Journal of Applied Research in Memory and Cognition 7 (2018) 199–208

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Are Forensic Scientists Experts?

Alice Towler*
David White
University of New South Wales, Australia

Kaye Ballantyne
Victoria Police Forensic Services Department, Australia

Rachel A. Searston
University of Melbourne, Australia
The University of Adelaide, Australia

Kristy A. Martire
Richard I. Kemp
University of New South Wales, Australia

Despite playing a critical role in our criminal justice system, very little is known about the expertise of forensic scientists. Here, we review three disciplines where research has begun to investigate such expertise: handwriting analysis, fingerprint examination, and facial image comparison. We assess expertise against the scientific standard, but conclude that meeting this standard does not provide a sufficiently high benchmark for the forensic sciences. Forensic scientists must demonstrate a minimum standard of performance, the ability to defer judgement in cases at high risk of error, and the ability to effectively communicate the strength of their evidence to factfinders. We discuss the limitations of current forensic science expertise research to adequately capture factors affecting operational accuracy and outline crucial differences between studies assessing perceptual skill and operational accuracy. Finally, we identify key areas for future research and encourage cognitive scientists to engage in forensic science research.

General Audience Summary

Forensic scientists provide investigators and courts with information about the source of traces left at crime scenes, such as fingerprints, hair, and blood. However, with the exception of DNA, there is limited scientific evidence that the methods used by forensic scientists can link evidence to a source with high levels of certainty, or whether forensic scientists themselves are experts at making those decisions. Here, we describe research in three forensic disciplines—handwriting analysis, fingerprint examination, and facial image comparison—and consider whether forensic scientists in those disciplines should be considered experts. We identify key issues related to how we define and measure expertise in the forensic sciences, the adequacy of current research to assess expertise in real-world settings, and key areas for future research.

Keywords: Forensic science, Expertise, Handwriting analysis, Fingerprint examination, Facial image comparison

---

Author Note

Alice Towler, David White, Kristy A. Martire, Richard I. Kemp: School of Psychology, University of New South Wales, Australia; Kaye Ballantyne: Office of the Chief Forensic Scientist, Victoria Police Forensic Services Department, Australia; Rachel A. Searston: Melbourne Graduate School of Education, Melbourne Centre for the Study of Higher Education, University of Melbourne, Australia; School of Psychology, The University of Adelaide, Australia.

* Correspondence concerning this article should be addressed to Alice Towler, School of Psychology, University of New South Wales, NSW 2052, Australia. Contact: a.towler@unsw.edu.au

In 1992, three-year-old Christine Jackson was abducted from her home, brutally raped, and murdered. Her body was found in a nearby creek two days later. Attention quickly turned to the victim's stepfather Kennedy Brewer who had been looking after Christine in the hours before she went missing. An autopsy revealed suspected bite-marks on the victim's body and a forensic odontologist testified these were inflicted by Brewer. Brewer was found guilty of capital murder and sexual battery and sentenced to death.

But Kennedy Brewer was innocent. Advances in DNA analysis allowed archived biological evidence to be examined. As a result, Brewer was exonerated before he could be executed, but not before serving 13 years on death row (Innocence Project, 2017).

In the years since the Brewer case, forensic bite-mark analysis has been classified as "junk science" (see PCAST, 2016). Indeed, studies have shown that forensic odontologists cannot reliably determine whether a bite-mark was left by *a human*, let alone identify *which* human (Freeman & Pretty, 2016; Page, Taylor, & Blenkin, 2012).

Forensic bite-mark analysis is not the only forensic science discipline with questionable reliability. In 2009, a scathing report by the National Research Council (hereafter, NRC Report) stated,

> "With 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." (p. 7)

More recently, President Barack Obama commissioned the Presidents' Council of Advisors on Science and Technology (PCAST) to conduct a comprehensive investigation into the status of forensic pattern-matching disciplines (see PCAST, 2016). Pattern-matching disciplines are those where a forensic scientist compares two samples "by eye" to determine if they have the same or different origin, for example fingerprint examination and hair comparison. PCAST reported that in five of the seven examined disciplines—bite-marks, firearms, footwear, complex-mixture DNA and hair analysis—there was little evidence that forensic scientists were able to reliably link samples of unknown origin to their source. In some forensic science disciplines research has revealed poor accuracy and reliability in forensic scientists' judgements, but more commonly, foundational research establishing their expertise has simply not been carried out.[1]

There is limited scientific evidence supporting the validity and reliability of the techniques forensic scientists' use. In addition, very few studies have examined the question of whether forensic scientists show expert-level performance. Despite this, forensic scientists regularly provide their opinions in court as expert witnesses. Ordinarily, witnesses are only permitted to testify to their first-hand experiences relevant to the facts at issue. However, a common legal exception in many countries permits opinion evidence if the opinion is based on "specialised knowledge" acquired through training, study, or experience (e.g., s79 Evidence Act, 1995). It is exceptions of this kind that allow forensic scientists to share their "expertise" with the court and which have established precedent for future admissions. However, simply having training, study, or experience in a particular forensic discipline is insufficient to guarantee expertise (Edmond, 2016; Edmond & Martire, 2017; PCAST, 2016).

Cognitive scientists have studied expert performance for many decades, and as such, are well-placed to examine the question of whether forensic scientists are experts. Prominent researchers in this field have defined expertise as "consistently superior performance on a specified set of representative tasks for a domain" (Ericsson & Lehmann, 1996). It is curious that this definition is not the one used by forensic scientists to benchmark their abilities. Instead, they rely on the presence of "specialised knowledge" together with legal assent as evidence of expert status. Furthermore, the court's willingness to accept "expertise" in unvalidated forensic science disciplines has been credited as a source of serious miscarriages of justice (see Edmond, 2016; Edmond, Found, et al., 2017; Edmond & Martire, 2017; Edmond et al., 2014; Edmond & San Roque, 2016; Koehler, 2016; Martire & Edmond, 2017; Mnookin et al., 2011; PCAST, 2016; Saks & Koehler, 2005). In fact, the Innocence Project estimates that nearly half of all wrongful convictions overturned by DNA evidence involved unvalidated or improper forensic science evidence (Innocence Project, 2017).

The NRC and PCAST reports prioritise empirical validation (or "black box") studies to establish (a) whether methods routinely used by forensic scientists allow them to make accurate determinations of the source of questioned samples, and (b) whether forensic scientists demonstrate expertise in using these methods compared to untrained novices. But forensic scientists do not necessarily know how to design, run, and analyse human performance studies, and thus may lack the skills necessary to undertake this critical research (Martire & Kemp, 2016; Mnookin et al., 2011). There is also a conflict of interest for forensic scientists who wish to establish the validity and reliability of their discipline's methods to provide evidence that they and their colleagues are in fact experts. We argue that cognitive scientists possess the skills needed to design, administer, and statistically analyse fair tests of human performance without being invested in the results. As such, cognitive scientists are particularly well-suited to conducting research on human expertise in the forensic sciences (see also Edmond, Towler, et al., 2017; Koehler, 2013, 2016; Martire & Kemp, 2016; Mnookin et al., 2011).

Here, we review research from three forensic science disciplines—handwriting analysis, fingerprint examination, and facial image comparison—where efforts to assess expertise have already begun. We draw on this research to determine whether forensic scientists in those disciplines should be considered experts. We then discuss some of the broad issues related to establishing expertise in the forensic sciences.

---

[1] See Koehler (2016) for a discussion of the reasons why this kind of basic research in forensic science disciplines has not been conducted.

Handwriting Analysis

Forensic handwriting examiners determine the authorship and manner of production of handwriting and signatures. Authorship decisions require examiners to decide whether the known writer of a handwriting sample was also the author of a questioned sample (e.g., a signature on a legal document; see Figure 1). Production decisions require examiners to decide *how* the writing was produced. Handwriting may be genuine, forged by someone attempting to mimic another person's handwriting, disguised in an attempt to deny authorship, or in some cases, written to look as though someone else has forged their handwriting.

Many studies have investigated handwriting examiners' expertise in making authorship and production decisions. Unlike other forensic science disciplines where cognitive scientists have led research efforts to establish expertise, research in handwriting examination has been largely led by practitioners. As a result, studies are designed to reflect casework as closely as possible, and participants are often not required to make a definitive decision. Instead, they are able to make "inconclusive" responses to indicate that they are not prepared to make a decision, usually because of insufficient information in the samples.

Kam, Fielding, and Conn (1997) investigated the accuracy of authorship decisions in 105 handwriting examiners. Participants were given six reference handwriting samples and asked to identify samples authored by the same person in a comparison set of 24. Handwriting examiners and novices made a similar number of correct decisions, correctly identifying 87.9% and 87.7% of matching samples in the comparison set. There were, however, large differences in false positive errors between the groups. Handwriting examiners made 6.5% false positive errors, where they incorrectly declared one of the comparison samples as matching a reference sample. In contrast, novices were far more error-prone, making false positive errors on 38.3% of trials.

Bird, Found, and Rogers (2010) investigated the accuracy of production decisions in 11 handwriting examiners. Participants were given 140 pairs of handwriting samples and were asked to determine which of each pair was genuine and which was disguised. Handwriting examiners made correct decisions on 73.4% of trials whereas novices made correct decisions on 80.1%. However, handwriting examiners made fewer errors (3.4%) compared to novices (11.4%), and were more likely to declare samples as inconclusive (23.1% vs. 8.4%).

Critically, research across the discipline consistently shows that the difference between handwriting examiner and novice performance is not due to the examiners making a greater proportion of *correct* decisions. Instead, group differences lie in the frequency of inconclusive and incorrect decisions, whereby examiners avoid making many of the errors novices make (see Bird, Found, Ballantyne, & Rogers, 2010; Found & Rogers, 2003; Found & Rogers, 2008; Kam, Abichandani, & Hewett, 2015; Kam, Gummadidala, Fielding, & Conn, 2001; Kam, Wetstein, & Conn, 1994; Sita, Found, & Rogers, 2002). Although handwriting examiners are not necessarily more accurate than novices, we argue that effective use of inconclusive decisions to avoid errors is an important component of forensic science expertise. We return to this point in the discussion.

Fingerprint Examination

Fingerprint examiners judge whether or not a print found at a crime scene (i.e., a *latent*) was left by a particular known person. In a typical case, a fingerprint examiner will systematically compare a latent print side-by-side with a list of suspect prints by eye on a computer screen (see Figure 2). The examiner will arrive at one of three decisions: the prints are a *match* (i.e., they originate from the same finger), the prints are a *non-match* (i.e., the prints come from different fingers), or *inconclusive*.

Fingerprint examination is a challenging perceptual task because matching prints can look very different, due to variation in surface, positioning, pressure, movement, moisture, distortion of the finger or substrate, and general wear and tear. Conversely, non-matching prints can look very similar, particularly because computer algorithms are increasingly being used to rapidly search very large fingerprint databases and return lists of highly similar suspect prints for examiners to compare (Dror & Mnookin, 2010). As a result of this within-finger variability and between-finger similarity, even experienced examiners make mistakes.

A study by Ulery, Hicklin, Buscaglia, and Roberts (2011), for example, measured examiners' ability to match challenging fingerprints representative of casework. The majority of the 169 examiners made at least one false negative error where they

Figure 1. Handwriting examiners compare handwriting samples side-by-side to determine if they were authored by the same person or different people. They may also decide whether the writing is genuine, forged, or disguised. Here, the reference signature of a known person (left) is accompanied by a signature written by the same person (top right) and a signature forged by a different person (bottom right).

Figure 2. Fingerprint examiners compare fingerprints side-by-side to determine if they originated from the same finger or two different fingers. Here, a latent print (left) is accompanied by a matching fingerprint (middle) and a non-matching fingerprint (right). These example fingerprints were sourced from the Forensic Informatics Biometric Repository (FIB-R; Tangen & Thompson, n.d.).

incorrectly declared matching prints as non-matches (a false negative error rate of 7.5%), but only five made a more serious false identification error where they incorrectly declared non-matching prints as matches (a false positive error rate of 0.1%). Just as in handwriting analysis, fingerprint examiners made many inconclusive decisions, judging 23% of the latent fingerprints to be of no value for comparison.

Tangen, Thompson, and McCarthy (2011) sought to establish fingerprint examiners' expertise by comparing them to novices. Tangen et al. did not allow participants to make inconclusive judgements, and thus their study provides a direct measure of examiners' and novices' perceptual sensitivity to matching and non-matching fingerprints. Examiners were significantly more accurate at matching fingerprints than were novices. Examiners correctly declared 92% of matching prints as matches (compared to 75% for novices), and 99% of highly similar non-matching prints as non-matches (compared to 45% for novices).

These studies, alongside other work by cognitive scientists, provide compelling converging evidence that trained fingerprint examiners are demonstrably more accurate than untrained novices at judging whether or not two fingerprints were left by the same person (Busey & Vanderkolk, 2005; Searston & Tangen, 2017a, 2017b, 2017c; Thompson & Tangen, 2014; Thompson, Tangen, & McCarthy, 2014; Vogelsang, Palmeri, & Busey, 2017). However, they also show low intra-examiner repeatability and, as a group, demonstrate a surprisingly wide range of performance (Dror et al., 2011; Ulery, Hicklin, Buscaglia, & Roberts, 2012).

Facial Image Comparison

Forensic facial examiners compare photographs of faces to decide if they show the same person or different people (see Figure 3). Despite decades of research showing novices make large proportions of error on these tasks when faces are unfamiliar (see Burton, White, & McNeill, 2010; Henderson, Bruce, & Burton, 2001; Megreya & Burton, 2006; Megreya & Burton, 2007), it is only very recently that researchers have begun to examine the accuracy of trained forensic scientists on this task.

Norell et al. (2015) tested 17 forensic facial examiners on a matching task that modelled forensic casework. Participants decided whether a high-quality mugshot style photograph, taken under controlled lighting conditions, matched a low-quality still image taken from CCTV footage. Facial examiners made correct decisions on 76% of trials, inconclusive decisions on 21% and errors on 3%. By comparison, novices made correct decisions on 73% of trials, inconclusive decisions on 7% and errors on 20%. These results parallel those found in handwriting examiners, showing that examiners' reduced error rate can mostly be attributed to them making more inconclusive decisions.

In the largest evaluation to date, White, Phillips, Hahn, and O'Toole (2015) tested 27 forensic facial examiners on three challenging face matching tasks that modelled the types of decisions they encounter in their daily work. In each task, examiners decided if two high-quality face photographs showed the same person or different people. Examiners showed superior accuracy to novice participants, making an average of 5% errors on a standardised test of face matching ability, compared to novices who made 20% errors (Glasgow Face Matching Test; Burton et al., 2010). Interestingly, on one test, the difference between examiners and the novice participants was most apparent when given 30 s to make their decision compared to 2 s. It seems that, given sufficient time, examiners were able to extract additional diagnostic information from faces. This suggests that the source of examiners' superior ability relative to novices originates in a slow, deliberate comparison strategy rather than quick, intuitive judgements (see also Towler, White, & Kemp, 2017).

Consistent with White, Phillips, et al. (2015), more recent work has also found superior accuracy in forensic facial examiners compared to untrained novices (Towler, White, & Kemp, 2017; White, Dunn, et al., 2015), suggesting that they are indeed experts. Importantly however, all of these studies focused on group-level differences between novices and examiners. When comparing individual examiners on these tasks, large differences in their performance emerge, with some examiners making 25% errors and others achieving almost perfect accuracy. These individual differences reflect similar levels of inter-individual variation in the novice population, suggesting that underlying skill in these tasks can arise not only from forensic training, but also from an

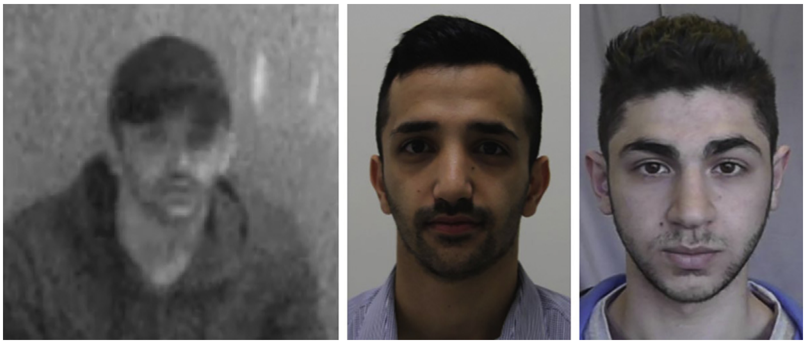

Figure 3. Forensic facial examiners compare face images to determine if they show the same person or different people. Here, an image from CCTV (left) is accompanied by a photograph of the same person (middle) and a photograph of a different person (right).

individual's natural talent in face identification (Balsdon et al., in press; Noyes, Phillips, & O'Toole, 2017; Russell, Duchaine, & Nakayama, 2009).

Discussion

At the beginning of this paper we argued that forensic scientists' expertise should be assessed against the conventional scientific standard, which requires experts to demonstrate superior performance relative to novices (see Ericsson & Lehmann, 1996). If we consider superior performance to mean higher levels of accuracy or fewer errors compared to novices, forensic scientists in all three of the reviewed disciplines meet the criteria for expertise. Nevertheless, forensic scientists still make substantial numbers of errors and show high levels of variability across different examiners. Given the significant consequences of errors, is this conventional criteria a sufficiently high benchmark for forensic scientists? Should we consider a forensic scientist an expert, and allow them to testify in court, just because they can perform better than novices? We do not think so.

Rethinking the Definition of Expertise in the Forensic Sciences

A forensic scientist who outperforms a novice may still make many errors and could be wrong more often than they are right. Given the pervasiveness of forensic science evidence and the serious risk of wrongful convictions, we argue that courts should only allow forensic scientists to testify if both the discipline, and the individual forensic scientist, have met a minimum standard of performance.[2] This minimum standard should be determined by the courts in collaboration with forensic and cognitive scientists.

In addition to meeting this minimum standard of performance, we argue that expertise in the forensic sciences requires two additional skills: knowing when *not* to make a decision because of increased risk of error, and the ability to communicate evidence to factfinders (e.g., judges, jurors) in a manner that enables them to understand the probative weight of that evidence.

Knowing when not to make a decision. Expert forensic scientists need to be able to determine when a reliable decision is impossible, either because there is insufficient information in the samples (e.g., because of ageing, smudging, wear, low resolution imagery), or because the difficulty of the comparison exceeds their abilities. In fact, some forensic scientists would argue that the ability to defer judgement is the crux of expertise in the forensic sciences.

There is, however, mixed evidence that forensic scientists possess sensitivity to these limits. For example, when given the same 42 signatures, handwriting examiners made inconclusive judgements on 0–100% of trials (Found & Rogers, 2008). Similarly, Ulery et al. (2011) found fingerprint examiners made inconclusive judgements on 5–64% of matching fingerprints and 7–96% of non-matching fingerprints. Such large disagreement between examiners suggests that assessing the suitability of a comparison for analysis is highly subjective. The extent to which forensic scientists are able to determine which cases are likely to lead them to make errors, and the consistency with which they can make these judgements, are important questions for future research to examine.

Communication of evidence. Expert forensic scientists also need to effectively communicate the meaning of their evidence (Martire, 2018). For instance, many forensic science bodies have advocated the use of standardised conclusion scales to convey their evidence to jurors (e.g., Association of Forensic Science Providers, 2009). However, research by Martire, Kemp, Watkins, Sayle, and Newell (2013) shows that the verbal label attached to one point of one scale (i.e., "Weak or limited support") conveys the opposite meaning to that intended by experts, with jurors interpreting weak evidence in favour of the hypothesis as evidence for the alternative proposition. Forensic scientists must be able to communicate their evidence in a way that facilitates factfinder understanding and allows them to properly weigh and combine pieces of evidence. Cognitive scientists are particularly well-placed to help forensic scientists develop better ways of conveying the meaning of their conclusions

---

[2] Some would argue that routine proficiency testing of forensic scientists ensures a minimum standard of performance. However, current proficiency tests are entirely inadequate to assess expertise. See Koehler (2013, 2016) for a detailed discussion.

and aiding comprehension of the evidence by judges and jurors.

Adequacy of Current Research to Assess Expertise in Forensic Science

Research into forensic scientists' expertise has largely ignored the fact that they are analysts operating within a large, complex decision-making system (e.g., see Towler, Kemp, & White, 2017). Studies usually investigate expertise using tightly controlled experimental methodology and model one or two real-world aspects of the task, such as by using stimuli representative of casework (e.g., Tangen et al., 2011; White, Dunn, et al., 2015) or introducing contextual information about a case (e.g., Dror, Charlton, & Péron, 2006). However, this approach does not help us understand how forensic scientists' expertise interacts with the complex and high-stakes environments in which they operate. The lack of research investigating decision-making in these complex environments highlights the naiveté of current research in the forensic sciences. Below, we outline some of the system-level factors that are often neglected in research and thus limit our ability to make conclusions regarding the expertise of forensic scientists in operational settings.

Realistic target prevalence. Unlike most experiments, the ratio of targets to non-targets in forensic contexts is almost certainly not 50:50. For instance, forensic scientists often compare crime scene samples against samples taken from *suspects*. Investigators submit samples from suspects for analysis because, based on other evidence, they have reason to believe these individuals are implicated in the crime. This process means that forensic scientists analyse samples that are more likely to be matches than non-matches. We know that base rates can affect performance on perceptual tasks (e.g., Wolfe, Horowitz, & Kenner, 2005), so studies that do not adequately model realistic rates of target-prevalence may underestimate or overestimate operational levels of accuracy.

Contextual information. Forensic scientists are rarely blind to case information. They often know details about the nature of the crime, the history of the suspect, and the presence or absence of other evidence. This contextual information can lead to *confirmation bias*, whereby forensic scientists seek out information consistent with their expectations and prior similar experiences (see Kassin, Dror, & Kukucka, 2013; Saks, Risinger, Rosenthal, & Thompson, 2003). Compared to other system-level factors, considerable research attention has focused on identifying and mitigating confirmation bias (e.g., Dror & Charlton, 2006; Dror et al., 2006; Kukucka & Kassin, 2014). However, confirmation bias is usually studied in isolation, without other system-level factors such as those described here.

Limited resources. Many forensic departments report they are under growing time pressure with large casework backlogs and diminishing resources to meet these demands (Kobus, Houck, Speaker, Riley, & Witt, 2011). Working under these conditions is likely to impair performance. There has been some investigation into the effects of time pressure on facial image comparison (e.g., Bindemann, Fysh, Cross, & Watts, 2016), but the effects of limited resources across the forensic sciences are largely unexplored.

Decision chains. In practice, decisions are not necessarily made by just one forensic scientist. Sometimes a forensic scientist will refer difficult cases to senior analysts or specialist teams, or a colleague will conduct a peer-review (see Ballantyne, Edmond, & Found, 2017; Towler, Kemp, & White, 2017). These processes may serve to catch errors made by individual forensic scientists, and may help calibrate decisions within disciplines. However, little is known about how these procedures work or whether some types of review detect errors—or errors of a particular type—more effectively than others.

Knowledge of being tested. In most studies forensic scientists are aware they are being tested and this is likely to affect their performance (see Risinger, Saks, Thompson, & Rosenthal, 2002). Covert testing, whereby ground-truth-known cases are surreptitiously inserted into routine casework, is therefore necessary to avoid testing effects (Dror & Cole, 2010; Mnookin et al., 2011). This approach is regularly used to maintain acceptable levels of accuracy in other domains such as airline baggage screening (see Wolfe, Brunelli, Rubinstein, & Horowitz, 2013), and is just starting to be used in the forensic sciences (see Kerkhoff et al., 2015).

Measuring Accuracy in the Forensic Sciences

The fact that research does not yet capture the full range of factors that are likely to affect a forensic scientists' ability to complete their job accurately causes problems when we attempt to use that research to report operational accuracy and error rates. We therefore need to be very clear about the purpose of individual studies and what they allow us to conclude about expertise in the forensic sciences. Studies assessing the expertise of forensic scientists fall into two distinct categories: those assessing perceptual skill and those assessing operational accuracy.

Perceptual skill. Perceptual skill refers to a person's raw ability to accurately classify or discriminate between samples representative of casework. For example, in fingerprint examination, this skill refers to an examiner's ability to discriminate between fingerprints left by the same finger and fingerprints left by different fingers. Studies of this kind typically test performance in controlled lab-based experiments, under conditions that do not reflect forensic scientists' daily work. For example, in these tests examiners are typically not allowed access to the tools, procedural documentation, or response scales they use in casework (e.g., White, Phillips, et al., 2015). Studies of perceptual skill thereby enable direct comparison of underlying skill in forensic scientists and novices, but do not provide valid tests of operational accuracy.

Operational accuracy. Operational accuracy refers to a forensic scientists' ability to accurately classify or discriminate between samples representative of casework *in operational settings*. For example, in fingerprint examination, operational accuracy refers to an examiner's ability to discriminate between fingerprints left by the same finger and fingerprints left by different fingers, in cases where they have decided there is sufficient information in the samples, and in the context of the aforementioned system-level factors. Operational accuracy experiments allow us to assess forensic scientists' expertise as it applies in practice.

There is a tendency for those interested in operational accuracy to undervalue tests of perceptual skill (e.g., see review of existing fingerprint studies in PCAST, 2016). However, we believe this is shortsighted. A thorough understanding of expertise in a discipline cannot be achieved without both types of studies, and each informs the other. For example, studies of perceptual skill demonstrated that identifying unfamiliar peoples' faces is error-prone, even for trained professionals (e.g., White, Kemp, Jenkins, Matheson, & Burton, 2014). Operational accuracy studies then sought to model the working conditions of facial examiners and found strikingly low levels of accuracy in real-world conditions (i.e., 50%), even without taking into consideration the full complement of system-level factors mentioned earlier (e.g., White, Dunn, et al., 2015). Researchers have now returned to tightly controlled, lab-based perceptual skill experiments to investigate ways of boosting accuracy through various procedures (e.g., Towler, White, & Kemp, 2017; White, Burton, Kemp, & Jenkins, 2013). Once these techniques are refined they will be implemented in operational settings and tested to determine their effectiveness in practice. Our knowledge in this discipline, like many others, is advanced through a series of interrelated studies of both perceptual skill and operational accuracy.

However, it is important to understand that perceptual skill and operational accuracy studies cannot be substituted for each other. Perceptual skill studies should not be used to make specific claims about operational accuracy. Similarly, operational accuracy studies should not be used to make claims about perceptual skill. Furthermore, operational accuracy studies that model only part of the "real world" should not be used to make definitive claims about operational accuracy; the generalisability of claims is constrained by the factors included in the study. The purpose of, and differences between, perceptual skill and operational accuracy studies will be particularly important moving forward as we seek to understand expertise in the forensic sciences.

The Future of Expertise Research in the Forensic Sciences

The research efforts reported here represent a tiny fraction of the research needed in the forensic sciences. For many disciplines, there have been no empirical assessments of forensic scientists' ability to perform the tasks they undertake on a daily basis. For disciplines where some research exists, such as the three reviewed here, we know far less than we should given the proliferation, importance, and impact of the evidence.

Because expertise research in the forensic sciences is in its infancy, there is tremendous scope for psychology-driven basic and applied research (see Koehler & Meixner, 2016). For example, we still have a very poor understanding of how expertise develops in individuals—whether it be through training, experience, or predetermined by genetics (see Searston & Tangen, 2017a; Shakeshaft & Plomin, 2015; White et al., 2014). We also know very little about individual differences in the development of expertise in forensic science and about the cognitive mechanisms underlying expertise in each discipline. Indeed, we should expect that the critical differences between experts and novices may be accounted for by different cognitive mechanisms in different disciplines. For example, preliminary work in fingerprint examination suggests that expertise is characterised by fast, automatic and holistic processing of fingerprints (Busey & Vanderkolk, 2005; Thompson & Tangen, 2014), as is typical of experts in domains outside forensic science (see Kahneman, 2011; Klein, 1998; Kundel, Nodine, Conant, & Weinstein, 2007). However, facial image comparison shows the exact opposite pattern. What appears to distinguish forensic facial examiners' cognitive processing from novices' is their slow, deliberate and piecemeal processing strategy (Towler, White, & Kemp, 2017; White, Phillips, et al., 2015).

Many important and interesting research questions remain to be addressed in the forensic sciences. We encourage cognitive and forensic scientists to build the collaborations necessary to tackle important research questions facing the forensic sciences. This of course does not come without challenges. Building interdisciplinary relationships and mutual trust takes time. It is also important to manage the expectations of partners who have a vested interest in study outcomes and may be reluctant to publish unfavourable results. However, overcoming these minor challenges is a worthwhile endeavour given the profound societal impact of a productive collaboration between cognitive and forensic scientists.

Author Contributions

Introduction (AT), handwriting (KB, KM), fingerprints (RS), faces (DW), discussion (RK, AT), with AT acting as coordinating author. All authors then contributed to the production of the final manuscript.

Conflict of Interest Statement

The authors declare no conflict of interest.

References

Association of Forensic Science Providers. (2009). Standards for the formulation of evaluative forensic science expert opinion. *Science & Justice*, *49*, 161–164.

Ballantyne, K., Edmond, G., & Found, B. (2017). Peer review in forensic science. *Forensic Science International*, *277*, 66–76.

Balsdon, T., Summersby, S., Kemp, R.I., & White, D. Improving face identification with specialist teams. *Cognitive Research: Principles and Implications* (in press).

Bindemann, M., Fysh, M., Cross, K., & Watts, R. (2016). Matching faces against the clock. *i-Perception*, *7*(5) http://dx.doi.org/10.1177/2041669516672219

Bird, C., Found, B., Ballantyne, K., & Rogers, D. K. (2010). Forensic handwriting examiners' opinions on the process of production of disguised and simulated signatures. *Forensic Science International*, *195*(1), 103–107.

Bird, C., Found, B., & Rogers, D. K. (2010). Forensic document examiners' skill in distinguishing between natural and disguised handwriting behaviors. *Journal of Forensic Sciences*, *55*(5), 1291–1295.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods*, *42*(1), 286–291. http://dx.doi.org/10.3758/BRM.42.1.286

Busey, T. A., & Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision Research*, *45*, 431–448.

Dror, I. E., Champod, C., Langenburg, G., Charlton, D., Hunt, H., & Rosenthal, R. (2011). Cognitive issues in fingerprint analysis: Inter-and intra-expert consistency and the effect of a 'target' comparison. *Forensic Science International*, *208*, 10–17.

Dror, I. E., & Charlton, D. (2006). Why experts make errors. *Journal of Forensic Identification*, *56*(4), 600–616.

Dror, I. E., Charlton, D., & Péron, A. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*, *156*, 174–178. http://dx.doi.org/10.1016/j.forsciint.2005.10.017

Dror, I. E., & Cole, S. A. (2010). The vision in "blind" justice: Expert perception, judgment, and visual cognition in forensic science pattern recognition. *Psychonomic Bulletin & Review*, *17*(2), 161–167.

Dror, I. E., & Mnookin, J. L. (2010). The use of technology in human expert domains: Challenges and risks arising from the use of automated fingerprint identification systems in forensic science. *Law, Probability and Risk*, *9*, 47–67. http://dx.doi.org/10.1093/lpr/mgp031

Edmond, G. (2016). Legal versus non-legal approaches to forensic science evidence. *International Journal of Evidence and Proof*, *20*(1), 3–28.

Edmond, G., Found, B., Martire, K. A., Ballantyne, K., Hamer, D., Searston, R. A., . . . & Roberts, A. (2017). Model forensic science. *Australian Journal of Forensic Sciences*, *57*(2), 144–154. http://dx.doi.org/10.1016/j.scijus.2016.11.005

Edmond, G., & Martire, K. A. (2017). Knowing experts? Section 79, forensic science evidence and the limits of 'training, study or experience'. In A. Roberts, & J. Gans (Eds.), *Critical perspectives on the uniform evidence law* (p. 80). Sydney: Federation Press.

Edmond, G., Martire, K. A., Kemp, R. I., Hamer, D., Hibbert, B., Ligertwood, A., . . . & White, D. (2014). How to cross-examine forensic scientists: A guide for lawyers. *Australian Bar Review*, *39*, 174–197.

Edmond, G., & San Roque, M. (2016). Before the High Court. Honeysett v The Queen: Forensic science, 'specialised knowledge' and the uniform evidence law. *Sydney Law Review*, *36*, 323–344.

Edmond, G., Towler, A., Growns, B., Ribeiro, G., Found, B., White, D., . . . & Martire, K. A. (2017). Thinking forensics: Cognitive science for forensic practitioners. *Science and Justice*, *57*, 144–154.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaption to task constraints. *Annual Review of Psychology*, *47*, 273–305.

Found, B., & Rogers, D. K. (2003). The initial profiling trial of a program to characterise forensic handwriting examiners' skill. *Journal of the American Society of Questioned Document Examiners*, *6*, 77–81.

Found, B., & Rogers, D. K. (2008). The probative character of forensic handwriting examiners' identification and elimination opinions on questioned signatures. *Forensic Science International*, *178*(1), 54–60.

Freeman, A., & Pretty, I. (2016). Construct validity of bitemark assessments using the ABFO decision tree. In *Paper presented at the 2016 Annual Meeting of the American Academy of Forensic Sciences*.

Henderson, Z., Bruce, V., & Burton, A. M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology*, *15*, 445–464. http://dx.doi.org/10.1002/acp.718

Innocence Project. (2017). *The innocence project.*. Retrieved from www.innocenceproject.org.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kam, M., Abichandani, P., & Hewett, T. (2015). Simulation detection in handwritten documents by forensic document examiners. *Journal of Forensic Sciences*, *60*(4), 936–941.

Kam, M., Fielding, G., & Conn, R. (1997). Writer identification by professional document examiners. *Journal of Forensic Sciences*, *42*(5), 1–33.

Kam, M., Gummadidala, K., Fielding, G., & Conn, R. (2001). Signature authentication by forensic document examiners. *Journal of Forensic Science*, *46*(4), 884–888.

Kam, M., Wetstein, J., & Conn, R. (1994). Proficiency of professional document examiners in writer identification. *Journal of Forensic Science*, *39*(1), 5–14.

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, *2*(1), 42–52.

Kerkhoff, W., Stoel, R. D., Berger, C. E. H., Mattijssen, E. J. A. T., Hermasen, R., Smits, N., & Hardy, H. J. J. (2015). Design and results of an exploratory double blind testing program in firearms examination. *Science & Justice*, *55*(6), 514–519.

Klein, G. A. (1998). *Sources of power: How people make decisions*. MIT Press.

Kobus, H., Houck, M. M., Speaker, P., Riley, R., & Witt, T. (2011). Managing performance in the forensic sciences: Expectations in light of limited budgets. *Forensic Science Policy and Management*, *2*(1), 36–43.

Koehler, J. J. (2013). Proficiency tests to estimate error rates in the forensic sciences. *Law, Probability & Risk*, *12*, 89–98.

Koehler, J. J. (2016). Forensics or fauxrensics? Ascertaining accuracy in the forensic sciences. *Arizona State Law Journal*, *49*(4), 1–41.

Koehler, J. J., & Meixner, J. B. (2016). An empirical research agenda for the forensic sciences. *Journal of Criminal Law and Criminology*, *106*(1), 1–34.

Kukucka, J., & Kassin, S. M. (2014). Do confessions taint perceptions of handwriting evidence? An empirical test of the forensic confirmation bias. *Law and Human Behavior*, *38*, 256–270.

Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic component of image perception in mammogram interpretation: Gaze-tracking study. *Radiology*, *242*(2), 396–402.

Martire, K. A. (2018). Clear communication through clear purpose: Understanding statistical statements made by forensic scientists. *Australian Journal of Forensic Sciences*, 1–9. http://dx.doi.org/10.1080/00450618.2018.1439101

Martire, K. A., & Edmond, G. (2017). Re-thinking expert opinion evidence. *Melbourne University Law Review*, *40*, 967–998.

Martire, K. A., & Kemp, R. I. (2016). Considerations when designing human performance tests in the forensic sciences. *Australian Journal of Forensic Sciences*, 1–17. http://dx.doi.org/10.1080/00450618.2016.1229815

Martire, K. A., Kemp, R. I., Watkins, I., Sayle, M. A., & Newell, B. R. (2013). The expression and interpretation of uncertain forensic science evidence: Verbal equivalence, evidence strength, and the weak evidence effect. *Law and Human Behavior*, *37*(3), 197–207.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory and Cognition*, *34*(4), 865–876.

Megreya, A. M., & Burton, A. M. (2007). Hits and false positives in face matching: A familiarity-based dissociation. *Perception and Psychophysics*, *69*(7), 1175–1184.

Mnookin, J. L., Cole, S. A., Dror, I. E., Fisher, B. A. J., Houck, M. M., Inman, K., … & Stoney, D. A. (2011). The need for a research culture in the forensic sciences. *UCLA Law Review*, *58*, 725–779.

National Research Council. (2009). *Strengthening forensic science in the United States: A path forward*.

Norell, K., Brorsson Lathen, K., Bergstrom, P., Rice, A., Natu, V., & O'Toole, A. J. (2015). The effect of image quality and forensic expertise in facial image comparisons. *Journal of Forensic Sciences*, *60*(2), 331–340. http://dx.doi.org/10.1111/1556-4029.12660

Noyes, E., Phillips, P. J., & O'Toole, A. J. (2017). What is a super-recogniser? In M. Bindemann, & A. M. Megreya (Eds.), *Face processing: Systems, disorders and cultural differences*. Nova Science.

Page, M., Taylor, J., & Blenkin, M. (2012). Context effects and observer bias – Implications for forensic odontology. *Journal of Forensic Sciences*, *57*(1), 108–112. http://dx.doi.org/10.1111/j.1556-4029.2011.01903

PCAST. (2016). *Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods*.

Risinger, D. M., Saks, M. J., Thompson, W. C., & Rosenthal, R. (2002). The Daubert/Kumho implications of observer effects in forensic science: Hidden problems of expectation and suggestion. *California Law Review*, *90*(1), 1–56.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin and Review*, *16*(2), 252–257. http://dx.doi.org/10.3758/PBR.16.2.252

Saks, M. J., & Koehler, J. J. (2005). The coming paradigm shift in forensic identification science. *Science*, *309*(5736), 892–895. http://dx.doi.org/10.1126/science.1111565

Saks, M. J., Risinger, D. L., Rosenthal, D., & Thompson, B. (2003). Context effects in forensic science: A review and application of the science of science to crime laboratory practice in the United States. *Science & Justice*, *43*, 77–90.

Searston, R. A., & Tangen, J. M. (2017a). The emergence of perceptual expertise with fingerprints over time. *Journal of Applied Research in Memory and Cognition*, *6*(4), 442–451.

Searston, R. A., & Tangen, J. M. (2017b). Expertise with unfamiliar objects is flexible to changes in task but not changes in class. *PLoS ONE*, *12*(6), e0178403.

Searston, R. A., & Tangen, J. M. (2017c). The style of a stranger: Identification expertise generalizes to coarser level categories. *Psychonomic Bulletin & Review*, *24*(4), 1324–1329.

Shakeshaft, N. G., & Plomin, R. (2015). Genetic specificity of face recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(41), 12887–12892. http://dx.doi.org/10.1073/pnas.1421881112

Sita, J., Found, B., & Rogers, D. K. (2002). Forensic handwriting examiners' expertise for signature comparison. *Journal of Forensic Sciences*, *47*(5), 1117–1124.

Tangen, J. M., Thompson, M. B., & McCarthy, D. J. (2011). Identifying fingerprint expertise. *Psychological Science*, *22*(8), 995–997. http://dx.doi.org/10.1177/0956797611414729

Thompson, M. B., & Tangen, J. M. (2014). The nature of expertise in fingerprint matching: Experts can do a lot with a little. *PLoS ONE*, *9*(12), 1–23. http://dx.doi.org/10.1371/journal.pone.0114759

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (2014). Human matching performance of genuine crime scene latent fingerprints. *Law and Human Behavior*, *38*(1), 84–93. http://dx.doi.org/10.1037/lhb0000051

Towler, A., Kemp, R. I., & White, D. (2017). Unfamiliar face matching systems in applied settings. In M. Bindemann, & A. M. Megreya (Eds.), *Face processing: Systems, disorders and cultural differences*. New York: Nova Science Publishers, Inc.

Towler, A., White, D., & Kemp, R. I. (2017). Evaluating the feature comparison strategy for forensic face identification. *Journal of Experimental Psychology: Applied*, *23*(1), 47–58. http://dx.doi.org/10.1037/xap0000108

Ulery, B. T., Hicklin, R. A., Buscaglia, J., & Roberts, M. A. (2011). Accuracy and reliability of forensic latent fingerprint decisions. *Proceedings of the National Academy of Science of the United States of America*, *108*(19), 7733–7738.

Ulery, B. T., Hicklin, R. A., Buscaglia, J., & Roberts, M. A. (2012). Repeatability and reproducibility of decisions by latent fingerprint examiners. *PLoS ONE*, *7*(3), 1–12. http://dx.doi.org/10.1371/journal.pone.0032800

Vogelsang, M. D., Palmeri, T. J., & Busey, T. A. (2017). Holistic processing of fingerprints by expert forensic examiners. *Cognitive Research*, *2*(1), 1–12.

White, D., Burton, A. M., Kemp, R. I., & Jenkins, R. (2013). Crowd effects in unfamiliar face matching. *Applied Cognitive Psychology*, *27*, 769–777.

White, D., Dunn, J. D., Schmid, A. C., & Kemp, R. I. (2015). Error rates in users of automatic face recognition software. *PLoS ONE*, *10*(10), 1–14. http://dx.doi.org/10.1371/journal.pone.0139827

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PLoS ONE*, *9*(8), 1–6.

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015). Perceptual expertise in forensic facial image comparison. *Proceedings of the Royal Society of London B: Biological Sciences*, *282*, 1814–1822.

Wolfe, J. M., Brunelli, D. N., Rubinstein, J., & Horowitz, T. S. (2013). Prevalence effects in newly trained airport checkpoint screeners: Trained observers miss rare targets, too. *Journal of Vision*, *13*(3), 1–9 (33).

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature*, *435*(7041), 439–440.

Received 5 December 2017;
received in revised form 30 March 2018;
accepted 30 March 2018