Science and Justice 57 (2017) 144–154

Contents lists available at ScienceDirect

Science and Justice

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

Review

Thinking forensics: Cognitive science for forensic practitioners

Gary Edmond [a,*], Alice Towler [b,c], Bethany Growns [c], Gianni Ribeiro [d], Bryan Found [e], David White [c], Kaye Ballantyne [e], Rachel A. Searston [d], Matthew B. Thompson [d,f], Jason M. Tangen [d], Richard I. Kemp [c], Kristy Martire [c]

[a] School of Law, UNSW, Sydney 2052, NSW, Australia
[b] Department of Psychology, The University of York, Heslington, York YO10 5DD, United Kingdom
[c] School of Psychology, UNSW, Sydney 2052, NSW, Australia
[d] School of Psychology, The University of Queensland, Saint Lucia 4072, QLD, Australia
[e] Office of the Chief Forensic Scientist, Victoria Police, 31 Forensic Drive, MacLeod 3085, VIC, Australia
[f] School of Psychology and Exercise Science, Murdoch University, 90 South Street, Murdoch 6150, WA, Australia

ARTICLE INFO

Article history:
Received 18 September 2016
Received in revised form 9 November 2016
Accepted 29 November 2016

Keywords:
Human factors
Expert
Performance
Bias
Psychology
Experience
Training

ABSTRACT

Human factors and their implications for forensic science have attracted increasing levels of interest across criminal justice communities in recent years. Initial interest centred on cognitive biases, but has since expanded such that knowledge from psychology and cognitive science is slowly infiltrating forensic practices more broadly. This article highlights a series of important findings and insights of relevance to forensic practitioners. These include research on human perception, memory, context information, expertise, decision-making, communication, experience, verification, confidence, and feedback. The aim of this article is to sensitise forensic practitioners (and lawyers and judges) to a range of potentially significant issues, and encourage them to engage with research in these domains so that they may adapt procedures to improve performance, mitigate risks and reduce errors. Doing so will reduce the divide between forensic practitioners and research scientists as well as improve the value and utility of forensic science evidence.

Contents

* Corresponding author at: School of Law, UNSW, Sydney 2052, Australia.
E-mail addresses: g.edmond@unsw.edu.au (G. Edmond), a.towler@unsw.edu.au (A. Towler), b.growns@unsw.edu.au (B. Growns), g.ribeiro@uqconnect.edu.au (G. Ribeiro), bryan.found@police.vic.gov.au (B. Found), david.white@unsw.edu.au (D. White), kaye.ballantyne@police.vic.gov.au (K. Ballantyne), r.searston@uq.edu.au (R.A. Searston), matthew.thompson@murdoch.edu.au (M.B. Thompson), jtangen@psy.uq.edu.au (J.M. Tangen), richard.kemp@unsw.edu.au (R.I. Kemp), k.martire@unsw.edu.au (K. Martire).

http://dx.doi.org/10.1016/j.scijus.2016.11.005
1355-0306/

1. Introduction

Over the last decade the forensic sciences have begun to engage with issues relating to human cognition and bias. One of the most conspicuous and important influences has been the recognition that research in experimental psychology (or the cognitive sciences) has serious implications for the organisation, production, reporting and evaluation of forensic science evidence [1]. This article reviews a range of mainstream research findings from the cognitive sciences and presents them in an accessible way to illustrate their significance to forensic practitioners and those relying on their evidentiary products. In this way it represents a contribution to the emerging discipline of *cognitive forensics* [2,3]. This article is intended to help forensic practitioners familiarise themselves with relevant research in order to encourage individuals and institutions to consider how procedures, professional practice and evidentiary products might all be improved.

Scientists who have studied and reviewed the forensic sciences have recommended the need to engage in formal evaluation, attend to human factors and present written and verbal evidence in ways that both fairly represent the results and facilitate comprehension. High profile cases such as Mayfield and McKie [4,5], notorious experiments by Dror and colleagues [6], and reports by the NAS [7], NIST [1,8], PCAST [9] and others [10,11], have drawn unprecedented attention to the need for forensic practitioners to engage with cognitive science and human factors. Concerns voiced in relation to the forensic sciences represent the most recent and conspicuous manifestations of a more widespread social trend involving the application of specialist knowledge to human activities in order to minimise risk. The practice of medicine, air traffic control, and managing nuclear power stations are examples of activities confronted with serious risks that have benefitted from the integration of cognitive science (knowledge and studies) into standard organisation, procedures and practice [12]. Unfortunately, change tends to occur following public failures or catastrophes. For the forensic sciences, problems emerged through the detection of high profile errors (many following the emergence of DNA profiling), the findings of innocence projects, the results of the Criminal Case Review Commission and systematic reviews by agencies such as the FBI (e.g. bullet-lead comparison and microscopic hair comparison) [13].

Notorious mistakes have provided the much-needed impetus to encourage forensic practitioners, their institutions and regulators to contemplate the benefits of incorporating psychological and cognitive science research into their practices. Until recently, this knowledge had been largely overlooked in conventional thinking and practice across the forensic sciences. Following important reviews (e.g. the NAS [7] and NIST [1] reports), forensic practitioners can no longer credibly ignore mainstream scientific research, nor legitimately contend that training and experience, or knowing something about risks to cognition, in some way insulates their cognitive processes. The findings reported in this article are scientifically robust, many developed from decades of research in domains unrelated to forensic science or law.

Legal institutions are, by and large, oblivious to the significance of many of these human factors, both the threats posed as well as their potential utility. Courts, for example, have been unwilling to exclude forensic science evidence where serious threats to conclusions were not addressed or even disclosed. There has, admittedly, been some interest in the expression of results (e.g. *Tang v The Queen* [14], *R v T* [15], *Aytugrul v The Queen* [16] and *US v Monteiro* [17]), although legal resolution has not been informed by mainstream scientific knowledge or consistent. Lack of systematic engagement with scientific research tends to be a hallmark of legal decision-making. Nevertheless, peak scientific and technical organisations have repeatedly recommended that cognitive processes in the production of forensic science evidence be studied and reformed in order to enhance probative value and assist with lay comprehension [1,7].

Courts in most jurisdictions maintain a strong and perhaps exaggerated confidence in the ability of trial safeguards, operated by lawyers and judges, to identify and effectively convey potential risks and dangers with scientific, medical and technical forms of evidence [18]. This article is designed to help forensic practitioners, as well as lawyers and judges, obtain a clearer sense of findings flowing from scientific research. Regardless of what courts might require in terms of admissibility and procedural rules, scientific recommendations and professional obligations would seem to converge around increased engagement with mainstream scientific research and practices [19]. Insights provided by cognitive scientists will likely be an enduring feature of modern forensic science. Part of this legacy will be to reduce exposure to potentially biasing information, inform reporting, communication, selection and training procedures, and minimise the risk of contamination and error by improving workflows.

Here we provide a variety of findings from mainstream scientific endeavours and present them in a form that is readily accessible to a wide audience. In doing so, we have tried to maintain a level of generality in order to avoid excessive didacticism or paternalism. We believe the kinds of insights explained in the following pages warrant attention. Forensic practitioners should not only be aware of this (and other) research, but they and their institutions should be thinking about how they might incorporate applicable insights to improve their performances and the quality of their evidentiary products. In conjunction with formal evaluation and rigorous proficiency testing, such responses will enhance the ability to provide impartial evidence that is demonstrably reliable. Engaging with the insights of experimental psychology will help to insulate forensic practitioners from challenges and criticisms. Simultaneously it will contribute toward broader criminal justice goals and values.

2. We do not experience the world as it really is

- Human perception is not like a video camera. We do not experience the world as it is – so-called verisimilitude.
- Our impression of the world is the result of an interpretive process, and depends on our attention, prior beliefs, expectations, experiences and knowledge.

Human perception does not operate like a video camera. We do not experience a literal reproduction of the world, though we tend to think and act as though we do [20,32]. Even our perception of basic properties like size [21,22], shape [23,24] and colour [25,26] is malleable, easily distorted by context, and often without conscious awareness. For example, spoken syllables can sound completely different when accompanied by footage of someone speaking [27], and identical chess pieces can appear black against one background and white against another [28]. These perceptual distortions occur frequently in everyday life, often with little consequence. However, in forensic science disciplines requiring a human to make perceptual judgments of similarity, for example, these effects could be problematic. Importantly, simply knowing about the existence of these and other perceptual distortions does not prevent the perceiver from experiencing them, and forensic scientists are no exception.

In the case of fingerprint comparison, examiners face the difficult task of comparing two unfamiliar and always different impressions side-by-side, and determining whether they were left by the same finger or two different fingers [29]. It is tempting to think that comparing such unfamiliar images involves an explicit and deliberative perceptual process, detached from memory or prior experience. However, empirical research suggests that what we, and what forensic practitioners see, is very much shaped by prior experience. This experience can be beneficial. For example, through training and on-the-job experience, latent fingerprint examiners can discriminate between most fingerprints at a glance, in visual noise, and when spaced briefly in time, more accurately than novices [33,34]. This prior experience however, can also be detrimental to performance. For instance, the similarity of a particular case to prior episodes (e.g., prior similar fingerprints or prior similar case information) influences image comparison judgments [35]. Similarly, fingerprint examiners' judgments are not always consistent with themselves or others, demonstrating that what we perceive is not purely a reflection of the visual information that we are presented with, but the product of a complex interaction between this visual information and external, often extraneous, inputs [36–39]. Forensic practitioners should be aware that, like all other humans, what they perceive and experience when viewing, comparing and interpreting samples, data or other results will be shaped by context, and their prior knowledge, biases, and expectations.

3. Human memory is unreliable

- Despite our best intentions, memory often fails without awareness.
- Many factors affect memory.
- Forensic scientists should use documentation and empirical information rather than rely on memory or impressions.

We cannot trust our memories. They are, at best, impoverished interpretations of reality. In a simple demonstration of the fallibility of memory, Roediger and McDermott [40] presented people with a list of 12 words (e.g., "bed", "rest", "awake") and immediately afterward asked them to recall as many words as possible. On average, participants recalled 65% of the presented words. Most notably though, 40% of participants reported seeing a word that had not been presented – the word "sleep". Confidence in this false memory was extremely high. Despite participants' best intentions on this basic memory task, their memories failed, and critically, they were none the wiser. The respondents' introduction of "sleep" was of course not coincidental, but due to the thematic associations with the other words.

Importantly, demonstrations of memory failures are not limited to highly controlled low-stakes laboratory experiments. The Innocence Project estimates that 72% of wrongful convictions are associated with failures of eyewitness memory [41]. One such case is that of Ronald Cotton, who was convicted of rape and burglary in 1985 based primarily on the testimony of the victim, Jennifer Thompson [41]. During the attack, Thompson made a deliberate effort to remember as much as she possibly could about her attacker, and impressed the police with her vivid and detailed account. When presented with a photo-lineup she declared with 100% confidence that Cotton was the man who raped her. However, Ronald Cotton was innocent, and served 10 years in prison before being exonerated by DNA evidence in 1995. This case is one of many, and is a poignant illustration that memories can be completely unreliable despite deliberate effort to remember details of the event, considerable time to commit these details to memory, high levels of motivation to remember and recall the event correctly, and complete confidence in the accuracy of the memory.

The notion that memory can be so catastrophically flawed is likely to be surprising to many. However, psychologists have studied human memory since the late nineteenth century and have found that memory errors can, and often do, occur in any of the three stages of memory. Errors can occur during the event (encoding), while the event is stored in long-term memory (storage), and when the memory is recalled at a later time (retrieval) [42].

3.1. Encoding

As explained in Section 2, how we perceive and experience the world around us is shaped not only by visual, auditory and other sensory input, but by our expectations, experiences, emotions, beliefs, attention and a number of other seemingly irrelevant factors. We do not perceive things exactly as they occur, consequently our memories do not resemble a veridical "video recording" of what happened (see the Storage section). For example, people typically overestimate the duration of unpleasant events [43,44], and when very similar events are repeated multiple times, it is very difficult to remember specific details of each occasion [45–47].

3.2. Storage

Events stored in long-term memory do not remain there permanently, unaltered until such time as they are recalled. Memories decay over time according to a logarithmic function [48], and they are frequently updated, altered and reconstructed based on new experiences, information and beliefs [49,50].

3.3. Retrieval

Even if a memory remains relatively intact up until the point of recall, the way in which memories are retrieved can influence their accuracy. For example, recollections of how we felt at a particular time are affected by our current knowledge and feelings about that event [51,52]. We can unconsciously incorporate aspects of someone else's account of a shared event into our own [53,54], and even the way a question is asked can reduce the accuracy of the recalled memory. In a series of experiments Loftus and colleagues [55,56] found that asking subjects "Did you see *the* broken headlight?" instead of "Did you see *a* broken headlight?" resulted in significantly more – and significantly more confident – reports of having observed a broken headlight even though, in reality, neither group saw one.

The research is clear: despite our best intentions, human memory is fallible. Forensic practitioners should be aware of the limitations of memory, and take steps to mitigate the risk of memory errors. In practice, the exact procedures and processes used in casework, and any subsequent conclusions, should be thoroughly and contemporaneously documented. Moreover, when asked about how common something is – such as a feature in a fingerprint or a type or set of wounds – forensic practitioners should be reluctant to make recourse to their non-systematic experiences. Having performed ten thousand autopsies, for example, might not enable a forensic pathologist to accurately recall the frequency of a particular type of stab wound (see *Gilham v R* [57]). A latent fingerprint examiner probably cannot recall the frequency or interrelatedness of features, notwithstanding having observed hundreds of thousands of fingerprints. Similarly, a podiatrist might not accurately recall the percentage of patients – a group that is not a representative sample of the general population – exhibiting eversion [58]. Forensic practitioners should be wary of placing too much reliance on, or confidence in, their memories.

4. Contextual information alters our decisions

- Contextual information can affect the decisions we make, without our awareness.
- Contextual information can lead forensic practitioners to make mistakes and even reverse decisions.
- Knowing the dangers does not enable forensic practitioners to take them into account or transcend them.

We routinely use contextual information to assist us in making decisions. Factors such as mood [59], prior experiences [60] and peripheral information [61] can all influence, and in some cases improve, decision-making. However, under some circumstances, making decisions in the presence of contextual information can lead to *confirmation bias*, where we deliberately seek out and interpret information in a manner that is consistent with our pre-existing beliefs or expectations [62]. For example, doctors tend to seek out evidence to confirm a diagnosis rather than formulate and investigate alternatives [63–65]. The tendency to rely on contextual information is an automatic and natural part of human perception and decision-making [62], and generally operates without conscious awareness [66].

In the criminal justice system however, this effect can be problematic if forensic practitioners are exposed to extraneous information (such as crime facts or details about suspects) while evaluating evidence [67]. In this situation, a practitioner may form an initial belief about the guilt or innocence of a suspect, or about an expected or desired outcome of their analysis. Under these conditions, particularly when the evidence is ambiguous or of poor quality [68], the practitioner may unconsciously evaluate evidence in a manner that confirms their initial belief; that is, they may seek out information consistent with their pre-existing belief and pay less attention to inconsistent information.

A now notorious study by Dror, Charlton and Péron [6] illustrates how forensic practitioners are vulnerable to contextual information. Experienced latent fingerprint examiners were presented with fingerprints in a way that suggested they did not match. However, the examiners had already assessed and judged the fingerprints to match in previous casework. When the examiners unwittingly re-evaluated the fingerprints with the suggestive information, the majority (four out of five) made decisions that contradicted their previous decisions (i.e. either that the fingerprints did not match or were inconclusive). This study clearly demonstrates that contextual information can influence the interpretations of forensic evidence, even where practitioners use otherwise robust procedures.

Further empirical studies across a range of disciplines have shown the large effect contextual information can have on evaluative opinions. For instance, knowledge of a suspect's confession resulted in significantly more match than non-match opinions on handwriting samples [69]. Presentation of skeletons in a mass grave made forensic anthropologists more likely to report the presence of trauma compared to other less suggestive archaeological or control contexts [70]. Viewing highly emotive crime scene photographs resulted in fewer match judgments on human bitemark comparison tasks, compared to when this contextual information was not provided [71]. Even software designed to assist decision-making, such as the relative ranking of highly similar fingerprints in AFIS, significantly influences examiners' judgments [72].

The influence of contextual information on decision-making is an unconscious and generally adaptive strategy. However, in the forensic sciences it is necessary to protect against bias to ensure decisions are based only on information relevant to the analysis [67,73]. Procedural mechanisms for managing the influence of contextual information, such as sequential unmasking (at a case or discipline level) or the introduction of blind analytic procedures, can help to minimise the problem [74–77]. If methods such as these are implemented, and forensic practitioners are not exposed to domain irrelevant (i.e. extraneous) information before evaluating evidence, the impact of contextual bias can be reduced. Although this may be difficult to implement in practice, forensic practitioners should be blinded to extraneous information for as long as possible [78,79].

5. Expertise is domain- and task-specific

- An expert is someone who has demonstrated superior performance relative to the performance of novices.
- Expertise does not simply transfer from one task to another.
- Demonstrations of claimed expertise should be directly related to the specific skill.

Experts are those who produce consistently superior performance to non-experts or laypersons. *Expertise*, on the other hand, refers to the mechanisms underlying this superior performance [80]. Cognitive scientists have studied human expertise across a range of seemingly disparate domains, including chess [81], dermatology [87] and music [82]. As a result, the general nature of expertise and its development is well understood.

It is common for experts to have had several years of experience, or to have engaged in thousands of hours of deliberate practice [82,88]. Compared to novices, experts have a larger number of effective strategies for performing their work accurately [89–91]. Experts rapidly retrieve, from memory, previous instances and decisions relevant to the current situation (but see Section 3), while novices rely on formal rules and procedures [92,93]. Experts can feel – in their "gut" – what the answer is, and they are often right. Novices, on the other hand, cannot rely on their intuition and are often wrong [83,94].

Some groups of people are capable of extraordinary feats of visual categorisation. For example, fire commanders can quickly, automatically and accurately determine when a building is about to collapse [85], and radiologists can detect an abnormal mammogram in less than 1 s [86]. Many forensic practitioners claim to possess similar levels of perceptual expertise (e.g. ballistics, fingerprints and facial comparison). In relation to these, and other claims of expertise, the first step is to establish, under controlled conditions (where "ground truth" is known), whether the claims of superior performance are justified. Somewhat counter-intuitively, highly trained, experienced, and qualified "experts" may fare no better than novices. Prominent examples include stockbrokers choosing profitable stocks [83], and passport officers matching photographs of faces [84] (see Section 8).

Critically, superior performance in a particular domain does not guarantee superior performance in another, even when the domains seem similar [95]. Similarly, prowess at a particular task does not necessarily transfer to other seemingly similar tasks. An expert in human anatomy, for example, may not have expertise in comparing human faces or bodies for purposes of identification [96]. Expertise, therefore, is often described as "domain"- and "task-specific". The specialised nature of expertise can, unfortunately, also render experts inflexible and prone to influence from irrelevant external information [89,97] (see Section 4).

A thorough understanding of the nature of expertise, and how experts differ from novices, can help to: (i) inform the design of training procedures that turn novices into experts more efficiently; (ii) identify people who are likely to make good experts so they can be recruited over others; and, (iii) inform the design of work environments that promote optimal work performance. In forensic practice and criminal proceedings, the term "expert" should be reserved for practitioners who can demonstrate, by means of independent empirical evidence, superior performance to novices.

6. We have limited insight into how we actually make decisions

- Experts often make decisions automatically and without conscious effort.
- Experts do not always possess insight into how they came to their decision, and so may not be able to explain the real reasons for their interpretation.
- Experts may provide reasons, but these might be retrospective rationalisations that do not correspond with the actual reasons for a decision.
- Asking experts to explain how they came to a decision can reduce performance.

Many of the activities that we perform everyday happen automatically and with little cognitive effort. Reading, carrying on a conversation, or riding a bicycle all require very little insight into the psychological processes that produce these complex behaviours. Squiggles on a page or verbal utterances are instantly rendered into fully formed concepts. Continuous tiny adjustments to one's centre of mass, handlebar position, and pedal stroke happen effortlessly. The reason that these tasks feel automatic and effortless is because we have accumulated countless similar experiences, so under normal circumstances, these everyday behaviours happen “to us” and demand very little attention and control [83].

When asked to articulate or explain the nature of these behaviours, we have almost no awareness of their operations. As your eyes move across each of the twisted little marks on this page, you are probably not thinking about word structure, grammar, and syntax. You may not remember learning to distinguish one mark from another. You simply open your eyes and comprehend the meaning of the various shapes as though they themselves are units of thought. Indeed, focusing too much on the details of language, a conversation, or one's own motor movements can cause a conversation (or bicycle ride) to come to a screeching halt. But under unusual circumstances – reading a doctor's handwriting, conversing with someone with a strong accent in a noisy pub, or riding a unicycle – our experience may be of limited utility. On these tasks we may revert to the condition of novices, once again having to devote significant concentration and effort to the task.

There are several activities—like reading—that most adults have mastered and perform easily and automatically. Similarly, most experts, by definition, have accumulated years of experience in their respective domains: distinguishing normal from abnormal symptoms [98], one bird species or wine variety from another [99,100] or distinguishing fingerprint patterns produced by the same person or two different people [101]. To these experts, processing the objects in their domain feels just as simple as reading feels to us. A radiologist just opens her eyes and sees a mammogram as cancerous or asymptomatic [86], and a fingerprint examiner can almost instantly classify a pair of fingerprints as matching or not matching [34].

Notwithstanding automaticity and effortlessness, these experts possess no special insight into the psychological processes that underpin these complex behaviours. Indeed, there has been a great deal of work in experimental psychology on the unreliability of introspection in everyday decision-making (e.g., [102]; see [103] for a review). This evidence suggests that we have no direct access to the cognitive processes that determine the choices we make, even though it often feels like we do. This impression has been described as the *introspection illusion* [104]. A particularly vivid example of this illusion was illustrated by Johansson and colleagues, who demonstrated that people can be misled into confabulating details about a choice that they did not in fact make. For example, in deciding which of two people is more attractive, a participant might select Photo A. Using sleight of hand, the experimenter swaps the two photographs and asks the participant to elaborate on why they selected Photo B (the photo they did not select). Oblivious to the switch, participants readily provide detailed justification for why the person they did not select is more attractive [105]. More generally, dozens of experiments on self-assessment have concluded that the relationship between actual performance and self-rated performance is weak at best (see [106] for a review).

These basic findings and failures of introspection have direct implications for expert testimony. Experts see objects and situations differently and deploy their skills automatically, largely outside conscious awareness [83,89,107]. Expertise in a specific domain does not necessarily include the ability to articulate the basis of that expertise or the reasons for a decision or action. Asking experts to describe what they are doing, for example, can hurt performance [108,109]. Despite the expectations and requirements of legal reports and testimony, experts may not have access to the actual basis for their decision-making. Explanations may be post facto rationalisations, and responses to examination-in-chief and cross-examination might be misleading [94]. Given the potential for gaps between the information and processes that experts think they rely on and those actually used, legal questioning may be of limited value in exposing the basis of decision-making. This reinforces the need for robust, objective, and independent measures and demonstration of genuine expert performance.

7. What experts say might not be what lay audiences hear

- Communicating expert evidence is difficult and error prone.
- Even when experts believe they are communicating clearly, non-experts will often interpret their statements in ways that are inconsistent with the intended message.

Expert opinions in the forensic sciences are always uncertain [110]. The expert almost never knows if a particular finger made a latent fingerprint or if a particular gun fired a particular bullet. They must estimate the likelihood of the observations if this had occurred, and compare that with the likelihood of the observations under alternative explanations (e.g., some other finger or gun was involved) [111]. Fact finders are therefore placed in a position where they must interpret and evaluate expert expressions (often in conjunction with other evidence) in order to reach a final determination regarding the guilt or innocence of a suspect.

There is a large literature relating to decision-making under uncertainty and the interpretation of probabilistic and other expressions [112,113]. For example, McQuiston-Surrett & Saks [114] were interested in how lay persons interpreted the verbal expressions utilized by forensic odontologists, namely: “reasonable scientific certainty”, “probable”, “consistent with” and “match” [9]. The odontologists intended “reasonable scientific certainty” to communicate the highest degree of certainty, with no reasonable probability of error. The level of certainty was meant to decrease from there such that “match”, their lowest level, communicated only “some similarity”. When lay persons were asked to rate the strength of each expression on a scale from 0 (low) to 100 (high), responses revealed discrepancies – clear evidence of misunderstanding and miscommunication. Contrary to the goals of the odontologists, “match” was rated the highest on average ($M = 86.0\%$) and “probable” was rated the lowest ($M = 57.4\%$), indicating that lay people did not understand the hierarchy in the manner intended.

Evidence of miscommunication has also been found in cases where numerical rather than verbal expressions of uncertainty are relied upon [115]. A probabilistic weather forecast, such as: “There is a 30% chance of rain tomorrow” is one example of an uncertain numerical expression. Gigerenzer and colleagues [116] investigated whether people had a shared understanding of what a 30% chance of rain tomorrow means by asking respondents from five capital cities to choose from the following three options: a) it will rain tomorrow for 30% of the time; b) it will rain tomorrow in 30% of the region; or c) it will rain on 30% of the days like tomorrow. The results of the experiment revealed that the most accurate answer “c” was regarded as the least appropriate option for respondents from Amsterdam, Berlin, Milan and Athens. Only those from New York considered it the most appropriate option. These results demonstrate that agreed upon numerical expressions do not guarantee consistent interpretation across audiences. Interpretations may vary between locations and be influenced by training, education, experience and other information.

There is also evidence to suggest that jurors will fail to consider alternative explanations for events unless they are made explicit. In general, people will search for evidence to support their working hypothesis, while neglecting information which is inconsistent [62] (see Section 4). This confirmation bias can be heightened by the presentation of only one explanation – individuals cannot consider alternatives if they do not know they exist. Such tunnel vision has been attributed to investigators, forensic practitioners and jurors in numerous wrongful convictions [66]. Even where jurors are aware of, and strive to avoid, such biases, there is an additional danger that they may construct alternative explanations for forensic evidence – such as how DNA was deposited or why a defendant had glass on their clothing. Such explanations may not be relevant to the case or supported by the evidence [117]. The best method of avoiding both confirmation bias and misattribution of evidence is for experts to explicitly present the propositions that were considered, and explain how and why the evidence does or does not support each. In this manner, the expert is forced to consider the evidence in light of the various alternatives, and jurors may consider valid explanations for the evidence, while discounting invalid hypotheses [118].

These and many other studies suggest that the communication of probabilities and uncertainty is a challenging process, and that consistency between intentions and interpretations cannot be assumed. It is vital that attention is given to the development of a shared vocabulary between experts and fact finders in order to maximise communication accuracy and efficacy, and to ensure the value of evidence as understood by the expert is not lost in translation.

8. Experience does not necessarily translate into expertise

- Extensive experience doing some task does not necessarily mean performance will be superior to that of a person with less or no experience.
- Experts should demonstrate superior accuracy relative to novices.

Courts have traditionally used experience as a proxy to determine expertise within specific domains or “fields”. Often this took place in the absence of empirical evidence of an individual's claimed skill, though courts have generally been reluctant to require that information even when available [119,120]. The opinions of forensic practitioners who have worked in a domain for many years, attended hundreds of crime scenes, performed thousands of autopsies, or provided evidence in numerous trials, will generally be admissible and may be assigned higher probative value than the opinions of practitioners who have not worked for as long, or on as many cases [121]. Numerous studies across a wide range of fields have demonstrated that exposure to, or experience with, a given procedure or activity does not in itself confer expertise [122].

In general, the amount of training and experience displays only a weak relationship with objective measures of performance. For example, superior proficiency in software design is not associated with experience [123,124]. The performance of wine experts, detecting, describing and discriminating between characteristics of wines, is only slightly better than regular wine drinkers on blind trials [125,126]. Treatment success and efficiency is not related to a clinical psychologist's length of training and professional experience [127]. The outcomes of stock investment decisions are not demonstrably superior for financial advisers compared to novices [128,129]. Novice drivers often demonstrate superior safe driving skills when compared to normal, experienced drivers for aspects which require explicit tuition and feedback [130].

While some studies demonstrate that experienced forensic practitioners are more accurate than novices [101,131–133], it is not clear that higher levels of professional experience necessarily equate to higher levels of accuracy. Within the forensic domain, few studies have examined the relationship between experience and expertise or performance (cf. [84,101]). Those that have show the same patterns found in other disciplines – experience does not predict performance success. For example, no relationship was found between expert forensic document examiners' experience and the number of correct, incorrect or inconclusive opinions provided on questioned signatures where ground truth was known [134]. The study found substantial variation between examiners; uncorrelated with years of experience as examiners.

What might explain this apparent dissociation between a person's experience with a given task, and their ability to perform it accurately? As outlined below in Section 11, a critical factor in producing learning through experience is the provision of feedback on the accuracy of our decisions. It appears that, in many domains, practice *without* appropriate feedback (operating in so-called “wicked” environments, see [135]) does not enhance expertise or success [135,136]. For example, unless doctors and nurses are provided with continuing training, they do not improve with extended experience [137,138], and experienced passport officers perform no better than recent recruits on tests of face identification ability ([84], see Fig. 1).

Experience alone is insufficient as a predictor of expertise and performance. Unless practitioners are provided with continuous training, involving deliberate practice on domain-specific tasks with associated feedback regarding their performance, levels of expertise are unlikely to change relative to the levels achieved during initial training [88]. The impact of training and available feedback on performance in most forensic science domains is currently unknown. It is possible that expertise could increase with years of continued practice and feedback, rendering experience a pertinent measure of expertise. However, opportunities to examine this relationship are rare in the forensic sciences, as many disciplines have either not developed rigorous expertise testing programs, based on known samples with appropriate feedback to practitioners, or access to such programs is limited. Bare experience tends to be relied on too heavily, especially by courts and decision-makers. Its value predicting expertise is limited at best.

9. Unless genuinely independent, review (and verification) might not be effective

- Forensic scientists sometimes make decisions in groups.
- Group decision-making may introduce problems and biases (e.g., contextual information and confirmation bias mentioned in Section 4, transference of errors, group concurrence seeking, conformity, deference etc.).
- Independent decisions avoid many of these threats while conferring the benefit of allowing wisdom of the crowd analyses.
- Forensic practitioners should strive for independent reviews and explain the nature of their review processes.

Forensic practitioners and forensic pathologists often reach decisions collectively or consult one another during the processing of information. In the forensic sciences it is common to hold laboratory meetings, or to conduct peer review on forensic science decisions, to ensure consensus and try to prevent costly errors. Although peer interactions provide potentially valuable safeguards, some types of

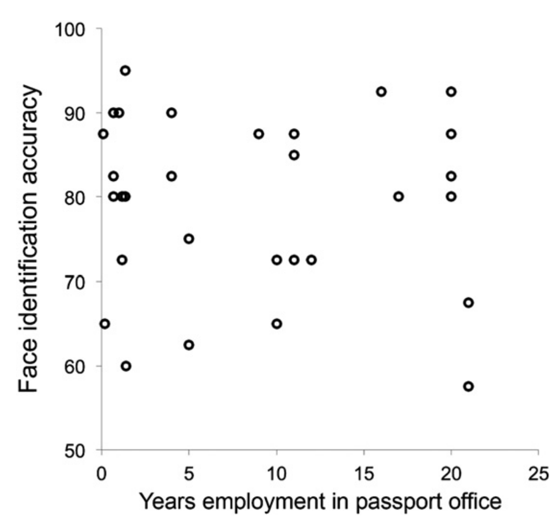

Fig. 1. Experience does not equal expertise. The y-axis shows accuracy of Australian Passport Office staff on a face matching task resembling the face comparison decisions made in their daily work [84]. Researchers observed large individual differences in accuracy – some passport officers performed with near perfect accuracy, while others performed close to chance. Critically, these differences were not predicted by the number of years employed as a passport officer (x-axis).

interaction and review processes can be problematic. When reviews are not conducted independently, errors from the initial examination can be adopted by (or influence) the reviewer [67].

It is however possible to ensure that reviews or collective decision-making is beneficial for the forensic sciences. A large body of empirical research shows that group decisions are often most accurate when the *independent* decisions of individual group members are combined [139]. Indeed, aggregating the independent responses of many individuals tends to produce a remarkably accurate decision. This observation dates back to Sir Francis Galton [140] who calculated the average response in a "guess the ox's weight" competition at a local country fair. Remarkably, the average estimate was within a few pounds of the 1200 lb. animal. This phenomenon has been replicated many times, across a diverse range of decisions, and is popularly referred to as the *Wisdom of Crowds* [141].

Recently, studies have begun to examine the benefit of aggregating responses in forensic pattern matching decisions. Experiments with university students, aggregating independent facial image comparison decisions, produced very substantial gains in accuracy [142]. Further, professional facial examiners approached maximum performance on levels of identification accuracy when tasked with challenging facial comparison decisions once their judgments were averaged across examiners [31,131,132].

Technological advances provide digital platforms that are very well suited to the aggregation of individual examiner responses. In fingerprint examination, aggregating manual fingerprint mark-ups made by multiple examiners prior to submitting the fingerprint AFIS system produces 10% gains in the accuracy of AFIS hits [143]. This is because individual examiners often disagree on the locations and number of fingerprint minutiae [37], and so aggregating responses produces a more reliable template for the biometric system.

These encouraging results suggest that forensic science evidence can be improved by designing intelligent processes for group decision-making. In this context, disagreement between forensic practitioners can be viewed as a strength of the discipline, providing fertile ground for robust collective decisions. If the wisdom of crowds can be harnessed intelligently, while abiding by the golden rule of independence, 'group' decision-making can make many forensic science procedures more reliable.

As for conventional peer review, generally it is desirable for reviewers to be blind to the results of the initial analysis. This facilitates independent and blind peer review. Where, because of resources or workflows, this is not available, forensic practitioners should clearly explain what they mean by peer review and whether the review was independent or undertaken in the suggestive shadow of the initial analysis [19].

10. Confidence and confidence hardening

- Confidence is a poor predictor of accuracy.
- Confidence in a decision tends to increase over time, especially if the person who made the decision receives some sort of confirmation – so-called confidence hardening.

When evaluating experts, jurors tend to believe that confidence is a reliable indicator of an expert's accuracy and credibility. This is known as the Confidence Heuristic Model [144], whereby confidence acts as a heuristic cue that an expert is knowledgeable and believable. Jurors are therefore sensitive to the level of confidence reported by an expert and use this to evaluate whether the testimony should be believed or discounted. Eyewitness testimony research suggests that witness confidence can account for up to 50% of the variance in jurors' decisions of whether or not to believe the witness [145]. This figure is likely to be similar for expert witnesses. For example, an expert with low confidence may exhibit verbal and nonverbal cues characteristic of nervousness, such as a trembling voice and fixed eye contact. A highly confident expert, however, might not exhibit these nervous cues due to their belief in their scientific prowess or the correctness of their conclusion [146]. As a result, jurors may regard the highly confident expert as more credible and believable than the expert with low confidence. However, jurors are not alone in their reliance on confidence as an indicator of an expert's accuracy. Judges and lawyers tend to prefer highly confident experts, even when they proffer less qualified conclusions [147].

Despite popular belief in a strong relationship between witness confidence and accuracy, research demonstrates this correlation is weak. The absence of a confidence-accuracy relationship can be seen in many domains and tasks, including physicians' confidence in their ability to make an accurate diagnosis [148], people's confidence in their ability to detect deception [149], nurses' confidence in their knowledge of basic life support tasks [150], and eyewitnesses' confidence in their identification [151].

In one study physicians made a diagnosis for two easy and two difficult medical cases, and rated their confidence in the decision. Although confidence was only slightly lower for the difficult cases, performance was far worse for these cases (5.8%) than the easier cases (55.3%) [148]. Similarly, a meta-analysis of 31 studies on eyewitness testimony found the average correlation between eyewitness confidence and accuracy was only barely better than chance ($r = 0.07$) [151]. Although the association between confidence and accuracy for forensic practitioners is poorly understood (due to a lack of accuracy data), there is no reason to believe a significant confidence-accuracy relationship exists.

Eyewitness testimony research has shown that the confidence-accuracy relationship is slightly stronger when confidence is assessed at the same time as the identification [152]. As time passes, eyewitnesses often become more certain and confident in their judgment, a process referred to as "confidence hardening". For example, between an identification and trial, an eyewitness may receive feedback from police that their identification is "correct", learn of other evidence that implicates the defendant, and be prepared by the prosecution for trial [153]. Similarly, Oskamp [154] found that clinicians' confidence increased as a function of the amount of information available to them, however there was no corresponding increase in the accuracy of their judgments. Taken together, these findings indicate that a forensic practitioner's confidence will be most predictive of accuracy when recorded at the time of analysis (see Section 3). Confidence may change if they receive gratuitous information (about the case or accused), or receive feedback, regardless of its probative value.

Heavy reliance on confidence when evaluating an expert's testimony might lead decision-makers to pay insufficient attention to factors that *do* predict an expert's accuracy. Kahneman and Klein [155] propose that the best way to evaluate the accuracy of a specific judgment is to consider the validity of the environment in which the judgment was made, as well as the decision-maker's history of learning the rules of the environment. Jurors should therefore be provided with information about the validity and reliability of forensic science procedures so that they may evaluate the evidence.

11. Feedback is essential for learning

- Feedback provides a learner with information about the accuracy of their decisions.
- The provision of accurate feedback aids learning in a variety of situations, including learning to interpret complex visual patterns.
- Receiving feedback on varied and demanding examples is more likely to lead to robust learning, which generalizes to novel stimuli.

Feedback helps us to learn new skills or hone existing ones by providing information about the accuracy of our responses. Positive feedback informs us when a decision was correct, and negative feedback indicates when a decision was incorrect. The provision of feedback allows us to adapt and shape our behaviour to increase the number of correct decisions. Feedback should not be confused with reinforcement, which increases the likelihood that a certain behaviour will occur, irrespective of whether it is correct. Giving a dog a biscuit every time it barks will increase the frequency of barking regardless of whether that is the desired (correct) response. When we provide feedback we usually do so in the hope that it will serve as a positive reinforcement and increase the frequency of correct responding, but this is not inevitably the case.

Although learning can occur in the absence of feedback, learning in the presence of accurate feedback generally occurs more quickly and is more robust and long-lasting. The provision of inaccurate, selective or unreliable feedback that is not directly related to actual performance can hamper learning, and false feedback (the provision of misleading feedback) may increase error rates [156].

The absence of accurate, timely feedback is to blame for poor performance in a variety of situations. For example, forensic psychologists' clinical "gut-instinct" judgments of the risk of re-offending are less accurate than actuarially-based predictions, which are based on empirical data and standardized assessment criteria. This may, in part, be because the psychologist is unlikely to receive accurate and/or timely feedback about which clients committed another offence and which ones did not [157].

The effect of feedback has been demonstrated in a variety of fields. In education, the provision of feedback is one of the most powerful influences on classroom learning [158], and researchers have studied the importance of feedback on elite sports performance [159]. More relevant to forensic science, is the study of the role of feedback in perceptual learning tasks. Perceptual learning involves changes to an individual's perceptual systems (e.g. vision, hearing and taste) that increase sensitivity to stimuli, for example increasing our ability to identify a stimulus or to differentiate between two similar stimuli [160]. Perceptual learning can be demonstrated with simple visual stimuli as well as with more complex visual patterns, such as faces.

A relevant example of feedback training leading to perceptual learning for more complex patterns can be seen in the work of White and colleagues [161], who demonstrated that feedback increased accuracy in an unfamiliar face matching task. This task requires participants to decide whether two images are of the same unfamiliar person. This is the basis of many identity verification processes, for example at border crossings and in the identification of suspects from CCTV images, but has been shown to be surprisingly difficult and error-prone [162]. Critically, it has also been shown that experience alone is not enough to improve performance, with passport staff with up to 20 years of experience sometimes performing no better than recent recruits [84] (see Fig. 1 in Section 8). However, when participants were given trial-by-trial feedback their accuracy improved, and this improvement was maintained after feedback was removed. Furthermore, this improvement generalised to images not used in training, and the feedback effect was largest for those participants who had initially performed most poorly.

Other studies have also shown that feedback can lead to learning which generalizes to novel cases, and there is some evidence that generalization is more likely when the training set is highly variable. For example, participants given feedback training in a mock luggage-screening task, which required them to detect dangerous objects in bags, showed greater generalization of learning when the training involved more varied targets [163]. Interestingly, a more uniform training set resulted in faster and more accurate responses during training, but did not generalize to novel test images; suggesting that although variability in training materials can make initial training more difficult, it is likely to lead to more robust learning effects.

Inappropriate feedback can sometimes have undesirable effects. Eyewitnesses to a crime are sometimes asked to identify the perpetrator from a lineup. If after making their selection the witness receives feedback suggesting their decision was correct (for example being told "Well done, you identified the suspect") this is likely to lead to increased confidence in the decision regardless of its accuracy, together with an inflated estimate of the ease of the identification decision and of the quality of the original viewing experience of the crime event [164]. In order to counter this effect psychologists recommend that the witness' decision and confidence be recorded prior to the provision of feedback (see Section 10).

Many areas of forensic science require practitioners to make difficult discriminations between complex visual patterns. These include the analysis of fingerprints, tool marks, voices, CCTV images, bullet casings, and tire and tread marks. The psychological research reviewed above suggests that repeated exposure to these stimuli over many years will not be sufficient to improve performance. However, training involving the timely provision of accurate feedback is likely to increase accuracy, and if the training set is varied and challenging, the resultant learning is more likely to generalize to novel stimuli. One implication of this research for the forensic sciences is that practitioners should routinely analyse cases where ground truth is known, so that they can be given meaningful feedback on their decision once it has been recorded.

12. Discussion

We have synthesized a range of mainstream research findings of particular relevance to forensic practitioners. Our goal has been to educate, and hopefully stimulate greater awareness of these effects, so that practitioners might engage with these studies and conclusions to improve their procedures and practices. Institutions and individuals should be thinking about what this kind of research might mean for traditional practices, and how conventional assumptions and practices might be revised to enhance performance and avoid criticism [19]. It seems likely, as publications by multi-disciplinary committees of the U.S. Commission on Forensic Sciences indicate, that forensic scientists will need to work hand-in-hand with cognitive scientists and experimental psychologists to reform and enhance their selection processes, procedures, work practices, workflows and resulting evidentiary products.

Rather than prescribe wholesale reform, we have preferred to present research and make a few suggestions where implications appear compelling. What forensic practitioners and their institutions ultimately do is an issue for them; however, they should be aware that these sorts of issues are gradually coming to the attention of lawyers and judges. Though, we also note that few judges have been willing to exclude forensic science evidence on the basis of human factors, at this stage. It is difficult to anticipate what judges might do over time. Several jurisdictions have adopted admissibility and procedural rules directly concerned with reliability (e.g. most US jurisdictions, Canada and recently England) and issues of bias have historically gained the attention of judges (e.g. [30,165]). Bias is a subject on which the judiciary believes it possesses expertise; even if largely limited to interests, conflicts of interest, and perceived conflicts of interest.

We also appreciate that legal categories do not necessarily align with scientific definitions and orientations. The way English and Australian courts place emphasis on training and experience in a "field", as opposed to demonstrated ability relative to non-experts, is a good example [119,121]. Similarly, English courts might be considered insufficiently sensitive to the risks posed by human factors and the ability of traditional trials safeguards, such as cross-examination, to explore unconscious influences on cognitive processes remote in time and place from events [166]. While accepting that forensic practitioners cannot ignore admissibility standards and procedural rules, we recommend caution in relying too heavily on legal (i.e. non-scientific) approaches as the basis for practice and justification. As professionals, forensic practitioners should look to high quality empirical research to support their procedures, practices and evidence. This research should be undertaken by scientists, including cognitive scientists. To look to courts for epistemic support is a mistake. As one eminent Australian judge and scholar noted [167]:

"… the last thing I would wish to encourage … is obsequiousness towards lawyers, either practitioners or judges. There are good social reasons for treating the legal system's normative and adjudicatory authority with respect, but none for endowing it with intellectual authority."

Regardless of what courts and lawyers do, we recommend that forensic practitioners engage with cognitive science research and cognitive scientists. As this article illustrates, there is scope for helping forensic practitioners to avoid cognitive pitfalls (e.g. in exposure to gratuitous information or relying on memory) and improving performance in ways that might assist with both accuracy (e.g. using the wisdom of independent forensic practitioners) and the provision of comprehensible evidence. Such insights might also assist with epistemic humility and the need to take very seriously the dangers of mis- or non-communication of complex and technical forms of evidence [19]. It is unlikely that forensic practitioners can resolve these sorts of issues on their own, but they are precisely the kinds of issues that cognitive scientists understand and might assist with.

Most of the risks to the forensic sciences, forensic science institutions (whether public or commercial) and their social legitimacy are associated with non-engagement with mainstream scientific research and methods. By reading into cognitive science and experimental psychology forensic practitioners might better understand their procedures and abilities, along with their limitations. Simultaneously they might be able to enhance performance and generate improved ways of producing and presenting evidentiary products in ways that accurately embody and convey what is known. Such responses would seem to be consistent with the kinds of expectations that a modern society has of both state-employed forensic practitioners and independent forensic science providers.

References

[1] 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, 2012.
[2] I.E. Dror, R. Stoel, Cognitive forensics: human cognition, contextual information and bias, Encyclopaedia of Criminology and Criminal Justice, Springer 2014, pp. 353–363.
[3] B. Found, Deciphering the human condition: The rise of cognitive forensics, Australian Journal of Forensic Sciences 47 (2015) 386–401.
[4] U.S. Department of Justice: Office of the Inspector General, A review of the FBI's progress in responding to the recommendations in the Office of the Inspector General Report on the fingerprint misidentification in the Brandon Mayfield Case, 2011.
[5] A. Campbell, The Fingerprint InquiryScotland 2011.
[6] I.E. Dror, D. Charlton, A. Péron, Contextual information renders experts vulnerable to making erroneous identifications, Forensic Sci. Int. 156 (2006) 174–178.
[7] National Research Council, Strengthening forensic science in the United States: a path forward, National Academy of Science, 2009.
[8] National Commission of Forensic Science, Ensuring that forensic analysis is based upon task-relevant information, 2012.
[9] Presidents' Council of Advisors on Science and Technology, Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-comparison Methods, 2016.
[10] Government Chief Scientific Adviser, Forensic Science and Beyond: Authenticity, Provenance and Assurance, 2015.
[11] Forensic Science Regulator, Cognitive bias effects relevant to forensic science examinations, 2015.
[12] L.T. Kohn, J.M. Corrigan, M.S. Donaldson, To Err Is Human: Building a Safer Health System, National Academy Press, 1999.
[13] National Research Council, Forensic analysis weighing bullet lead evidence, National Academies Press, 2004.
[14] R v Tang, 65 NSWLR 681, [2006] NSWCCA 167, (2006).
[15] R v T, EWCA 2439, 2010.
[16] Aytugrul v The Queen, 247 CLR 170, 2012.
[17] US v Monteiro, 407 F·Supp.2d 351, (2006).
[18] G. Edmond, M. San Roque, The cool crucible: forensic science and the frailty of the criminal trial, Current Issues in Criminal Justice 24 (2012) 51.
[19] G. Edmond, B. Found, K. Martire, K. Ballantyne, D. Hamer, R. Searston, M. Thompson, E. Cunliffe, R. Kemp, M. San Roque, J. Tangen, R. Dioso-Villa, A. Ligertwood, B. Hibbert, D. White, G. Ribeiro, G. Porter, A. Towler, A. Roberts, Model forensic scienceAustralian Journal of Forensic Sciences 2016.
[20] E. Pronin, D.Y. Lin, L. Ross, The bias blind spot: perceptions of bias in self versus others, Personal. Soc. Psychol. Bull. 28 (2002) 369–381.
[21] R.N. Shepard, Mind Sights: Original Visual Illusions, Ambiguities, and Other Anomalies, with a Commentary on the Play of Mind in Perception and Art, W.H. Freeman & Company, New York, 1990.
[22] New World Encyclopaedia, Ebbinghaus Illusion, 2013.
[23] R.G. Carraher, J.B. Thurston, Optical Illusions and the Visual Arts, Reinhold, New York, 1966.
[24] R.L. Gregory, P. Heard, Border locking and the cafe wall illusion, Perception 8 (1979) 365–380.
[25] M. Bach, Hinton's "Lilac Chaser": Visual Phenomena and Optical Illusions, in, 2005.
[26] E.H. Adelson, Checkershadow Illusion, 1995.
[27] H. McGurk, J. MacDonald, Hearing lips and seeing voices, Nature 264 (1976) 746–748.
[28] B.L. Anderson, J. Winawer, Image segmentation and lightness perception, Nature 434 (2005) 79–83.
[29] M.B. Thompson, J.M. Tangen, R.A. Searston, Understanding Expertise and Non-Analytic Cognition in Fingerprint Discriminations Made by Humans, Frontiers in Psychology, Cognition, 5, 2014.
[30] J.L.B. Allsop, Practice Note CM 7: Expert Witnesses in Proceedings in the Federal Court of Australia, 2013.
[31] Facial Identification Scientific Working Group, Guidelines for Facial Comparison Methods, 2012.
[32] M.H. Segall, T.C. Donald, J.H. Melville, The Influence of Culture on Visual Perception, Bobbs-Merrill, Indianapolis, IN, 1966.
[33] T.A. Busey, J.R. Vanderkolk, Behavioral and electrophysiological evidence for configural processing in fingerprint experts, Vis. Res. 45 (2005) 431–448.
[34] M.B. Thompson, J.M. Tangen, The nature of expertise in fingerprint matching: experts can do a lot with a little, PLoS One 9 (2014) 1–23.
[35] R.A. Searston, J.M. Tangen, K.W. Eva, Putting bias into context: the role of familiarity in identification, Law Hum. Behav. 40 (2015) 50–64.
[36] I.E. Dror, R. Rosenthal, Meta-analytically quantifying the reliability and biasability of forensic experts, J. Forensic Sci. 53 (2008) 900–903.
[37] I.E. Dror, C. Champod, G. Langenburg, D. Charlton, H. Hunt, R. Rosenthal, Cognitive issues in fingerprint analysis: inter- and intra-expert consistency and the effect of a 'target' comparison, Forensic Sci. Int. 208 (2011) 10–17.
[38] B.T. Ulery, R.A. Hicklin, J. Buscaglia, M.A. Roberts, Accuracy and reliability of forensic latent fingerprint decisions, PNAS 108 (2011) 7733–7738.
[39] B.T. Ulery, R.A. Hicklin, J. Buscaglia, M.A. Roberts, Repeatability and reproducibility of decisions by latent fingerprint examiners, PLoS One 7 (2012) 1–12.
[40] H.L. Roediger, K.B. McDermott, Creating false memories: remembering words not presented in lists, J. Exp. Psychol. Learn. Mem. Cogn. 21 (1995) 803–814.
[41] The Innocence Project, Eyewitness misidentification, 2016.
[42] R.L. Klatzky, Human Memory: Structures and Processes, W. H. Freeman & Co., San Francisco, 1975.
[43] E.F. Loftus, Eyewitness Testimony, Harvard University Press, Cambridge, MA, 1979.
[44] C. Stetson, M.P. Fiesta, D.M. Eagleman, Does time really slow down during a frightening event? PLoS One 2 (2007) 1–3.
[45] J. Hudson, K. Nelson, Repeated encounters of a similar kind: effects of familiarity on children's autobiographic memory, Cogn. Dev. 1 (1986) 253–271.
[46] M. Linton, Memory for real-world events, in: D.A. Norman, D.E. Rumelhart (Eds.), Explorations in Cognition, Freeman, San Francisco, 1975.
[47] M. Linton, Transformations of memory in everyday life, in: U. Neisser (Ed.), Memory Observed, Freeman, San Francisco 1982, pp. 77–91.
[48] H. Ebbinghaus, Memory: A Contribution to Experimental Psychology, 1885.
[49] M.R. Leippe, Effects of integrative memorial and cognitive processes on the correspondence of eyewitness accuracy and confidence, Law Hum. Behav. 4 (1980) 261–274.
[50] J.C. Yuille, A critical examination of the psychological and practical implications of eyewitness research, Law Hum. Behav. 4 (1980) 335–346.
[51] L.J. Levine, Reconstructing memory for emotions, J. Exp. Psychol. Gen. 126 (1997) 165–177.
[52] M.A. Safer, L.J. Levine, A.L. Drapalski, Distortion in memory for emotions: the contributions of personality and post-event knowledge, Personal. Soc. Psychol. Bull. 28 (2002) 1495–1507.
[53] C.B. Harris, H.M. Paterson, R.I. Kemp, Collaborative recall and collective memory: what happens when we remember together? Memory 16 (2008) 213–230.
[54] H.L. Roediger, M.M. Meade, E.T. Bergman, Social contagion of memory, Psychon. Bull. Rev. 8 (2001) 365–371.
[55] E.F. Loftus, J.C. Palmer, Reconstruction of auto-mobile destruction: an example of the interaction between language and memory, J. Verbal Learn. Verbal Behav. 13 (1974) 585–589.
[56] E.F. Loftus, G. Zanni, Eyewitness testimony: the influence of the wording of a question, Bull. Psychon. Soc. 5 (1975) 86–88.
[57] Gilham v R, NSWCCA 131, 2012.
[58] E. Cunliffe, G. Edmond, Gaitkeeping in Canada: mis-steps in assessing the reliability of expert testimony, 92, Canadian Bar Review, 2014 327–368.
[59] N. Schwarz, Emotion, cognition, and decision making, Cognit. Emot. 14 (2000) 433–440.
[60] G. Klein, Naturalistic decision making, Hum. Factors 50 (2008) 456–460 The Journal of the Human Factors and Ergonomics Society.
[61] R.E. Petty, J.T. Cacioppo, The Elaboration Likelihood Model of Persuasion, Springer, New York, 1986.
[62] R.S. Nickerson, Confirmation bias: a ubiquitous phenomenon in many guises, Rev. Gen. Psychol. 2 (1998) 175–220.
[63] J.J.J. Christensen-Szalanski, J.B. Bushyhead, Physicians' use of probabilistic information in a real clinical setting, J. Exp. Psychol. Hum. Percept. Perform. 7 (1981) 928–935.

[64] G.-M. Joseph, V.L. Patel, Domain knowledge and hypothesis generation in diagnostic reasoning, Med. Decis. Mak. 10 (1990) 31–44.
[65] B.H. Bornstein, A.C. Emler, Rationality in medical decision making: a review of the literature on doctors' decision-making biases, J. Eval. Clin. Pract. 7 (2001) 97–107.
[66] K.A. Findley, M.S. Scott, The multiple dimensions of tunnel vision in criminal cases, Wisconsin Law Review 2 (2006) 291–397.
[67] S.M. Kassin, I.E. Dror, J. Kukucka, The forensic confirmation bias: problems, perspectives, and proposed solutions, Journal of Applied Research in Memory and Cognition 2 (2013) 42–52.
[68] M. Page, J. Taylor, M. Blenkin, Context effects and observer bias – implications for forensic odontology, J. Forensic Sci. 57 (2012) 108–112.
[69] J. Kukucka, S.M. Kassin, Do confessions taint perceptions of handwriting evidence? An empirical test of the forensic confirmation bias, Law Hum. Behav. 38 (2014) 256–270.
[70] S. Nakhaeizadeh, I. Hanson, N. Dozzi, The power of contextual effects in forensic anthropology: a study of biasability in the visual interpretations of trauma analysis on skeletal remains, Journal of Forensic Science 59 (2014) 1177–1183.
[71] N.K. Osborne, S. Woods, J. Kieser, R. Zajac, Does contextual information bias bitemark comparisons? Sci. Justice 54 (2014) 267–273.
[72] I.E. Dror, K. Wertheim, F.-M. P., W. J., The impact of human-technology cooperation and distributed cognition in forensic science: biasing effects of AFIS contextual information on human experts, J. Forensic Sci. 57 (2012) 343–352.
[73] D.M. Risinger, M.J. Saks, W.C. Thompson, R. Rosenthal, The Daubert/Kumho implications of observer effects in forensic science: Hidden problems of expectation and suggestion, California Law Review 90 (2002) 1–56.
[74] W.C. Thompson, What role should investigative facts play in the evaluation of scientific evidence? Australian Journal of Forensic Sciences 43 (2011) 123–134.
[75] I.E. Dror, Practical solutions to cognitive and human factor challenges in forensic science, Forensic Science Policy and Management 4 (2014) 105–113.
[76] I.E. Dror, W.C. Thompson, C.A. Meissner, I. Kornfield, D. Krane, M.J. Saks, M. Risinger, Letter to the editor - context management toolbox: a linear sequential unmasking (LSU) approach for minimising cognitive bias in forensic decision making, Journal of Forensic Science 60 (2015) 1111–1112.
[77] D.E. Krane, S. Ford, J.R. Gilder, A. Jamieson, R. Koppl, I.L. Kornfield, D.L. Risinger, N. Rudin, M.S. Taylor, W.C. Thompson, Sequential unmasking: a means of minimising observer effects in forensic DNA interpretation, Journal of Forensic Science 53 (2008) 1006–1007.
[78] R. Rosenthal, Experimenter Effects in Behavioural Research, Appleton-Century-Crofts, New York, 1966.
[79] Human Factors Subcommittee, National Commission on Forensic Science: ensuring that forensic analysis is based upon task-relevant information, National Institute of Standards and Technology, 2015.
[80] K.A. Ericsson, N. Charness, Expert performance: its structure and acquisition, Am. Psychol. 49 (1994) 725–747.
[81] F. Gobet, N. Charness, Expertise in chess, in: K.A. Ericsson, N. Charness, P.J. Feltovich, R.R. Hoffman (Eds.), Cambridge Handbook of Expertise and Expert Performance, Cambridge University Press, New York 2006, pp. 523–538.
[82] K.A. Ericsson, R.T. Krampe, C. Tesch-Romer, The role of deliberate practice in the acquisition of expert performance, Psychol. Rev. 100 (1993) 363–406.
[83] D. Kahneman, Thinking, fast and Slow, Farrar, Straus and Giroux, New York, 2011.
[84] D. White, R.I. Kemp, R. Jenkins, M. Matheson, A.M. Burton, Passport officers' errors in face matching, PLoS One 9 (2014) 1–6.
[85] G.A. Klein, Sources of Power: How People Make Decisions, MIT Press, 1998.
[86] K.K. Evans, D. Georgian-Smith, R. Tambouret, R.L. Birdwell, J.M. Wolfe, The gist of the abnormal: above-chance medical decision making in the blink of an eye, Psychon. Bull. Rev. 20 (2013) 1170–1175.
[87] G.R. Norman, D. Rosenthal, L.R. Brooks, S.W. Allen, L.J. Muzzin, The development of expertise in dermatology, Arch. Dermatol. 125 (1989) 1063–1068.
[88] K.A. Ericsson, The influence of experience and deliberate practice on the development of superior expert performance, in: K.A. Ericsson, N. Charness, P.J. Feltovich, R.R. Hoffman (Eds.), The Cambridge Handbook of Expertise and Expert Performance, Cambridge University Press, New York, NY, US 2006, pp. 683–703.
[89] M.T.H. Chi, Two approaches to the study of experts' characteristics, in: K.A. Ericsson, N. Charness, P.J. Feltovich, R.R. Hoffman (Eds.), The Cambridge Handbook of Expertise and Expert Performance, Cambridge University Press 2006, pp. 21–30.
[90] H.G. Schmidt, G.R. Norman, H.P. Boshuizen, A cognitive perspective on medical expertise: theory and implication, Acad. Med. 65 (1990) 611–621.
[91] J. Shanteau, Psychological characteristics and strategies of expert decision makers, Acta Psychol. 68 (1988) 203–215.
[92] L.R. Brooks, The blossoms and the weeds, Can. J. Exp. Psychol. 59 (2005) 62–74.
[93] G.R. Norman, M. Young, L.R. Brooks, Non-analytical models of clinical reasoning: the role of experience, Med. Educ. 41 (2007) 1140–1145.
[94] G.R. Norman, K.W. Eva, Diagnostic error and clinical reasoning, Med. Educ. 44 (2010) 94–100.
[95] K.A. Ericsson, A.C. Lehmann, Expert and exceptional performance: evidence of maximal adaption to task constraints, Annu. Rev. Psychol. 47 (1996) 273–305.
[96] A. Towler, Match me if you can: Evaluating professional training for facial image comparisonunpublished PhD thesis UNSW Australia, Sydney, Australia, 2016.
[97] I.E. Dror, G. Hampikian, Subjectivity and bias in forensic DNA mixture interpretation, Sci. Justice 51 (2011) 204–208.
[98] L.R. Brooks, N.G. R, S.W. Allen, Role of specific similarity in a medical diagnostic task, J. Exp. Psychol. Gen. 120 (1991) 278–287.
[99] J.W. Tanaka, T. Curran, D.L. Sheinberg, The training and transfer of real-world perceptual expertise, Psychol. Sci. 16 (2005) 145–151.
[100] A.L. Hughson, R.A. Boakes, Perceptual and cognitive aspects of wine expertise, Aust. J. Psychol. 53 (2001) 103–108.
[101] J.M. Tangen, M.B. Thompson, D.J. McCarthy, Identifying fingerprint expertise, Psychol. Sci. 22 (2011) 995–997.
[102] R. Nisbett, T.D. Wilson, Telling more than we can know: verbal reports on mental processes, Psychol. Rev. 84 (1977) 231–259.
[103] T.D. Wilson, Strangers to ourselves, Harvard University Press, 2004.
[104] E. Pronin, The introspection illusion, Adv. Exp. Soc. Psychol. 41 (2009) 1–67.
[105] P. Johansson, L. Hall, S. Sikström, A. Olsson, Failure to detect mismatches between intention and outcome in a simple decision task, Science 310 (2005) 116–119.
[106] D. Dunning, C. Heath, J. Suls, Flawed self-assessment: Implications for health, education, and the workplace, Psychol. Sci. Public Interest 5 (2004) 69–106.
[107] R.W. Proctor, A. Dutta, Skill Acquisition and Human Performance, Sage, Thousand Oaks, CA, 1995.
[108] C. Kulatunga-Moruzi, L.R. Brooks, G.R. Norman, Using comprehensive feature lists to bias medical diagnosis, J. Exp. Psychol. Learn. Mem. Cogn. 30 (2004) 563–572.
[109] C. Kulatunga-Moruzi, L.R. Brooks, G.R. Norman, Teaching posttraining: Influencing diagnostic strategy with instructions at test, J. Exp. Psychol. Appl. 17 (2011) 195–209.
[110] C. Aitken, C.E.H. Berger, J.S. Buckleton, C. Champod, J. Curran, A. Dawid, I.W. Evett, P. Gill, J. Gonzalez-Rodriguez, G. Jackson, Expressing evaluative opinions: A position statement, Sci. Justice 51 (2011) 1–2.
[111] G. Jackson, D.H. Kaye, C. Neumann, A. Ranadive, V.F. Reyna, Communicating the Results of Forensic Science Examinations, 2015.
[112] K.A. Martire, R.I. Kemp, B.R. Newell, The psychology of interpreting expert evaluative opinions, Australian Journal of Forensic Sciences 45 (2013) 305–314.
[113] T.S. Wallsten, D.V. Budescu, A review of human linguistic probability processing: General principles and empirical evidence, Knowl. Eng. Rev. 10 (1995) 43–62.
[114] D. McQuiston-Surrett, M. Saks, Communicating Opinion Evidence in the Forensic Identification Sciences: Accuracy and Impact, Hastings LJ, 2007.
[115] G. Gigerenzer, A. Edwards, Simple tools for understanding risks: from innumeracy to insight, BMJ: British Medical Journal (2003) 741–744.
[116] G. Gigerenzer, R. Hertwig, E. van den Broek, B. Fasolo, K. Katsikopoulos, "A 30% chance of rain tomorrow": how does the public understand probabilistic weather forecasts? Risk analysis 25 (2005) 623–629 an official publication of the Society for Risk Analysis.
[117] N. Pennington, R. Hastie, A cognitive theory of juror decision making: the story model, Cardozo Law Review 13 (1991) 5001–5039.
[118] C. Aitken, P. Roberts, G. Jackson, Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses, Royal Statistical Society, London, 2010.
[119] G. Edmond, Legal versus non-legal approaches to forensic science evidence, International Journal of Evidence and Proof 20 (2016) 3–28.
[120] Honeysett v The Queen, 253 CLR 122, 2014.
[121] K.A. Martire, G. Edmond, Re-thinking expert opinion evidence, Melbourne University Law Review, 40 (in press).
[122] K.A. Ericsson, N. Charness, P.J. Feltovich, R.R. Hoffman, The Cambridge Handbook of Expertise and Expert Performance, Cambridge University Press, 2006.
[123] M.B. Rosson, The role of experience in editing, Proceedings of INTERACT '84 IFIP Conference on Human-Computer Interaction, 1985 45–50.
[124] S. Sonnentag, Expertise in professional software design: a process study, The Journal of Applied Psychology 83 (1998) 703–715.
[125] R. Gawel, The use of language by trained and untrained experienced wine tasters, J. Sens. Stud. 12 (1997) 267–284.
[126] D. Valentin, M. Pichon, V. de Boishebert, H. Abdi, What's in a wine name? When and why do wine experts perform better than novices? Abstracts of the Psychonomic Society 5 (1994) 36.
[127] R.M. Dawes, House of Cards: Psychology and Psychotherapy Built on Myth, "The" Free Press, 1994.
[128] C.F. Camerer, E.J. Johnson, The process-performance paradox in expert judgment - How can experts know so much and predict so badly? in: K.A. Ericsson, J. Smith (Eds.), Toward a General Theory of Expertise: Prospects and Limits, Cambridge University Press, Cambridge 1991, pp. 195–217.
[129] J. Shanteau, T.R. Stewart, Why study expert decision making? Some historical perspectives and comments, Organ. Behav. Hum. Decis. Process. 53 (1992) 95–106.
[130] J. Duncan, P. Williams, I. Brown, Components of driving skill: experience does not mean expertise, Ergonomics 34 (1991) 919–937.
[131] D. White, P.J. Phillips, C.A. Hahn, M. Hill, A.J. O'Toole, Perceptual expertise in forensic facial image comparison, Proc. R. Soc. Lond. B Biol. Sci. 282 (2015) 1814–1822.
[132] A. Towler, D. White, R.I. Kemp, Evaluating the feature comparison strategy for forensic face identification, Journal of Experimental Psychology: Applied, (in press).
[133] M. Kam, G. Fielding, R. Conn, Writer identification by professional document examiners, J. Forensic Sci. 42 (1997) 1–33.
[134] J. Sita, B. Found, D.K. Rogers, Forensic handwriting examiners' expertise for signature comparison, J. Forensic Sci. 47 (2002) 1117–1124.
[135] R.M. Hogarth, T. Lejarraga, E. Soyer, The two settings of kind and wicked learning environments, Curr. Dir. Psychol. Sci. 24 (2015) 379–385.
[136] R.M. Hogarth, Educating Intuition, The University of Chicago Press, Chicago, IL, 2001.
[137] N.K. Choudhry, R.H. Fletcher, S.B. Soumerai, Systematic review: The relationship between clinical experience and quality of health care, Ann. Intern. Med. 142 (2005) 260–273.
[138] K.A. Ericsson, Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains, Academic Medicine: Journal of the Association of American Medical Colleges 79 (2004) S70–S81.
[139] L.J. Gunn, F. Chapeau-Blondeau, M. McDonnell, B. Davis, A. Allison, D. Abbott, Too good to be true: When overwhelming evidence fails to convince, Proceedings of

the Royal Society A: Mathematical, Physical and Engineering Sciences, 472, 2016 1–15.
[140] F. Galton, Vox populi, Nature 75 (1907) 450–451.
[141] J. Surowiecki, The Wisdom of Crowds: Why the Many Are smarter than the Few, Little Brown, New York, 2004.
[142] D. White, A.M. Burton, R.I. Kemp, R. Jenkins, Crowd effects in unfamiliar face matching, Appl. Cogn. Psychol. 27 (2013) 769–777.
[143] S.S. Arora, K. Cao, A.K. Jain, G. Michaud, Crowd Powered Latent Fingerprint Identification: Fusing AFIS with Examiner Markups, Proceedings of the 8th International Conference on Biometrics (ICB), 2015.
[144] P.C. Price, E.R. Stone, Intuitive evaluation of likelihood judgement producers: evidence for a confidence heuristic, J. Behav. Decis. Mak. 17 (2004) 39–57.
[145] G.L. Wells, R.C.L. Lindsay, T.J. Ferguson, Accuracy, confidence, and juror perceptions in eyewitness identification, J. Appl. Psychol. 64 (1979) 440–448.
[146] R.J. Cramer, S.L. Brodsky, J. DeCoster, Expert witness confidence and juror personality: their impact on credibility and persuasion in the courtroom, Journal of the American Academy of Psychiatry and the Law 37 (2009) 36–74.
[147] A. Champagne, D. Shuman, E. Whitaker, An empirical examination of the use of expert witnesses in American courts, Jurimetrics Journal 31 (1992) 375–392.
[148] A.N.D. Meyer, V.L. Payne, D.W. Meeks, R. Rao, H. Singh, Physicians' diagnostic accuracy, confidence, and resource requests, JAMA Intern. Med. 173 (2013) 1952–1958.
[149] B.M. DePaulo, K. Charlton, H. Cooper, J.J. Lindsay, L. Mulhenbruck, The accuracy-confidence correlation in the detection of deception, Personal. Soc. Psychol. Rev. 1 (1997) 346–357.
[150] T.M. Marteau, M. Johnston, G. Wynne, T.R. Evans, Cognitive factors in the explanations of the mismatch between confidence and competence in performing basic life support, Psychol. Health 3 (1989) 173–182.
[151] G.L. Wells, D.M. Murray, Eyewitness Confidence, in: G.L. Wells, E.F. Loftus (Eds.), Eyewitness Testimony: Psychological Perspectives, Cambridge University Press, New York, NY 1984, pp. 155–170.
[152] J.T. Wixted, L. Mickes, S.E. Clark, S.D. Gronlund, Initial eyewitness confidence reliably predicts eyewitness identification accuracy, The American Psychologist 70 (2015) 515–526.
[153] J.C. Brigham, L.B. Bennett, C.A. Meissner, T.L. Mitchell, R.C.L. Lindsay, D.F. Ross, J.D. Read, M.P. Toglia, The influence of race on eyewitness memory, Handbook of Eyewitness Psychology 2 Volume Set, Psychology Press, 2014.
[154] S. Oskamp, Overconfidence in case-study judgements, J. Consult. Psychol. 29 (1965) 261–265.
[155] D. Kahneman, G. Klein, Conditions for intuitive expertise: A failure to disagree, Am. Psychol. 64 (2009) 515–526.
[156] M.H. Herzog, M. Fahle, The role of feedback in learning a Vernier discrimination task, Vis. Res. 37 (1997) 2133–2141.
[157] R.M. Dawes, D. Faust, P.E. Meehl, Clinical versus actuarial judgment, Science 243 (1989) 1668–1674.
[158] J. Hattie, H. Timperley, The power of feedback, Rev. Educ. Res. 77 (2007) 81–112.
[159] L. Baudry, D. Leroy, R. Thouvarecq, D. Chollet, Auditory concurrent feedback benefits on the circle performed in gymnastics, Journal of Sports 24 (2006) 149–156.
[160] R.L. Goldstone, Perceptual learning, Annu. Rev. Psychol. 49 (1998) 585–612.
[161] D. White, R.I. Kemp, R. Jenkins, A.M. Burton, Feedback training for facial image comparison, Psychon. Bull. Rev. 21 (2014) 100–106.
[162] R.I. Kemp, N.A. Towell, G.E. Pike, When seeing should not be believing: photographs, credit cards and fraud, Appl. Cogn. Psychol. 11 (1997) 211–222.
[163] C. Gonzalez, P. Madhavan, Diversity during training enhances detection of novel stimuli, J. Cogn. Psychother. 23 (2011) 342–350.
[164] N.K. Steblay, G.L. Wells, A.B. Douglass, The eyewitness post identification feedback effect 15 years later: theoretical and policy implications, Psychol. Public Policy Law 20 (2014) 1–18.
[165] NSWLRC, NSW Law Reform Commission, Report 109: Expert WitnessesSydney 2005.
[166] Law Commission, Expert Evidence in Criminal Proceedings in England and Wales, in, 2009.
[167] H. Wootten, Conflicting imperatives: Pursuing truth in the court, in: I. McCalman, A. McGrath (Eds.), Proof & Truth: The Humanist as Expert, Australian Academy of the Humanities, Canberra 2003, pp. 15–50.