Document Title: **Communicating Forensic Science**

Author(s):   **N.J. Schweitzer**

Document No.:   **249804**

Date Received: **May 2016**

Award Number: **2008-DN-BX-0003**

Communicating Forensic Science

NIJ Final Report / Executive Summary

Project 2008-DN-BX-0003

N. J. Schweitzer, Principal Investigator
Arizona State University
PO BOX 37100 | Mail Code 3051
Phoenix, AZ 85069-7100
tel: 602-543-8133
fax: 602-543-6004
email: njs@asu.edu

Background and Purpose

The broad purpose of this project was to contribute to the sparse knowledge that currently exists regarding how jurors respond to forensic identification expert testimony presented at trial. We pursued this goal by conducting a series of controlled studies involving jury simulations or mock trials. Jury simulations are experimental studies in which the researcher attempts to construct a setting that mirrors, in fundamental ways, a jury decision-making environment. Jury simulations vary widely in terms of participants, materials, physical settings, realism, methods, independent variables, dependent measures, and other experimental features. As we describe later, we produced a series of mini-trials for our studies, and went to great lengths to incorporate a variety of realistic features into those trials and subsequent mock juror deliberations.

The law expects witnesses to present, accurately and completely, relevant information of which they have personal or expert knowledge. Jurors, in turn, are expected to understand the testimony and to assign it appropriate weight in deliberations when they integrate it with other relevant evidence. But what factors affect how a juror will understand and weigh complex forensic identification expert testimony? For example, will jurors assume that experienced experts are more trustworthy (or less trustworthy) than less experienced ones? Will they assume that forensic results that are the product of highly sophisticated technology are more probative than equivalent results that emerge from less sophisticated technology? Will they assume that forensic methods that have been scientifically tested are more reliable than those that have not been tested? We suspected that matters might become more complicated when forensic methods have some potentially desirable properties (e.g., high technology methods that have been scientifically tested) but not others (e.g., expert is not experienced). Under these conditions, how will jurors weigh the forensic evidence that they hear?

2

We also set out to examine whether the forensic expert's expressed level of subjective certainty and his/her degree of willingness to concede the possibility of error could impact jurors' views of the strength of the forensic evidence. Consider subjective confidence levels. On the one hand, and for obvious reasons, extreme levels of professed certainty (e.g., "I'm 100% certain") might bolster a factfinder's confidence in the forensic result. But if such certainty seems unwarranted or comes across as hyperbole or bias, jurors might treat those forensic results as *less* reliable than they otherwise would.

Finally, we considered the issue of exculpatory evidence offered by the defense. Defendants sometimes vigorously challenge damaging forensic evidence either by drawing attention to weaknesses in the forensic evidence or by pointing to other forensic evidence that is favorable to the defendant's case. Little is known about how jurors respond to such concessions or counter-evidence when it is elicited from prosecution experts.

In sum, we conducted a series of experimental studies to gain insight into the effects of a variety of real world factors on how jurors think about and use expert forensic identification evidence.

Project Methods and Data Collection

The research we conducted used a "mock-trial paradigm" in which we created fictitious criminal trials and asked a sample of laypeople to serve as "jurors" in the cases. As a starting point, we created two completely different cases: one, a rape case, involved a defendant who allegedly bit the victim's shoulder, leaving a bitemark behind. The second case, an attempted murder, included fingerprints left behind on a mug that the suspect had touched during the crime. Both trials were strategically designed such that the strength of the forensic evidence – bitemark or fingerprint – would be the key piece of evidence on which jurors would base their decisions.

3

Within each case, we changed various aspects (described above) about the expert testimony that accompanied the forensic evidence.

Each of these mock trials was tested in two different ways: A preliminary test using a US-wide sample of online mock-jurors, and a formal test using groups of in-person jurors who deliberated in order to reach a verdict. The number of mock jurors who participated in this research along with the variables and conditions tested in each experiment is outlined in Table 1 below.

Table 1 - Variables, Conditions, and Participant Numbers across Three Experiments

Study	Variables Tested	# of US-Wide Online Jurors	# of In-Person Jurors
1	- Level of Technology (High vs Low) - Experience of Expert (High vs Low) - Scientific Testing of Technique (Tested vs. Not)	441	318
2	- Strength of Exculpatory Evidence (Strong vs Weak) - Subjective Certainty of Expert (High vs Low)	222	308
3	- Scientific Testing of Technique (Tested vs Not) - Concession that an Error is Possible (Yes vs No)	241	250
Total		904	876

Creation of the Mock Trials

One of the goals of this research was to examine forensic expert testimony using a highly realistic experimental method – videotaped "trials," loosely based on trial transcripts in cases that included forensic science testimony, viewed by real "jurors." However, before undergoing such an expensive task, we first created a written version of both trials so that we could collect some preliminary data on laypeople's perception of the trial and evidence contained therein. Using a custom online survey system, we presented samples of demographically-diverse US citizens with a two-page summary of either our assault / bitemark or attempted murder / fingerprint case. These summaries contained background information, expert testimony from a forensic scientist, and cross-examination of that expert. Within the expert testimony were the manipulations we

4

hoped to employ in the trial videos. Following the trial summary were measures of the perceived credibility of the expert, credibility of the evidence, case strength, and various other items designed to help us understand how our participants were comprehending the case. Over the course of several months, 904 mock jurors were obtained via Amazon Mechanical Turk (an online system that, among other things, provides survey respondents to researchers) with each of the jurors viewing and responding to one randomly-chosen combination of our different trials and variables.

The findings from our preliminary online studies (details of which are in our Results section below) indicated that our trial materials were being accurately understood and perceived in a neutral and unbiased fashion. Further, "manipulation checks" on our variables indicated that the participants were indeed attentive to our manipulations (e.g., high-tech methods were seen as more sophisticated than low-tech methods; our "experienced" expert was seen as more experienced than our "inexperienced" expert). As such, we proceeded to create a set of highly-realistic trial videos.

We began by adapting the trial summaries used in the preliminary research into the form of a realistic trial transcript, including dialogue from the judge, opening statements, direct and cross exam of three witnesses, closing statements, and final jury instructions. To further enhance the reality of the trial, objections, additional questions, and re-direct examinations were also included, and we ensured that the attorneys' closing statements referred back to any of the manipulations included in the forensic expert testimony. These transcripts were then compared to actual trial transcripts and reviewed by attorneys to verify their realism. Once the transcripts were complete, we obtained the use of a mock-courtroom at the Arizona State University College of Law, and cast several law faculty members with trial litigation experience as the

attorneys, a forensic science expert as our key witness, and other faculty members and age-appropriate student actors as a judge, victim, and police officer witness. The screenshots below are from the actual trial video and depict the judge, forensic expert, defense attorney, and prosecutor (clockwise from upper left).

Over the course of several days, we filmed the footage necessary to create 32 different trial videos – each corresponding to a specific combination of the base trial (attempted murder fingerprint / rape bitemark) and the various manipulated variables included in the expert testimony (e.g., experienced expert vs. inexperienced expert). Table 1 above lists the experimental manipulations and conditions within each of the experiments. This footage was then professionally edited to create our final set of trial videos. These videos ran between 35-40 minutes, depending on the specific variation.

Experimental Procedure

As noted in the various interim reports submitted for this project, we collected data in two different settings. At the outset of the project, we partnered with the City of Phoenix, AZ Municipal Court to recruit our participants from among jurors who had been excused from service. During this time, a team of research assistants was stationed at the jury waiting room each day and signed up individuals who were interested in participating. Once a group of 6-8 people had signed up, that group was led to a room in a nearby building where they were seated around a large table. These jurors were then given some basic introductory instructions, after which a randomly chosen trial video from among the 32 variations was played on a large screen. During the time that the trial video was playing, the participants were allowed to take written notes as allowed in the Arizona courts. Once the trial video was complete, a preliminary questionnaire was administered in order to obtain each juror's initial impression of the case. When all jurors had completed that questionnaire, a jury verdict sheet was provided to the group and the jurors were asked to deliberate and reach a unanimous verdict. The jurors were not given a specific time limit but if the jury had not reached a verdict after 30 minutes we considered the jury hung (this occurred in only a small fraction of the cases – most juries decided within 15-20 minutes). After a verdict was reached we administered a final questionnaire that again measured the jurors' perceptions of the case as well as various attitude measures about science and personal demographic information. By collecting impressions of the case both pre- and post-deliberation, we are able to ascertain how the process of deliberation influences the jurors' understanding and impressions of the forensic expert testimony.

After nearly a year of collecting data at the Phoenix Municipal Court, it became apparent that there simply weren't enough jurors available to participate in order to complete this project in a timely manner—during some months, only a single jury was able to be assembled. In

7

consultation with NIJ, we turned to an alternate plan: We partnered with a jury consultant and a market research firm to recruit individuals from the Phoenix, AZ metropolitan area that would match the demographic distribution of typical jury respondents. Once recruited, these individuals were scheduled into groups of 6-8 that would attend an experimental session at one of two locations in the Phoenix area. From that point, the procedure was identical to that described above. As it worked out, the participants in our first experiment were nearly exclusively from the Phoenix Municipal court, while those from the other two experiments were exclusively from our jury consultant pool. We conducted some cursory analyses on the demographic variables between the two participant sources and found no significant differences on age, sex, education, or political views (ps .16 - . 59). Because the experimental materials differed between these sources, we are not able to directly compare their responses on our dependent measures.

Project Findings

Experiment 1: Technology, Experience, and Scientific Testing

Our first experiment examined how an expert's experience level, the "tech-factor" of the methods used (high-tech vs low-tech), and the extent to which the expert's methods have been scientifically validated combine to influence jurors' judgments about the forensic evidence. In both the preliminary online sample as well as the actual juror sample, the most powerful indicator of the perceived credibility of the forensic evidence was the experience level of the expert. In fact, our in-person jurors were only minimally swayed by the tech-factor of the expert's methods. Most concerning, however, is the fact that the jurors were not overall affected by whether the expert's methods had been scientifically validated. The only place in which this variable came into play was in that the expert's experience level mattered more when his technique had been scientifically validated, and less when it had not been validated. These

patterns persisted even after deliberation. Interestingly, while the perceptions of the forensic evidence were the primary predictor of the jurors' verdict choices, the expert's experience level did not directly influence verdicts.

Experiment 2: Exculpatory Evidence and Subjective Certainty

Our second experiment examined the forensic expert's subjective level of certainty over his findings combined with the presence or absence of any exculpatory forensic evidence. Across both trials and our online and in-person samples, the expert's subjective level of certainty was the primary determinant of the perceived strength of the evidence. The presence of exculpatory forensic evidence did little to influence the jurors. This pattern appeared in the jurors' individual verdicts, in which a "certain" expert (saying the defendant was definitely the source of the evidence) was over three times as likely to yield a guilty verdict as a less-certain expert (saying the evidence was merely "consistent" with the defendant). Interestingly, deliberation appeared to highlight the importance of the exculpatory evidence: After deliberating, the presence of the exculpatory evidence became nearly as influential on the jurors' judgments as the expert's level of certainty. This finding helps to clarify the important function of deliberation and how an individual juror's impressions of forensic evidence may be swayed by other jurors.

Experiment 3: Scientific Testing and Error Concessions

Our third experiment examined the scientific validity of a forensic technique combined with whether the forensic expert concedes that forensic identifications may (generally speaking) be made in error. In this experiment, we found that the form of evidence used in the trial interacted with our experimental manipulations. While there was again no effect of the scientific testing variable, the forensic expert's error concession had a paradoxical effect between the two trials: When the case involved bitemark examination, the expert's error concession decreased the

jurors' confidence in the forensic evidence (and their resulting verdicts). However, when the case involved fingerprint examination, the error concession actually somewhat strengthened the evidence. This suggests that pre-existing beliefs about the type of forensic evidence being presented influences how jurors react to testimony related to that evidence. This pattern was consistent both pre- and post-deliberation. This finding demonstrates the importance of testing multiple forms of forensic evidence when evaluating how that evidence may be used by legal decisions makers.

Additional Data Notes

It is important to note that our original proposal included five rather than three experiments. At the advice of NIJ, the scope of this project was narrowed to only the first three experiments in order to conduct a more thorough examination of those variables and to ensure that the project could be completed in a timely manner. We do have some preliminary online data gathered from the fourth and fifth experiments that we can provide upon request. In addition, while the results reported here pertain to the main goals of our project, we do plan to conduct other analyses to more deeply examine secondary topics such as individual / demographic differences in decision making, methodological differences (e.g., comparing our online experiments to the in-person counterpart), and attitude shifts that emerged during deliberation.

Recommendations for Practice

Our research has obvious implications for criminal justice policy and practice. At a general level, it provides an empirical basis for reforms in forensic science training and practice. The findings suggest that jurors tend to over-value some attributes of forensic science expert testimony and under-value other aspects. The most persistent finding is that jurors rely

10

heavily on the "experience" of the testifying expert and the expert's asserted certainty in his conclusions. By contrast, jurors appear insensitive to variables indicative of the actual validity and accuracy of a forensic technique, including testimony concerning whether a technique has been empirically validated (or not) and whether evidence is offered that the technique is known to make errors (or not).

Those findings raise some concerns: Nothing could be more vital to a scientific claim than whether or not it has been empirically validated, and yet jurors appear to be insensitive to this variable. Instead, jurors are more interested in and sensitive to a forensic scientist's experience such as the number of cases in which he or she has appeared. This is troubling because there is no evidence that more experienced forensic scientists are any more competent or accurate than less experienced ones. Indeed, research on expertise across a variety of domains indicates that greater experience generally does not even correlate with greater judgmental accuracy. Experience does seem to predict confidence, but not accuracy. Thus, it appears that jurors who hear forensic science testimony may be attending to irrelevant, correlated cues (examiner experience and confidence) while ignoring a key cue (scientific validity).

Furthermore, the finding that the examiners' expressed certainty in a finding influences jurors' beliefs about the facts of the cases calls to mind another line of research. That research has found that factfinders who are given the same underlying (probability) data that the expert possesses are reluctant to find liability (in a civil case). But when the expert is permitted to assert his conclusion (asserting that the defendant as the source of the evidence pattern) then jurors are far more willing to find liability – even though the expert knew nothing the jurors did not already know.

Our findings suggest that jurors many not always assess scientific evidence on the meaning and strength of the evidence, but lean heavily on the people telling them what lesson to take from the evidence. If jurors tend not to appreciate the actual strengths and actual weaknesses of expert testimony, they will not accurately evaluate the testimony they are presented, and they will not draw the best value from such evidence. In the worst circumstances, they might reach incorrect verdicts.

Forensic science organizations and broader scientific and legal authorities might choose to take note of our findings, in part, to identify standards for the conduct and presentation of expert witness testimony at trial. These could inform those with oversight authority, contribute to the training of experts regarding their courtroom presentations, and ultimately contribute to more meaningful trial presentations and jury decisions. Similarly, courts, legislatures, rules committees, and scientific bodies charged with making policy recommendations (e.g., future National Research Council panels) might also take into account parts of our research when developing admissibility guidelines for trial judges pertaining to contested identification evidence. Or our research might contribute to the development of judicial instructions on the meaning and limits of identification evidence.

Another example might be that our findings underscore the importance of the judicial gatekeeping function as identified in *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993). If jurors have great difficulty evaluating the quality of the *evidence* offered by forensic scientists at trial, and instead allow themselves to be led to conclusions by superficial and ultimately uninformative attributes of the *witness*, then the court's gatekeeping duties are that much more important to ensuring a fair trial and a correct verdict.