Science, Technology, or the Expert Witness: What Influences Jurors’ Judgments About Forensic Science Testimony?

Jonathan J. Koehler  
Northwestern University

N. J. Schweitzer  
Arizona State University

Michael J. Saks  
Arizona State University and University of Haifa

Dawn E. McQuiston  
Wofford College

The impact of forensic science evidence on jurors’ judgments is critically important to the criminal justice system. The assignment of low or high weight to such testimony can be the difference between acquittal or conviction. Many of the traditional forensic sciences (e.g., fingerprints and bitemarks) draw their strength largely from the subjective judgments of examiners who testify about whether evidentiary prints or other markings are consistent with (or “match”) known markings from a person or object. In an online experiment (Experiment 1) and a realistic jury simulation using actual jurors or jury-eligible adults (Experiment 2), this article investigates 3 factors that might affect how jurors think about and use forensic science evidence. These factors are (a) whether the forensic science method had been scientifically tested, (b) the forensic scientist’s background and experience, and (c) the sophistication of the forensic science technology. The results show a strong and consistent effect for examiner background and experience on evidence strength judgments, no effect for forensic technology sophistication, and a limited and inconsistent effect for scientific testing (present in the online experiments, absent in the realistic jury simulation). These findings raise concerns about potential undue influence of examiner background and experience on jurors’ judgments and lack of clear influence of scientific testing. The implications of our findings for criminal justice practices and policies are considered.

Keywords: evidence, experience, forensic science, jury decision making, testing

A paradigm shift is under way in the forensic sciences (Fabricant & Carrington, in press; Saks & Koehler, 2005). Traditional forensic sciences, like fingerprint and bitemark analysis, that once were admitted at trial with no hesitation or scientific evaluation are now being questioned and scrutinized. In the wake of a 2009 National Academy of Sciences (NAS) report that found many forensic sciences to be sorely lacking in scientific support and practices (NAS, National Research Council, 2009), a National Commission on Forensic Science was established by the U.S. government in 2014 “to enhance the practice and improve the reliability of forensic science” (http://www.justice.gov/ncfs). At present, a small band of researchers are testing the foundational assumptions of some forensic disciplines, while the Department of Justice and the National Institute of Standards and Technology are endeavoring to create more scientific and professional standards to guide forensic scientists.

Although a focus on improving the soundness of the forensic sciences is welcome and long overdue, the success of the forensic science enterprise ultimately turns on how jurors and other legal decision makers interpret the information they receive. What are the factors that affect jurors’ judgments about the strength and significance of forensic science evidence? Are jurors considering the right factors when evaluating this evidence?

Psychological research on belief formation and updating provides a starting point for investigating how jurors value forensic science testimony. People generally are motivated to acquire accurate beliefs about all kinds of evidence, and to weigh the evidence appropriately. Under ideal conditions, people would seek and more or less rationally analyze a broad array of factors that would help them with this type of task. But people often do not have the time, inclination, or cognitive resources to do so, even for important tasks. Instead, studies show that people commonly em-
ploy mental shortcuts, or heuristics, en route to determining the meaning and value of the evidence and arguments that they receive (Chaiken, 1980; Kahneman, 2011; Simon, 1955).

Legal decision-making is difficult. In courtroom trials, untrained jurors process an array of unfamiliar and often contradictory items of evidence (Saks & Spellman, 2016). In most jurisdictions, jurors hear instructions, evidence, and arguments with little or no opportunity to ask clarifying or substantive questions. Moreover, expert testimony is often complex and presented in technical language. The task that awaits jurors is capable of pulling them in opposite psychological directions. On the one hand, their role would motivate them to want to understand the evidence and seek correct answers. In fact, research indicates that jurors are focused on the substantive content of the evidence, including expert testimony (Diamond & Rose, 2005; Vidmar & Diamond, 2001). But as the testimony becomes complex or draws on techniques whose validity is unknown or difficult to understand, jurors may rely more on external cues such as the expert’s credentials (Cooper & Neuhaus, 2000) or use of language (McKimmie, Newton, Schuller, & Terry, 2013) to evaluate the testimony. Such an approach is consistent with social psychological theory that indicates that people rely on both content-based cues and peripheral (superficial) cues when evaluating evidence and arguments, but they rely more on peripheral cues when they lack the ability or expertise to focus on content (Cacioppo & Petty, 1982; Petty, Cacioppo, & Schumann, 1983; see also Cooper & Neuhaus, 2000).

Little is known about the cues jurors use to assess the value of forensic science testimony. On the one hand, it would seem that jurors should and will rely on findings from the scientific community when it comes to assigning weight to forensic evidence. Evidence that is based on techniques that have been scientifically validated or otherwise approved by the broader scientific community would likely be given more weight than those that are not. On the other hand, evidence from the heuristic and peripheral processing literatures hints that the weight that jurors give to forensic science evidence may be influenced by features of the expert or technology that are easier to process, though potentially less diagnostic. In this article, we identify three cues that might affect the weight jurors assign to forensic science testimony. We then test the impact of these factors on mock jurors in an online study and a realistic jury simulation.

Three Potential Influences

When a forensic scientist testifies that some kind of impression evidence (e.g., fingerprint, bitemark) recovered from a crime scene matches a suspect, jurors have reason to believe the matching suspect may be the source of that evidence. This, in turn, gives jurors some reason to believe that the suspect may be guilty of the crime. But how do jurors go about assessing the diagnosticity of the forensic science match evidence? We identify three potential influences on those judgments.

Has the Forensic Method Been Scientifically Tested?

Under the admissibility standard articulated in Daubert v. Merrell Dow Pharmaceuticals, Inc. (1993), all scientific evidence (including forensic science evidence) that a party wishes to introduce in court must be derived from a sound method. A key determinant of a method’s soundness is whether it has undergone scientific testing and has been validated. Normatively, forensic science evidence produced from a method that has been thoroughly and scientifically tested is more trustworthy than evidence produced using an untested method. Unfortunately, scientific testing and validation are not priorities for many forensic sciences: “Much forensic science evidence . . . is introduced in criminal trials without any meaningful scientific validation, determination of error rates, or reliability testing to explain the limits of the discipline” (NAS, National Research Council, 2009, pp. 107–108).

One might wonder how the results of a technique or expertise that has not been scientifically validated can be admitted into evidence through a legal filter that places a high value on validation. The answer is that, notwithstanding the requirements of Daubert and Federal Rule of Evidence 702, many if not most judges are more comfortable repeating past practice (what has been admitted in the past continues to be admitted) than they are rigorously following a new test to unfamiliar destinations. Judges have varied from creative to flagrant in evading their gatekeeping duties where traditional forensic sciences have been concerned (see, e.g., Edwards, 2010; Faigman et al., 2014, Chapter 32; NAS, National Research Council, 2009; Saks & Votruba, 2015).

Weak judicial gatekeeping makes the study of jurors’ decisions in reaction to forensic science expert testimony all the more important. Research by Schweitzer and Saks (2009) found that jurors generally presume that scientific evidence that has been admitted has been soundly screened by the court. If jurors learn that their assumptions about how thoroughly a forensic method has been vetted are contradicted by trial testimony to the contrary, they may be inclined to discount the value of the forensic science evidence.

A very different possibility is that jurors will not appreciate the importance of a lack of testing and validation in the forensic science context. Upon hearing that a forensic method has not been scientifically validated they might respond with a collective shrug. This possibility receives some indirect support from evidence that mock jurors are insensitive to a wide range of factors that affect the validity of scientific testimony (McAuliff & Kovera, 2008; McAuliff, Kovera, & Nunez, 2009). McAuliff and Duckworth (2010) reported that mock jurors were unaffected by the internal validity of an expert witness’s study in a child sexual abuse case. Bornstein (2004) found that mock jurors in a personal injury case assigned more weight to expert testimony that relied on anecdotal evidence than to expert testimony that relied on a large amount of scientific data. Parrott, Neal, Wilson, and Brodsky (2015) found that mock jurors were insensitive to an expert’s degree of relevant knowledge when evaluating the expert’s opinion in the sentencing phase of a hypothetical capital murder case. Thus, jurors called upon to evaluate and employ forensic science expert testimony might not make use of evidence indicating whether a forensic method has been validated or not.

Is the Forensic Science Expert Experienced?

Social psychologists have long known that source credibility influences persuasion. Information sources who have more expertise in the topic at issue are more persuasive than those who lack such expertise, even when offering identical information (Chaiken
When it comes to expert testimony, a witness’s background and experience are presumably important sources of expertise. Thus, when jurors receive forensic “match” evidence, their evaluations of the strength and value accorded that evidence may be influenced by the background and experience of the examiner. Forensic witnesses vary in such ways as the kind and amount of training they received, how long they have been working with the kind of evidence at issue in the case at bar, their professional association memberships, how many times they have testified in court, and so on. One might expect this relation to be especially pronounced in cases where the underlying method used by the forensic scientist is complex or otherwise difficult for jurors to evaluate substantively. There is little direct evidence on this issue, though a study by Cooper, Bennett, and Sukel (1996) is suggestive. Cooper et al. examined how mock jurors responded to technical expert testimony in a civil case. These jurors relied heavily on a scientific expert’s credentials when weighing the expert’s testimony, but only when the testimony was linguistically complex. This result hints that the background and experience of a forensic scientist might affect jurors’ judgments about the strength of the evidence they provide.

Is the Forensic Science Technology Sophisticated?

All forensic sciences rely on technology to some extent. Both within and across forensic disciplines, examiners often rely on different methods, some of which are more technologically sophisticated than others. For example, fingerprints might be detected using a relatively high-technology technique involving ion lasers (Dalrymple, Duff, & Menzel, 1977) or a lower technology technique involving the manual application of powders. An as yet unexamined issue is how jurors respond to forensic science evidence as a function of the underlying technological sophistication of the method used.

One possibility is that jurors will be more impressed by sophisticated forensic technologies because they will presume that technological sophistication represents a qualitative advance over older methods. Direct evidence for this proposition is scarce. To the extent that sophisticated technologies are hard for lay people to comprehend, jurors and others might judge the evidence produced by these methods using peripheral cues that favor yielding to this impressive-sounding evidence (Cooper & Neuhaus, 2000). The finding in Cooper et al. (1996) that linguistically complex testimony was more persuasive than simpler testimony among high credibility witnesses favors this prediction as well. However, it is also possible that jurors will actually be less persuaded by sophisticated technologies if those technologies are so hard to understand that jurors simply disregard the product along with the process of the incomprehensible technology.

The Present Experiments

The research presented here seeks to discover how lay jurors are influenced by forensic science expert testimony that varies along the three dimensions described above. Normatively, the first factor would seem to be the most important one. The results of a scientific method that have not been tested are generally less trustworthy than those produced by a validated method. This is why the 2009 NAS report on forensic science emphasizes controlled scientific testing across the forensic sciences as a key reform. The background and experience of the examiner would seem to be the next most important factor. On the one hand, intuition suggests that examiners who have less impressive backgrounds or who are relatively inexperienced would be more likely to err. On the other hand, research on trained experts generally shows that those with the most experience are no more accurate in their judgments and decisions than those with much less experience (Camerer & Johnson, 1991). Consistent with this result, studies of forensic pattern-comparison examiners do not find accuracy differences among experts as a function of experience (Bowers, 2010 [bitemarks]; Pacheco, Cerchiai, & Stoloff, 2014 [fingerprints]; Risinger, Denbeaux, & Saks, 1989 [handwriting]; Sita, Found, & Rogers, 2002 [handwriting]; Vernon, 2009 [gait]). Finally, the complexity of the forensic technology should be the least important factor. There is nothing inherent in complex technologies that make them superior to simpler technologies that are used for the same purpose. In some instances, a complex technology might be able to achieve greater precision or consistency under ideal conditions. But the increased sensitivity of higher technology forensic techniques can sometimes produce greater error (Cale, 2015). Further, jurors (and the mock jurors in our studies) are unlikely to receive the information they need to evaluate these subtle matters.

As noted in the paragraphs above, three primary variables are of interest here: method testing, examiner experience, and technological sophistication of method. To explore the possible effects of these variables, we conducted two experiments, first online (Experiment 1) and then using a videotaped trial with deliberating groups of community members (Experiment 2). To increase the external validity of our online results, we provided the simulated jurors in Experiment 1 with one of two types of forensic science evidence in the context of a criminal case: fingerprint evidence or bitemark evidence. The public’s familiarity with (and presumably confidence in) the dependability of fingerprint identification exceeds its familiarity with bitemark identification (Lieberman, Miethe, Carrell, & Krauss, 2008). For both experiments, we examined the impact of scientific testing of a forensic method, an examiner’s experience, and the technological sophistication of the method on simulated jurors’ judgments about the forensic science testimony.

**Experiment 1**

**Overview**

Experiment 1 utilized written materials administered to an online sample of community member participants. We presented one of two separate trial summaries (a sexual assault case involving bitemark evidence and an attempted murder case involving fingerprint evidence) to a sample of participants obtained via Amazon Mechanical Turk. Within each summary, we manipulated three independent variables for a 2 (Forensic Technique Scientifically...
Tested: yes or no) × 2 (Expert’s Experience: high or low) × 2 (Technology Level of the Forensic Technique: high or low) between-subjects design. We analyze these variables separately for each of our two trial summaries, because those trials differ in a number of ways that make direct comparisons difficult.

**Method**

**Participants.** The participants were 441 jury-eligible U.S. residents ("jurors") obtained via Amazon Mechanical Turk. A total of 510 responses were received, with 69 responses excluded for failing a basic screening item (recall of two important case facts) and/or completing the task in less than 5 min. Our sample approximated the jury-eligible population: Nineteen percent of the jurors were non-White, the mean age was 32, 51% completed 2 or more years of college, and 58% were women. Jurors received $1.25 in exchange for their participation.

**Materials and procedure.** In an effort to examine the effect of our manipulations on multiple types of forensic evidence, we prepared materials for two separate trials: a hypothetical attempted sexual assault case that included bitemark evidence that implicated the defendant (the bitemark case) and a hypothetical attempted murder case that included fingerprint evidence that implicated the defendant (the fingerprint case). For both trials, we created a ~700-word summary that included background information, direct examination testimony from a forensic scientist, and cross-examination of that expert.

In both trials, we experimentally manipulated the amount of scientific testing the examiner’s technique had undergone, the experience of the examiner, and the technology used by the examiner. The general public has little knowledge of which forensic techniques have been validated through scientific testing and which have not. Consequently, when our mock jurors learn from the expert’s testimony in our mock trials that a technique has been subjected either to a great deal of testing or to no testing, we expect them to accept that testimony or to at least be pushed in the direction asserted by the testimony. Our manipulation of the expert’s experience is similarly strong: extensive experience versus zero experience (training, job history, case experience, witness experience) that commonly arise in voir dire when attorneys attempt to bolster or denigrate the credibility of a forensic witness. We described the less experienced forensic dentist as having worked as a dental hygienist prior to receiving forensic science training, having assisted with just a few previous forensic dentistry cases, and having testified in only one other case. We described the more experienced forensic dentist as having worked as an orthodontist prior to receiving forensic science training, having participated in hundreds of cases, and having testified many times. By varying several dimensions of experience simultaneously to create a more and less experienced expert, we created a realistic experience gestalt of the sort that commonly arises at trial.

For the scientific testing variable, the expert in the not tested conditions testified that his technique had not been subjected to any actual scientific testing, whereas the expert in the tested conditions testified that his technique had been subjected to a great deal of scientific testing. For the technology variable, the expert in the low-technology conditions described using an old-fashioned photographic technique, whereas the expert in the high-technology conditions describing using high-technology ultraviolet photographic equipment.

**Fingerprint case.** Participants who received the fingerprint case read that the defendant was charged with attempted murder. The victim was shot by an intruder as she descended to the bottom of the staircase in her house. She could not identify her assailant. As part of the subsequent investigation, police searched for and obtained latent fingerprints in the victim’s house. At trial, a fingerprint examiner testified about the fingerprinting process and identified the defendant as the source of a print that he recovered from a mug on victim’s kitchen table. Actual fingerprint examiners generally rely on the analyze, compare, evaluate, and verify (ACE-V) framework for analyzing fingerprints. However, according to the 2009 NAS report, ACE-V “is not specific enough to qualify as a validated method” (NAS, National Research Council, 2009, p. 142) and is a highly subjective process. The fingerprint testimony provided in this experiment reflected one such subjective ACE-V analysis, and was modeled on actual courtroom testimony (see, e.g., State of Minnesota v. Columbus, 2006).

As in the bitemark case, we provided sharp, but realistic, contrasts between the two levels of each of our three primary independent variables in the fingerprint case. Paralleling the manipulations in the bitemark case, the less experienced fingerprint examiner had worked as a security guard prior to receiving foren-
sic training, had been on the job as a police officer for less than one year, and had testified in just one other case. The more experienced fingerprint examiner was a long-time police officer who had been conducting fingerprint analyses for more than 20 years and had testified in many cases. The expert in the not tested conditions testified that his technique had not been subjected to any actual scientific testing, whereas the expert in the tested conditions testified that his technique had been subjected to a great deal of scientific testing. Finally, the examiner in the low-technology conditions described using an old-fashioned powder technique, whereas the examiner in the high-technology conditions described using what he said was a state-of-the-art ultrasound technique called optical coherence tomography.

Dependent measures. After reading the one version of the one case to which they had been randomly assigned, jurors answered a series of questions using a 7-point Likert-type scale. These questions included four items about the strength of the forensic evidence, two questions about the guilt of the defendant, and three manipulation check questions. We also asked jurors to answer a few secondary questions that explored their understanding of the case as well as a set of questions pertaining to their own background/demographic characteristics.

Results

Preliminary analyses. Before analysis, we created two primary dependent variables (DVs): The first DV, evidence strength, was a composite of four items we designed to measure the overall perceived quality and usefulness of the forensic evidence (Cronbach’s α = .94 in the bitemark trial and .92 in the fingerprint trial).

The second DV, guilt, was a composite of two items measuring participants’ confidence in the proposition that the defendant committed the act of which he was accused (Cronbach’s α = .94 in the bitemark trial and .95 in the fingerprint trial). While these two composite items were substantially correlated with each other (.90 in the bitemark trial and .71 in the fingerprint trial), each DV measures a different construct so we analyzed them separately. The list of DVs is included as the Appendix.

We then analyzed responses to three manipulation check items in each trial that explicitly measured participants’ perceptions about scientific testing of the forensic technique, the expert’s experience, and the technology level of the forensic technique. We found strong effects for all three manipulations in the expected directions; all ps < .001; r2 range from .33 to .63. The manipulation check items are provided in the Appendix.

We initially included our two trials (fingerprint case and bitemark case) as a fourth independent variable, and found that the trial type did not interact with any other variables on either of our DVs. This result might lead some to recommend collapsing across trial type when analyzing the data. However, because our two sets of trial materials differed along a number of dimensions over and above the basic type of forensic evidence present (different crime, different witnesses, etc.), they are best thought of as replications under different factual circumstances. For the sake of providing thorough and complete analyses, we analyze each trial separately.

Bitemark case. We conducted a 2 (Expert’s Experience) × 2 (Technology Level) between-subjects analysis of variance (ANOVA) on each of our primary DVs in the bitemark case. On the evidence DV, participants were sensitive to whether or not the bitemark identification technique used in the attempted sexual assault case had been scientifically tested. The evidence was rated as stronger and more convincing when the underlying technique had been scientifically tested (4.92 vs. 4.55), F(1, 207) = 4.87, p = .028, r2 = .03. In addition, the evidence was rated as stronger when it was provided by a more experienced examiner (5.16 vs. 4.32), F(1, 207) = 25.55, p < .001, r2 = .11. Likewise, ratings of perceived guilt were higher when the expert witness was more experienced (5.53 vs. 4.59), F(1, 205) = 17.64, p < .001, r2 = .079. However, the effect of the scientific testing did not translate into differing guilt ratings, F(1, 205) = 2.71, p = .10, r2 = .013. Participants were not sensitive to the technology variable on either the evidence or guilt DVs, and there were no interactions among the three independent measures on either DV (ps range from .12 to .89). Means and standard deviations are reported in top two rows of Table 1.

Fingerprint case. We similarly conducted a 2 (Scientifically Tested) × 2 (Expert’s Experience) × 2 (Technology Level) between-subjects ANOVA on each of our primary DVs in the fingerprint case. The results were substantially similar to those in the bitemark case: On the evidence DV, participants were sensitive to whether or not the fingerprint technique had been scientifically tested. A method that had been scientifically tested was rated as more convincing than an untested method (5.41 vs. 4.86), F(1, 214) = 21.06, p < .001, r2 = .09. As in the bitemark case, participants were also sensitive to the expert’s amount of experience: Participants rated the evidence provided by the more experienced examiner as more persuasive on the evidence DV (5.52 vs. 4.70), F(1, 214) = 36.03, p < .001, r2 = .15. We also detected an Experience × Technology interaction, F(1, 214) = 11.85, p = .003, r2 = .040, on the evidence DV, illustrated in Figure 1. Among fingerprint examiners with lower levels of experience, the use of high-tech tools reduced the perceived strength of the evidence, F(1, 111) = 1.75, p = .19, r2 = .02. No other significant effects or interactions on the evidence strength DV were found, and there were no significant effects or interactions of any type on the rating of defendant guilt (ps range from .08 to .76). Means and standard deviations are reported in rows 3 and 4 of Table 1.

Discussion

The results for the bitemark and fingerprint cases were similar in two different criminal case scenarios. In both cases, on the evidence DV, jurors were sensitive to whether the forensic technique in question had been scientifically tested. Techniques that had been scientifically validated were viewed as providing stronger evidence of identity. However, this effect did not carry through in either case to the guilt DV, which may be influenced by a broader range of considerations (e.g., the amount of nonforensic evidence against the defendant, standard of proof, etc.). Indeed, in both
cases, a majority of jurors voted to acquit in all experimental conditions.

Examiner experience was important to our jurors on both the evidence and guilt DVs. This result supports the possibility that experience is a powerful, central trait for people when they are called upon to assess the value of information that another person or a process has provided. Experience may function as a kind of marketplace-driven “proof” that those who have given expert testimony surely must be reliable sources or else they would not be in the position that they are in at this time. Experience may also serve as an antidote to the uncertainty that exists when people tell us things that we cannot evaluate or interpret for ourselves. In our study, it appears that jurors leaned on the experience of the testifying forensic scientist to guide their assessments of the soundness of his findings.

In the fingerprint case, the impact of examiner experience on the evidence measure was greatest when the technology was complex. One reason for this could be that the high-technology fingerprint technique described (optical coherence tomography, a state-of-the-art ultrasound technique) was unfamiliar to most people and therefore potentially risky in the wrong (i.e., inexperienced) hands. By contrast, it is understandable that jurors would pay less attention to level of technology when examiner experience is high if jurors trust that an experienced fingerprint examiner would be capable of interpreting the results from any type of fingerprint technology. The absence of this interaction in the bitemark case may be due to people’s unfamiliarity with any bitemark techniques: If people are not confident about any bitemark technology, they might fall back on examiner experience as an indicator of reliability.

**Experiment 2**

**Overview**

The online experiments described in Experiment 1 point to the importance of examiner experience as an influence on jurors’ assignment of weight to forensic science evidence, as well as whether or not a technique had been scientifically tested. Though these experiments are suggestive, the experimental setting, materials and procedures were less rich than those that actual jurors

---

<table>
<thead>
<tr>
<th>Evidence strength and Guilt ratings</th>
<th>Expert’s experience</th>
<th>Technology level</th>
<th>Scientifically tested</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bitemark evidence</td>
<td>5.16 (1.24)**</td>
<td>4.32 (1.32)**</td>
<td></td>
</tr>
<tr>
<td>Bitemark guilt</td>
<td>5.35 (1.36)**</td>
<td>4.59 (1.47)**</td>
<td></td>
</tr>
<tr>
<td>Fingerprint evidence</td>
<td>5.52 (1.14)**</td>
<td>4.70 (1.30)**</td>
<td></td>
</tr>
<tr>
<td>Fingerprint guilt</td>
<td>4.81 (1.59)</td>
<td>4.47 (1.64)</td>
<td></td>
</tr>
<tr>
<td>Experiment 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predeliberation evidence</td>
<td>4.68 (1.62)**</td>
<td>3.74 (1.74)**</td>
<td></td>
</tr>
<tr>
<td>Predeliberation guilt</td>
<td>3.54 (1.98)</td>
<td>3.41 (2.02)</td>
<td></td>
</tr>
<tr>
<td>Postdeliberation evidence</td>
<td>4.30 (1.67)**</td>
<td>3.27 (1.70)**</td>
<td></td>
</tr>
<tr>
<td>Postdeliberation guilt</td>
<td>3.23 (1.66)</td>
<td>3.03 (1.73)</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05.  **p < .01.
encounter when making decisions about criminal matters. Jury researchers are well aware of this so-called ecological validity issue, and the matter has been discussed extensively in the literature (e.g., Bornstein, 1999; Vidmar, 2008). On the one side are those who have criticized jury simulations for using unrepresentative participants, providing unrealistic and attenuated stimuli, and failing to provide participants with such common trial features as expert testimony, opportunities for cross-examination, objections, judicial rulings, judicial instructions, group deliberation, and realistic measures (Koehler & Meixner, in press; Wiener, Krauss, & Lieberman, 2011). On the other side are those who have argued that ecological validity concerns are overstated because there is little evidence that the results obtained from these types of studies differ from those that employ richer and more realistic stimuli and procedures (Bornstein, 1999; Bray & Kerr, 1979, p. 117). Still, the gaping chasm that frequently exists between the methods used in relatively impoverished jury simulation studies such as our online study above and the experience of empaneled jurors in criminal trials throughout the United States provides reason enough to try to replicate our findings in a more ecologically valid setting. In our second experiment, we went to great lengths to provide mock jurors with a more ecologically valid experience that included realistic trial procedures, video testimony, and group deliberations.

Method

Participants. A total of 315 jury-eligible community members were recruited from one of two sources: the jury pool at the Phoenix, Arizona, municipal court (n = 193) and from community members through an email solicitation to jury-eligible nonacademic staff members of Arizona State University along with a mailing list provided by a market research firm (Survey Sampling International; n = 122). Mock jurors were compensated $50 for their participation. The participants were recruited over the course of several months to appear in groups of approximately six to eight in one of two study locations in the Phoenix metropolitan area resulting in 46 individual juries. Our sample approximated the jury-eligible population: The mean age was 44 years, the sample was 64% female, 74% White, and 58% of the sample possessed at least a 2-year college degree. A full listing of demographics separated by recruitment source is provided in Table 2.

Materials and procedure. Using the same conceptual framework as the materials presented to the online participants, we created a ~7,500-word trial transcript of the fingerprint case used in Experiment 1. We developed the fingerprint case rather than the bitemark case for the video because fingerprint evidence is a more common, relatively well-understood forensic science, and because the presence of a match in this case is not necessarily dispositive on the matters of guilt due to potential ambiguities about when the print was deposited. As in Experiment 1, the victim in this case was shot in her home by an intruder she could not identify. Within minutes after the shooting, the police picked up an individual from a nearby street who matched the general physical description of the assailant. Following questioning, the police identified this individual as a suspect in the shooting. A fingerprint that matched the suspect was found in the victim’s home, and he was charged with attempted murder. A criminal trial followed.

The sequence of events in the 7,500-word transcript followed that of the typical courtroom trial (post jury selection). The transcript opened with preliminary instruction from the presiding judge. The judge briefly described the importance of jury service, the nature of the case, and the likely order in which jurors would hear from the various parties. Next, there were opening statements from the prosecution and defense that provided jurors with a roadmap for the evidence and arguments they were about to hear. The prosecution then presented its case through testimony from the victim, an investigating police officer, and a fingerprint examiner. Following direct examination of each prosecution witness, the defense cross-examined the witnesses. The direct and cross-examinations of witnesses were condensed relative to most of those that would occur at an actual trial. However, care was taken to include the important elements of these exchanges and to make the exchanges as realistic as possible. For example, witnesses answered preliminary questions before offering their substantive testimony, attorneys made occasional objections, and the trial judge ruled on those objections. Secondary material of this sort was held constant across experimental conditions. Direct and cross-examination testimony from the fingerprint examiner was modeled on transcripts that we collected from actual criminal cases (see, e.g., State of Minnesota v. Columbus, 2006). This testimony included a detailed description of the examiner’s experience, the method he used to analyze the fingerprint, and comments on the extent to which the method had been subjected to scientific validation. Following cross-examination of the prosecution’s last witness, the defense declined to present witnesses of its own. Next, the prosecution and defense presented closing arguments and the judge provided final instructions. The instructions were modeled on federal pattern jury instructions and included an explanation of reasonable doubt.

The written transcript served as the script for the filming of the mock trial. Those who played the various roles in the videos were age-appropriate, dressed appropriately, and spent time learning and practicing their testimony. To minimize unwanted influences for gender, the judge, attorneys and prosecution witnesses were all men. Those who played attorneys were actual attorneys, and the forensic expert was knowledgeable about forensic matters. To ensure consistent presentation of the verbal content across experi-

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Jury pool</th>
<th>E-mail list</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (M)</td>
<td>46.5</td>
<td>39.0</td>
<td>43.7</td>
</tr>
<tr>
<td>Sex (% female)</td>
<td>59.9%</td>
<td>71.2%</td>
<td>64.2%</td>
</tr>
<tr>
<td>Ethnicity (%)</td>
<td>White 72.7%</td>
<td>Hispanic 75.4%</td>
<td>73.8%</td>
</tr>
<tr>
<td>African American</td>
<td>3.2%</td>
<td>4.2%</td>
<td>3.6%</td>
</tr>
<tr>
<td>Other</td>
<td>8.0%</td>
<td>8.5%</td>
<td>8.2%</td>
</tr>
<tr>
<td>Education (%)</td>
<td>No college 12.5%</td>
<td>Some college 38.6%</td>
<td>9.4%</td>
</tr>
<tr>
<td>2- or 4-year degree</td>
<td>34.4%</td>
<td>21.6%</td>
<td>29.5%</td>
</tr>
<tr>
<td>Some graduate school</td>
<td>7.8%</td>
<td>10.3%</td>
<td>8.8%</td>
</tr>
<tr>
<td>Graduate degree</td>
<td>16.6%</td>
<td>25.9%</td>
<td>20.1%</td>
</tr>
<tr>
<td>Political views (M)</td>
<td>2.5</td>
<td>2.0</td>
<td>2.4</td>
</tr>
</tbody>
</table>

*Rated from 1 (liberal) to 4 (conservative).
imental conditions, participants delivered their lines with the aid of scripts—either reading from paper when that was natural (e.g., lawyers working from their notes) or from a teleprompter-like device (for witnesses). Eight 35- to 40-min versions of the video were created. As in Experiment 1, the eight trial videos differed in whether the forensic method had or had not been tested, whether the examiner was highly experienced or not, and whether the technology employed by the examiner was sophisticated or simple. The study employed a 2 (Forensic Technique Scientifically Tested: yes or no) × 2 (Expert’s Experience: high or low) × 2 (Technology Level of the Forensic Technique: high or low) between-subjects design. The experimental manipulations involved the same general language described in Experiment 1, and occurred during direct and cross-examination of the forensic scientist. The Scientifically Tested manipulation was delivered as a direct response to a question about whether the technique had been tested. The Experience manipulation was delivered in several lines of dialog pertaining to the examiner’s training and professional background. The Technology manipulation was delivered in several lines of dialog explaining the name of and technological processes behind the technique. For increased realism, the defense attorney repeated shortcomings (e.g., lack of scientific testing, lack of experience, a low-tech methodology), if there were any, in both cross examination (e.g., “So you’re telling us that the method that you used in the case before us today has not been subjected to a thorough scientific examination?”) and again in closing arguments. The amount of time in the videos that was spent on each of the three variables was similar in all conditions and amounted to a total of three to four minutes during the video trial.

After providing consent, a group of jurors (typically six to eight) was seated in a conference room. The jury group was then randomly assigned to watch one of the eight versions of the trial video. Participants were instructed to watch the video as if they were jurors in the case. During the trial video, the jurors were provided with pads and pens with which to take notes.

When the video concluded, jurors were asked to complete a predeliberation questionnaire individually and anonymously. Jurors were allowed to refer to their written notes and to a written copy of the judge’s jury instructions. The first question on the questionnaire was a verdict (not guilty or guilty). The remaining questions, which were answered using a 7-point Likert-type scale, included four questions about the strength of the forensic evidence, two questions about the guilt of the defendant, and various questions that explored jurors’ understanding of the case. A series of Likert-type questions that served as manipulation checks pertaining to the three independent variables were also included.

After jurors completed their predeliberation questionnaires, they were instructed to begin deliberating about the case with their fellow jurors. Juries deliberated until they reached a unanimous verdict or for 30 min, whichever came first.3 Seventy-eight percent of jurors reached a verdict. Next, individual jurors answered a second packet of written questions about the case (“postdeliberation questionnaire”). The first half of these questions was identical to the set of predeliberation questions. The second half included an 18-question Need for Cognition (NFC) scale (Cacioppo, Petty, Feinstein, & Jarvis, 1996; Cacioppo & Petty, 1982) and a series of background items. The NFC scale, which identifies how much people enjoy engaging in effortful cognitive activities, has been used elsewhere in the law and psychology literature to predict which jurors will be most sensitive to identifying flawed scientific evidence (Bornstein, 2004; McAuliff & Kovera, 2008). The background questions included a series of demographic questions as well as items pertaining to classes completed in “science, math and logic,” political views, and exposure to forensic detective TV shows. After completing this questionnaire, jurors were debriefed and paid.

Results

Preliminary analyses. Before analysis, we again created two primary DVs. The first DV, evidence strength, was a composite of the same four items used in the fingerprint case in Experiment 1 (provided in the Appendix) designed to measure perceived quality and usefulness of the forensic evidence (Cronbach’s α = .94 predeliberation; .92 postdeliberation). The second DV, guilt, was a composite of the same two items used in Experiment 1 to measure the participant’s confidence in the fact that the defendant committed the act of which he was accused (Cronbach’s α = .96 predeliberation; .94 postdeliberation). The correlation between the two items was .82 prior to deliberation and .72 postdeliberation.

We again analyzed responses to three manipulation check items that directly measured participants’ subjective judgments of scientific testing of the forensic technique, expert’s experience, and the technology level of the forensic technique. We found large significant effects for all manipulations, both pre- and postdeliberation, in the expected directions; all ps < .001; η² range from .21 to .66.

Primary analyses. To examine the effects of our independent variables both pre- and postdeliberation we analyzed each portion using a different analysis. For predeliberation items, we conducted a 2 (Scientifically Tested) × 2 (Expert’s Experience) × 2 (Technology Level) between-subjects ANOVA on our two primary DVs. For postdeliberation items, we conducted a two-level hierarchical linear model (using the SPSS “MIXED” procedure) with the independent variables and their interactions (full factorial) at the group/jury level and the postdeliberation questionnaire responses at the individual level. This allows us to test the effects of our independent variables on jurors’ individual postdeliberation responses while controlling for the fact that they were nested within groups (juries) and, thus, not independent.

Consistent with Experiment 1, we found a significant effect of the examiner’s experience level on the perceived strength of the forensic evidence. Predeliberation, the evidence was rated as stronger when presented by the more experienced examiner than by the less experienced examiner (4.68 vs. 3.74; F(1, 300) = 24.93, p < .001, η² = .077. This effect persisted postdeliberation (4.30 vs. 3.26); t(36.37) = 2.41, p = .02. Further, consistent with Experi-
ment 1, the examiner’s experience level had no effect on perceptions of defendant guilt either pre-deliberation, $F(1, 300) = 0.40$, $p = .52$ or post-deliberation, $t(300.97) = 0.57$, $p = .57$.

Participants in this videotaped experiment were less sensitive than their counterparts in Experiment 1 to whether the forensic method had been tested: Both before and after deliberation, participants who were told that the fingerprint method had been scientifically validated did not think the evidence was stronger than those who were told that the fingerprint method had not been validated. Likewise, the testing manipulation did not significantly affect or interact with any other independent variable on either the pre- or post-deliberation composite guilt measure ($p$s range from .09 to .95; all $\eta^2$s < .01).

No pre- or post-deliberation differences emerged for the technology factor (nor any of its interactions, as was found in Experiment 1) on either the evidence or the guilt measures ($p$s range from .10 to .91; all $\eta^2$s < .01). Pre- and post-deliberation means and standard deviations are listed in rows 5–8 of Table 1.

Secondary analyses. We examined whether the demographic variables of age, education, or gender, or the 18-question NFC variables of age, education, or gender, or the 18-question NFC did not yield any effects, we would not expect to find them on verdicts.

Discussion

Contrary to scientific norms, we found little indication that the scientific validation or lack of such validation of the fingerprint technique made any difference to jurors who watched the video trial. The testing variable manipulated did not exert a main effect or interactive effect on any dependent measure. This null result did not arise because jurors simply missed the testing manipulations in the trial. Our manipulation check showed that jurors picked up on whether the fingerprint method had been scientifically tested or not. However, it may be that jurors simply didn’t perceive a connection between the scientific validation of a forensic technique and its accuracy. Support for this idea also appears in McQuiston-Surratt and Saks (2009) who found that admission by a microscopic hair analyst that the assumptions underlying his opinion had been subject to little scientific testing had almost no impact on jurors’ judgments.

As in the online experiment, jurors assigned more weight to the forensic science evidence when the testifying witness was a highly experienced examiner. Why do we find effects for experience but not for testing? One possibility is that people believe fingerprint analyses are quite reliable when performed by the right person. Under this belief system, an examiner who has a great deal of experience is also one who has great expertise and competence. Extending this logic, an expert examiner wouldn’t use a method that isn’t reliable, and so it really doesn’t matter whether the method has been tested or not. In this manner, a characteristic of the examiner, experience, serves as a proxy for scientific validity. Further, people’s prior beliefs about the soundness of fingerprint methods and outcomes may be so strong that those beliefs do not budge upon learning that the method has not yet been scientifically tested. The fact that trial judges who conduct Daubert hearings on forensic science evidence are rarely affected by learning that fingerprint methods have minimal (if any) empirical validation provides some support for this idea (Faigman et al., 2014, Chapter 32).

There were no main effects or interactions involving the technology variable. Jurors did not think that matches detected using the higher technology fingerprint method provided stronger (or weaker) proof of source or guilt than matches detected using a simpler, “old-fashioned” method. Though we entertained the possibility that jurors would be more persuaded by sophisticated-sounding technology because such technology is presumed to be “better” than more primitive methods, that idea was not supported by this null result.

Deliberation had little impact on jurors’ responses to our key dependent measures. For the most part, jurors felt the same way about the evidence and the likely guilt of the defendant following deliberation as they had heading into deliberation. This result is consistent with the empirical findings from Hans, Kaye, Dann,
Farley, and Albertson (2011) and Thompson, Kaasa, and Peterson (2013). One reason why deliberation has little impact may be that people think through the implications of the evidence on their own prior to deliberation, particularly when expressly invited (by our predeliberation questionnaire) to do so. At the same time, we note that deliberation did change the verdicts of nearly one out of five jurors, almost exclusively moving them from a predeliberation verdict of guilty to a postdeliberation verdict of not guilty. This result is consistent with a few other studies that indicated that deliberation decreases conviction rates (Kaasa, Peterson, Morris, & Thompson, 2007; MacCoun & Kerr, 1988). Though not of central concern in the present study, continuing research is needed to identify the conditions under which deliberation alters individuals’ predeliberation preferences.

Our analyses of secondary considerations (demographics and need for cognition) did not yield many insights. Whereas Thompson et al. (2013) found higher conviction rates for women, non-native English speakers, and people who are high in self-professed “ability to evaluate scientific evidence,” we did not find differences in conviction rates as a function of gender, years of education, or number of science-oriented classes completed.

General Discussion

At the outset, we noted that little is known about factors that influence jurors’ assessments of the value of forensic science testimony. Whereas a rational juror might focus on the extent of scientific testing and validation (or lack thereof) of the technique used to generate the forensic evidence, an actual juror might rely on more peripheral cues to make this assessment. In two experiments, we conducted jury simulations that examined the impact of two features of the forensic science techniques themselves (whether they had been tested and whether they were low or high tech), and one feature of the testifying forensic scientist (his degree of experience). Though the research is designed to uncover empirical relationships that might exist, rather than using the trial context as a setting in which to test theoretical propositions, the empirical phenomena that are found can be better understood by drawing on relevant explanatory theory. The results of our experiments are notable for what we found, what we did not find, and what we found in one experiment but not the other.

In general, we found a strong and consistent main effect for examiner experience on the evidence composite measure. Forensic science match evidence presented by a highly experienced examiner with an impressive background was judged to be stronger and more persuasive than identical evidence presented by a less experienced examiner with a less impressive background. We observed this effect in an online bitemark case, an online fingerprint case, and a videotaped attempted murder case that featured fingerprint evidence. Examiner experience was the only variable we examined that consistently influenced jurors’ evidence strength judgments.

This finding suggests that jurors use the background and experience of an expert as a proxy for the value of the evidence the expert provides. Such a process is consistent with heuristic models that are premised on the idea that decision makers who lack the motivation or ability to assess the value of available substantive information rely on easily understood peripheral cues (such as an expert witness’s experience) to judge the quality of that information (Cacioppo & Petty, 1982; Chaiken, 1980; Kahneman & Fredrick, 2002). This process is also consistent with literature that shows an expert witness’s perceived credibility is influenced by the witness’s status and experience level (Swenson, Nash, & Roos, 1984; Wechsler, Kehn, Wise, & Cramer, 2015), and such noncontent cues as likability and confidence (Brody, Griffin, & Cramer, 2010).

Various features of the trial process invite jurors to draw on an expert’s background and experience. Direct examination of experts usually begins with a discussion of their credentials and experience, attorneys frequently refer to experts’ qualifications or lack thereof in closing arguments, and jury instructions often advise jurors to attend to experts’ “qualifications and experience” (Vidmar & Diamond, 2001, p. 1131). But from an epistemic standpoint, these cues are problematic because it is not clear that a more impressive sounding background or more case experience provide a valid indicator of greater expertise or accuracy. As others have noted, experience alone is unlikely to improve judgmental accuracy unless the judgments made are followed up with a good deal of timely, unambiguous outcome feedback (Einhorn, 1980; Einhorn & Hogarth, 1981). Armstrong (1980) has argued that expertise and accuracy are almost entirely unrelated beyond a minimal level, presumably a level that is achieved by forensic experts following their training in combination with minimal case experience. In at least several forensic science domains, Armstrong (1980) has some support: There appears to be little or no relationship between experience and accuracy in identification by bite-marks (Bowers, 2010), fingerprints (Pacheco et al., 2014), handwriting (Risinger et al., 1989; Sita, Found, & Rogers, 2002) or gait (Vernon, 2009).

A word of caution about our findings related to experience: we varied a number of facets of the forensic examiner’s experience simultaneously to provide study participants with an examiner who would be seen as either very experienced or very inexperienced. Our data do not speak clearly to the amount of experience required to persuade jurors to abide by the word of the examiner. Nor do our data speak to which facets of experience (e.g., amount of training, cases worked, courtroom appearances) are more and less influential. This refinement is left for another study. Our goal was simply to determine whether examiner experience writ large matters, and our data are consistent with the inference that it does.

Regardless of what we did not find, our jurors did not assign greater weight to forensic science results generated from impressive-sounding high-technology methods relative to those generated from low-technology methods that were described as “old-fashioned.” This null result is encouraging as it suggests that jurors may not easily be influenced by mere flash.

A second notable absence in our results concerns the guilt composite measure. With the exception of a main effect for examiner experience in the bitemark case in Experiment 1, there were no main effects or interactive effects on the guilt measure. One possible explanation for this absence is that jurors recognized that the probative value of the fingerprint match for estimating the chance of the defendant’s guilt is diminished by the possibility that although the defendant is the source of the print, he might have innocently left his prints on the item at some time other than when the crime occurred. A related explanation is that jurors based their assessments of guilt on a wide range of considerations, some of which were simply more important to them than the forensic science evidence.
An important inconsistent effect in our studies concerns the effects of the scientific testing variable. In the online bitemark and fingerprint studies, jurors thought that the forensic evidence was somewhat stronger when the expert indicated that the method he used had been scientifically tested. However, this effect did not carry over to the guilt dependent measure, and there was no indication that jurors were sensitive to whether the fingerprint method had been scientifically validated in the more ecologically valid video study either before or after deliberating. Importantly, the absence of effects for scientific testing in the video study did not arise because jurors didn’t process or recall this bit of information; a manipulation check confirmed that jurors were aware of whether the fingerprint method used had or had not been scientifically tested.

A finding that jurors are insensitive to whether a forensic method has been scientifically tested or not in a realistic jury simulation is worrisome. To a scientist, empirical testing of a purported technique is essential. In the scientific world, a method that has not been scientifically tested is less trustworthy than a method that has been tested. The lack of scientific testing of forensic techniques was the most profound of the concerns that troubled the NAS committee in its landmark review of the forensic sciences (NAS, National Research Council, 2009).

An interesting question to consider is why testing appeared to have mattered to our online jurors given that it did not matter to our video jurors. The answer may lie in the difference between the amounts of information provided in the two contexts. Whereas people may have the cognitive resources to attend to scientific validation in a relatively simple information environment such as what we provided in the online study, the sheer number of considerations that are introduced in a longer, richer video study may have pushed aside this variable.

Our research could have implications for criminal justice policies and practices. First, if jurors are responsive to peripheral cues rather than content cues when evaluating forensic evidence, then courts need to be especially vigilant about enforcing their gatekeeping duties under Daubert v. Merrell Dow Pharmaceuticals, Inc. (1993). The danger in failing to apply Daubert’s tough reliability test on the front end is that jurors will presume that admitted forensic evidence is accurate evidence, and cross-examination that exposes substantive scientific weaknesses will be ignored. Second, jurors’ apparent belief that evidence provided by more experienced forensic scientists is more accurate and valuable than the identical evidence provided by less experienced forensic scientists has no basis in reality. As noted earlier, research on expertise finds that experience and credentials are a poor predictor of accuracy, and forensic science does not appear to be an exception.

If future research confirms jurors’ lack of sensitivity to something so important as scientific validation or lack thereof, along with the excessive value jurors place on examiners’ experience and qualifications, then courts, legislatures, rules committees, and scientific bodies charged with making policy recommendations would be wise to promote policies, rules and instructions that minimize the significance of a forensic expert’s background, and emphasize the quality and extent of scientific testing of the forensic techniques being used. Examples exist of courts that sought to compensate for a lack of rigor in gatekeeping by instructing jurors concerning important weaknesses of the forensic science testimony (e.g., United States v. Starzecpyzel, 1995). Moreover, attorneys could play a more active role in regard to these issues by emphasizing these matters in their closing arguments and by asking courts to provide jurors with appropriate instruction.

Caution should be exercised when drawing inferences from the experimental data that we present here to actual trials and sitting jurors. Despite our best efforts in Experiment 2 to produce an experiment with high ecological validity, shortcomings are inevitable. Unlike an actual trial, our mock jurors did not hear live testimony, did not sit through a full length trial, did not deliberate with 11 others, and did not render consequential verdicts. Further, our stimuli were confined to different versions of an attempted assault case that included bitemark evidence and an attempted murder case that included fingerprint evidence. Future research should attempt to replicate and refine our findings using a broader range of forensic science evidence, participant populations, cases, and research designs. Furthermore, it may be desirable to use stimuli that provide a stronger case against the defendant and to use examiners whose experience levels are at several intermediate points across the spectrum of expert training.

A recent study of fingerprint error rates indicated that experienced, examiners certified by the International Association for Identification, were no more or less likely to commit false negative errors when comparing pairs of prints than inexperienced, noncertified examiners. Similarly, examiners certified by the International Association for Identification committed false positive errors at the same rate as uncertificated examiners. However, very inexperienced examiners (i.e., those who “reported they either had no training and experience or less than a year experience” as a fingerprint examiner) reportedly had a higher false positive error rate than experienced examiners (Pacheco, Cerchiae, & Stoiloff, 2014).

References


Appendix
A List of Primary Dependent Variables

Evidence Strength Individual Items

1. In comparison with other possible kinds of evidence, how strong would you say the [fingerprint/bitemark] evidence is? (1 = not at all strong; 7 = extremely strong)

2. How convincing is the testimony from the expert? (1 = not at all convincing; 7 = extremely convincing)

3. How confident are you that the [fingerprints/bitemarks] recovered at the crime scene were left by the defendant? (1 = not at all confident; 7 = extremely confident)

4. What would you say is the probability that the [fingerprints/bitemarks] recovered at the crime scene were left by the defendant? (0–100%; rescaled to 1–7 for composite)

Manipulation Check Items

1. Expert’s experience: How experienced would you say the [fingerprint/bitemark] expert was in this case? (1 = not at all experienced; 7 = extremely experienced)

2. Technology level: How would you describe the level of technology that the expert used in this case? (1 = extremely low tech; 7 = extremely high tech)

3. Scientifically tested: Based on the expert’s testimony, how much scientific testing has the method used by the expert been subjected to? (1 = none; 7 = an extremely large amount)

Guilt Individual Items

1. How confident are you that the defendant shot the victim in this case? (1 = not at all confident; 7 = extremely confident)

2. What would you say is the probability that the defendant shot the victim? (0–100%; rescaled to 1–7 for composite)

Received July 18, 2016
Revision received August 31, 2016
Accepted September 3, 2016