Target article

The forensic confirmation bias: Problems, perspectives, and proposed solutions

Saul M. Kassin [a,*], Itiel E. Dror [b], Jeff Kukucka [a]

[a] *John Jay College of Criminal Justice, United States*
[b] *University College London (UCL), United Kingdom*

ARTICLE INFO

ABSTRACT

As illustrated by the mistaken, high-profile fingerprint identification of Brandon Mayfield in the Madrid Bomber case, and consistent with a recent critique by the National Academy of Sciences (2009), it is clear that the forensic sciences are subject to contextual bias and fraught with error. In this article, we describe classic psychological research on primacy, expectancy effects, and observer effects, all of which indicate that context can taint people's perceptions, judgments, and behaviors. Then we describe recent studies indicating that confessions and other types of information can set into motion *forensic confirmation biases* that corrupt lay witness perceptions and memories as well as the judgments of experts in various domains of forensic science. Finally, we propose best practices that would reduce bias in the forensic laboratory as well as its influence in the courts.

© 2013 Society for Applied Research in Memory and Cognition. Published by Elsevier Inc. All rights reserved.

1. The problem

On March 11, 2004, a coordinated series of bombs exploded in four commuter trains in Madrid. The explosions killed 191 people, wounded 1800 others, and set into motion a full-scale international investigation. On the basis of a latent fingerprint lifted from a bag containing detonating devices, the U.S. Federal Bureau of Investigation (FBI) positively identified Brandon Mayfield, an American Muslim from the state of Oregon. Subsequent to 9–11, Mayfield had been on an FBI watch list. Following standard protocol, a number of FBI fingerprint examiners independently concluded that the fingerprint was definitely that of Mayfield. After being arrested and appearing in court, Mayfield requested to have a fingerprint examiner on the defense team examine the prints. That fingerprint examiner concurred with the judgment that the print was Mayfield's. Soon thereafter, however, the Spanish authorities matched the prints to the real Madrid bomber, an Algerian national by the name of Ouhnane Daoud. Following an internal investigation at the FBI and a report by the Office of the Inspector General (OIG, 2006), "confirmation bias" was listed as a contributing factor to the erroneous identification. At that point, the U.S. government issued a formal apology, and paid two million dollars in compensation.

The FBI has rigorous standards of training and practice and highly competent forensic examiners. It is considered one of the best, if not *the* best forensic laboratories in the U.S., if not in the entire world. Thus, it was not easy to dismiss the error and

claim it to be the product of mere "bad apples." The Mayfield case (preceded by a decade in which the U.S. Supreme Court had sought to curb the introduction at trial of experts in junk science—see *Daubert v. Merrell Dow Pharmaceuticals, 1993; Kumho Tire Co. v. Carmichael*, 1999), along with improprieties discovered in various state laboratories, have come together to draw attention to forensic science and to the fact that is not infallible. Forensic science errors have also surfaced with alarming frequency in DNA exoneration cases and other wrongful convictions (Garrett, 2011; http://www.innocenceproject.org/fix/Crime-Lab-Oversight.php). In "The genetics of innocence," Hampikian, West, and Akselrod (2011) found that several types of forensic science testimony had been used to wrongfully convict innocent individuals. In cases where trial transcripts or reliable forensic science data were available for review, 38% contained incorrect serology testimony, which is highly regarded. In addition, 22% involved hair comparisons; 3% involved bite mark comparisons; and 2% involved fingerprint comparisons.

The National Academy of Sciences (NAS, 2009) published a scathing assessment of a broad range of forensic disciplines. Included in this critique were toolmarks and firearms; hair and fiber analysis; impression evidence; blood spatter; fibers; handwriting; and even fingerprints—until recently considered infallible. NAS concluded that there are problems with standardization, reliability, accuracy and error, and the potential for contextual bias. Specifically, the NAS report went on to advise that: "These disciplines need to develop rigorous protocols to guide these subjective interpretations and pursue equally rigorous research and evaluation programs. The development of such research programs can benefit significantly from other areas, notably from the large body

---

* Corresponding author.
  E-mail address: Skassin@jjay.cuny.edu (S.M. Kassin).

of research on the evaluation of observer performance in diagnostic medicine and from the findings of cognitive psychology on the potential for bias and error in human observers" (p. 8).

The criticisms of the forensic sciences are twofold. First is the realization that too often the stimulus does not compel a perceptual judgment that is objective and, hence, there is a concern both for inter-rater reliability across experts and for intra-test reliability over time within experts. In many forensic disciplines, the human examiner is the main instrument of analysis. It is the forensic expert who compares visual patterns and determines if they are "sufficiently similar" to conclude that they originate from the same source (e.g., whether two fingerprints were made by the same finger, whether two bullets were fired from the same gun, or whether two signatures were made by the same person). However, determinations of "sufficiently similar" have no criteria and quantification instruments; these judgments are subjective. Indeed, a recent study has shown that when the same fingerprint evidence is given to the same examiners, they reach different conclusions approximately 10% of the time (Ulery, Hicklin, Buscaglia, & Roberts, 2012). Dror et al. (2011) have shown not only that the decisions are inconsistent but that even the initial perception of the stimulus, prior to comparison, lack inter- and intra-expert consistency.

Following from this realization about the lack of reliability is a corollary concern that forensic experts' judgments are "biasable"—that is, they are significantly influenced by psychological factors (Dror & Cole, 2010; Dror & Rosenthal, 2008). The biasability of forensic science is a particular concern because forensic experts work within a variety of contextual influences: Knowing the nature and details of the crime, being pressured by detectives; working within—and as part of—the police; the use of computer-generated lists that feature some suspects ahead of others; appearing in court within an adversarial criminal justice system. Describing the various sources of bias, Saks, Risinger, Rosenthal, and Thompson (2003) note that examiners often receive direct communications from police (e.g., in transmittal letters that accompany submitted evidence, in person, and by phone), that there is often cross-communication among different examiners involved in a case (e.g., via informal channels or as mandated in "peer review" processes designed to ensure the reasonableness of conclusions), and that police and prosecutors sometimes respond to non-supportive test results by requesting a re-examination. In short, the contextual influences that impinge on forensic examiners are numerous and they come in many forms, some of which are subtle. The erroneous identification in the Madrid bomber case illustrated a number of psychological factors at work (e.g., the latent fingerprint was examined against a pre-existing "target," without first being properly analyzed in isolation; the examiners were pre-armed with contextual information, leading them to be suspicious of their target; and the case was high in profile and time-urgent, increasing the need for closure).

In this article, we overview prior critiques of the forensic sciences and specific cases in which experts have rendered judgments that were fraught with bias and error. Then we consider classic psychological research on primacy, expectancy effects, and observer effects, and the various confirmation biases that can taint people's perceptions, judgments, and behaviors. Then we examine recent empirical work on confirmation biases in various domains of forensic science. Finally we use psychology to propose best practices that would minimize such effects—both in the crime laboratory and in the courtroom.

2. The forensic sciences: accuracy and error

For over 100 years forensic science disciplines have produced evidence used both to prosecute and convict criminals as well as to exonerate and release those who are innocent. The domains of forensic science are varied and include judgments of fingerprints, firearms examinations, toolmarks, bite marks, tire and shoe impressions, bloodstain pattern analysis, handwriting, hair, coatings such as paint and chemicals—including drugs and such materials as fibers, fluids, fire and explosive analysis, digital evidence, and serological analysis.

Since the 1990s, advances in DNA technology have proved particularly useful in these regards. Many previously unsolved crimes have been solved because of DNA samples left in hair, semen, blood, skin, and saliva. Often, however, these DNA cases have revealed that faulty forensic sciences have contributed to the wrongful convictions of innocent people. As exposed by more than 300 DNA exonerations identified by the Innocence Project, two sets of problems have come to light: (1) Forensic science judgments are often derived from inadequate testing and analysis, if not outright fabrication; and (2) Experts often give imprecise or exaggerated testimony, drawing conclusions not supported by the data—in some cases drawing charges of misconduct. Indeed, some form of invalid or improper forensic science was a contributing factor in the original convictions of more than half of all DNA exonerees (Garrett, 2011; http://www.innocenceproject.org/understand/Unreliable-Limited-Science.php).

In cases that are not subject to bias, certain forensic sciences—such as latent fingerprint identifications—offer a potentially powerful tool in administering justice (e.g., Tangen, Thompson, & McCarthy, 2011; Ulery, Hicklin, Buscaglia, & Roberts, 2011). In most domains, however, there are no quantitatively precise objective measures and no instruments of measurement—just partial samples from a crime scene to be compared against a particular suspect. No two patterns are identical, so an examiner invariably must determine whether they are "sufficiently similar" (a term that has yet to be defined or quantified) to conclude that they originate from the same source. The absence of objective standards is reflected in the lack of consistency not only between examiners but within examiners over time. Hence, not only do inter-variations exist, but intra-variations show that the same examiner inspecting the same data on multiple occasions may reach different conclusions (Ulery et al., 2012). The lack of reliability indicates that the identification process can be subjective and that judgments are susceptible to bias from other sources. This is especially problematic in cases that contain complex forms of forensic evidence, as is often the case in evidence gathered in crime scene.

Popular TV programs, such as *CSI*, communicate a false belief in the powers of forensic science, a problem that can be exacerbated when forensic experts overstate the strength of the evidence. Such occurrences are common when you consider the following: (1) Across many domains, experts are often overconfident in their abilities (e.g., Baumann, Deber, & Thompson, 1991); (2) the courts, for the most part, have blindly accepted forensic science evidence without much scrutiny (Mnookin et al., 2011); (3) errors are often not apparent in the forensic sciences because ground truth is often not known as a matter of certainty; (4) many forensic examiners work for police and appear in court as advocates for the prosecution; and (5) many forensic examiners consider themselves objective and immune to bias. As stated by the Chair of the Fingerprint Society: "Any fingerprint examiner who comes to a decision on identification and is swayed either way in that decision making process under the influence of stories and gory images is either totally incapable of performing the noble tasks expected of him/her or is so immature he/she should seek employment at Disneyland" (Leadbetter, 2007).

3. Classic confirmation biases: a psychological perspective

Over the years, research has identified a number of confirmation biases by which people tend to seek, perceive, interpret, and create new evidence in ways that verify their preexisting beliefs. Confirmation biases are a pervasive psychological phenomenon. Classic studies showed that prior exposure to images of a face or a body, an animal or a human, or letters or numbers, can bias what people see in an ambiguous figure. More recent research shows that our impressions of other people can similarly be tainted.

Recognition of confirmation bias as a human phenomenon is not new. Julius Caesar is cited to have said that "Men freely believe that which they desire" (e.g., Hochschild, 2008). References can also be found in the writings of William Shakespeare and Francis Bacon (Risinger, Saks, Thompson, & Rosenthal, 2002). Indeed, Nickerson (1998) notes that confirmation biases may be implicated in "a significant fraction of the disputes, altercations, and misunderstandings that occur among individuals, groups, and nations"—including, among others, the witch trials of Western Europe and New England, the continuation of ineffective medical treatments, inaccurate medical diagnoses, and adherence to erroneous scientific theories (p. 175).

3.1. Perceptual and cognitive effects

Contemporary work on confirmation biases began with classic research suggesting that the perception of a stimulus is not solely a function of the stimulus itself (i.e., "bottom-up" processing), but is also shaped by the qualities of the observer (i.e., "top-down" processing). For example, Bruner and Goodman (1947) asked children to estimate the size of coins from memory and found that children of low-SES overestimated the size of the coins to a greater degree than did children of high SES. Bruner and Potter (1964) demonstrated that one's expectations can also interfere with visual recognition. Participants were shown photographs of common objects (e.g., a dog, a fire hydrant, etc.) that had been blurred to various degrees, and then watched as the pictures were gradually brought into focus. The blurrier the photographs were at the start, the less able participants were to correctly recognize the objects later. Bruner and Potter explained these results by noting that participants readily generated hypotheses about the blurry images and then maintained these beliefs even as the pictures came into focus. Using simple ambiguous ("reversible") figures, other research as well showed that expectations shape perception (Boring, 1930; Leeper, 1935; for a compendium of such figures, see Fisher, 1968).

Recent studies have demonstrated similar effects using more complex stimuli. For example, Bressan and Dal Martello (2002) showed participants photographs of adult-child pairs and asked them to rate their facial resemblance. When led to believe that the adult and child were genetically related (e.g., parent and offspring), participants rated their facial similarity as higher – even when the two were not truly related. Other studies have similarly shown that people perceive more similarity between a suspect and a facial composite when led to believe the suspect is guilty (Charman, Gregory, & Carlucci, 2009); and people hear more incrimination in degraded speech recordings when the interviewee was thought to be a crime suspect (Lange, Thomas, Dana, & Dawes, 2011).

To sum up: A wealth of evidence indicates that an observer's expectations can impact visual and auditory perception. Although similar effects can be driven by motivation (Balcetis & Dunning, 2006, 2010; Radel & Clement-Guillotin, 2012), confirmation biases are a natural and automatic feature of human cognition that can occur in the absence of self-interest (Nickerson, 1998) and operate without conscious awareness (Findley & Scott, 2006; Kunda, 1990).

3.2. Social perception effects

Strong expectancy effects can also contaminate the processes of social perception. This research literature can be traced to Asch's (1946) initial finding of primacy effects in impression formation by which information about a person presented early in a sequence is weighed more heavily than information presented later which is ignored, discounted, or assimilated into the early-formed impression. Illustrating the process of assimilation, or "change of meaning" hypothesis, later research revealed that depending on one's first impression of a person, the word "proud" can mean self-respecting or conceited; "critical" can mean astute or picky; and "impulsive" can mean spontaneous or reckless (Hamilton & Zanna, 1974; Watkins & Peynircioglu, 1984). As a result of these processes, additional research has shown that beliefs, once they take root, can persist even after the evidence on which they were based has been discredited (Anderson, Lepper, & Ross, 1980). In fact, the presence of objective evidence that can be selectively interpreted may *exacerbate* the biasing effects of pre-existing beliefs (Darley & Gross, 1983).

Research on confirmatory hypothesis testing also explains the power and resistance to change of first impressions. In a classic experiment, Wason (1960) gave participants a three-number sequence, challenged them to discern the rule used to generate the set, and found that very few discovered the correct rule because once they seized upon a hypothesis they would search only for confirming evidence (see also Klayman & Ha, 1997). In a social-interactional context, Snyder and Swann (1978) brought together pairs of participants for a getting-acquainted interview. In each pair, interviewers were led to believe that their partner was either introverted or extroverted. Expecting a certain kind of person, participants unwittingly sought evidence that would confirm their expectations: Those in the introverted condition chose to ask mostly introvert-oriented questions ("Have you ever felt left out of some social group?"); those in the extroverted condition asked extrovert-oriented questions ("How do you liven up a party?"). In doing so, interviewers procured support for their beliefs, causing neutral observers who later listened to the tapes to perceive the interviewees as introverted or extroverted on the basis of their randomly assigned condition.

The fact that people can be jaded by existing beliefs is a phenomenon of potential consequence in forensic settings. In one study, participants reviewed a mock police file of a crime investigation that contained weak circumstantial evidence pointing to a possible suspect. Some participants but not others were asked to form and state an initial hypothesis as to the likely offender. Those who did so proceeded to search for additional evidence and interpret that evidence in ways that confirmed their hypothesis. Hence, a weak suspect became the *prime* suspect (O'Brien, 2009). In another study, Kassin, Goldstein, and Savitsky (2003) had some participants but not others commit a mock crime, after which all were questioned by interrogators who by random assignment were led to presume guilt or innocence. Interrogators who presumed guilt asked more incriminating questions, conducted more coercive interrogations, and tried harder to get the suspect to confess. In turn, this more aggressive style made the suspects sound defensive and led observers who later listened to the tapes to judge them as guilty, even when they were innocent. Follow-up research has confirmed variants of this latter chain of events in the context of suspect interviews (Hill, Memon, & McGeorge, 2008; Narchet, Meissner, & Russano, 2011).

An individual's prior beliefs can produce dramatic behavioral consequences as well, often setting into motion a three-step behavioral confirmation process by which a perceiver forms an impression of a target person, interacts in a manner that is consistent with that impression, and causes the target person unwittingly

to adjust his or her behavior. The net result: a process that transforms expectations into reality (Darley & Fazio, 1980; Rosenthal & Jacobson, 1966; Snyder & Swann, 1978).

In an early demonstration of this phenomenon, Rosenthal and Fode (1963) reported on an experimenter expectancy effect, whereby an experimenter who is aware of the hypothesis of a study and the condition to which a participant is assigned can unwittingly produce results consistent with the expected outcome. Thus, when students were led to believe that the rats they would be training at maze learning were bright or dull, those rats believed to be bright learned more quickly (for an overview of this research, see Rosenthal, 2002). In subsequent research on *teacher* expectancy effects, Rosenthal and Jacobson (1966) extended these findings to human participants and found that when elementary school teachers were led to believe that certain of their students, randomly assigned, were on the verge of an intellectual growth spurt, those selected students exhibited greater improvement in academic tests eight months later. Whether training rats or teaching students, it appears that people unwittingly act upon their beliefs in ways that produced the expected outcomes. Although the interpretation of the teacher expectancy effect is a source of some controversy (Jussim, 2012), self-fulfilling prophecies have amply been demonstrated not only in the laboratory but in schools and other types of organizations as well (for reviews, see Kierein & Gold, 2000; McNatt, 2000).

3.3. Cognitive and motivational sources of bias

It is clear that belief-confirming thought processes are an inherent feature of human cognition. In their classic studies, Tversky and Kahneman (1974) demonstrated that people naturally rely on various cognitive *heuristics* – and that heuristic thinking, while generally beneficial, can also produce systematic errors in judgment, especially where strong prior expectations exist. Over time, and across a range of domains, basic psychological research has shown that strong expectations provide a sufficient and unwitting trigger of our tendency to seek, perceive, interpret, and create new evidence in ways that verify preexisting beliefs.

At times, confirmation biases can be fueled by *motivational goals*. Kunda (1990) argued that motivation influences reasoning indirectly as a result of two types of goals: *accuracy goals*, where individuals strive to form an accurate belief or judgment, and *directional goals*, where individuals seek a particular desired conclusion. In the latter case, people maintain an "illusion of objectivity" that prevents them from recognizing that their cognition has been tainted by preference or desire (Kunda, 1990, p. 483). Motivated reasoning is pervasive. Hence, people exhibit a ubiquitous self-serving positivity bias in the attributions they make for their own successes and failures (Mezulis, Abramson, Hyde, & Hankin, 2004). Likewise, people's attributions for external events are influenced by their political ideologies (Skitka, Mullen, Griffin, Hutchinson, & Chamberlin, 2002).

Recent empirical research supports the notion that directional goals can unconsciously guide perception. In a series of studies, Balcetis and Dunning (2006) showed participants an ambiguous figure that could be readily perceived as either of two different stimuli (e.g., the letter "B" or the number "13"). Depending on which stimulus they perceived, participants were assigned either to drink orange juice or a foul-smelling beverage. For those told that a letter would assign them to the orange juice condition, 72% saw the letter B. For those told that a number would assign them to the orange juice, 61% saw the number 13. Using an array of methods, follow-up studies showed that these results were not due to selective reporting but rather that motivation had a genuine unconscious effect on perception. In additional research on "wishful seeing," Balcetis and Dunning (2010) found that people judged objects that they want as

physically closer than more neutral objects (e.g., participants who were thirsty compared to those who were quenched estimated that a bottle of water across a table was closer to them).

Perceptions of form and distance are not limitlessly malleable, even among people who are highly motivated. As Kunda (1990) noted, "people do not seem to be at liberty to conclude whatever they want to conclude merely because they want to" (p. 482). To some extent, reality constrains perception. Evidence in favor of one's biased judgment must be sufficient to allow for the construction of that judgment; a desired outcome cannot be rationalized in the face of irrefutable evidence to the contrary. This is precisely why ambiguous stimuli prove particularly susceptible to confirmation biases. It is also why many forensic judgments are subject to bias.

4. The forensic confirmation bias

Nearly 40 years ago, Tversky and Kahneman (1974) reasoned that confirmation bias effects could extend to the legal system insofar as "beliefs concerning the likelihood of. . . the guilt of a defendant" could impact judicial decision-making (p. 1124). They further speculated that the operation of such biases would affect not only the layperson but also experienced professionals. These statements proved quite prescient. Empirical and anecdotal evidence now suggests that pre-judgment expectations can indeed influence interrogators (Hill et al., 2008; Kassin, Goldstein, & Savitsky, 2003; Narchet et al., 2011), jurors (Charman et al., 2009; Lange et al., 2011), judges (Halverson, Hallahan, Hart, & Rosenthal, 1997), eyewitnesses (Hasel & Kassin, 2009), and experts in a range of forensic domains (e.g., see Dror & Cole, 2010; Dror & Hampikian, 2011).

Thus, we use the term *forensic confirmation bias* to summarize the class of effects through which an individual's preexisting beliefs, expectations, motives, and situational context influence the collection, perception, and interpretation of evidence during the course of a criminal case. As Findley and Scott (2006) have noted, the pernicious result produces a form of "tunnel vision"—a rigid focus on one suspect that leads investigators to seek out and favor inculpatory evidence, while overlooking or discounting any exculpatory evidence that might exist. A growing body of literature has begun to identify the ways in which such biases can pervade the investigative and judicial processes.

4.1. Context effects on forensic judgments

In an 1894 treatise on distinguishing genuine from forged signatures, William Hagan wrote: "There must be no hypothesis at the commencement, and the examiner must depend wholly on what is seen, leaving out of consideration all suggestions or hints from interested parties. . . Where the expert has no knowledge of the moral evidence or aspects of the case. . . there is nothing to mislead him" (p. 82). With this statement, Hagan was among the first scholars to acknowledge the potential biasing effect of expectation and context on perceptual judgments made by forensic examiners. It was not until recently, however, that empirical data emerged to support Hagan's admonition.

A growing body of work now suggests that confessions, a highly potent form of incrimination (Kassin, 1997; Kassin et al., 2010)—and other strong contextual cues—may bias forensic judgments in the criminal justice system, producing an effect that Kassin (2012) has called "corroboration inflation." Saks et al. (2003) note that the resulting non-independence among items of evidence can create an "investigative echo chamber" in which certain items reverberate and seem stronger and more numerous than they really are. Simon (2011) notes that coherence-based reasoning promotes

false corroboration among different witnesses, resulting in trials that are limited in their diagnostic value. Dror (2012) notes that the overall effect on judgments can increase as a result, creating a "bias snowball effect."

To our knowledge, the first study to examine this effect was by Miller (1984), who explored the impact of contextual information on the judgments of 12 college students trained to identify forged signatures. Miller found that participants who were exposed to additional inculpatory evidence formed a belief in the suspect's guilt, which skewed their perceptions. More recent work builds upon this finding. Kukucka and Kassin (2012) found that knowledge of a recanted confession can taint evaluations of handwriting evidence. In this study, lay participants read a bank robbery case in which the perpetrator gave a handwritten note to a bank teller. Soon afterward, they were told that a suspect was apprehended and interrogated, at which point he gave a handwritten *Miranda* waiver. Participants were asked to compare the handwriting samples taken from the perpetrator (bank note) and the defendant (Miranda waiver). When told that the defendant had confessed—even though he later retracted his confession, claiming it was coerced—participants perceived the handwriting samples as more similar and were more likely to conclude, erroneously, that they were authored by the same individual.

Other research indicates that interpretations of polygraph tests may also be shaped by preexisting beliefs. Elaad, Ginton, and Ben-Shakhar (1994) noted two ways in which expectations can impact the outcome of a polygraph test: By influencing the way examiners conduct their interviews and the questions they ask, and by influencing the conclusions they draw from the test results. To test the latter hypothesis, these investigators asked ten polygraph examiners from the Israeli Police to analyze 14 records from polygraph examinations of criminal suspects, all of whom had been judged inconclusive by independent raters. Each chart was accompanied by biasing information—for half of the charts, examiners were told that the interviewee had later confessed; for the remaining half, they were told that someone else had later confessed. Although most charts were judged inconclusive in the absence of biasing information, the charts were more likely to be scored as deceptive in the suspect-confession condition and as truthful in the other-confession condition. This effect was obtained with both experienced and inexperienced examiners—but not when the charts were conclusive. Thus, the conclusions drawn from ambiguous polygraph results were influenced by prior expectations.

Additional studies suggest that even fingerprint judgments may be subject to bias. In one study, Dror, Charlton, and Peron (2006) asked five experienced fingerprint experts to assess pairs of fingerprints that, unbeknownst to them, they had examined years earlier and declared to be a match. Before the stimuli were re-presented, these examiners were told that the fingerprints were taken from a high-profile case of erroneous identification, implying that they were not a match. Given this biasing information, only one of the five experts judged the fingerprints to be a match, indicating that context undermined reliability. This study is particularly troubling because the change as a function of context was obtained among experienced examiners, in a highly trusted forensic science, and in a within-subject experimental design.

In a followup study, Dror and Charlton (2006) presented six latent fingerprint experts with eight pairs of prints from a crime scene and suspect in an actual case in which they had previously made a match or exclusion judgment. The participants did not know they were taking part in a study, believing instead that they were conducting routine casework. The prints were accompanied either by no extraneous information, information that the suspect had confessed, suggesting a match; or information that the suspect was in custody at the time, suggesting exclusion. The results showed

that contextual information in the custody condition produced an overall change in 17% of the originally correct match decisions.

Based on a meta-analysis of these two studies, Dror and Rosenthal (2008) estimated that the reliability of fingerprint experts' judgments over time likely falls in the range of 0.33–0.80, implying a considerable degree of subjectivity. Similarly, effect size estimates of biasability were 0.45 and 0.41, respectively, for the two studies. These findings are likely to extend to other forensic science domains that are based on visual similarity judgments, such as firearms; microscopic hair and fiber analysis; bite marks; impression evidence involving shoeprints, bite marks, tire tracks, and handwriting; and bloodstain pattern analysis (Dror & Cole, 2010).

Additional research suggests that confessions can also influence the testimony of lay witnesses. Looking at the possible effects of confession on eyewitnesses themselves, Hasel and Kassin (2009) staged a theft and took photographic identification decisions from eyewitnesses who viewed a culprit-absent lineup. Two days later, individual witnesses were told that the person they had identified denied guilt during a subsequent interrogation, or that he confessed, or that a specific other lineup member confessed. Among those who had made a selection but were told that another lineup member confessed, 61% changed their identifications—and did so with confidence. Among those who had correctly not made an initial identification, 50% went on to select the confessor.

The biasing effect of confessions can have grave consequences. The criminal justice system presupposes that suspects, eyewitnesses, forensic experts, and others offer information that is independent—not subject to taint from outside influences. But does this presupposition describe the reality of criminal investigation? Both basic psychology and forensic psychology research suggest otherwise—and, in particular, suggest the possibility that confessions can corrupt other evidence. To determine if this phenomenon might occur in actual cases, Kassin, Bogart, and Kerner (2012) conducted an archival analysis of DNA exonerations from the Innocence Project case files. Testing the hypothesis that confessions may prompt additional evidentiary errors, they examined whether other contributing factors were present in DNA exoneration cases containing a false confession. They found that additional errors were present in 78% of these cases. In order of frequency, false confessions were accompanied by invalid or improper forensic science (63%), mistaken eyewitness identifications (29%) and snitches or informants (19%). Consistent with the causal hypothesis that the false confessions had influenced the subsequent errors, the confession was obtained first rather than later in the investigation in 65% of these cases.

As a result of improprieties in U.S. laboratories, the frequency with which forensic science errors have surfaced in wrongful convictions, and the scathing critique from the National Academy of Sciences (2009)—which concluded that there are problems with standardization, reliability, accuracy and error, and the potential for contextual bias—it is not surprising that the most common means of corroboration for false confessions comes from bad forensic science (http://www.innocenceproject.org/). When coupled with recent laboratory studies, this presence of numerous forensic errors in Innocence Project confession cases suggests that confession evidence constitutes the kind of contextual bias that can skew expert judgments in many domains.

Confession is not the only form of evidence that can bias people's judgments. Mistaken eyewitness identifications constitute the most common contributing factor in DNA exoneration cases (Brewer & Wells, 2011; Wells et al., 1998). In fact, many Innocence Project cases contained two or more mistaken eyewitnesses who expressed high levels of certainty in their identifications. In some instances, these multiple errors can occur independently—as in the highly publicized mistaken identification of Ronald Cotton by

Jennifer Thompson, where Cotton physically resembled the perpetrator (Thompson-Cannino, Cotton, & Torneo, 2009). In other instances, however, the eyewitnesses may have influenced one another, a phenomenon demonstrated in numerous co-witness experiments (Gabbert, Memon, & Allan, 2003; Skagerberg, 2007). To further complicate matters, eyewitnesses who have been tainted by extrinsic information cannot accurately estimate the extent of the influence, suggesting that self-reports cannot be used to diagnose the corruption once it occurs (Charman & Wells, 2008).

4.2. Elasticity of forensic evidence

It is not surprising that expectations can taint questioned document examination (QDE), the discipline pertaining to documents, the authenticity or source of which are in dispute. QDE has been criticized for being a subjective domain of forensic science (Miller, 1984; Risinger et al., 2002; Risinger & Saks, 1996; *U.S. v. Hines, 1999*). In accordance with the research described earlier, examiners are more likely to exhibit bias when evaluating evidence that is ambiguous. This is consistent with Ask, Rebelius, and Granhag's (2008) assertion that some types of evidence are more "elastic"—i.e., more vulnerable to extraneous influence—than others.

Not all evidence is equally malleable or subject to confirmation bias. Paralleling classic research indicating that expectations can color judgments of stimuli that are ambiguous but not those that compel a particular perception, forensic research indicates that ambiguity is a moderating condition. Asked to make an identification decision on the basis of a memory trace that cannot be recovered for a side-by-side comparison to a stimulus face, eyewitnesses are particularly malleable when informed of a confession (Hasel & Kassin, 2009). Prior expectations can also bias interpretations of sensory stimuli such as auditory speech—but only when the recordings are degraded in quality and the stimuli are phonologically ambiguous, such as the words *gum* and *gun* or *ripped* and *raped* (Lange et al., 2011). The same is true of the judgments of polygraph examiners—again, when the physiological test data are ambiguous but not when the charts are strongly indicative of truth or deception (Elaad, Ginton, & Ben-Shakhar, 1994).

Still, within the forensic domains critiqued by the National Academy of Sciences (2009), the potential for bias is greater than previously imagined. In "The vision in 'blind' justice," Dror and Cole (2010) noted that many forensic judgments involve matching a visual pattern left at a crime scene with a sample taken from a suspect (e.g., shoe prints, tool marks, bite marks, tire marks, handwriting). The prototype is fingerprint identification, a forensic science long considered near-perfect (Cole, 2001). No two fingerprint impressions are totally identical because of variations in skin elasticity, the amount of pressure applied, the material on which the print was left, how the prints were recovered and other variables. And in criminal cases, where prints are lifted from crime scenes, many such latent fingerprints are partial and distorted. Hence, an impressive body of research now indicates that the judgments made by latent fingerprint experts are sensitive to biasing contextual information (Charlton, Fraser-Mackenzie, & Dror, 2010; Dror & Charlton, 2006; Dror, Charlton, & Peron, 2006; Dror, Peron, Hind, & Charlton, 2005).

Even when it comes to DNA testing—commonly considered the "gold standard" of forensic evidence (Lieberman, Carrell, Miethe, & Krauss, 2008; Lynch, 2003; Saks & Koehler, 2005)—the interpretation of certain complex DNA mixtures requires judgment that is subject to bias. To illustrate the risk, Dror and Hampikian (2011) described an actual gang rape case in which one of the assailants had accepted a plea bargain in exchange for testimony against other suspects. In order for the testimony of the cooperating assailant to be admissible, evidence was needed to corroborate his

identifications. Aware of the situation, expert DNA analysts were asked to analyze the complex DNA mixture, and they concluded that the forensic evidence implicated those identified in the plea bargain. However, one of the alleged assailants repeatedly denied any involvement in the rape. To test the potential for contextual bias, Dror and Hampikian later took the same sample from this case and presented it, devoid of the biasing contextual information, to 17 neutral DNA analysts. Only one agreed with the original analysts; four deemed the sample inconclusive; 12 concluded that the DNA excluded the suspect in question. Despite the claim that DNA evidence is "inelastic" (e.g., Ask et al., 2008), it thus appears that confirmation biases may influence even the work of DNA analysts.

4.3. Bias and self-insight

Although confirmation bias typically operates outside of conscious awareness, forensic examiners may have some insight into the cognitive, motivational, and emotional factors that guide their job performance. Charlton et al. (2010) conducted semi-structured interviews of 13 experienced fingerprint examiners and identified a number of recurrent themes in their experiences. While describing their methodology in an objective manner, examiners expressed a personal interest in catching criminals and solving crimes, which some reported as more pronounced in serious and high-profile cases. They also expressed a strong need for closure, indicating a desire to provide definitive conclusions as a result of their work, and the feeling of joy that accompanies the discovery of a fingerprint match. At the same time, these experts consistently expressed a fear of making erroneous judgments, and in particular, a fear of committing a false-positive error that would implicate an innocent person. Thus, perhaps some experts deliberately endeavor to be conservative in their judgments to avoid such errors.

Furthermore, mere awareness of the type of crime being investigated may not be sufficient to bias fingerprint expert judgments. Utilizing an experimental paradigm, Hall and Player (2008) asked experienced examiners to judge pairs of fingerprints either in the context of a forgery case or a murder case—but no emotionally-arousing crime scene photos were included. Results indicated that examiners in the murder condition were more likely to self-report feeling influenced by this context, but the type of case had no overall effect on their conclusions. Perhaps for context to influence judgments, participants must really believe it. In short, there may be a dissociation between forensic examiners' insight into their own biases and the actual manifestation of bias in their actual judgments; they can be biased and unaware, or they can be relatively objective despite the self-perception of bias.

4.4. Null effects from the Netherlands

It is important to note that two additional studies from a single lab failed to replicate confirmation bias effects on forensic experts. First, Kersholt, Paashuis, and Sjerps (2007) recruited twelve Dutch officers trained in forensic shoe print examinations and asked them to evaluate eight pairs of shoes and prints. Each pair was presented in the context of a fictional criminal investigation, which either did or did not contain biasing information to suggest that the shoe had created the print. This manipulation had no effect on evaluations. Similarly, Kersholt et al. (2010) had six Dutch firearms examiners judge six pairs of bullets that were presented twice, several months apart. Each pair of bullets was presented twice—once with, and once without, a biasing case description—to be categorized as a match, a non-match, or inconclusive. Overall, 10 out of 36 judgments of the same pair of bullets changed from one presentation to the next, indicating a problem with intra-examiner reliability and subjectivity. However, the bias manipulation did not have a significant effect on judgments.

To account for these replication failures, Kersholt et al. (2010) note that each of these forensic sciences utilizes a highly-standardized procedure in the Netherlands (e.g., shoe print examiners follow a protocol whereby they assign numerical values to various features of the shoe print and then sum the values to obtain a total score). Perhaps examiners in these studies were not biased by expectations precisely because they are trained in evaluation procedures that are well-defined. Alternatively, perhaps the biasing information used in these studies failed to create a strong expectation of guilt. As far as we can tell, the expectation manipulations had not been pilot tested nor was their effectiveness confirmed through manipulation checks. In the study of shoe print examinations, one of the fictional cases described the burglary of an electronics store. To raise the expectation that the print matched the suspect's shoe, examiners were told that the suspect had been found selling electronics on the street and that he owned a van (which would presumably be needed to transport large electronics). Each of these facts arguably constitutes a necessary but not sufficient condition to imply guilt, and thus may not have cultivated an a priori belief that the shoe would match the print. As in Hall and Player (2008), these studies also involved examiners who knew they were taking part in a study—not involved in actual criminal casework (Dror, 2009b).

4.5. Forensic confirmation bias in actual cases

The biases set into motion by confessions and other guilt-presumptive sources of information are not without consequence. A growing number of real-world wrongful convictions, as seen in the opening story about the Madrid bombing case in which Brandon Mayfield was misidentified and as reported in many cases from the Innocence Project, provide ample real world instantiation of this hypothesis.

In one case, in Pennsylvania, suspect Barry Laughman was induced to confess during an unrecorded interrogation to the rape and murder of his elderly neighbor. The next day, serology tests showed that Laughman had Type B blood; yet the semen recovered from the victim was from a Type A secretor. Aware that Laughman had confessed, the state forensic chemist went on to propose four "novel" theories, none grounded in science, to dismiss the mismatch. On the basis of his confession, Laughman was convicted. Sixteen years later, he was exonerated by DNA and set free (http://www.innocenceproject.org/Content/Barry_Laughman).

Another example can be found in the 2004 trial of Mississippi v. Tyler Edmonds. In that case, 13 year-old Edmonds was induced to confess that he had physically assisted his older half-sister in the shooting and killing of her husband. Supporting what had become disputed confession, the state's medical pathologist who conducted the autopsy of the victim's body and submitted his report after the confession was taken testified without any basis in science that the gunshot wound suggested a bullet fired by two persons pulling the trigger simultaneously. Edmonds was convicted at trial and sentenced to life in prison. Highly critical of this expert's "speculative" and "scientifically unfounded" opinion, the state Supreme Court overturned the conviction (Tyler Edmonds v. State of Mississippi, 2007). The following year, Edmonds was retried and acquitted. After an investigation by the state's medical board, the pathologist in question was removed from the state's designated list of pathologists.

4.6. Implications for accuracy and error

The fact that confessions and other strong bases for a presumption of guilt can bias the search, collection, perception, and interpretation of subsequently obtained evidence undermines a silent but basic tenet of the judicial system—namely, that the items

of evidence presented at trial are independent of one another. When one witness influences another, then a strong bias is created, creating what Kassin (2012) described as "corroboration inflation" and a gathering momentum for more and more bias, or what Dror (2012) referred to as a "bias snowball effect." The influence of one witness or item of evidence on another witness or item of evidence constitutes a biasing process of confirmation, one that can increase the likelihood of error. In the Texas arson-murder case against Cameron Todd Willingham, for example, eyewitnesses changed their account once told about forensic evidence suggesting that the fire was not accidental. Although this forensic conclusion was later found to be erroneous, Willingham was found guilty and executed (Grann, 2009).

Just as forensic science is subject to bias, so too are suspects pressed for confession and eyewitnesses pressed for identification. Many of the studies described above focused on how confessions can spawn other incriminating evidence. This influence can be bidirectional; just as confessions can taint other evidence, other evidence can taint confessions as well. Indeed, numerous studies and case anecdotes support the fact that innocent people can be induced to confess by the true or false presentation of an eyewitness, physical evidence, failed polygraph, or other incriminating evidence (e.g., Gudjonsson, 2003; Kassin, 1997; Kassin & Gudjonsson, 2004; Kassin & Kiechel, 1996; Nash & Wade, 2009; Perillo & Kassin, 2011). In one case, for example, Dwayne Jackson confessed to a crime he did not commit after he was erroneously identified in DNA testing by Las Vegas forensic examiners (Mower & McMurdo, 2011).

Forensic examiners are aware of and trained to avoid physical contamination in an effort to protect the integrity of the evidence. However, "psychological contamination" has not received similar attention and is prevalent throughout the criminal justice system. The sources of psychological contamination are numerous (e.g., knowing the context of the crime, police pressure to influence a forensic evaluation, information about a prior confession or eyewitness identification). Biasing context can take on other subtle forms as well. For example, forensic examiners work with a variety of technologies—including computerized systems that suggest a list of candidates for the human examiner to consider. In a recent study, Dror, Wertheim, Fraser-Mackenzie, and Walajtys (2012) independently varied the order of the candidates on the list and found that examiners spent less time on the same candidate when it was placed further down the list. Examining 55,200 forensic decisions, these investigators also found that examiners are more likely to make false positive errors on candidates on the top of the list and false negative errors on those near the bottom. This result illustrates how meta-data provided by computerized systems can also bias forensic examiners.

5. How to reduce bias: proposed reforms

As detailed earlier, forensic confirmation biases may be particularly problematic in the forensic sciences—where stimulus ambiguity, context-driven expectations, and motivations conspire to create fertile conditions for psychological contamination and bias to operate. There are two levels at which it is necessary to reduce this bias and its consequences: The first level is in the forensic laboratory, and even at the crime scene, where evidence is collected and sometimes analyzed; the second level is at the trial and appellate courts, where that evidence is evaluated. Hence, we offer a number of suggestions.

5.1. Reducing bias in the crime laboratory

In a study of four crime laboratories, Peterson, Mihajlovic, and Gilliland (1984) discovered that very few reports excluded the

known suspect from the crime scene or from a connection to the victim. It is not clear whether this result indicates that police manage to identify actual perpetrators for suspicion at high levels of accuracy or that forensic examiners have strong and biasing base-rate expectations that lab results will prove incriminating. As a result of numerous DNA exonerations since that time, however, it is clear that the forensic sciences have contributed to wrongful convictions (Hampikian et al., 2011)—especially in cases that featured other flawed evidence, most notably mistaken eyewitness identifications and false confessions that chronologically preceded the forensic errors (Kassin et al., 2012). To minimize the problem, we suggest the following:

- Examiners should work "linear" rather than "circular," thus initially examining the evidence from the crime scene and documenting their findings before making comparisons against a target. This will eliminate the potential influence of the target on how information is processed and the weight assigned to it (Dror, 2009a).

  It is conceivable that forensic examiners sometimes "re-assess" the evidence to fit the target. If the initial assessment is done in isolation of the target, then such potential influences are eliminated. Indeed, the FBI recently revised its Standard Operating Procedures (SOPs) to "include some steps to avoid bias: examiners must complete and document analysis of the latent fingerprint before looking at any known fingerprint" and to "instruct examiners conducting analysis of a latent fingerprint to analyze it for evidence of distortion, determine whether it is 'of value,' and document the data used during analysis" (OIG, 2011, p. 27).

  Initial analysis in isolation lacks the direction guided by the comparison to a target. For example, when examining a latent print from a crime scene, it may be hard to know where to look for minutia—the important characteristics in a print. Having a suspect's print can guide the examiner as to where such characteristics may be found on the latent print. It is therefore suggested that examiners be allowed to revisit the analysis stage but document their inquiry, justify it, and limit it (for example, for features that were inconclusive during the initial analysis). Although this revisit may open the door to some bias, we believe it is important to use reasonable procedures that both balance the need to avoid bias but facilitate examiners in doing their work. The Office of the Inspector General (OIG, 2011) supports this cognitively informed approach in its report: "a solution to bias may be requiring initial analysis of the latent fingerprint in isolation from the known fingerprints, but also permitting, with clear and detailed documentation, some 're-analysis' of the latent print after comparison" (p. 28). A recent Expert Group set up by the National Institute of Standards and Technology made a similar recommendation (NIST, 2012, Recommendation 3.2).

- The simplest way to protect against the biasing effects of contextual variables is to conduct blind testing. Too often, examiners are exposed to extraneous information from various sources that may taint their conclusions. It is important to shield them from this information. There is no reason why examiners should receive information that is *not* relevant to their work and that they do not need. Thus, we recommend, as much as possible, that forensic examiners be isolated from undue influences such as direct contact with the investigating officer, the victims and their families, and other irrelevant information—such as whether the suspect had confessed.

- Blind testing can shield the forensic examiner from a confession, eyewitness identification, and other information about an investigation that is irrelevant to their forensic work. But it does not protect against the simple base-rate assumption that any individual identified as a suspect is the likely perpetrator. In

current forensic practice, examiners often compare a sample of material to that of a target, presumably belonging to the suspect, in an effort to determine if the two samples derive from the same individual. This protocol is structurally identical to the eyewitness "showup" in which a witness is asked to make a memory-based identification decision via exposure to a single individual. Research shows that showups result in more false positive errors when the suspect and comparison are generally similar to one another (Steblay, Dysart, Fulero, & Lindsay, 2003). Modeled after the extensive scientific literature on best way to collect eyewitness identifications (Wells et al., 1998), which forms the basis for a set of best practice guidelines adopted by U.S. Department of Justice (Technical Working Group for Eyewitness Evidence, 1999), we agree with Saks et al. (2003) in proposing, when possible, the use of an *evidence lineup*.

Modeled after the practice of administering a photograph eyewitness lineup, often called a "six pack," we would recommend that a target-blind examiner be presented with six samples—one belonging to the suspect and five plausible fillers (for the importance of having lineup identifications conducted by a blind administrator, see Canter, Hammond, & Youngs, 2012). From that array, he or she would then seek to determine which, if any, constitutes a match to the evidence found at the crime scene or on the victim. In the only test of the effects of an evidence lineup, Miller (1987) presented students trained in human hair identification with hair samples recovered from a crime scene, which they compared against either a singular innocent suspect sample or a "target-absent lineup" of five innocent samples. Results indicated that the use of a lineup produced a significantly lower error rate than the traditional method (3.8% vs. 30.4%, respectively). Given that none of the samples presented was a true match, all of the errors committed were false positives.

- The *verification* of forensic decisions should be a more controlled process in which blind and double-blind procedures are used whenever possible. Such procedures would require that the verifier is not informed of the initial conclusion; if possible, that the verifier does not know who the examiner was; and that the examiner does not select the verifier (a common practice in many laboratories). Cross-laboratory verifications are also advisable to provide an independent means of checking on the propriety of the initial forensic work (Koppl, Kurzban, & Kobilinsky, 2008).

- Technology plays an increasing and effective tool in solving crimes, enabling the speedy examination of large databases. As noted earlier, however, such technology that examines millions of potential suspects can also lead to error because the likelihood of finding incidental close non-matches is increased (Dror & Mnookin, 2010). This technology can also unwittingly provide meta-data, such as a ranking of potential candidates, which can bias expectations and cause examiners to miss matches or make incorrect identifications (Dror et al., 2012).

  To minimize this problem, careful consideration should be given to deploying these technologies. When a list of potential candidates is provided, that list should be reasonable in length and the order of entries should be randomized as a way to keep examiners from developing a strategy that considers candidates according to their ordinal position on the list. This simple safeguard will enable human examiners to evaluate each candidate fully, equally, and without bias.

- Finally, we believe that it would be useful for forensic science education and certification to include training in basic psychology that is relevant to forensic

work—for example, psychology coursework that addresses experimental methods as well as aspects of perception, judgment and decision making, and social influence, all illustrated through the use of forensic case materials.

5.2. Reducing bias in the courts

The forensic confirmation bias spawns three problems. The first is that it can corrupt the conclusions and testimony of forensic examiners. The second problem is that these conclusions, once corrupted, can have grave consequences—influencing other lines of evidence, be it other forensic examiners, eyewitnesses, and even inducing false confessions among the suspects themselves. The third problem is that these biased sources of information are presented to judges, juries, and appeals courts, which heavily rely on forensic science evidence in their decision-making.

To address these problems, we believe it is important that legal decision makers be educated with regard to the procedures by which forensic examiners reached their conclusions and the information that was available to them at that time. In particular, both trial and appellate courts should be trained to ask "What did the examiner know and when did he or she know it?" and probe routinely for the possibility of contamination across items of evidence that are allegedly independent and corroborative. In cases in which a forensic examiner was unduly exposed and possibly biased by extraneous information, such forensic evidence should be subject to a pretrial reliability hearing aimed at determining if the judgment was tainted and should be excluded rather than admitted into evidence.

At the trial level, judges and juries need to know that forensic science conclusions that appear to corroborate a confession or eyewitness identification may, in fact, have been influenced by these previously collected forms of evidence. This problem has relevance at the appellate level as well. In the U.S., appeals courts may determine that flawed evidence (e.g., a coerced confession or suggestive eyewitness identification) was erroneously admitted at trial but that this trial error was "harmless" (the implication of which is to affirm a defendant's conviction) based on an assessment whether that error had contributed to the jury's verdict in light of all of the evidence presented (for a history of the harmless error rule, see Bilaisis, 1983; as applied to confessions, see *Arizona v. Fulminante, 1991*; Kassin, 2012).

This harmless error doctrine—that an erroneously admitted confession can prove harmless if other evidence is sufficient to support conviction—rests on the tacit and often incorrect assumption that the alleged other evidence was independent of the erroneously admitted item, say, a coerced confession. Indeed, according to Garrett (2011), appellate courts that conducted post-conviction reviews of several confessors who were later exonerated had affirmed the convictions by citing the "overwhelming nature of the evidence against them" (p. 1107). In light of classic psychology research on perceptual and cognitive confirmation biases and the more recent studies of psychological contamination of forensic evidence, we now believe that the courts must consider the proposition on a case-by-case basis that the erroneous evidence presented at trial had corrupted the very forensic examinations that were used to make the error appear harmless.

Going forward, therefore, we believe that the research reviewed in this article has far reaching implications not only for how forensic examinations are conducted but for how the evidence, once gathered, is later presented and evaluated in the courts. It is clear that forensic science evidence often involves subjective judgments that may be biased in a variety of ways. Such influences are psychological in nature, and therefore an area ripe for further empirical research. This research will not only enhance forensic work and

the administration of justice but also provide insights and a testing ground for psychological theory.

References

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology*, 39, 1037–1049.

Arizona v. Fulminante, 499 U.S. 279 (1991).

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, 41, 258–290.

Ask, K., Rebelius, A., & Granhag, P. A. (2008). The elasticity of criminal evidence: A moderator of investigator bias. *Applied Cognitive Psychology*, 22, 1245–1259. http://dx.doi.org/10.1002/acp.1432

Balcetis, E., & Dunning, D. (2006). See what you want to see: Motivational influences on visual perception. *Journal of Personality and Social Psychology*, 91, 612–625. http://dx.doi.org/10.1037/0022-3514.91.4.612

Balcetis, E., & Dunning, D. (2010). Wishful seeing: More desired objects are seen as close. *Psychological Science*, 21, 147–152.

Baumann, A. O., Deber, R. B., & Thompson, G. G. (1991). Overconfidence among physicians and nurses: The micro-certainty, macro-uncertainty phenomenon. *Social Science and Medicine*, 32, 167–174.

Bilaisis, V. (1983). Harmless error: Abettor of courtroom misconduct. *Journal of Criminal Law & Criminology*, 74, 457–475.

Boring, E. G. (1930). A new ambiguous figure. *The American Journal of Psychology*, 42, 444–445. http://dx.doi.org/10.2307/1415447

Bressan, P., & Dal Martello, M. F. (2002). 'Talis pater, talis filius': Perceived resemblance and the belief in genetic relatedness. *Psychological Science*, 13, 213–218. http://dx.doi.org/10.1111/1467-9280.00440

Brewer, N., & Wells, G. L. (2011). Eyewitness identification. *Current Directions in Psychological Science*, 20, 24–27.

Bruner, J. S., & Goodman, C. C. (1947). Value and need as organizing factors in perception. *The Journal of Abnormal and Social Psychology*, 42, 33–44. http://dx.doi.org/10.1037/h0058484

Bruner, J. S., & Potter, M. C. (1964). Interference in visual recognition. *Science*, 144, 424–425. http://dx.doi.org/10.1126/science.144.3617.424

Canter, D., Hammond, L., & Youngs, D. (2012). Cognitive bias in line-up identifications: The impact of administrator knowledge. *Science and Justice*, http://dx.doi.org/10.1016/j.scijus.2012.12.001

Charlton, D., Fraser-Mackenzie, P. A. F., & Dror, I. E. (2010). Emotional experiences and motivating factors associated with fingerprint analysis. *Journal of Forensic Sciences*, 55, 385–393. http://dx.doi.org/10.1111/j. 1556-4029.2009.01295.x

Charman, S. D., Gregory, A. H., & Carlucci, M. (2009). Exploring the diagnostic utility of facial composites: Beliefs of guilt can bias perceived similarity between composite and suspect. *Journal of Experimental Psychology: Applied*, 15, 76–90. http://dx.doi.org/10.1037/a0014682

Charman, S. D., & Wells, G. L. (2008). Can eyewitnesses correct for external influences on their lineup identifications? The actual/counterfactual assessment paradigm. *Journal of Experimental Psychology: Applied*, 14, 5–20.

Cole, S. A. (2001). *Suspect identities: A history of fingerprinting and criminal identification*. Cambridge, MA: Harvard University Press.

Darley, J. M., & Fazio, R. H. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist*, 35, 867–881. http://dx.doi.org/10.1037/0003-066X.35.10.867

Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44, 20–33. http://dx.doi.org/10.1037/0022-3514.44.1.20

Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579. (1993).

Dror, I. E. (2009a). How can Francis Bacon help forensic science? The four idols of human biases. *Jurimetrics: The Journal of Law, Science, and Technology*, 50, 93–110.

Dror, I. E. (2009b). On proper research and understanding of the interplay between bias and decision outcomes. *Forensic Science International*, 191, 17–18.

Dror, I. E. (2012). Cognitive bias in forensic science. In *The 2012 yearbook of science & technology*. New York: McGraw-Hill, pp. 43–45.

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. http://dx.doi.org/10.1016/j.forsciint.2010.10.013

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

Dror, I. E., Charlton, D., & Peron, 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 pattern recognition. *Psychonomic Bulletin & Review*, 17, 161–167. http://dx.doi.org/10.3758/PBR.17.2.161

Dror, I. E., & Hampikian, G. (2011). Subjectivity and bias in forensic DNA mixture interpretation. *Science & Justice*, 51, 204–208. http://dx.doi.org/10.1016/j.scijus.2011.08.004

Dror, I. E., & Mnookin, J. (2010). The use of technology in human expert domains: Challenges and risks arising from the use of automated fingerprint identification systems in forensics. *Law, Probability and Risk*, 9(1), 47–67.

Dror, I. E., Peron, A. E., Hind, S.-L., & Charlton, D. (2005). When emotions get the better of us: The effect of contextual top-down processing on matching fingerprints. *Applied Cognitive Psychology*, 19, 799–809. http://dx.doi.org/10.1002/acp.1130

Dror, I. E., & Rosenthal, R. (2008). Meta-analytically quantifying the reliability and biasability of forensic experts. *Journal of Forensic Sciences*, 53, 900–903. http://dx.doi.org/10.1111/j. 1556-4029.2008.00762.x

Dror, I. E., Wertheim, K., Fraser-Mackenzie, P., & Walajtys, J. (2012). The impact of human-technology cooperation and distributed cognition in forensic science: Biasing effects of AFIS contextual information on human experts. *Journal of Forensic Sciences*, 57(2), 343–352.

Edmonds v. Mississippi, 955 So.2d 787 (2007).

Elaad, E., Ginton, A., & Ben-Shakhar, G. (1994). The effects of prior expectations and outcome knowledge on polygraph examiners' decisions. *Journal of Behavioral Decision Making*, 7, 279–292. http://dx.doi.org/10.1002/bdm.3960070405

Findley, K. A., & Scott, M. S. (2006). The multiple dimensions of tunnel vision in criminal cases. *Wisconsin Law Review*, 2, 291–397.

Fisher, G. H. (1968). Ambiguity of form: Old and new. *Attention, Perception, & Psychophysics*, 4, 189–192. http://dx.doi.org/10.3758/BF03210466

Gabbert, F., Memon, A., & Allan, K. (2003). Memory conformity: Can eyewitnesses influence each other's memories for an event? *Applied Cognitive Psychology*, 88, 341–347.

Garrett, B. L. (2011). *Convicting the innocent: Where criminal prosecutions go wrong.* Cambridge, MA: Harvard University Press.

Grann, D. (2009). Trial by Fire: Did Texas execute an innocent man? *The New Yorker*, (September).

Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions: A handbook.* Chichester: Wiley.

Hagan, W. E. (1894). *A treatise on disputed handwriting and the determination of genuine from forged signatures.* New York, NY: Banks & Brothers.

Hall, L. J., & Player, E. (2008). Will the introduction of an emotional context affect fingerprint analysis and decision-making? *Forensic Science International*, 181, 36–39. http://dx.doi.org/10.1016/j.forsciint.2008.08.008

Halverson, A. M., Hallahan, M., Hart, A. J., & Rosenthal, R. (1997). Reducing the biasing effects of judges' nonverbal behavior with simplified jury instruction. *Journal of Applied Psychology*, 82, 590–598. http://dx.doi.org/10.1037/0021-9010.82.4.590

Hamilton, D. L., & Zanna, M. P. (1974). Context effects in impression formation: Changes in connotative meaning. *Journal of Personality and Social Psychology*, 29, 649–654.

Hampikian, G., West, E., & Akselrod, O. (2011). The genetics of innocence: Analysis of 194 U.S. DNA exonerations. *Annual Review of Genomics and Human Genetics*, 12, 97–120.

Hasel, L. E., & Kassin, S. M. (2009). On the presumption of evidentiary independence: Can confessions corrupt eyewitness identifications? *Psychological Science*, 20, 122–126. http://dx.doi.org/10.1111/j. 1467-9280.2008.02262.x

Hill, C., Memon, A., & McGeorge, P. (2008). The role of confirmation bias in suspect interviews: A systematic evaluation. *Legal and Criminological Psychology*, 13, 357–371. http://dx.doi.org/10.1348/135532507X238682

Hochschild, J. L. (2008). How ideas affect actions. In R. E. Goodin, & C. Tilly (Eds.), *The Oxford Handbook of Contextual Political Analysis* (p. 284). Oxford University Press.

Jussim, L. (2012). *Social perception and social reality: Why accuracy dominates bias and self-fulfilling prophecy.* New York: Oxford University Press.

Kassin, S. M. (1997). The psychology of confession evidence. *American Psychologist*, 52, 221–233.

Kassin, S. M. (2012). Why confessions trump innocence. *American Psychologist*, 67, 431–445.

Kassin, S. M., Bogart, D., & Kerner, J. (2012). Confessions that corrupt: Evidence from the DNA exoneration case files. *Psychological Science*, 23, 41–45.

Kassin, S. M., Drizin, S. A., Grisso, T., Gudjonsson, G. H., Leo, R. A., & Redlich, A. D. (2010). Police-induced confessions: Risk factors and recommendations. *Law and Human Behavior*, 34, 3–38. http://dx.doi.org/10.1007/s10979-009-9188-6

Kassin, S. M., Goldstein, C. C., & Savitsky, K. (2003). Behavioral confirmation in the interrogation room: On the dangers of presuming guilt. *Law and Human Behavior*, 27, 187–203. http://dx.doi.org/10.1023/A:1022599230598

Kassin, S. M., & Gudjonsson, G. H. (2004). The psychology of confessions: A review of the literature and issues. *Psychological Science in the Public Interest*, 5, 33–67. http://dx.doi.org/10.1111/j.1529-1006.2004.00016.x

Kassin, S. M., & Kiechel, K. L. (1996). The social psychology of false confessions: Compliance, internalization, and confabulation. *Psychological Science*, 7, 125–128.

Kerstholt, J., Eikelboom, A., Dijkman, T., Stoel, R., Hermsen, R., & van Leuven, B. (2010). Does suggestive information cause a confirmation bias in bullet comparisons? *Forensic Science International*, 198, 138–142. http://dx.doi.org/10.1016/j.forsciint.2010.02.007

Kerstholt, J., Paashuis, R., & Sjerps, M. (2007). Shoe print examinations: Effects of expectation, complexity and experience. *Forensic Science International*, 165, 30–34. http://dx.doi.org/10.1016/j.forsciint.2006.02.039

Kierein, N. M., & Gold, M. A. (2000). Pygmalion in work organizations: A meta-analysis. *Journal of Organizational Behavior*, 21, 913–928.

Klayman, J., & Ha, Y.-W. (1997). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Bulletin*, 94, 211–228. http://dx.doi.org/10.1037/0033-295X.94.2.211

Koppl, R., Kurzban, R., & Kobilinsky, L. (2008). Epistemics for forensics. *Episteme*, 5(2), 141–159.

Kukucka, J., & Kassin, S. M. (2012, April). Do confessions taint juror perceptions of handwriting evidence? *Paper presented at the Annual Meeting of the American Psychology-Law Society* San Juan, Puerto Rico, March 14–17.

Kumho Tire Co. v. Carmichael, 526 U.S. 137. (1999).

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480–498. http://dx.doi.org/10.1037/0033-2909.108.3.480

Lange, N. D., Thomas, R. P., Dana, J., & Dawes, R. M. (2011). Contextual biases in the interpretation of auditory evidence. *Law and Human Behavior*, 35, 178–187. http://dx.doi.org/10.1007/s10979-010-9226-4

Leadbetter, M. (2007). Letter to the Editor. *Fingerprint World*, 33, 231.

Leeper, R. (1935). A study of a neglected portion of the field of learning: The development of sensory organization. *The Pedagogical Seminary and Journal of Genetic Psychology*, 46, 41–75.

Lieberman, J. D., Carrell, C. A., Miethe, T. D., & Krauss, D. A. (2008). Gold versus platinum: Do jurors recognize the superiority and limitations of DNA evidence compared to other types of forensic evidence? *Psychology, Public Policy, & Law*, 14, 27–62. http://dx.doi.org/10.1037/1076-8971.14.1.27

Lynch, M. (2003). God's signature: DNA profiling, the new gold standard in forensic evidence. *Endeavor*, 27, 93–97. http://dx.doi.org/10.1016/S0160-9327(03)00068-1

McNatt, D. B. (2000). Ancient Pygmalion joins contemporary management: A meta-analysis of the result. *Journal of Applied Psychology*, 85, 314–322.

Mezulis, A., Abramson, L., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin*, 130, 711–746.

Miller, L. S. (1984). Bias among forensic document examiners: A need for procedural changes. *Journal of Police Science and Administration*, 12, 407–411.

Miller, L. S. (1987). Procedural bias in forensic science examinations of human hair. *Law and Human Behavior*, 11, 157–163. http://dx.doi.org/10.1007/BF01040448

Mnookin, J. L., Cole, S. A., Dror, I. E., Fisher, B. A., Houck, M., Inman, K., et al. (2011). The need for a research culture in the forensic sciences. *UCLA Law Review*, 58(3), 725–779.

Mower, L., & McMurdo, D. (2011, July). Las Vegas police reveal DNA error put wrong man in prison. *Las Vegas Review Journal*.

Narchet, F. M., Meissner, C. A., & Russano, M. B. (2011). Modeling the influence of investigator bias on the elicitation of true and false confessions. *Law and Human Behavior*, 35, 452–465. http://dx.doi.org/10.1007/s10979-010-9257-x

Nash, R. A., & Wade, K. A. (2009). Innocent but proven guilty: Using false video evidence to elicit false confessions and create false beliefs. *Applied Cognitive Psychology*, 23, 624–637.

National Academy of Sciences. (2009). *Strengthening Forensic Science in the United States: A Path Forward.* Washington, DC: National Academies Press.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2, 175–220. http://dx.doi.org/10.1037/1089-2680.2.2.175

NIST. (2012). Expert Working Group on human factors in latent print analysis. *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach.* U.S. Department of Commerce, National Institute of Standards and Technology. http://www.nist.gov/customcf/get_pdf.cfm?pub_id=910745

O'Brien, B. (2009). Prime suspect: An examination of factors that aggravate and counteract confirmation bias in criminal investigations. *Psychology, Public Policy and Law*, 15, 315–334.

OIG. (2006). A review of the FBI's handling of the Brandon Mayfield case. Office of the Inspector General, Oversight & Review Division, US Department of Justice.

OIG. (2011). *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.* <http://www.justice.gov/oig/special/s1105.pdf>.

Perillo, J. T., & Kassin, S. M. (2011). Inside interrogation: The lie, the bluff, and false confessions. *Law and Human Behavior*, 35, 327–337. http://dx.doi.org/10.1007/s10979-010-9244-2

Peterson, J. L., Mihajlovic, S., & Gilliland, M. (1984). *Forensic evidence and the police.* Washington, DC: National Institute of Justice.

Radel, R., & Clement-Guillotin, C. (2012). Evidence of motivational influences in early visual perception: Hunger modulates conscious access. *Psychological Science*, 23, 232–234.

Risinger, D. M., & Saks, M. J. (1996). Science and nonscience in the courts: Daubert meets handwriting identification expertise. *Iowa Law Review*, 82, 21–74.

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, 1–56.

Rosenthal, R. (2002). Covert communication in classrooms, clinics, courtrooms, and cubicles. *American Psychologist*, 57, 839–849. http://dx.doi.org/10.1037/0003-066X.57.11.839

Rosenthal, R., & Fode, K. (1963). The effect of experimenter bias on performance of the albino rat. *Behavioral Science*, 8, 183–189.

Rosenthal, R., & Jacobson, L. (1966). Teachers' expectancies: Determinants of pupils' IQ gains. *Psychological Reports*, *19*, 115–118. http://dx.doi.org/10.2466/pr0.1966.19.1.115

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

Saks, M. J., Risinger, D. M., Rosenthal, R., & Thompson, W. C. (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.

Simon, D. (2011). The limited diagnosticity of criminal trials. *Vanderbilt Law Review*, *64*, 143–223.

Skagerberg, E. M. (2007). Co-witness feedback in line-ups. *Applied Cognitive Psychology*, *21*, 489–497.

Skitka, L. J., Mullen, E., Griffin, T., Hutchinson, S., & Chamberlin, B. (2002). Dispositions, scripts, or motivated correction? Understanding ideological differences in explanations for social problems. *Journal of Personality and Social Psychology*, *83*, 470–487.

Snyder, M., & Swann, W. B., Jr. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology*, *36*, 1202–1212. http://dx.doi.org/10.1037/0022-3514.36.11.1202

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2003). Eyewitness accuracy rates in police showup and lineup presentations: A meta-analytic comparison. *Law and Human Behavior*, *27*, 523–540. http://dx.doi.org/10.1023/A:1025438223608

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

Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement [booklet].* Washington, DC: United States Department of Justice, Office of Justice Programs. Document Number NCJ-178240.

Thompson-Cannino, J., Cotton, R., & Torneo, E. (2009). *Picking Cotton: Our memoir of injustice and redemption.* New York: St. Martin's Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131. http://dx.doi.org/10.1126/science.185.4157.1124

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 made by latent fingerprint examiners. *PLoS ONE*, *7*, 1–12. http://dx.doi.org/10.1371/journal.pone.0032800

U.S. v. Hines, 55 F. Supp. 2d 62. (1999).

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129–140. http://dx.doi.org/10.1080/17470216008416717

Watkins, M. J., & Peyrincioglu, Z. F. (1984). Determining perceived meaning during impression formation: Another look at the meaning change hypothesis. *Journal of Personality and Social Psychology*, *46*, 1005–1016.

Wells, G. L., Small, M., Penrod, S., Malpass, R., Fulero, S., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law & Human Behavior*, *22*, 603–647.