices at comparing matching and non-matching prints. This high level of performance was indeed impressive; especially when considering the corresponding performance of novices as well as the accuracy of experts in other expert domains, such as face recognition[18], diagnostic medicine[19], radiology and pathology[20].

Experiments such as ours permit qualified fingerprint examiners to legitimately claim specialised knowledge, which may satisfy legal admissibility criteria. These findings will go a long way in addressing the critics of fingerprint identification and signify that fingerprint examiners are trying to characterise their error rate. Our results make it less likely that examiners themselves will suffer unfounded attacks on their practice. If someone wants to challenge their expertise, then our methodology and experimental design ought to be the target of criticism rather than the examiners themselves (assuming, of course, that their testimony does not extend beyond what these limited tests can support).

We have demonstrated that the training and experience that fingerprint examiners have with matching prints provides a real benefit, but the competency of examiners in most other areas of forensic science has not been demonstrated empirically. Similar experiments that demonstrate expertise ought to be conducted across other areas in order to establish specialised knowledge. Research programs of this sort are relatively

inexpensive and, if they are well designed, will minimise the risk of paying the cost of unreliable evidence.

Beyond bad apples

The field of medical decision making has developed over the last several years under similar conditions as forensics, as there was an increasing awareness that physicians' decisions too often resulted in adverse consequences for patients. In November of 1999, the National Institute of Medicine released a report entitled: *To err is human: building a safer health system*[21]. The report indicated that roughly 5% of autopsies reveal lethal diagnostic errors for which a correct diagnosis coupled with treatment could have averted death, and an estimated 40,000 to 80,000 US hospital deaths result from misdiagnosis annually[19,21]. Research into clinical expertise has matured over the last 14 years and offers a wealth of experience and clues about the nature of decision making in forensics and ways of enhancing forensic practices. Significant progress has already been made in measuring expertise, validating procedures and diagnostic information, and developing resilient systems.

One of the main conclusions of the Institute of Medicine report was that 'the majority of medical errors do not result from individual recklessness or the actions of a particular group – this is not a "bad apple" problem' (p. 2). Rather than focussing on the active errors that occur at the level of the frontline operator, attention should be on the errors that are latent in the system (e.g. poor design, faulty maintenance, bad management decisions, poorly structured organisations, and so on). The ultimate goal is to build resilient systems. To 'make it harder for people to do something wrong and easier for them to do it right' (p. 2).

What can forensic managers learn from medicine about how to handle errors? For example, they might find out that a junior examiner incorrectly matched two prints from different people. This error might have occurred during a training exercise, during casework when it was being verified by a senior examiner, or even in a criminal trial. One option for addressing the issue is to scold the examiner, tell her to be more careful, send her back to basic training, have a senior examiner watch over her carefully, or even give her the sack. That is, get rid of the bad apple. Another option that is in line with the recommendations of the Institute of Medicine report is to build up a repository of these errors (or close calls) to pin down particular patterns and underlying causes. One might find that several examiners consistently make errors when prints are lifted using a particular agent, when they are left on a particular surface, when the examiners are working night shift, or after lunch. Take an analogous case in medicine, which remains a significant problem: illegible physician handwriting in medical prescriptions[22]. One solution might be to scold the doctor or retrain him in proper penmanship. A more efficient solution, however, might be to introduce computer-based prescription writing[19]. Addressing the problem at the blunt end is much more effective than eliminating errors at the sharp end. In fact, eliminating errors at the level of the frontline operator allows the latent failures to remain and accumulate in the system, making the system more prone to future failure[21].

In the wake of the Institute of Medicine report in 1999, anonymous near miss reporting systems have been developed across the sector to help health professionals learn from experience[23], and identify accidents waiting to happen. In fact, schemes for reporting 'close calls', have been institutionalised across many safety-critical sectors including aviation, nuclear power technology, petrochemical processing, steel

production, military operations, and air transportation[24]. Incident reporting systems, however, are only effective if staff participate in the process and if timely corrective actions effectively address vulnerabilities[23]. Building a repository of close calls and errors in forensic science would certainly be a welcome shift from the current culture of blame to a culture that acknowledges mistakes and learns from them.

More and better research

The NAS and others[25] have called for the development of a research culture within forensic science, which would fundamentally change examiners' relationship with empirical data and affect how evidence is understood and reported. Forensic science would benefit from adopting a similar approach to diagnostic medicine in handling and reporting errors, conducting experiments, and building resilient systems centred on human decision making. The recognition of error in forensic practice is an excellent first step in strengthening the field, but 'more and better research' still needs to be conducted to better understand the nature of expertise in identification to improve training and the value of expert testimony.

Expertise

Much has been learned in medical decision making about novice and expert differences, the influence of cognitive biases, and how best to incorporate such knowledge into practice. There are many interesting and important research questions about forensic expertise that would be well adapted to a similar and sustained program of research. What sets an expert apart from a novice? How does forensic expertise develop over time? Can we measure expertise physiologically? How does memory for forensic information relate to matching accuracy? Addressing questions about the nature of expertise would provide a scientific basis for demonstrating the validity of forensic methods. This line of research would promote scholarship that supports, challenges, and improves the state of forensic science while allowing examiners to interpret evidence more effectively and efficiently.

Training

Research into the nature of expertise in identification would clearly have important implications for training, assessment, and recruitment. We often assume that through repeated efforts, we will learn from our experience, improve, and eventually succeed. But feedback is critical for continuous improvement. Without corrective feedback, good practice is not reinforced, poor performance is not corrected, and the path to improvement is not identified[26]. A problem that may be unique to identification, however, is that the ground truth of the source is often uncertain. Whereas doctors often have mortality and morbidity meetings where diagnoses are compared with actual patient outcomes (e.g. ground truth determined by autopsy), there is no way to be certain that an identification is correct in forensics. As Mnookin et al.[25] note, if an examiner has no independent knowledge of whether or not her conclusions in any given case are actually correct, how can she learn from her experience? If she never knows when or if she makes an error, how can she adjust her practices to increase accuracy? In order for research and training in this field to progress, it is critical that researchers have free

access to databases of standardised forensic materials where the ground truth of the source is built into the system[14].

Longitudinal studies on the development of expertise could potentially reduce training time without compromising performance. For example, we could track the progress of novice examiners entering a training program and test them on a range of measures throughout their training. We could then determine how cognitive processing develops, as novices become experts, and how quickly these capabilities develop. These findings would help to determine the factors that best predict expertise and potentially reduce the amount of time that it takes to turn novices into experts.

Testimony

The language that is typically used in current models of expert testimony does not acknowledge human judgement in forensic practice – the 'personification of identification'. Given the widespread criticism of a 'zero error rate' in fingerprints, and the excessively low probabilities reported in DNA profiling, a contemporary model for presenting expert testimony is needed[8]. Following developments in the US and Canada, Edmond[27] has suggested that Australia should adopt a reliability standard, and the UK Law Commission[28] has recently announced similar recommendations for admissibility practice in England and Wales[29]. Indeed, the authoritative reports listed above have called for empirical information about accuracy and performance, along with details about the performance of the entire forensic decision making process – including human performance.

Examiners are obliged to acknowledge the possibility of error throughout the entire decision making process from collecting specimens at the crime scene to the presentation of the evidence in court. Claims about the uniqueness of features (in DNA, fingerprints, or any trace evidence) sidestep the role that humans play in the decision-making process. A wrongful conviction could result from being in the numerator of an extremely small probability (e.g. one in a million or one in a billion), or from the many potential errors that might arise from human involvement (e.g. clerical errors, bias, distraction, corruption, continuity, and so on). Given the likelihood of error across all other areas of human expertise, mistakes that occur on the order of one in a hundred or one in a thousand times would severely deflate the very small probabilities that are commonly reported in DNA[30]. An ideal model of expert testimony would reflect the very real possibility of error for the entire system rather than the probability of error for a particular methodology or technique.

The way that forensic examiners present their testimony is an area of future research and should be discussed and debated. Several factors must be considered in the development of an acceptable model. How should fact finders interpret experiments on expertise and accuracy? How should examiners express their opinions about a match, the risk of error, and exactly what they mean by an identification? If a person's confidence in an identification is a poor indicator of their accuracy[31], should examiners express their confidence or certainty in the strength of the fingerprint match? Should juries be permitted to examine and consider the strength of forensic evidence? Regardless of the answers to these questions, a contemporary model of expert testimony must not extend beyond the capabilities of examiners or beyond what the results of the latest experimental findings on expertise and accuracy can support.

Conclusion

Fingerprint examiners have traditionally cited longstanding use, the application of precise methods, and the uniqueness of the specimen as evidence for the reliability and validity of their discipline and justification for claims of individualisation and infallibility. The same logic persists across the analysis of all trace evidence, including forensic DNA, where promoters typically cite vanishingly small random match probabilities, which tend to be regarded by juries as virtually infallible[32]. Given the recent experiments on expertise and the reports issued by the NAS, Lord Campbell's inquiry into the McKie case in Scotland, and the Expert Working Group on Human Factors in Latent Print Analysis, the field is beginning to acknowledge the central role that humans play in the identification process, and that methodologies cannot operate without practitioners; and practitioners, like all of us, make mistakes.

Given the inevitability of errors, then, the question that needs to be addressed is: how difficult is the task of matching trace evidence to its source, and what are the factors that influence this difficulty? Significant progress has been made in reporting accuracy and dealing with error in diagnostic medicine and many of these lessons and pitfalls can be tailored to forensic identification. The issue is no longer whether forensic examiners make errors, but rather how to acknowledge and learn from them, how to move beyond the adversarial system currently impeding advancement of the field, and how to develop a culture of cooperation between researchers and experts.

Acknowledgements

The research for this paper was supported by an Australian Research Council Linkage Grant (LP120100063) and was presented at the Impressions and Expressions Conference in Sydney, December 2011. Thanks to Gary Edmond and an anonymous reviewer for comments on the paper.

References

1. Edmond G, Roach K. A contextual approach to the admissibility of the state's forensic science and medical evidence. Univ Toronto Law J. 2011;61(3):343–409.
2. Cole SA. More than zero: accounting for error in latent fingerprint identification. J Crim Law Crim. 2005;95(3):985–1078.
3. Cole SA. Who speaks for science? A response to the National Academy of Sciences report on forensic science Law Prob Ris. 2010;9:25–46.
4. Dror I, Cole S. The vision in 'blind' justice: expert perception, judgment, and visual cognition in forensic pattern recognition. Psychon B Rev. 2010;17(2):161–167.
5. Loftus EF, Cole SA. Contaminated evidence. Science. 2004;304(5673):959.
6. Saks MJ, Koehler JJ. The coming paradigm shift in forensic identification science. Science. 2005;309(5736):892–895.
7. Koehler JJ. Fingerprint error rates and proficiency tests: what they are and why they matter. Hastings Law J. 2007;59:1077.
8. National Research Council. Committee on identifying the needs of the forensic science community. Strengthening forensic science in the United States: a path forward. Washington, DC: The National Academies Press; 2009.
9. Campbell A. The fingerprint inquiry report. Edinburgh (Scotland): APS Group Scotland; 2011.
10. Expert working group on human factors in latent print analysis, latent print examination and human factors: Improving the practice through a systems approach. US Department of Commerce, National Institute of Standards and Technology, National Institute of Justice; 2012.

11. Ross A. Forensic science in Australia – can we learn from international reports? Aust J Forensic Sci. 2011;43(2–3):135–145.
12. Robertson J. Forensic science – a true profession? Aust J Forensic Sci. 2011;43(2–3):105–122.
13. Edmond G. Actual innocents? Legal limitations and their implications for forensic science and medicine Aust J Forensic Sci. 2011;43(2–3):177–212.
14. Tangen JM, Thompson MB, McCarthy DJ. Identifying fingerprint expertise. Psychol Sci. 2011;22(8):995–997.
15. Thompson MB, Tangen JM, McCarthy DJ. Expertise in fingerprint identification. J Forensic Sci. Forthcoming 2013.
16. Ulery BT, Hicklin RA, Buscaglia J, Roberts MA. Accuracy and reliability of forensic latent fingerprint decisions. P Natl Acad Sci USA 2011;108(19):7733–38.
17. Ulery BT, Hicklin RA, Buscaglia J, Roberts MA. Repeatability and reproducibility of decisions by latent fingerprint examiners. PLoS ONE. 2012;7(3):e32800.
18. Megreya AM, Burton AM. Matching faces to photographs: poor performance in eyewitness memory (without the memory). J Exp Psychol-Appl. 2008;14(4):364–372.
19. Newman-Toker DE, Pronovost PJ. Diagnostic errors – the next frontier for patient safety. JAMA-J Am Med Assoc. 2009;301(10):1060–1062.
20. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. Am J Med. 2008;121(5):S2–S3.
21. Institute of Medicine. To err is human: building a safer health system. Washington (DC): National Academy Press; 2000.
22. Daly P, Moloney FJ, Doyle M, O'Mahony JB. Legibility of doctor's signatures: Novel approaches to improving an age-old problem. Ir Med J. 2006;99(7):214.
23. Benn J, Koutantji M, Wallace L, Spurgeon P, Rejman M, Healey A, Vincent C. Feedback from incident reporting: information and action to improve patient safety. Qual Saf Health Care. 2009;18(1):11–21.
24. Barach P, Small SD. Reporting and preventing medical mishaps: lessons from non-medical near miss reporting systems. BMJ. 2000;320(7237):759–763.
25. Mnookin JL, Cole SA, Dror IE, Fisher BAJ, Houck M, Inman K, Kaye DH, Koehler JJ, Langenburg G, Risinger DM, et al. The need for a research culture in the forensic sciences. UCLA Law Rev. 2011;58(3):725–779.
26. Cantillon P, Sargeant J. Giving feedback in clinical settings. BMJ. 2008;337:a1961.
27. Edmond G. Specialised knowledge, the exclusionary discretions and reliability: Reassessing incriminating expert opinion evidence. UNSW Law J. 2008;31:1–55.
28. Law Commission of England and Wales. Expert evidence in criminal proceedings in England and Wales. London: Stationery Office; 2011.
29. Ligertwood A, Edmond G. Expressing evaluative forensic science opinions in a court of law. Law Prob Ris. Forthcoming 2013.
30. Koehler JJ. Fingerprint error rates and proficiency tests: what they are and why they matter. Hastings Law Journal. 2007;59:1077.
31. Krug K. The relationship between confidence and accuracy: Current thoughts of the literature and a new area of research. Appl Psychol Crim Justice. 2007;3(1):7–41.
32. Thompson WC. Forensic DNA evidence: the myth of infallibility. In: Krimsky S, Gruber J, editors. Genetic explanations: sense and nonsense. Harvard University Press; 2013.