2-21-2019

# How Cross-Examination on Subjectivity and Bias Affects Jurors' Evaluations of Forensic Science Evidence

William C. Thompson
*University of California, Irvine*

Nicholas Scurich
*University of California, Irvine*

# How Cross-Examination on Subjectivity and Bias Affects Jurors' Evaluations of Forensic Science Evidence


## Abstract

Contextual bias has been widely discussed as a possible problem in forensic science. The trial simulation experiment reported here examined reactions of jurors at a county courthouse to cross-examination and arguments about contextual bias in a hypothetical case. We varied whether the key prosecution witness (a forensic odontologist) was cross-examined about the subjectivity of his interpretations and about his exposure to potentially biasing task-irrelevant information. Jurors found the expert less credible and were less likely to convict when the expert admitted that his interpretation rested on subjective judgment, and when he admitted having been exposed to potentially biasing task-irrelevant contextual information (relative to when these issues were not raised by the lawyers). The findings suggest, however, that forensic scientists can immunize themselves against such challenges and maximize the weight jurors give their evidence by adopting context management procedures that blind them to task-irrelevant information.



## Disciplines

Forensic Science and Technology

## Comments

This is a manuscript of an article published as Thompson, William C., and Nicholas Scurich. "How cross-examination on subjectivity and bias affects jurors' evaluations of forensic science evidence." *Journal of forensic sciences* 64, no. 5 (2019): 1379-1388. Posted with permission of CSAFE.

*William C. Thompson,[1] J.D., Ph.D.; and Nicholas Scurich,[1,2] Ph.D.*

# How Cross-Examination on Subjectivity and Bias Affects Jurors' Evaluations of Forensic Science Evidence[†]

**ABSTRACT:** Contextual bias has been widely discussed as a possible problem in forensic science. The trial simulation experiment reported here examined reactions of jurors at a county courthouse to cross-examination and arguments about contextual bias in a hypothetical case. We varied whether the key prosecution witness (a forensic odontologist) was cross-examined about the subjectivity of his interpretations and about his exposure to potentially biasing task-irrelevant information. Jurors found the expert less credible and were less likely to convict when the expert admitted that his interpretation rested on subjective judgment, and when he admitted having been exposed to potentially biasing task-irrelevant contextual information (relative to when these issues were not raised by the lawyers). The findings suggest, however, that forensic scientists can immunize themselves against such challenges and maximize the weight jurors give their evidence by adopting context management procedures that blind them to task-irrelevant information.

**KEYWORDS:** forensic science, bias, task-relevant, context management, blind, juror decision-making, bite mark evidence

People generally view forensic science evidence as trustworthy (1). They think the chances of a false or misleading finding are extremely low (2). It is debatable, however, whether forensic science is always worthy of that trust. A number of distinguished scientific bodies have criticized forensic scientists for relying on inadequately validated methods, over-stating findings, and making insufficient efforts to avoid bias (3–5).

Lawyers sometimes challenge forensic science evidence in court, but judges rarely rule it inadmissible under the *Daubert* or *Frye* standards (3,6,7). That means that lawyers who wish to attack forensic science generally must do so in front of the jury through cross-examination or presentation of opposing witnesses. As the Supreme Court explained in *Daubert v. Merrell Dow Pharmaceuticals* ([8], p. 576), "vigorous cross-examination, presentation of contrary evidence, and careful instructions on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence."

Although cross-examination has been called "the greatest legal engine ever invented for the discovery of truth (9)," it is not clear how effective cross-examination is for challenging scientific evidence (10). Kovera and her colleagues found that the strength of the cross-examination had no influence on mock jurors' evaluations of expert testimony in child sexual abuse cases (10). Kovera, McAuliff, and Hebert (11) found that cross-examination failed to sensitize mock jurors to problems with the validity of scientific evidence presented in a gender discrimination case. But cross-examination sometimes makes a difference. Austin and Kovera (12) found that a "scientifically informed" cross-examination that focused on the requisites of experimental design sensitized mock jurors to a methodological flaw in a study presented by an eyewitness expert. Lieberman and his colleagues (1) found that an "evidence-based" cross-examination that focused on things that could go wrong in DNA testing reduced the influence of DNA evidence, but only when the DNA testing was done by a "low-reliability" laboratory that lacked accreditation and had never undergone proficiency testing.

The experiment reported here also examined the effect of cross-examination in a case involving forensic science. The cross-examination was designed to highlight two common features of forensic pattern-matching evidence—the expert's reliance on subjective judgment when interpreting data and the expert's exposure to potentially biasing task-irrelevant contextual information. Although there is a growing literature on jurors' reactions to forensic science evidence (13–15), relatively little is known about jurors' reactions when the evidence is challenged by an opposing party (for an exception see [16]). This article examines one aspect of that question, looking at how members of a courthouse jury pool evaluate expert testimony of a forensic bite mark examiner, and how their evaluations are affected by cross-examination about the potential for subjectivity and bias in the expert's analysis.

### *Subjectivity and Bias in Forensic Science*

In 2009, as part of a broad examination of the forensic sciences, the National Academy of Sciences expressed concern that

[1]Department of Criminology, Law, & Society, University of California-Irvine, Irvine, CA.
[2]Department of Psychological Science, University of California-Irvine, 4312 Social and Behavioral Sciences Gateway, Irvine, CA 92697.
Corresponding author: Nicholas Scurich, Ph.D. E-mail: nscurich@uci.edu

[†]Supported by the Center for Statistical Applications in Forensic Evidence (CSAFE), which is funded through Cooperative Agreement #70NANB15H176 with the National Institute of Standards and Technology (NIST).

Received 18 Nov. 2018; and in revised form 15 Jan. 2019; accepted 15 Jan. 2019.

"forensic science experts are vulnerable to cognitive and contextual bias" that "renders experts vulnerable to making erroneous identifications." ([3], p. 4, note 8). This conclusion is supported by a growing body of studies showing that forensic examiners can be influenced by contextual factors that are irrelevant to the task the examiners are supposed to be performing (17). For example, after learning that the main suspect had a solid *alibi*, fingerprint examiners were less likely to identify a latent print found at a crime scene as having come from that particular suspect (18–20). Similar types of contextual bias have been demonstrated in studies of document examination (21,22), bite mark analysis (23), bloodstain pattern analysis (24), forensic anthropology (25), and even DNA analysis (26,27). Contextual bias has been cited as a contributing factor to documented casework errors that have occurred in latent print analysis (28,29), bite mark analysis (30), and DNA testing (31,32).

Contextual information that is irrelevant to the examiner's scientific assessment, but potentially biasing, may reach the examiner in several ways (33,34). Dror and his colleagues (34) present a five-level taxonomy that distinguishes task-irrelevant information that may be conveyed by the trace evidence itself (Level 1), the reference samples (Level 2), the case information (Level 3), the examiners' base-rate expectations (Level 4), and organizational and cultural factors (Level 5). In the experiment reported here, we focused on Level 3—information about the case that the examiner may learn through communication with detectives and other investigators or through access to police reports that provide information about such matters as the identity of suspects, the nature of the evidence against them, witness statements, and police theories (35,36). Commentators have suggested that, without even realizing it, forensic examiners may be influenced by such information, creating an unconscious contextual bias (17,35,36).

Concerns about contextual bias are supported by a large psychological literature showing that human judgment can be influenced by contextual factors, that people can be biased without being aware of it, and that even well-trained experts are susceptible to bias (for reviews, *see* [17,35,37]). This literature suggests, however, that the degree of bias in forensic science is likely to depend on the nature of the forensic examination and the circumstances of a particular case. The influence of context tends to be stronger when the data being interpreted are ambiguous and weaker when the correct interpretation is more obvious (17). Jurors should therefore be less concerned about contextual bias when forensic examiners interpret data using objective standards that produce clear-cut results (e.g., reading data from a computer print-out and interpreting it according to a quantitative standard) than when they interpret data according to more subjective standards that may be vague, idiosyncratic, or flexible.

It is doubtful, however, that jurors understand the degree of subjectivity entailed in forensic pattern-matching methods, such as bite mark examination. Television programs often make it appear that pattern-matching tasks are automated (the computer screen rather dramatically flashes the word "match"), when in reality the conclusion that two items with similar patterns have a common source is a decision reached by a human examiner following a painstaking, subjective analysis. Consequently, we hypothesized that cross-examination that reveals that an examiner relied, in part, on subjective judgment, when interpreting data will reduce the weight jurors give to the examiners' conclusions.

### *Context Management Procedures*

Whether jurors should be concerned about contextual bias may also depend on the nature of the contextual information and how the expert deals with it. Forensic scientists can greatly reduce or even eliminate the potential for contextual bias from Level 3 information by adopting context management procedures that shield them from exposure to contextual information that is irrelevant to their scientific judgment (33,36,38). For example, some laboratories make use of a "case manager" who intervenes between laboratory examiners and criminal investigators (36,39,40). The case manager works with investigators to determine what evidence needs to be examined and then passes evidentiary items along to examiners for examination, testing, and comparison. This division of duties allows the case manager to be fully informed about the background of the case while examiners are given only the information necessary to perform the requested examination or test (36). The examiners can eventually learn the background details about the case, but only after they have recorded their conclusions.

One of the issues addressed in the experiment presented here is whether jurors appreciate the importance of blinding procedures. Do they give more weight to forensic evidence when it is the product of a blind examination than when it is the product of a nonblind (and therefore potentially biased) examination? Previous research has raised doubts about whether people appreciate the importance of using blind procedures in social science research. Two studies have found that mock jurors' evaluations of the quality of a social science experiment were not affected by whether the experimenter was blind to the expected results, even though blinding would have been appropriate to prevent experimenter bias (41,42). It is possible, however, that participants in these studies thought that contextual bias was unlikely (or unimportant) in a social science experiment. People may take the potential for bias more seriously when it could influence the interpretation of forensic science evidence and thereby affect the results of a criminal investigation. Moreover, their appreciation of the importance of blinding may increase if it becomes apparent that the forensic examiner relied heavily on subjective judgment to reach conclusions. Hence, we hypothesized that jurors would give more weight to forensic evidence when the examiner was blind to task-irrelevant contextual information than when the examiner was not blind. We also hypothesized that the effect of blinding would increase when jurors were also exposed to cross-examination that revealed that the forensic examination required subjective judgment.

### *Arguments Against Blinding*

Although forensic scientists show a growing awareness of the problem of contextual bias (43), the need for blinding procedures has been questioned (44–46). Some commentators have argued that blinding is unnecessary because forensic scientists can resist or overcome any contextual influences simply by being committed to objectivity and professionalism. A prominent forensic scientist explained this perspective as follows:

> I reject the insinuation that we do not have the wit or the intellectual capacity to deal with bias, of whatever sort. If we are unable to acknowledge and compensate for bias, we have no business in our profession to begin with, and certainly no legitimate plea to the indulgence of the legal system (47).

This perspective has been criticized as psychologically naïve in its failure to recognize that bias can occur without conscious awareness and can affect even well-meaning and conscientious experts (36,48,49). Research indicates that people generally have a "blind spot" when it comes to recognizing their own biases (50–52) and this appears also to be true of forensic scientists also (43). In the broader scientific community, it is widely accepted that individuals cannot be relied upon to correct their own biases, which is why blinding procedures are used in many scientific fields, particularly fields like clinical medicine where reported conclusions often rest, in part, on researchers' subjective judgments and evaluations (35,53).

A key goal of the experiment reported here was to learn how jurors view a forensic science expert who claims to have resisted or overcome contextual bias. Are jurors skeptical, or do they find such claims credible? To investigate these questions, we included an experimental condition (the "Ignored condition") in which the forensic examiner acknowledged being exposed to task-irrelevant contextual information, but claimed that he had intentionally ignored it and had not allowed it to influence his judgments. We wanted to find out whether jurors would view this expert as less credible than an expert who was blind to the task-irrelevant information.

Forensic scientists who oppose the use of blinding procedures sometimes advance a second argument: they deny that being influenced by contextual information constitutes a bias (36). According to this perspective, it is acceptable and even desirable for forensic examiners to look beyond the physical evidence they have been asked to evaluate in order to draw conclusions. If the examiner is uncertain whether a fingerprint found on a murder weapon is that of a suspect, for example, learning that the suspect confessed to handling the weapon may help the examiner reach the correct conclusion. As one examiner (quoted in [36]) put it: "how can it be a bias if it leads us to the truth."

The problem with this perspective is that it allows forensic scientists to be influenced by matters beyond their scientific expertise. The forensic examiner's conclusions no longer rest solely on interpretation of the physical evidence presented for examination but also are influenced by matters that have nothing to do with forensic science. That creates problems for the legal system because nonscientific matters, such as witness statements, confessions, and the like, are supposed to be evaluated by the trier of fact (i.e., the judge or jury) rather than by experts. Suppose, for example, that a fingerprint examiner is uncertain whether a latent print on a murder weapon is sufficient to support identification of the defendant as the source, but the examiner decides to report an identification after learning the defendant admitted to handling the weapon. To the trier of fact, it then appears that there are two independent pieces of evidence against the defendant—the admission and the fingerprint identification. Because the two pieces of evidence are not actually independent, however, they are not as powerful collectively as they appear to be. The tendency to over-estimate the collective strength of nonindependent evidence items is sometimes called "double-counting" of evidence and is recognized as a source of error in human decision-making (36,54,55). Page and his colleagues (56) explained the importance of this factor for forensic odontology: "if a forensic examiner reaches a conclusion that includes consideration of factors other than the evidence before them, their conclusions should not carry the independent weight that the trier of fact has assumed is inherent in such testimony." This creates a situation that Thompson (36,55) has called "the criminalist's paradox" in which the expert's use of task-irrelevant information (in a misguided effort to make the "right" call) undermines the value and independence of their scientific conclusions, making the forensic science evidence misleading for the trier of fact. By trying too hard to be "right," these experts may, paradoxically, cause the justice system to go wrong (36,55).

Recognizing this problem, the National Commission on Forensic Science issued an emphatic statement calling on forensic scientists to ensure that forensic analysis is based solely upon "task-relevant" information (38). According to the Commission, information is "task-relevant" only if it helps the examiner draw conclusions "*from the physical evidence that has been designated for examination*" and "*through the correct application of an accepted analytic method...*" ([38], p. 3). By this standard, the suspect's admission clearly is not "task-relevant" for a fingerprint examination and hence is not something the examiner should consider. According to the National Commission, contextual information is sometimes task-relevant and sometimes task-irrelevant. The Commission noted, for example, that "a fingerprint examiner may need information about the surfaces from which the prints were lifted in order to assess whether discrepancies between prints could have been caused by curvature or distortion of one of the surfaces." Hence contextual information about the surfaces is "task-relevant." The Commission offered several examples of contextual information that is "task-irrelevant" and hence should not be considered during a forensic examination (e.g., of latent prints), including information about the suspect's criminal history and statements to police, information about the suspect's *alibi*, and information about other evidence implicating the suspect ([38], p. 3-4). For additional commentary on task-relevance, see (55).

Another goal of the experiment reported here is to learn whether jurors recognize that it is problematic for a forensic scientist to rely on "task-irrelevant" contextual information when reaching conclusions. Do they give such conclusions less weight (as they should)? Or do they see nothing improper about examiners relying on such information? To investigate this matter, we included a condition (the "Used condition") in which the forensic examiner acknowledged being exposed to task-irrelevant information but claimed that it was perfectly appropriate for him to rely on that information when drawing conclusions, and stated that he had in fact done so. We wanted to find out whether jurors would view this expert as less or more credible than an expert who was blind to the task-irrelevant information.

### *Overview and Hypotheses*

To summarize, we asked members of a county jury pool to evaluate a hypothetical criminal case in which forensic bite mark evidence was the key evidence linking the defendant to the crime. A forensic odontologist testified that he had compared a bite mark on the victim's skin to the defendant's teeth and had determined "to a scientific certainty" that the bite mark was made by the defendant. A defense lawyer cross-examined the odontologist. The nature of the cross-examination was varied experimentally. Half of participants received a version of the case in which the defense lawyer began the cross-examination with a series of questions designed to highlight the subjectivity of bite mark comparison and the absence of formal standards for source determinations. For the other half of participants, this line of cross-examination was omitted. A second variable, which cut across the first, was the examiners' reaction to questions about his exposure to task-irrelevant information. The examiner either denied that he had been exposed to task-irrelevant information (Blind condition); or he admitted that he had been exposed to such information but claimed he had ignored it (Ignored condition); or admitted that he had been exposed to the information and acknowledged that he had been influenced by it (Used condition); in a fourth condition (Control condition), the defense lawyer did not inquire about whether the examiner had been exposed to (or had relied upon) any task-irrelevant information.

Based on our hypotheses, we expected a main effect of blinding. Specifically, we expected jurors to view the forensic science evidence as more credible and give it more weight, in the "Blind" condition than in the "Ignored," "Used," and "Control" conditions. We also expected a main effect the subjectivity variable—specifically, that jurors would find the evidence less credible and give it less weight if they heard the cross-examination about subjectivity than if they did not. Finally, we expected to find an interaction between the two independent variables. We expected the effect of blinding to be larger when the jurors heard the cross-examination about subjectivity than when they did not. Correspondingly, we expected the cross-examination about subjectivity to be less influential in the Blind condition than in the Ignored, Used, and Control conditions.

## Method

### *Participants and Procedure*

We recruited participants in the jury assembly room of a county courthouse in a suburban county in the Southwest United States. As jurors were released from service, the courthouse staff informed them about this study. The staff members explained that the study was being conducted by university researchers, participants would be paid $10, and participation was entirely voluntary. Interested individuals received a brief study information sheet, and those who consented to participate were then provided experimental booklets to read and complete. After they completed the materials, participants were debriefed, thanked for their participation, and paid $10. All procedures were approved by both the university Institutional Review Board and the county Jury Commissioner.

Two hundred twenty-five individuals participated in the study, of which 54% were female. The median age was 42 (IQR = 28). A majority of the sample (56%) had at least graduated from a 4-year college or university, and the median annual family income was $60,000–80,000. With regard to political beliefs, 27% identified as liberal, 36% as middle of the road, and 37% as conservative.

Participants read a synopsis of a criminal case in which a bar owner was being extorted by a local gang for "protection money." After the owner refused to pay, an individual entered the bar and discharged a firearm. The owner wrestled the assailant to the ground, at which time the assailant bit the owner and ran off. No DNA was recovered. Eyewitnesses did not get a good look at the assailant because he was wearing a bandana over his face. Police later located a suspect—a known affiliate of the gang—who agreed to provide a dental impression for comparison to the bite mark. An expert in forensic odontology (analysis of bite marks) compared a plastic model of the suspect's teeth (prepared from the dental impression) to the bite mark and concluded "to a scientific certainty" that the bite mark was made by the suspect.

The experimental materials included a summary of the case, a transcript of both the direct and cross-examination of the odontologist, and a summary of closing arguments by the prosecution and defense.

### *Experimental Design*

The nature of the cross-examination varied in a $2 \times 4$ between-participants factorial design. Participants were randomly assigned to one of eight possible cells. The first variable was whether the cross-examination included questions about subjectivity and lack of standards. For half of the participants (Subjectivity-Cross condition), the cross-examination included the following ten questions and answers; for the other half of participants (No Subjectivity-Cross condition), these questions and answers were omitted:

Q: You decided that Stanley Wilson's teeth are consistent with the bite mark?
A: Yes.
Q: But that required some interpretation, didn't it?
A: Well, yes, there is always an element of interpretation.
Q: It wasn't a perfect match, was it?
A: There were eight clear points of comparison on which there was consistency.
Q: But there were differences between the measurements you made on the bite mark and the measurements you made on the model of Mr. Wilson's teeth?
A: There are always some minor differences when you compare a model to a bite mark, but in my judgment the differences were not meaningful differences.
Q: You measured something you call the inter-canine distance on the bite mark?
A: Yes, that is the distance in millimeters between the marks made by the upper canine teeth.
Q: But when you measured the inter-canine distance on Mr. Wilson's teeth it was four millimeters shorter?
A: Yes, but in my judgment that could have been caused by stretching of Mr. Johnson's skin during the attack, so it is still consistent.
Q: And you measured the angle of the tooth impressions made by the incisors?
A: Yes.
Q: And again, there were differences?
A: As I said, one must make allowances for the elasticity of the skin. And some of the differences in measurement were due, I believe, to distortions caused by biting through the fabric of the bandana. But a number of the measurements were identical.
Q: There aren't any formal rules in your field about what constitutes a match? You just rely on your own personal judgment?
A: There is no formula for what constitutes a match because each case is unique. Like all forensic odontologists, I rely on my knowledge, training and experience in the field to make the right judgment.
Q: So it's a match because you say it's a match?
A: Yes. I'm a qualified expert in this field.

The second variable was the examiner's reaction to the defense lawyer's questions about exposure to potentially biasing task-irrelevant information. The examiner admitted that the case file maintained by the crime laboratory contained a phone log which included a handwritten notation saying:

> Suspect known Crips gang member. Keeps skating on charges. Never serves time. This time his gun jammed while trying to shake down a bar. Bit the bar owner before getting away. Murphy [the lead detective] wants to connect him to the bite mark.

The negative information about the defendant that is referenced in the notation had already been presented to the jury as part of the case synopsis, so the jurors already knew about it. The purpose of this cross-examination was to demonstrate that personnel in the crime laboratory also knew about it *before the bite mark examination was conducted*. In other words, the cross-examination demonstrated that crime laboratory personnel had been exposed to potentially biasing task-irrelevant contextual information.

What varied was the examiner's testimony about his exposure to and use of this information. In the *Blind condition,* the examiner testified that the annotation had been made by an evidence control manager and that, because the laboratory uses blinding procedures, he had not seen the annotation (and knew nothing about the underling details of the case) until after he had completed his analysis and reached his conclusions. In the *Ignored condition*, the examiner testified that he had made the annotation when speaking to the lead detective and hence that he was aware of the information contained in the annotation before performing the bite mark comparison, but he claimed that he had paid no attention to this information and that he had not allowed it to influence him in any way. In the *Used condition*, the examiner also testified that he had made the annotation when speaking to the lead detective and hence that he was aware of the information contained in the annotation before performing the bite mark comparison, but he admitted that he had taken this information into account when reaching his conclusions and offered the following explanation for why it was proper for him to have done so:

> It is important for a forensic scientist to take into account all the evidence when analyzing a case. You can't do forensic analysis in isolation. You have to consider the big picture. If you don't have all the facts you might miss something important.... In order to get to the truth, it is important to consider everything.

Finally, in the *Control condition* the defense lawyer did not bring up the annotation in the laboratory notes and there was no discussion of whether the examiner had been exposed to task-irrelevant information.

In closing arguments, the prosecutor called the bite mark testimony the key evidence and said:

> Dr. Krauss made careful measurements and found that the bite mark was completely consistent with Stanley Wilson's teeth. Dr. Krauss explained that bite marks are unique, like fingerprints. The bite mark on Mr. Johnson's arm was so distinctive that Dr. Krauss could determine to a scientific certainty that it was made by Stanley Wilson.

The defense lawyer asserted that the bite mark evidence cannot be trusted because bite mark comparison "is not an exact science." In the Subjectivity-Cross condition, the defense lawyer went on to say: "Dr. Krauss admitted that there are no standards for what is a match and what isn't a match. A match is whatever he decides to call a match. It is all a matter of interpretation..." (This argument was not included in the No Subjectivity-Cross Condition). In the Blind, Ignored, and Used conditions, the defense lawyer said:

> Detective Murphy...gave the lab a lot of irrelevant information about my client that had nothing to do with the bite mark. He basically told them that my client is guilty before they even did the comparison.

In the Ignored and Used Conditions, he went on to say: "Dr. Krauss admitted that he talked to Detective Murphy and heard all about his theories of the case before doing his analysis. He knew exactly what he was supposed to find."

In his rebuttal, the prosecutor responded to these arguments by saying:

> Detective Murphy didn't do anything wrong. He just told the lab about the facts of the case, which is exactly what he was supposed to do. How else would the lab know how to test the evidence? They need to know the facts of the case.

In the blind condition, the prosecutor reminded the jury that Dr. Krauss had been blind: "...he didn't even know whose teeth he was comparing to the bite mark. He didn't talk to Detective Murphy until after he had reached his conclusions. So how could Dr. Krauss have been biased?" In the Ignored condition, the prosecutor said: "Dr. Krauss testified that he paid no attention to what Detective Murphy said when doing his analysis. As a trained forensic scientist he knows how to maintain objectivity." In the Used Condition, he said:

> Dr. Krauss is a forensic scientist. He needs to consider all of the evidence in order to reach a valid opinion. Dr. Krause did what every good forensic scientist would do—he considered all the facts of the case. As a trained forensic scientist he knows how to find the truth.

In the Control Condition, the defense lawyer never raised the issue of the expert's exposure to task-irrelevant information so the prosecutor did not present these rebuttal arguments.

### *Dependent Measures*

After reading the experimental materials, participants were first asked to state the verdict they would return if instructed by the judge to find the defendant guilty only if the evidence convinced them beyond a reasonable doubt that the defendant was the perpetrator. Participants then were given seven questions that probed the credibility of forensic odontology and the expert himself. For example, "Do you think testimony about forensic odontology (bite mark comparison) is a trustworthy type of scientific evidence?" or "Did you find Dr. Krauss' testimony to be credible?" These items were rated on a 7-point Likert scale with higher values reflecting higher credibility. A reliability analysis revealed high inter-item correlations for four of the items (in addition to the two aforementioned items, the other two items were: "Do you think Dr. Krauss followed proper scientific procedures?" and "Do you think the bite mark analysis in this case was subjective [based on personal interpretation] or objective [based on fixed scientific standards]?"); the other three items were either uncorrelated or negatively correlated. A composite measure of the four items was created (yielding a Cronbach's $\alpha = 0.88$) and z-scored to facilitate interpretation. This measure is called "scientific credibility" hereafter. Participants finally provided demographic information.

## Results

### *Scientific Credibility*

A two-way analysis of variance (ANOVA) with the experimental conditions as the independent variables and participants' judgments of scientific credibility as the dependent variable detected a main effect of the cross-examination regarding exposure to task-irrelevant information ($F(3,196) = 5.29$, $p < 0.001$, $\eta_p^2 = 0.084$), a main effect of the cross-examination regarding subjectivity and lack of standards ($F(1,196) = 10.72$, $p < 0.001$, $d = 0.496$), and a significant two-way interaction ($F(3,196) = 2.80$, $p < 0.001$, $\eta_p^2 = 0.046$). Figure 1 depicts the mean scientific credibility ratings decomposed by experimental condition. Note that we report Bootstrapped (1000 samples) means and confidence intervals.

Bonferroni contrasts revealed that the Blind condition was deemed significantly more credible than the Ignored condition (mean difference = 0.57, Bootstrap 95% Confidence Interval (CI)[0.204, 0.945], $p = 0.02$), the Used condition (mean difference = 0.744, Bootstrap 95% CI[0.335, 1.12], $p < 0.001$), and the Control condition (mean difference = 0.58, Bootstrap 95% CI [0.190, 0.944], $p = 0.02$). Additionally, participants in the Control condition found the expert significantly less credible when they heard the cross-examination about the subjectivity of expert's method than when they did not hear this cross-examination (95% CI for No and Yes = [0.35, 0.794] and [−1.12, −0.358] respectively). None of the other contrasts were statistically different.

To summarize briefly, the expert was viewed as more credible when the expert was blind to the task-irrelevant contextual information than when the expert knew about task-irrelevant information and claimed to have ignored it or used it. Raising the issue of subjectivity on cross-examination only had a significant effect in the Control condition, in which the issue of task-irrelevant information was not mentioned. In that instance, scientific credibility was significantly reduced when the expert admitted that his analysis required making subjective judgments in the absence of formal standards. Although it appears that the expert's scientific credibility is reduced if the expert admits either to exposure to task-irrelevant information or to using a subjective method, we did not detect a significant additive effect on scientific credibility ratings when the expert admitted to both.

*Verdicts*

Overall, fifty percent of participants voted to convict the defendant at the conclusion of trial. Figure 2 depicts the proportion of participants voting to convict in each experimental condition.

A binary logistic regression with the experimental conditions as the independent variables and participants' verdict as the dependent variable was conducted to determine whether guilty verdicts varied across the experimental conditions. The model was significant ($\chi^2 = 23.34$, df = 7, $p < 0.001$), as was the main effect of the cross-examination on subjectivity (Wald = 11.77, df = 1, $p < 0.001$). Participants who did not hear the cross-examination about subjectivity were over nine times more likely to convict the defendant than those who heard it (Exp($B$) = 9.05, 95%CI = 2.57, 31.84). The main effect of the cross-examination about exposure to task-irrelevant information was also significant (Wald = 11.70, df = 3, $p = 0.01$); participants in the Blind condition were over six times (Exp($B$) = 6.01, 95%CI = 1.84, 20.32) more likely to convict than participants in the Control condition. Participants in the Ignored and the Used conditions were not significantly more likely to vote to convict than participants in the control conditions (both ps > 0.05).

These main effects were qualified, however, by a significant interaction (Wald = 8.44, df = 3, $p = 0.04$). As can be seen in Fig. 2, the cross-examination regarding subjectivity had a stronger effect on verdicts in the Control condition than in the Blind or Used conditions (Wald = 6.58, $p = 0.01$ and Wald = 5.84, $p = 0.016$, respectively); however, the effect of the cross-examination regarding subjectivity in the Ignored condition and the Control condition did not significantly differ (Wald = 1.61, $p = 0.204$). In other words, the cross-examination about subjectivity significantly reduced the likelihood of a guilty verdict in the Control and Ignored conditions but did not significantly affect verdicts in the Blind or Used conditions.

Another binary logistic regression was conducted, identical to the previous one, except that it included participants' judgments of scientific credibility. The model was significant ($\chi^2 = 137.19$, df = 8, $p < 0.001$). After taking judgments of credibility into account, there were no main effects of the cross-examination about exposure to task-irrelevant information or the cross-examination about subjectivity, nor was the interaction significant (all ps > 0.05). The scientific credibility ratings did, however, predict verdicts; for each unit increase in credibility, the odds of a guilty verdict increased by 14.8 (95% CI [6.88, 31.91], df = 1, $p < 0.001$). In sum, it appears that credibility is the key variable driving verdicts. Our experimental variables affected the perceived credibility of the scientific expert, and the perceived credibility of the expert is what affected the verdicts. There were

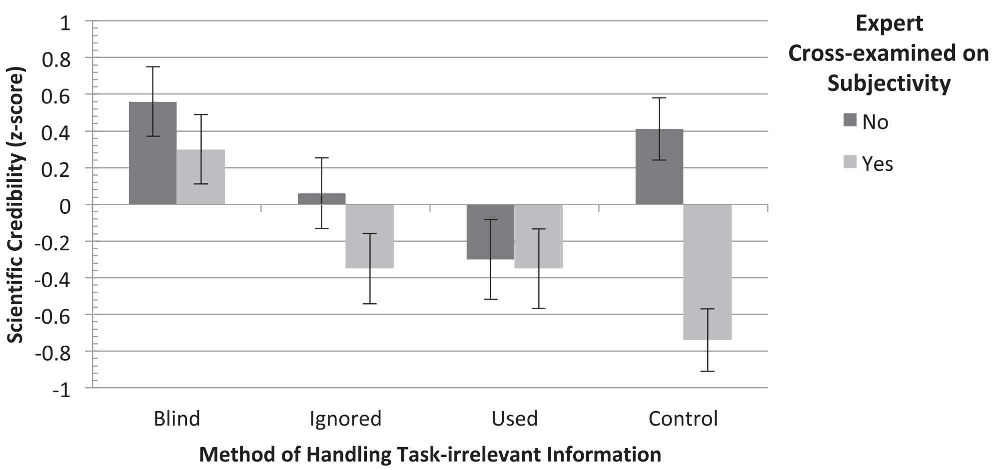


FIG. 1—*Ratings of "scientific credibility" (z-scored) decomposed by whether or not the expert was cross-examined about the subjectivity of bite mark comparison and how he handled the task-irrelevant information. Bootstrapped (based on 1000 samples) error bars reflect ±1SE.*

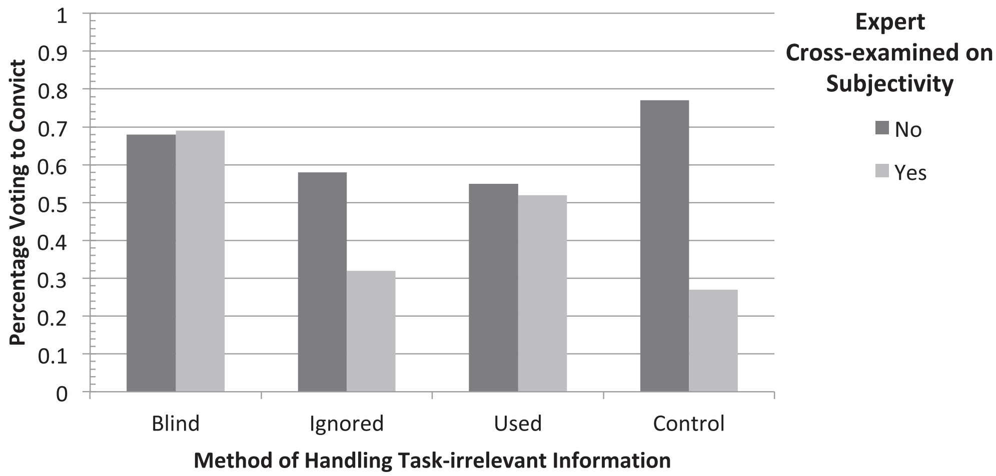


FIG. 2—*Percentage of guilty verdicts decomposed by whether or not the expert was cross-examined about the subjectivity of bite mark comparison and how he handled the task-irrelevant information.*

more guilty verdicts in conditions where the expert was more credible.

## Discussion

This experiment was designed to cast light on jurors' evaluations of forensic science evidence when that evidence is challenged through cross-examination. Specifically, it examined the effects of two lines of cross-examination on jurors' perceptions of the credibility of a forensic examiner (a forensic odontologist), and their willingness to convict in a hypothetical case that rested heavily on forensic evidence. One line of cross-examination concerned the examiner's reliance on subjective judgment when interpreting data. The second line of cross-examination concerned the examiner's exposure to task-irrelevant contextual information.

The cross-examination about subjectivity consisted of ten questions. The defense lawyer managed to get the expert to acknowledge two key points: first, that there were inconsistencies between some of the measurements made on the bite mark and the corresponding measurements made on the model (which the examiner attributed to elasticity of the victim's skin); and second that there are no formal rules regarding the degree of similarity or dissimilarity needed to reach a conclusion. We believe this line of examination made it clear to jurors that the examiner's conclusion rested on a subjective interpretation (one made without the use of formal standards for interpretation).

The second line of cross-examination concerned the examiner's exposure to information in a laboratory phone log concerning the defendant's previous criminal history and gang affiliations. It is similar to information from an actual case file discussed by Risinger and his colleagues (35) as an example of potentially biasing contextual information; it is also an example of what Dror and his colleagues (34) call Level 3 contextual information. The jury already knew about this information from other sources (it was part of the case summary that they had read), and it was all information that would typically be admissible in a criminal trial because it was relevant to the defendant's motives. The point of the cross-examination was to explore whether the examiner had been exposed to this information before performing the bite mark comparison. In the blind condition, the examiner stated that the information was known to a case manager but had not been passed on to him—hence, he was blind to it. In the Ignored condition, the examiner admitted being exposed to the information but claimed he had ignored it. In the Used condition, the examiner admitted both that he had been exposed to the information and that he had considered it in reaching a conclusion. Finally, in the control condition, this entire line of cross-examination was omitted and there was no discussion of whether the examiner had been exposed to contextual information about the case.

Jurors perceived the examiner to be less credible when he admitted having been exposed to the task-irrelevant information (Ignored and Used conditions) than when he employed context management procedures to shield himself from exposure (Blind condition). This finding indicates that jurors have at least some appreciation of the danger of contextual bias and give less weight to forensic evidence that may have been influenced by contextual bias. It suggests that the use of context management procedures (blinding) may enhance the credibility of a forensic expert in the eyes of a jury and lead the jurors to give more weight to the forensic evidence.

Jurors also found the bite mark examiner less credible when he admitted during cross-examination that his method required subjective interpretation, relative to the conditions in which the cross-examiner did not raise this issue. The strength of this effect varied, however, depending on what jurors' heard about the examiner's exposure to task-irrelevant information. The cross-examination about the subjectivity had the strongest effect in the control condition, where there was no mention of the examiner's exposure to task-irrelevant information. Just by asking a few questions designed to reveal the examiner's reliance on subjective judgment, and the absence of an objective standard for interpretation, the defense lawyer managed to significantly reduce the examiner's perceived credibility and dramatically reduce the conviction rate. This finding lends considerable support to the theory that jurors fail to appreciate the level of subjectivity entailed in forensic pattern-matching tasks (unless it is pointed out to them) and that this factor can be important to their evaluation of an examiner's credibility.

But the cross-examination about subjectivity did not significantly affect the examiner's credibility when the examiner was blind to task-irrelevant information (Blind Condition). Perhaps jurors' underlying concern is not about subjectivity *per se* but about the potential for bias that may arise when experts use subjective methods. Hence, jurors were not so concerned about subjectivity when they learned that the expert had taken other measures (blinding) to prevent bias.

Cross-examination about subjectivity also failed to reduce credibility when the examiner claimed that it was proper for him to rely upon task-irrelevant information (Used Condition). This finding fits the theory that the cross-examination about subjectivity invokes concerns about bias. By denying that the use of task-irrelevant information constitutes a bias, the examiner may have mitigated those concerns. Of course, this examiner (Used Condition) was still seen as less credible than the expert in the Blind condition, which suggests that jurors were at least a bit skeptical of the expert having been exposed to task-irrelevant information.

The cross-examination about subjectivity had a greater effect on the jurors when the expert claimed he had ignored the task-irrelevant information (Ignored Condition). The effect of this cross-examination on perceived credibility did not reach statistical significance, but it significantly reduced the conviction rate. Overall, the results indicate that jurors were skeptical of the expert's claim that he had ignored the task-irrelevant information, and this skepticism increased when the expert also admitted that his interpretation of the findings required subjective judgment in the absence of objective standards for interpretation. The examiner was less credible when he claimed he could avoid bias simply by ignoring potentially biasing information than when he had used context management procedures that prevent exposure to such information. As already noted, these findings indicate that jurors have at least some appreciation of the dangers of unintended contextual bias.

From a legal perspective, the finding suggests that lawyers can successfully challenge the credibility of a nonblind forensic expert in two ways: either by revealing the subjectivity of the expert's methods or by revealing the expert's exposure to task-irrelevant information. When the defense lawyer in our case pursued *neither* line of attack (in the control condition where there was no cross about subjectivity), jurors' perceptions of the expert's credibility were as high as in the blind condition. When the defense lawyer pursued *either* line of attack, the expert's perceived credibility was significantly lower. But this was true only for nonblind forensic experts. The expert was immune to both attacks in the Blind Condition, where the laboratory employed context management procedures to prevent exposure to task-irrelevant information.

The findings on expert credibility are important because our regression analyses suggest that jurors' verdicts largely followed from their credibility ratings. By raising concerns about exposure to potentially biasing task-irrelevant information, and concerns about subjectivity and lack of standards, attorneys can undermine the credibility of a forensic expert. And when the forensic expert is less credible, jurors are less likely to be persuaded by the expert's testimony when reaching a verdict. The connection between expert credibility and verdicts can be seen most easily in a case like the one used in our experiment where the forensic evidence is pivotal, but will likely be a factor in every case involving forensic evidence.

Our findings support the theory that cross-examination can sensitize jurors to the quality of scientific evidence (12). It is particularly noteworthy that jurors appreciated the importance of context management (blinding) procedures. Rather than producing general skepticism about the forensic science evidence, the cross-examination reduced the perceived credibility of an expert who had been exposed to potentially biasing task-irrelevant information, but not an expert was blind to it. With regard to blinding, our findings paint a more positive view of jurors' scientific sophistication than earlier work (41,42), which found mock jurors who evaluated expert testimony about a social science study were insensitive to whether the experimenter who conducted the study had been blind to the expected results. We suspect that lay people might find it easier to appreciate the biasing potential of information about a defendant's past criminal history than information about an experimental hypothesis, and that may be why jurors appreciated the importance of blinding in the present study. Whether this interpretation is correct is a worthy topic for future research.

Criminal defense lawyers may sometime face difficult choices about whether to cross-examine a forensic scientist about exposure to task-irrelevant information. Cross-examination of this type may reinforce in jurors' minds the existence or implications of damaging evidence about the defendant. In some cases, it might even require informing jurors of damaging information about the defendant that they would not otherwise hear. Suppose, for example, that the examiner was told that the defendant had confessed to the crime, but the confession was ruled inadmissible because it was involuntary. In order to inform the jurors of the examiner's exposure to the potentially biasing information about the confession, the defense lawyer would also need to inform the jury of the confession. We suspect that such a strategy would do more to undermine than to help the defendant's case.

Jurors had already been told about the task-irrelevant information about the defendant that was discussed during cross-examination in our experiment. Nevertheless, by questioning the examiner about his exposure to this information the defense lawyer might have reminded the jurors about the defendant's gang affiliation and reinforced that fact in their minds, which might have offset gains that the defendant accrued as a result of this line of examination. It is worth noting that the expert's credibility (and the conviction rate) was lowest when the defense lawyer cross-examined the expert about the subjectivity of his method but did not cross-examine him about exposure to task-irrelevant information (Control Condition with Subjectivity Cross). Future research might usefully examine circumstances under which cross-examination about exposure to task-irrelevant information is effective, and whether it sometimes does more to undermine than help the defendant's case.

### *Limitations and Directions for Future Research*

We have assumed that the cross-examination in this study produced its effects by conveying information—by revealing to jurors that the expert's conclusion rested on subjective judgment in the absence of formal standards for interpretation and by telling jurors whether the examiner had been exposed to potentially biasing task-irrelevant information. It should be noted, however, that our experimental manipulations confounded the information conveyed by the cross-examination with other factors, such as the length of the cross-examination and the number of questions asked. It should also be noted that our experimental manipulation included differences in the lawyers' closing arguments, although the closing arguments simply repeated and emphasized information conveyed by the cross-examination. The varying effects of the cross-examination across conditions support the notion that the manipulations were operating through the information conveyed, though it is also possible that other mechanisms were involved. For example, the cross-examination might have made the defense attorney look more masterful, or made the forensic expert look defensive or weak. Future studies on cross-examination might attempt to disentangle the effects of the style, length, and content of lawyers' questions from the content of the expert's responses.

Some of the terminology used by the expert and by the lawyers in this study is less than ideal from a scientific perspective, but was nevertheless included in the study to create a realistic simulation of what jurors might hear in court. For example, the expert testified (and the prosecutor emphasized) that he had concluded "to a reasonable scientific certainty" that the defendant was the source of the bite mark. The National Commission on Forensic Science (57) has urged lawyers and forensic scientists not to use the expression "reasonable scientific certainty" on grounds that the phrase has no established scientific meaning and is likely to be misunderstood in a legal context. The phrase was still widely used at the time we conducted this experiment, however, so we decided to include it for the sake of realism.

Another example of problematic terminology was the defense lawyer's use of the term "match." The difficulties associated with this term have been widely noted (58) and were illustrated here, as the lawyer sometimes used "match" to refer to the degree of consistency between the bite mark and a model of defendant's teeth ("It wasn't a perfect match was it?") and sometimes to mean a source identification ("So, it's a match because you say it's a match."). We recognize this use of terminology is problematic but believe, unfortunately, that it is a fair representation of what jurors often hear in court, and hence was reasonable for this study. Additional research on the way jurors react to such terminology would, of course, be helpful.

Future research could also usefully test the generality of these findings across different domains or forensic science. For example, it would be useful to know whether the kinds of cross-examination used against the bite mark expert in this case would have similar effects in other pattern-matching disciplines (e.g., latent print examination; tool mark analysis; footwear analysis), and when used more broadly to challenge forensic science evidence. We collected the data reported here before the publication of some recent journalistic exposes challenging the validity of bite mark evidence (59). Nevertheless, it is possible that our jurors were more skeptical of bite mark evidence than they would have been of other types of forensic evidence and this could have affected their reactions to the cross-examination.

Future research should also examine jurors' views regarding other forms of contextual bias. This study indicates that jurors appreciate the importance of using a case manager system to reduce Level 3 contextual bias arising from exposure to task-irrelevant case information. It would be useful to know, for example, whether jurors also appreciate the importance of using procedures such as sequential unmasking (60) or linear sequential unmasking (LSU) (34) to reduce the risk of Level 2 contextual bias, that may arise when an examiner is exposed to information about references before interpreting evidentiary specimens. Our understanding of such issues is extremely limited at this time.

Future research might also use richer stimulus materials, such as presenting jurors with videos of testimony rather than transcripts (*see* [61]). Replication with different cases, different stimuli, different experts (62), and a more ecologically valid methodology (e.g., deliberation) is important and should be conducted before strong policy pronouncements are made.

With those caveats, the policy implications of this study seem straightforward. Lawyers who wish to challenge forensic science evidence should investigate the degree of subjectivity entailed in the underlying methods and whether there was an objective standard for reaching the reported conclusion. They should also investigate whether the examiner was exposed to potentially biasing task-irrelevant information. Lawyers can sometimes find evidence of such exposure in examining laboratory records and bench notes (31,32,35,36). If examiners working in a crime laboratory communicate directly with criminal investigators, without any intervention by laboratory managers, then it is very likely that the examiners will be aware of task-irrelevant information about the case. Our findings suggest that jurors will view the examiners as less credible if the opposing lawyer can show through cross-examination either that the expert's interpretation relied on subjective judgment rather than objective standards or that the expert was exposed to potentially biasing task-irrelevant information.

Fortunately, it appears that forensic scientists can immunize themselves against such challenges by employing context management procedures that shield them from exposure to task-irrelevant information. Academic commentators have long suggested that forensic scientists make greater use of blinding procedures in order to reduce the potential for contextual bias (3–5,35,36,55). They believe context management makes forensic science more rigorous. Our findings suggest an additional reason for adopting such procedures—these procedures also make the testimony of forensic science experts more credible in the eyes of jurors.

*Acknowledgments*

The authors thank the Jury Commission and staff of the Superior Court of Orange County, California, for their assistance and cooperation in recruitment of participants for this research.

## References

1. Lieberman JD, Carrell CA, Miethe TD, Krauss DA. Gold versus platinum: do jurors recognize the superiority and limitations of DNA evidence compared to other types of forensic evidence. Psychol Public Policy Law 2008;14(1):27–62.
2. Koehler JJ. Intuitive error rate estimates for the forensic sciences. Jurimetrics 2017;57:153–68.
3. National Academy of Sciences. Strengthening forensic science in the United States: a path forward. Washington, DC: The National Academies Press, 2009.
4. President's Council of Advisors on Science and Technology (PCAST). Forensic science in criminal courts: ensuring scientific validity of feature-comparison methods. Washington, DC: Executive Office of the President, 2016; https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf (accessed January 15, 2019).
5. Thompson W, Black J, Jain A, Kadane J. Forensic science assessments: a quality and gap analysis – latent fingerprint examination. Washington, DC: American Association for the Advancement of Science, 2017.
6. Risinger DM. Navigating expert reliability: are criminal standards of certainty being left on the dock? Albany Law Rev 2000;64:99–152.
7. Thompson WC. The National Research Council's plan to strengthen forensic science: does the path forward run through the courts? Jurimetrics 2009;50:35–51.
8. Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).
9. Wigmore J. Evidence (J. Chadbourn rev., 1974) § 1367, p. 32.
10. Kovera MB, Levy RJ, Borgida E, Penrod SD. Expert testimony in child sexual abuse cases. Law Hum Behav 1994;18(6):653–74.
11. Kovera MB, McMuliff BD, Hebert KS. Reasoning about scientific evidence: effects of juror gender and evidence quality on juror decisions in a hostile work environment case. J Appl Psychol 1999;84(3):362–75.
12. Austin J, Kovera MB. Cross-examination educates jurors about missing control groups in scientific evidence. Psychol Public Policy Law 2015;21(3):252–64.
13. Thompson WC. How should forensic scientists present source conclusions? Seton Hall Law Rev 2018;48:773–813.

14. Jackson G, Kaye DH, Neumann C, Ranadive A, Reyna VF. Communicating the results of forensic science examinations: final technical report for NIST Award 70NANB12H014 (Penn State Law Research Paper No. 22-2015, 2015); https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2690899 (accessed January 19, 2019).
15. McQuiston-Surrett D, Saks MJ. Communicating opinion evidence in the forensic identification sciences: accuracy and impact. Hastings Law J 2008;89:1159–83.
16. Scurich N. The effect of numeracy and anecdotes on the perceived fallibility of forensic science. Psychia, Psychol, Law 2015;22(4):616–23.
17. Kassin SM, Dror IE, Kukucka J. The forensic confirmation bias: problems, perspectives, and proposed solutions. J Appl Res Mem Cogn 2013;2(1):42–52.
18. Dror IE, Charlton D. Why experts make errors. J Forensic Identif 2006;56:600–16.
19. Dror IE, Charlton D, Peron A. Contextual information renders experts vulnerable to making erroneous identifications. Forensic Sci Int 2006;156:174–8.
20. Dror IE, Rosenthal R. Meta-analytically quantifying the reliability and biasability of forensic experts. J Forensic Sci 2008;53(4):900–3.
21. Miller LS. Bias among forensic document examiners: a need for procedural change. J Police Sci Admin 1984;12(4):407–11.
22. Stoel RD, Dror IE, Miller LS. Bias among forensic document examiners: still a need for procedural changes. Aust J Forensic Sci 2014;46(1):91–7.
23. Osborne NK, Woods S, Kieser J, Zajac R. Does contextual information bias bitemark comparisons? Sci Justice 2014;54(4):267–73.
24. Taylor MC, Laber TL, Kish PE, Owens G, Osborne NK. The reliability of pattern classification in bloodstain pattern analysis, part 1: bloodstain patterns on rigid, non-absorbent surfaces. J Forensic Sci 2016;61(4):922–7.
25. Nakhaeizadeh S, Dror IE, Morgan RM. Cognitive bias in forensic anthropology: visual assessment of skeletal remains is susceptible to confirmation bias. Sci Justice 2013;54(3):208–14.
26. Dror IE, Hampikian G. Subjectivity and bias in forensic DNA mixture interpretation. Sci Justice 2011;51(4):204–8.
27. Thompson WC. Painting the target around the matching profile: the Texas sharpshooter fallacy in forensic DNA interpretation. Law Probab Risk 2009;8:257–76.
28. Stacey RB. Report on the erroneous fingerprint individualization in the Madrid train bombing case. J Forensic Identif 2004;54(6):706–18.
29. Office of the Inspector General. A review of the FBI's handling of the Brandon Mayfield case. Washington, DC: U.S. Department of Justice, 2006.
30. Pretty I, Sweet D. A paradigm shift in the analysis of bitemarks. Forensic Sci Int 2010;201:38–44.
31. Thompson WC. Beyond bad apples: analyzing the role of forensic science in wrongful convictions. Southwest Univ Law Rev 2008;37(4):1027–50.
32. Thompson WC. Forensic DNA evidence: the myth of infallibility. In: Krimsky S, Gruber J, editors. Genetic explanations: sense and nonsense. Cambridge, MA: Harvard University Press, 2013;227–55.
33. Stoel RD, Berger CE, Kerkhoff W, Mattijssen E, Dror I. Minimizing contextual bias in forensic casework. In: Strom K, Hickman MJ, editors. Forensic science and the administration of justice. New York, NY: Sage, 2015;67–86.
34. Dror IE, Thompson WC, Meissner CA, Kornfield I, Krane D, Saks MJ, et al. Context management toolbox: a Linear Sequential Unmasking (LSU) approach for minimizing cognitive bias in forensic decision making. J Forensic Sci 2015;60(4):1111–2.
35. Risinger DM, Saks MJ, Thompson WC, Rosenthal R. The Daubert / Kumho implications of observer effects in forensic science: hidden problems of expectation and suggestion. Calif Law Rev 2002;90:1–56.
36. Thompson WC. What role should investigative facts play in the evaluation of scientific evidence. Aust J Forensic Sci 2011;43(2–3):123–34.
37. Nickerson RS. Confirmation bias: a ubiquitous phenomenon in many guises. Rev Gen Psychol 1998;2(2):175–220.
38. Subcommittee on Human Factors, National Commission on Forensic Science (NCFS). Ensuring that forensic analysis is based upon task-relevant information, 2015; https://www.justice.gov/ncfs/file/818196/download (accessed January 15, 2019).
39. Found B, Ganas J. The management of domain irrelevant context information in forensic handwriting examination casework. Sci Justice 2013;53(2):154–8.
40. Osborne NK, Taylor MC. Contextual information management: an example of independent-checking in the review of laboratory-based bloodstain pattern analysis. Sci Justice 2018;58(3):226–31.
41. McAuliff BD, Kovera MB, Nunez G. Can jurors recognize missing control groups, confounds, and experimenter bias in psychological science? Law Hum Behav 2009;33(1):247–57.
42. McAuliff BD, Duckworth TD. I spy with my little eye: jurors' detection of internal validity threats in expert evidence. Law Hum Behav 2010;34(6):489–500.
43. Kukucka J, Kassin SM, Zapf PA, Dror IE. Cognitive bias and blindness: a global survey of forensic science examiners. J Appl Res Mem Cogn 2017;6(4):452–9.
44. Budowle B, Bottrell MC, Bunch SG, Fram R, Harrison D, Meagher S, et al. A perspective on errors, bias, and interpretation in the forensic sciences and direction for continuing advancement. J Forensic Sci 2009;54(4):798–809.
45. Wells JD. Commentary on: Krane DE, Ford S, Gilder JR, Inman K, Jamieson A, Koppl R, et al. Sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation. J Forensic Sci 2008;53(4):1006–7. J Forensic Sci 2009;54(2):500.
46. Butt L. The forensic confirmation bias: problems, perspectives, and proposed solutions—commentary by a forensic examiner. J Appl Res Mem Cogn 2013;2(1):59–60.
47. Thornton JI. Letter to the editor—a rejection of ''working blind'' as a cure for contextual bias. J Forensic Sci 2010;55(6):1663.
48. Mnookin JL, Cole SA, Dror IE, Fisher BA, Houck MM, Inman K, et al. The need for a research culture in the forensic sciences. UCLA Law Rev 2011;58:725–79.
49. Thompson WC, Ford S, Gilder JR, Inman K, Jamieson A, Koppl R, et al. Commentary on: Thornton JI. Letter to the Editor—A rejection of 'working blind' as a cure for contextual bias. J Forensic Sci 2011;56(2):562–3.
50. Pronin E, Yin DY, Ross L. The bias blind spot: perceptions of bias in self and others. Pers Soc Psychol Bull 2002;28:369–81.
51. Pronin E, Kugler MB. Valuing thoughts, ignoring behavior: the introspection illusion as a source of the bias blind spot. J Exp Soc Psychol 2007;43(4):565–78.
52. Nisbett RE, Wilson TD. Telling more than we can know: verbal reports on mental processes. Psychol Rev 1977;84:231–59.
53. Sheldrake R. How widely is blind assessment used in scientific research? Altern Ther 1999;5(3):88–91.
54. Simon D. In doubt: the psychology of criminal justice process. Cambridge, MA: Harvard University Press, 2013.
55. Thompson WC. Determining the proper evidentiary basis for an expert opinion: what do experts need to know and when do they know too much? In: Robertson C, Kesselheim A, editors. Blinding as a solution to bias: strengthening biomedical science, forensic science, and law. London: Elsevier, 2015;133–50.
56. Page M, Taylor J, Blenkin M. Context effects and observer bias—implications for forensic odontology. J Forensic Sci 2012;57(1):108–12.
57. National Commission on Forensic Science. Testimony using the term "reasonable scientific certainty", 2015; https://www.justice.gov/archives/ncfs/page/file/641331/download (accessed January 15, 2019).
58. Jackson G. Understanding forensic science opinions. In: Fraser J, Williams R, editors. Handbook of forensic science. Devon: Willan Publishing, 2009;419–45.
59. Balko R. Yet another bite-mark conviction is unraveling, 2018; https://www.washingtonpost.com/news/the-watch/wp/2018/05/21/yet-another-bite-mark-conviction-is-unraveling/?noredirect=on&utm_term=.6467da044498 (accessed January 15, 2019).
60. Krane DE, Ford S, Gilder J, Inman K, Jamieson A, Koppl R, et al. Sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation. J Forensic Sci 2008;53(4):1006–7.
61. Scurich N. What do experimental simulations tell us about the effect of neuro/genetic evidence on jurors? J Law Biosci 2018;5(1):204–7.
62. Brodsky SL, Neal TM, Cramer RJ, Ziemke MH. Credibility in the courtroom: how likeable should an expert witness be? J Am Acad Psychiatry Law 2009;37(4):525–32.