Evaluating Negative Forensic Evidence: When Do Jurors Treat Absence of Evidence as Evidence of Absence?

William C. Thompson, Nicholas Scurich, Rachel Dioso-Villa, and Brenda Velazquez*

Two jury simulation experiments tested participants' sensitivity to variations in the probative value of a piece of negative forensic evidence: failure to find gunshot residue (GSR) on a defendant alleged to have fired a gun. Experiment 1 found that if no GSR was detected, juries ($N = 115$) of undergraduates were appropriately less likely to convict a criminal defendant when the probability of detecting GSR was high than when it was low. Participants were unaffected by contextualizing expert testimony that emphasized either the value of the negative evidence for making inductive inferences or that the negative evidence was inconclusive for making deductions. Experiment 2 used a sample of venire jurors ($N = 420$) and manipulated the probability of detecting GSR (0 percent, 50 percent, 60 percent, 90 percent, or 100 percent) given that a gun was fired. Consistent with the first experiment, venire jurors were more likely to convict when the probability of detection was 0 percent or 50 percent than when it was 100 percent, but verdicts did not differ between the middle groups. This pattern of results suggests that jurors may evaluate negative evidence according to a fairly crude metric—giving it no weight if the probability of detection is zero, a great deal of weight if the probability of detection is 100 percent, and moderate weight if the probability of detection is somewhere in between.

Sherlock Holmes: [Consider] the curious incident of the dog in the night-time.
Inspector Gregory: The dog did nothing in the night-time.
Holmes: That was the curious incident.

A. Conan Doyle's, *The Silver Blaze*

---

*Address correspondence to Nicholas Scurich, University of California, Irvine, 4213 Social & Behavioral Sciences Gateway, Irvine, CA 92697-7085; email: nscurich@uci.edu. Thompson is Professor, Department of Criminology, Law and Society, University of California, Irvine; Scurich is Associate Professor, Department of Psychology and Social Behavior and Department of Criminology, Law and Society, University of California, Irvine; Dioso-Villa is Senior Lecturer, School of Criminology and Criminal Justice, Griffith University, Queensland, Australia; Velazquez is Director of Planning and Analysis for the Office of the Bronx District Attorney, Bronx, NY.

The research reported here was supported by National Science Foundation Grant No. SES-0617672 and by the Center for Statistics and Applications in Forensic Evidence (CSAFE). The first author thanks the Isaac Newton Institute for Mathematical Sciences for its hospitality during the program Probability and Statistics in Forensic Science, which was supported by EPSRC Grant No. EP/K032208/1. The authors also thank Mr. Alan Carlson, Chief Executive Officer and Jury Commissioner of the Superior Court of Orange County, California, and his staff, for their assistance in completing this research, and Nicci Bowman Fowler for assistance in data collection.

I. Introduction

Does the absence of a defendant's fingerprint on an alleged murder weapon mean he never touched the weapon? Does the absence of detectable drug metabolites in a person's blood mean the person was drug-free? Questions like these concern the value of *negative evidence*. Negative evidence may arise from the failure of an investigation to produce findings consistent with a particular hypothesis about what occurred. It may also arise from the nonoccurrence of an event that was likely to occur if a particular hypothesis is true. Negative evidence is generally admissible in jury trials when it is relevant to a pertinent issue in a case, although courts may exclude it if they find it more prejudicial than probative (e.g., *United States v. Hoffman* 1992).

In this article we will discuss two issues related to negative evidence. One issue is the probative value of such evidence: When and to what extent does the *failure to find evidence* of an event help prove that the event did *not* occur? The second issue is the effect of such evidence in jury trials: When and to what extent do jurors treat negative evidence as probative? In distinguishing the two issues we are, obviously, acknowledging a possibility that jurors will assess negative evidence incorrectly, giving such evidence more (or less) weight than it deserves.

It is important to distinguish *negative evidence* from *missing evidence*. Suppose that a defendant is accused of shooting a victim with a handgun. If the police fail to test for the presence of gunshot residue (GSR) on the defendant, then we will say that GSR evidence is *missing*. If the police test for GSR and do not find it, we will say the evidence is *negative*. According to some commentators, the CSI television series has conditioned jurors to expect forensic science evidence in virtually every case (Podlas 2006; Shelton 2009; Shelton et al. 2006; Tyler 2006). There has been much discussion of whether this "CSI effect" exists and whether it has made jurors reluctant to convict when forensic evidence is *missing* (Tyler 2006; Cole & Dioso-Villa 2007; Schweitzer & Saks 2007). Our focus here, however, is not on cases in which the parties have failed to conduct an inquiry or examination into a pertinent issue, but on cases in which such an inquiry has been conducted and the findings were negative.

Negative evidence, as we use that term here, is also distinguishable from evidence designed to explain or "contextualize" negative or missing findings. According to one legal scholar, concern about "CSI infection within the jury" has led many prosecutors to pursue a strategy of "defensive prosecution" in which they present expert testimony to explain "why certain types of tests were not done," and why other tests were negative, in order to "curb unrealistic expectations and misinformation regarding forensic science and its capabilities" (Lawson 2009). Some commentators have referred to this explanatory testimony as *negative evidence* (Lawson 2009:168; Cole & Dioso-Villa 2011). To avoid confusion, we will reserve the term *negative evidence* for evidence of nonoccurrence or nondetection, such as the failure of a toxicology test to find evidence of drugs. Following Jenkins and Schuller (2007), we will use the term *contextualizing evidence* to describe expert testimony designed to provide an explanatory context for negative or missing evidence.

How do jurors respond to negative evidence? We know that people find it difficult to draw inferences from the absence of information (Einhorn & Hogarth 1978; Ward &

Jenkins 1965). As Hearst explained, "human beings and other animals have trouble using the mere absence of something as a basis for efficient and appropriate processing of information" (1991:432). In studies of operant conditioning, for example, pigeons learn more quickly that they can obtain food by pecking a lever when the availability of food is signaled by the presence of a blue light than when it is signaled by the absence of the blue light—a phenomenon called the "feature positive effect" (Jenkins & Sainsbury 1969). Using slightly different stimuli (symbols rather than lights) and more species-appropriate rewards, Newman et al. (1980) observed the same effect in undergraduates.

But studies in this area typically require participants to realize without much guidance the significance of events that did not occur or evidence that was not found. Negative evidence may be easier to appreciate in the context of a jury trial because testimony about it typically recounts the investigation that yielded the negative findings. An expert might testify, for example, that a test for gunshot residue was conducted and that none was found. It may well be easier to see the implications of the GSR that was not found (after being told that someone looked for it) than to appreciate the implications of the dog that did not bark, particularly if no one has drawn attention to the dog.

Jenkins and Schuller were the first researchers to study reactions to negative evidence in a jury simulation study. Mock jurors evaluated a case in which the defendant was accused of administering a "date-rape" drug to a female acquaintance and then raping her. The defendant denied giving the woman a drug and claimed his sexual encounter with her was consensual. The key *negative evidence* came from toxicology tests on the woman's blood and urine, which failed to detect evidence of a date-rape drug. The mock jurors gave considerable weight to this negative evidence, but only when it was presented *without* contextualizing testimony. When an expert testified that the drugs in question cannot always be detected by toxicology tests, and, on that basis, suggested that the negative results are "not conclusive either way" on the question of whether the woman ingested the drugs, jurors gave no weight to the negative evidence. As Jenkins and Schuller explained, "the expert witness completely negated the impact of the negative forensic report on participants' decisions" (2007:377).

Jenkin and Schuller's findings are groundbreaking and provocative, but (as they recognized) their study had an important limitation. It is not possible to determine, based on the materials they presented, how much weight the negative evidence deserved. So we do not know whether the mock jurors gave the negative evidence too much or too little weight. We cannot tell, therefore, whether the expert's contextualizing testimony was beneficial because it persuaded jurors to disregard evidence that had little or no probative value; or was prejudicial because it persuaded jurors to disregard probative evidence that supported the defendant.

To assess the probative value of a negative forensic test, one must have information about the probability of positive and negative findings under the relevant hypotheses (Lempert 1977; Lyon & Koehler 1996).[1] In most cases, the key variable affecting the

[1]For example, the value of a negative drug test for proving the woman did not ingest drugs depends on the ratio of two conditional probabilities: the probability of a negative result if the woman did not ingest drugs and the probability of a negative result if she did. The probative value of the negative evidence varies directly in proportion to this ratio (which is known as a likelihood ratio).

value of negative evidence is what we will call the probability of detection. In the drug case, it would be the probability of obtaining a positive test result if the woman ingested drugs.[2] Statisticians sometimes call this factor the *sensitivity* of the test. It is also necessary to know the probability of a false positive, which (in the drug case) would be the probability of a positive test result if the woman did not ingest drugs. Because the probability of a false positive is likely to be low for most forensic tests, the probability of detection will typically be the most important variable affecting value of negative evidence.

Other factors being equal, the probative value of a negative test will vary directly with the probability of detection. If the probability of detecting a date-rape drug is zero, then a negative drug test has no value because the test would be negative even if the woman did ingest drugs. If the probability of detection is one (100 percent), then the negative drug test is highly probative: it essentially proves that the woman did not ingest drugs. If a juror truly believed that the probability of detection was 100 percent, and that drugs were not detected, then whatever the juror's prior estimate of the odds the woman ingested drugs, the juror should reduce that estimate to zero.

It is important to realize that negative evidence may have substantial probative value even if the probability of detection is less than one. Suppose, for example, that the probability of detection for the drug test is 0.50 (50 percent) and the false positive rate is zero. In this case, the occurrence of a negative drug test is twice as likely if the woman did *not* ingest drugs than if she did. Whatever a juror's prior estimate of the odds the woman ingested drugs, the juror should reduce that estimate *by half* in light of this evidence. An exposition of this conclusion appears in Appendix A.

When the probability of detection is less than 100 percent, forensic scientists sometimes dismiss negative evidence as unimportant. Failure to find the defendant's fingerprint on a murder weapon, for example, is said to have little or no evidentiary value because there are many possible reasons (other than innocence) for a negative finding: the defendant might have worn gloves, wiped the weapon, and so forth (Nielson 2005). When discussing such cases, forensic scientists often cite a well-worn aphorism: "absence of evidence is not evidence of absence" (Starrs 2006).

This broad dismissal of negative findings is similar to the treatment of negative findings in hypothesis testing. Scientists have traditionally approached truth using a deductive logic that involves ruling out hypotheses (Popper 1959). If an investigation produces findings that cannot be accounted for under a null hypothesis, then the null hypothesis is rejected (which has the effect of supporting alternative hypotheses). But what if an investigation fails to produce evidence that can rule out null hypothesis? As all beginning science students are taught, "negative" findings of this type do not prove the null hypothesis is true; nor do they prove any alternative hypothesis is false. There generally are many possible reasons for negative findings, including limitations or

---

[2]We will assume for this discussion that detection and nondetection are mutually exclusive and exhaustive events, and hence that the probability of nondetection is the complement of the probability of detection. Hence, if the probability of detection is high, then the probability of nondetection is low, and vice-versa.

inadequacies in the investigation. Consequently, negative findings have little or no value in the deductive, hypothesis-testing framework.

But evidence that has little value for ruling out a hypothesis may have considerable value for making other inferences. In daily life, and in the legal system, people typically evaluate evidence through *induction* rather than *deduction* (Hacking 2001). In an inductive framework, the absence of evidence can indeed be evidence of absence, even if the probability of detection is less than one. To the extent evidence is *likely* to be found if a particular event occurred, the failure to find that evidence supports an inference that the event did not occur.

Forensic scientists who give contextualizing testimony about negative findings sometimes appear to confuse the deductive logic of hypothesis testing with the inductive logic of evidence evaluation in the courtroom. Indeed, this confusion may be the basis of the common claim that "absence of evidence is not evidence of absence." The contextualizing testimony in the Jenkins and Schuller (2007) study was based on the testimony of a forensic scientist in the Canadian case of *Regina v. Alouache* (2003), where the expert explained that there are a number of reasons why a toxicology test may have failed to detect drugs in the woman's blood and urine and, on that basis, concluded that the test "was not conclusive one way or the other" on the issue of whether the woman had been drugged. The issue the jury needed to assess, however, was not whether the test was "conclusive" but whether it was probative. If the test was incapable of detecting drugs that the defendant had administered to the woman, then (as discussed above) the negative test result had no probative value. In that case, the prosecution should have objected to its admission in evidence and the judge should have excluded it on grounds of irrelevance. If there was at least some chance that the test could detect drugs the defendant had administered, however, the failure to detect those drugs supports the hypothesis that the woman was not drugged, even if the evidence could not conclusively rule out the alternative hypothesis. In that case, the contextualizing testimony may well have confused and misled the jury by persuading it to give no weight to evidence that actually supported the defendant's claim of innocence. Epistemologist Ian Hacking (2001) has noted that people can miss or ignore the inductive value of evidence when evaluating it in a deductive framework. That may well be a danger when forensic experts "contextualize" negative findings the way the expert did in *Alouache*.

Schuller et al. (2013) reported two additional studies of undergraduate mock jurors' reactions to a negative drug test in a case where the defendant was accused of administering a date-rape drug. Because the probability of detection decreases over time, they varied whether the test was conducted relatively soon after the complainant was alleged to have ingested the drug (five hours) or after a longer delay (12 hours), when the probability of detection would be lower. In Study 1, the time-delay variable had no effect on the weight jurors assigned to a negative drug test. The researchers suggested, however, that the expert's testimony may not have drawn "sufficient attention to the relationship between reporting time and the scientific veracity of the report." In Study 2, where the expert placed greater emphasis on this relationship, male jurors gave substantially more weight to the negative findings when the probability of detection was

higher (five-hour delay) than when it was lower (12-hour delay), whereas female jurors gave little weight to the negative findings in either condition. The researchers suggest that female jurors may have been more sympathetic to the complainant and may, as a result, have been skeptical of the negative evidence in either case, given that even in the short-delay condition the probability of detection was less than certain. A limitation of both studies is that it was impossible to estimate precisely the probability of detection at the two time intervals, and hence it is unclear how much this variable *should* have influenced jurors' judgments.

A similar limitation arose in another study of negative evidence. Eerland and Rassin (2012) examined the reactions of Dutch undergraduate law students to evidence that criminal investigators, after identifying a suspect, had searched unsuccessfully for additional incriminating or exculpatory evidence. They found that the law students gave no weight to the negative findings of these investigations. Because it is not clear what the probability of detection might have been, however, it is difficult to assess whether this reaction was normatively appropriate.

The experiments reported here explore mock jurors' reactions to negative evidence in a context in which information about the probability of detection is provided. We varied the probability of detection experimentally to see how jurors respond to this key variable. We hypothesized that the weight given to a negative finding would increase as the probability of detection increases. We also varied the nature of the contextualizing testimony that was offered to explain the negative finding. In the *deductive condition*, the expert emphasized that the negative evidence could not conclusively rule out a particular hypothesis. In the *inductive condition*, the expert focused on the value of the negative evidence for making inferences about the probability of the hypothesis. We expected that jurors would give less weight to negative evidence when it was accompanied by the deductive contextualizing testimony. We also predicted an interaction—we expected that participants' judgments would be more sensitive to the probability of detection when they received inductive contextual testimony than when they received deductive contextual testimony.

II. Experiment 1

A. Method

1. Participants and Procedure

Undergraduates ($N = 573$) recruited from a university human subjects pool participated in groups of four to seven. Nineteen percent of the participants were male and 81 percent were female, ranging in age from 18–52, with a median of 20. They read an eight-page description of the evidence in a hypothetical criminal case, including transcripts of key testimony by three witnesses, closing arguments by the prosecutor and defense lawyer, and the judge's instructions on the legal standard for conviction. They then responded individually to a one-page questionnaire that asked their "initial reactions to the case" based on "what you think right now." The questionnaire asked them to rate

the strength of the case against the defendant on a seven-point Likert scale and to indicate whether they thought they would vote guilty or not guilty if they were a juror in the case.

At that point, each group was left alone to deliberate on the case for up to 20 minutes (or until they agreed on a verdict). They were informed that the experimenter would be monitoring their discussion from behind a one-way window. After deliberation, the experimenter asked participants to respond individually to a postdeliberation questionnaire, which began with the same questions as the initial questionnaire, but also included questions designed as manipulation checks that tested recall of key facts in the case as well as demographic and background questions about participants. Upon finishing this second questionnaire, they were debriefed, thanked, and dismissed. The entire procedure took approximately one hour.

2. Experimental Design and Materials

The experimental materials described a hypothetical case in which a young man was charged with murdering the owner of a grocery store. An eyewitness heard the grocer arguing with a man, heard a shot fired, and saw a man running from the scene. About eight minutes later, the police apprehended the defendant, who matched the description provided by the eyewitness. The defendant was running down a street about a mile from the crime scene when police stopped him. He was not carrying a gun, but a handgun identified as the murder weapon was found in a trash can near the crime scene. The defendant was taken into custody and his hands and clothing were tested for the presence of gunshot residue (GSR). The test was negative; no GSR was detected. The defendant was nevertheless prosecuted based on testimony of the witness who saw the man running from the crime scene as well as a second witness who identified the defendant as the man who ran past him about a block from the crime scene just after the shooting. The prosecution also established that the defendant had previously been employed by the grocer to stock shelves and make deliveries but had been fired several weeks before the crime for stealing. The defendant denied any involvement in the crime and claimed he had been running near his home for exercise when he was apprehended.

After reporting the negative GSR test, a forensic scientist described a study that the laboratory had conducted to assess its ability to detect GSR from the murder weapon on the hands and clothing of a shooter. The laboratory asked 10 volunteers to fire the gun and then to engage in vigorous exercise by running around a track for eight minutes to simulate the defendant's alleged actions following the shooting. The lab then attempted to collect GSR from the hands and clothing of each of the volunteers. Each volunteer wore the same type of clothing that the defendant was wearing when apprehended.

In a 3×2 between-subjects factorial design, we varied the reported results of the GSR detection study: GSR was detected on 100 percent, 50 percent, or 0 percent of the volunteers. We also varied the forensic scientist's interpretation of the defendant's

negative GSR test (deductive or inductive interpretation). In the deductive condition, the expert claimed that the negative test was uninformative:

> Q: Does a negative GSR test prove that a person did *not* fire a gun?
> Expert: No. A negative test doesn't prove anything one way or the other. There are too many possible explanations for a negative test. Because you cannot rule out all the other explanations, you cannot say for sure that the person didn't fire a gun. It is a basic principle of science. In order to prove a theory you have to rule out all the alternative theories.
> Q: In your opinion, does the negative GSR test on Benjamin Everett prove that he was not the man who shot Mr. Kim?
> Expert: No. The negative test proves nothing because there are other possible explanations for the negative result.

In the inductive condition, the expert acknowledged that the negative test has probative value, but claimed it was not definitive:

> Q: Does a negative GSR test prove that a person did *not* fire a gun?
> Expert: It reduces the likelihood that the person fired a gun. If you expect to find GSR if the person fired a gun, and you don't find it, then you should be less confident that he fired a gun. It is a basic principle of science. If you get results that are unlikely if a hypothesis is true, then you should be less confident that the hypothesis is true.
> Q: In your opinion, does the negative GSR test on Benjamin Everett prove that he was not the man who shot Mr. Kim?
> Expert: It may reduce the likelihood that he fired a gun, but it does not rule out that possibility.

The complete experimental stimuli appear in Appendix B.

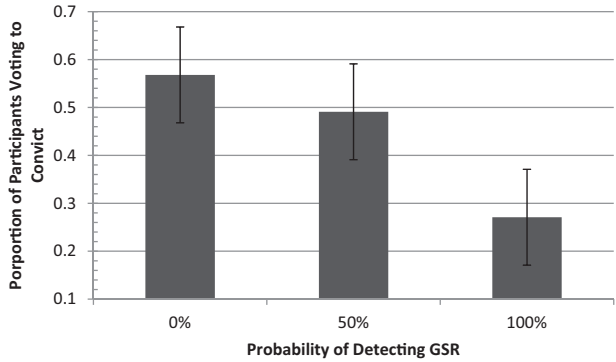
3. Manipulation Checks

The final questionnaire included a multiple-choice item designed to check whether participants correctly recalled the percentage of men in the GSR detection study who had detectable levels of GSR on their hands and clothing. More than 90 percent of participants chose the correct percentage, which indicates that this aspect of the evidence was salient and widely comprehended, and hence that the percentage manipulation was successful.

The majority of participants (96.3 percent in the inductive condition; 76.3 percent in the deductive condition) responded correctly to a second item that asked participants to choose which of four statements most closely reflected the forensic expert's interpretation of the GSR test results. This finding indicates that most participants correctly recalled the expert's statements about his interpretation of the negative GSR test, and hence that the interpretation manipulation was also successful.

The results reported below do not change depending on whether or not participants who responded incorrectly to the comprehension questions are included in the analyses. Therefore, we elected to the include all participants in the analyses to increase the statistical power of the inferential tests and for reasons of ecological validity, as actual jurors are not given comprehension check questions nor are they excluded from deliberation for failing to correctly answer such questions.

Figure 1: Proportion of participants within juries voting to convict as a function of the probability of detection. Note that error bars reflect 95 percent CIs.



B. Results

A total of 115 juries deliberated. One-third of the juries ($n = 39$) unanimously voted to acquit the defendant whereas less than one-quarter of the juries ($n = 26$) unanimously voted to convict the defendant. As noted, the size of each jury varied from four to seven participants with a median of five (IQR = 2) and mode of five participants per jury. The small number of participants within some of the juries (e.g., four) precluded the use of hierarchical models to analyze the data (Tabachnick & Fidell 2013). Consequently, an alternative approach was utilized to analyze the data: examining the proportion of jurors voting to convict within each jury. The proportion of jurors voting to convict within each jury was calculated by taking the number of jurors voting to convict and dividing it by the total number of jurors within the jury. Thus, values could range from 0 to 1. Note that this approach leaves the unit of analysis at the jury (group level), which is appropriate since individuals interacted within groups and thus created interdependencies among their responses.

A 2 × 3 ANOVA with probability of detection and inductive versus deductive interpretation as the independent variables and percentage voting to convict as the dependent variable detected a significant main effect for probability of detection ($F(2, 114) = 5.91$, $p = 0.004$, $\eta_p^2 = 0.098$). The main effect for inductive versus deductive was not significant ($F(1, 115) = 2.04$, $p = 0.156$), nor was the interaction ($F(2, 115) < 1$). The main effect for the probability of detection is plotted in Figure 1.

As is apparent, the proportion of jurors who voted to convict decreased as the probability of detection increased. Post-hoc comparisons (Bonferroni) revealed that the difference between the 0 percent and 100 percent conditions was statistically significant ($p = 0.004$), as was the difference between the 50 percent and 100 percent conditions ($p = 0.047$), but the difference between the 0 percent and 50 percent conditions was not significant ($p = 1.0$).

We next examined perceptions of the strength of the case against the defendant (i.e., the strength of the case against the defendant rated on a seven-point Likert scale), which potentially could be more sensitive to the experimental manipulations than binary verdicts. Because deliberation may have created intraclass correlations within juries, we aggregated the pre- and postdeliberation data by jury and performed a 2 × 3 analysis of variance on the mean strength of case rating for each jury (effectively treating jury as the unit of analysis), and found similar results. For predeliberation ratings, we found a main effect for probability of detection, $F(2,114) = 6.38$, $p = 0.002$, $\eta_p^2 = 0.105$, but no significant effect for inductive versus deductive interpretation ($F < 1$), and no significant interaction ($F(2,114) = 1.12$, $p = 0.33$). For postdeliberation ratings, we again found a significant main effect of probability of detection, $F(2,114) = 3.16$, $p < 0.05$, $\eta_p^2 = 0.055$, with no effect of interpretation and no interaction (both Fs < 1). Post-hoc comparisons found that mean postdeliberation ratings in the 0 percent condition were significantly higher than in the 100 percent condition ($p < 0.05$), while mean estimates in the 50 percent condition failed to differ significantly from those in the 0 percent and 100 percent conditions (both ps > 0.05). Thus, the main effect for the probability of detection holds whether one examines binary verdicts or the strength of case ratings.

Table 1 displays the group mean ratings of the strength of case in the six experimental conditions, both before and after group deliberation. In general, and as expected, the strength of case ratings decreased with increases in the probability of detection in the GSR study.

Table 1: Mean Pre- and Postdeliberation Strength of Case Ratings (with Standard Deviation) Decomposed by Experimental Condition

	Percent with Detectable GSR			
Contextualizing Testimony	*0%*	*50%*	*100%*	*Total*
Predeliberation				
Deductive				
Mean strength of case	5.23	4.90	4.85	5.01
(sd)	*(0.59)*	*(0.63)*	*(0.43)*	*(0.58)*
Inductive				
Mean strength of case	5.14	5.05	4.69	4.96
(sd)	*(0.31)*	*(0.61)*	*(0.52)*	*(0.53)*
Total	5.22	4.97	4.78	
	(0.48)	*(0.62)*	*(0.47)*	
Postdeliberation				
Deductive				
Mean strength of case	4.81	4.76	4.31	4.62
(sd)	*(1.14)*	*(1.07)*	*(0.83)*	*(1.03)*
Inductive				
Mean strength of case	4.91	4.72	4.38	4.68
(sd)	*(0.82)*	*(0.92)*	*(0.78)*	*(0.86)*
Total	4.86	4.74	4.34	
	(0.98)	*(0.99)*	*(0.80)*	

C. Discussion

The first question we sought to answer with this experiment is whether people consider the probability of detection (under relevant hypotheses) when evaluating a negative test result. It appears that they do. Strength of case ratings and the proportion of participants voting to convict were both significantly lower in the 100 percent condition, where the probability of detection was high, than in the 0 percent condition, where the probability of detection was low. As logic dictates, participants gave more weight to the negative evidence (which favored the defendant) when GSR was detected on *all* the test subjects who fired the murder weapon than when GSR was detected on *none* of them. When the probability of detection was high, participants clearly treated the absence of (GSR) evidence as evidence of (GSR's) absence—that is, as evidence that the defendant did not fire the gun and was therefore innocent.

Although participants in the 100 perent condition gave considerable weight to the negative GSR test, they obviously did not view it as fully exculpatory because the conviction rate in this condition was not zero. Comments during deliberation indicated that jurors had various doubts about the negative GSR test. They thought there might be reasons for the negative result that were not reflected in the results of the GSR detection study—for example, the defendant washed his hands, changed his clothes, wore gloves; the laboratory botched the test. While they clearly thought the probability of detection was higher in the 100 percent condition than in the 0 percent condition, those in the 100 percent condition did not accept that the chances of detection were really 100 percent.

The second question we sought to answer with this experiment was whether "contextualizing testimony" affects people's reactions to negative evidence. The testimony we presented did not significantly affect the weight given to the negative evidence. We had also expected that inductive testimony might increase sensitivity to the probability of detection, creating an interaction between probability of detection and contextualizing testimony, but the expected interaction was not statistically significant. We can only speculate about the reasons we failed to detect an effect of contextualizing testimony. Although the failure could conceivably have arisen from a lack of statistical power,[3] we see no trend in the data that would support that conclusion. The explanation may simply be that our participants paid more attention to the evidence itself (i.e., to the substance of the negative finding) than to the expert's commentary about how that finding should be interpreted. They may have ignored the expert's guidance on what the finding meant because they had the confidence to draw their own conclusions.

The third question we sought to answer with this experiment was how people evaluate negative evidence when there is a substantial chance of nondetection, as was the case in the 50 percent condition. Because there was only a 50 percent chance of detection, the negative GSR evidence in this condition cannot rule out the hypothesis that the defendant fired the gun. Consequently, this evidence deserves little or no weight in a deductive,

[3]Assuming a medium effect size and Type I error rate of 5 percent, Experiment 1 required more than 35 participants per cell to obtain power greater than 0.80 (Cohen 1992). Note that Experiment 1 had approximately 100 participants per cell.

hypothesis-testing framework. In an inductive framework, however, it should substantially decrease jurors' estimates of the odds the defendant fired the gun. We were interested to see whether estimates in the 50 percent condition would differ from those in the 0 percent condition, which would indicate whether or not jurors were giving weight to the negative evidence. We were also interested in whether the expert's contextualizing testimony would influence the weight (if any) given to negative evidence in this condition.

We found that regardless of the contextualizing testimony or the dependent variable (verdicts or strength of evidence ratings) there were no significant differences between the 50 percent condition and the 0 percent condition, which indicates that jurors gave little or no weight to the negative evidence in the 50 percent condition. While strength of case ratings and conviction rates were consistently a bit lower in the 50 percent condition (hinting that jurors might be giving a bit of weight to the negative evidence that this study lacked the sensitivity to detect), the results do not allow us to conclude that jurors gave any weight at all to negative evidence when the probability of detection (in the GSR detection study) was only 50 percent.

One possible interpretation of the findings of Experiment 1 is that jurors give weight to negative evidence only when the probability of detection is 100 percent. We have doubts about this interpretation, however, because jurors clearly thought the probability of detection was lower than 100 percent in the 100 percent condition, due to factors not taken into account by the GSR detection study, but they nevertheless were influenced by the negative GSR evidence in that condition. But it may be the case that a juror's subjective estimates of the probability of detection must pass some threshold value before the juror takes negative evidence seriously. Perhaps a 50 percent probability of detection did not meet that threshold.

One important limitation of Experiment 1 was that participants were undergraduates rather than actual members of a jury pool. We worried that bright undergraduates at a fairly selective university might be better at drawing conclusions from scientific and statistical evidence than many actual jurors. So we decided to replicate Experiment 1 using a sample drawn from a pool of actual jurors.

III. Experiment 2

A. Method

1. Procedure and Participants

Participants were members of the jury pool at a courthouse in a large, relatively affluent suburban county. On days when this experiment was conducted, the courthouse staff announced to the jurors, as they were about to be released, that they had the opportunity to remain in the jury assembly room to take part in a university study of juror decision making. They were offered $10 as an incentive for participation. As part of the consent process, those who expressed interest in the study were given a brief study information sheet. Those who chose to participate after reading the sheet were given experimental booklets to read and complete. The rate of participation among those being released

varied from day to day, but generally appeared to be about 50 percent. Participants read the experimental booklets and answered questions individually; because no room was available where discussions could take place, this study did not include jury deliberation. Most participants were able to read the materials and complete the questionnaire in 20–30 minutes. At that point, they returned their booklets to the researcher, who answered any questions they had about the study, thanked them, and paid them $10.

Forty-one percent of the participants ($N = 420$) were male and 59 percent were female, ranging in age from 19–77, with a median of 40. As a group they were relatively well-educated: 54 percent had graduated from a four-year college or university and 17 percent also had a graduate degree. When asked to place themselves on a spectrum of political beliefs, the modal answer (35 percent) was "moderate"; 23 percent placed themselves on the "liberal" side of the spectrum and 40 percent on the "conservative" side. About equal numbers were registered Republicans (33 percent) and Democrats (32 percent). A question designed as a manipulation check asked them to report what the expert had said about the results of the GSR detection study by circling either one of 11 percentages (0 percent, 10 percent, 20 percent, . . ., 100 percent) or the words "Don't know." Most participants (87 percent) accurately recalled what the expert had said.

2. Experimental Design and Materials

Experiment 2 was designed to be a replication of Experiment 1 with a different participant population. However, there were several differences in the experimental design. First, given that the inductive versus deductive interpretation manipulation was ineffective previously, it was omitted in this experiment. Second, the stimulus materials were modified so that the GSR evidence would be a more prominent and important part of the case. In the new version, there were only two witnesses: an eyewitness who confidently identified the defendant as the shooter, and a forensic scientist who testified that no GSR particles were detected in samples collected from the defendant's hands and clothing shortly after the crime. As in the previous experiment, the expert also presented the results of a GSR detection study. The independent variable in Experiment 2 was the reported results of this study. There were five experimental conditions in which the expert testified that 0 percent, 50 percent, 60 percent, 90 percent, or 100 percent of men tested under conditions similar to the crime had detectable levels of GSR on their hands and clothing. Although the expert testified that "there are a number of reasons you could get a negative finding,"[4] he did not offer an opinion on whether (or to what extent) the absence of GSR proved the defendant had not fired a gun. Hence, there was no testimony comparable to either the inductive or deductive testimony in Experiment 1. The purpose of manipulating the detection probability is to explore the possibility, raised by Experiment 1, that the detection probability must pass some threshold before jurors will give serious weight to a negative test result.

---

[4]The expert elaborated as follows: "One possibility is that the individual was not near a firearm that was discharged. Another possibility is that the GSR particles were removed before the test through physical activity, hand washing, or weather conditions. Another possibility is that the firearm failed to produce a sufficient quantity of GSR to be detectable by the test. A final possibility could be a problem in the collection process, such as improper use of the collection kit."

Table 2: Strength of Case Ratings and Verdicts (Experiment 2)

	Percent with Detectable GSR				
	0%	*50%*	*60%*	*90%*	*100%*
Mean strength of case	7.47^{a}	7.08^{a}	6.68^{ab}	6.66^{ab}	6.12^{b}
Percentage voting guilty	61.3^{a}	44.0^{b}	42.2^{b}	44.3^{b}	25.6^{c}
(n)	*(80)*	*(84)*	*(83)*	*(79)*	*(82)*

NOTE: Numbers that do not share the same superscript letter differ significantly, $p < 0.05$.

B. Results

Table 2 shows mean strength of case ratings and conviction rates in the five experimental conditions. (In Experiment 2, participants rated strength of case on a 10-point Likert scale.)

A one-way ANOVA detected significant differences between the experimental conditions in the strength of case ratings, $F(4,412) = 4.73$, $p = 0.001$, $\eta_p^2 = 0.044$. Post-hoc comparisons (Bonferroni) found that the 100 percent condition differed significantly from both the 0 percent condition ($p = 0.001$) and the 50 percent condition ($p = 0.033$), but no other differences were significant.

A binary logistic regression model with the experimental conditions as the independent variables and verdict (guilty or not guilty) as the dependent variable was conducted ($\chi^2 = 21.55$, df = 4, $p < 0.001$), and detected a significant main effect for the independent variable (Wald = 20.02, df = 4, $p < 0.001$). The odds of conviction were 4.59 times greater among participants who were told the probability of detection was 0 percent than among those told it was 100 percent (95 percent CI [2.35, 8.97] ($B = 1.52$, Wald = 19.91, $p < 0.001$). Similarly, the odds of conviction were over twice as high among participants told that the detection probability was 50 percent, 60 percent, and 90 percent detection probability compared to those told it was 100 percent detection probability ($\exp(B) = 2.29$, 95 percent CI [1.19,4.41], $B = 0.827$, Wald = 6.09, $p = 0.014$; $\exp(B) = 2.12$, 95 percent CI [1.10, 4.10], $B = 0.750$, Wald = 4.97, $p = 0.026$; $\exp(B) = 2.31$, 95 percent CI [1.89,4.50] $B = 0.838$, Wald = 6.08, $p = 0.014$; respectively). Deviation contrasts revealed that conviction rates in the 0 percent condition (61.3 percent) were significantly higher than in the other four conditions (all ps < 0.05) and conviction rates in the 100 percent condition (25.6 percent) were significantly lower than those in the other conditions (all ps < 0.05), but conviction rates in the 50 percent condition (44 percent), 60 percent condition (42.2 percent), and 90 percent condition (44.3 percent) did not differ significantly from one another.

C. Discussion

This experiment demonstrates that strength of case ratings and conviction rates varied across conditions in a manner that corresponded very roughly to the probative value of the negative evidence: conviction rates were highest in the 0 percent condition, were moderate in the 50 percent, 60 percent, and 90 percent conditions (which did not

significantly differ from one another), and were lowest in the 100 percent condition. This pattern of findings is similar to that in Experiment 1, despite the fact that Experiment 2 used an entirely different participant pool. Notably, this pool of participants—venire jurors—is quite high in external validity, uncharacteristically so for jury research (see Bornstein 1999).

Although the results are in the appropriate direction, assessments of the strength of the negative evidence do not appear to have been well calibrated. Jurors gave as much weight to the negative GSR evidence when the probability of detection in the GSR detection study was 50 percent as when it was 90 percent, but there was a big drop in the conviction rate when the percentage went from 90 percent to 100 percent. In other words, the difference in probability of detection between 90 percent and 100 percent had more of an impact on jurors' verdicts than the difference between 50 percent and 90 percent. It may be that jurors were particularly impressed by the 100 percent detection rate because it seemed to foreclose one possible explanation for the negative findings. When GSR is detected on 100 percent of the test subjects, then jurors must posit the existence of some factor not accounted for in the detection study to explain the negative test results (assuming the defendant is guilty). If GSR is detected on only 90 percent of test subjects, then jurors have a readier explanation for the absence of GSR on defendant—he may simply have been one of the 10 percent for whom GSR was undetectable. The pattern of results suggests that jurors may evaluate negative evidence according to a fairly crude metric—giving it no weight if the probability of detection is zero, a great deal of weight if the probability of detection is 100 percent, and moderate weight if the probability of detection is somewhere in between.

IV. General Discussion

Forensic scientists sometimes argue that negative evidence has no value because "absence of evidence is not evidence of absence" (Starrs 2006; Nielson 2005). We have shown that this argument is misleading. Negative evidence (the failure to find something) is indeed evidence that the thing sought was "absent" to the extent the probability of detection exceeds the probability of a false positive. The claim that such evidence has no value arises from a misapplication of the deductive logic of hypothesis testing in a legal context that requires inductive inferences.

To evaluate negative evidence, triers of fact need information about the probability of detection. If the probability of detection is unknown, the probative value of negative evidence is speculative and cannot be meaningfully assessed. Courts should consider carefully whether negative evidence should even be admitted under such circumstances. Courts should also consider excluding negative evidence when the probability of detection is known to be low. In such cases, negative evidence has little or no probative value but may carry substantial risk of prejudice.

Our findings suggest that "contextualizing testimony" can play an important role in helping the jury understand negative evidence *when it provides information about the probability of detection.* Our results indicate that jurors appreciate the general implications of the

probability of detection and respond in the appropriate direction to such data when evaluating negative evidence. We are skeptical, however, about the value of "contextualizing testimony" like that presented in *Regina v. Alouache* (2003), where the expert stated that negative results are inherently inconclusive *without providing data on the probability of detection.* As Jenkins and Schuller (2007) showed, such testimony can negate the value of negative evidence, leading jurors to give it no weight. While this result would be appropriate if the probability of detection was low, it could cause jurors to undervalue the negative evidence if the probability of detection was moderate or high. Consequently, we believe "contextualizing testimony" should be allowed only when the expert can provide information that helps the trier of fact assess the probability of detection. Testimony that consists solely of aphorisms ("absence of evidence is not evidence of absence") or conclusory statements ("negative evidence is not conclusive either way") may be misleading when the probability of detection is unknown, or when it is high.

Although jurors seem to understand the general implications of the probability of detection for the probative value of negative evidence, Experiment 2 found relatively poor calibration between probability of detection and the weight given to negative evidence. Poor calibration is hardly surprising given people's well-known tendency to make probabilistic judgments according to simplifying heuristics rather than Bayesian calculations (Tversky & Kahneman 1974; Kahneman et al. 1982; Koehler 2001; Koehler & Macchi 2004). Deviations between Bayesian norms and actual human performance have been a rich source of insight into human cognitive process. Future research into the deviations that arise when reasoning about negative evidence could contribute usefully to psychological theory. Research on this topic could also provide insights on constructing persuasive arguments about negative evidence and improving people's reasoning about it.

In future studies it would be interesting (and useful) to study negative evidence in cases where jurors take seriously the claim that the probability of detection is 100 percent. In the experiments reported here, jurors clearly did not accept that GSR was certain to be detected if the defendant fired the murder weapon, even when GSR was found on 100 percent of subjects in the detection study. Whether jurors will view negative evidence as conclusive in cases where they acknowledge that the probability of detection is certain remains unclear (and is worthy of further study). Another suggestion for future researchers is that they ask participants to indicate their subjective assessment of the probability of detection and the probability of a false positive. Information about jurors' subjective assessments of these key variables will make it easier to assess the appropriateness of their judgments, relative to Bayesian norms, than was possible in the experiments reported here. Future research should also test the generality of our findings to other types of negative evidence and other types of cases.

Expert testimony about how to interpret negative evidence is one possible way to improve the calibration of jurors' responses to such evidence. Proposals to allow experts to present "decision aids" that illustrate the implications of Bayes's rule have been debated for many years and raise a number of potential problems (Saks & Thompson 2003; Kaye 1988; Tribe 1971; Finkelstein & Fairley 1970). It is undoubtedly too early to decide, as a matter of policy, whether (and when) such testimony is desirable in cases

involving negative evidence, but it is not too early for researchers to begin examining the issue.

References

Bornstein, B. H. (1999) "The Ecological Validity of Jury Simulations: Is the Jury Still Out?" 23 *Law & Human Behavior* 75.

Cohen, J. (1992) "A Power Primer," 112(1) *Psychological Bulletin* 155.

Cole, S. A., & R. Dioso-Villa (2007) "CSI and its Effects: Media, Juries, and the Burden of Proof," 41 *New England Law Rev.* 435.

—— (2011) "Should Judges Worry About the 'CSI Effect'?" *Court Review: The J. of the American Judges Association* Paper 348. Available at: http://digitalcommons.unl.edu/ajacourtreview/348.

Eerland, A., & E. Rassin (2012) "Biased Evaluation of Incriminating and Exonerating (Non)-Evidence," 18(4) *Psychology, Crime & Law* 351.

Einhorn, H. J., & R. M. Hogarth (1978) "Confidence in Judgment: Persistence of the Illusion of Validity," 85 *Psychological Rev.* 395.

Finkelstein, Michael O., & William B. Fairley (1970) "A Bayesian Approach to Identification Evidence," 83 *Harvard Law Rev.* 489.

Hacking, I. (2001) *An Introduction to Probability and Inductive Logic*. Cambridge, UK: Cambridge Univ. Press.

Hearst, E. (1991) "Psychology and Nothing," 79 *American Scientist* 432.

Jenkins, G., & R. A. Schuller (2007) "The Impact of Negative Forensic Evidence on Mock Jurors' Perceptions of a Trial of Drug-Facilitated Sexual Assault," 31 *Law & Human Behavior* 369.

Jenkins, H. M., & R. S. Sainsbury (1969) "The Development of Stimulus Control Through Differential Reinforcement," in N. J. Mackintosh & W. K. Honig, eds., *Fundamental Issues in Associative Learning*, pp. 123–67. Halifax: Dalhousie University Press.

Kahneman, D., P. Slovic, & A. Tversky (1982) *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge Univ. Press.

Kaye, D. H. (1988) "Plemel as a Primer on Proving Paternity," 24 *Willamette Law J.* 867.

Koehler, J. J. (2001) "When Are People Persuaded by DNA Match Statistics?" 25 *Law & Human Behavior* 493.

Koehler, J. J., & L. Macchi (2004) "Thinking About Low-Probability Events: An Exemplar Cuing Theory," 15 *Psychological Science* 540.

Lawson, T. F. (2009) "Before the Verdict and Beyond the Verdict: CSI-Infection Within Modern Criminal Jury Trials," 41 *Loyola Univ. of Chicago Law J.* 119.

Lempert, R. O. (1977) "Modeling Relevance," 75 *Michigan Law Rev.* 1021.

Lyon, T. D., & J. J. Koehler (1996) "The Relevance Ratio: Evaluating the Probative Value of Expert Testimony in Child Sexual Abuse Cases," 82 *Cornell Law Rev.* 43.

Newman, J., W. T. Wolff, & E. Hearst (1980) "The Feature-Positive Effect in Adult Human Subjects," 6 *J. of Experimental Psychology: Human Learning & Memory* 630.

Nielson, J. P. (2005) "Rebutting the 'No Fingerprint' Defense," 39 *Prosecutor* 32.

Podlas, K. (2006) "The 'CSI Effect': Exposing the Media Myth," 16 *Fordham Intellectual Property, Media & Entertainment Law J.* 429.

Popper, Karl R. (1959) *The Logic of Scientific Discovery*. London: Hutchinson Education.

Regina v. Alouache (2003) CanLLI 9494 (ON S.C.).

Saks, M. J., & W. C. Thompson (2003) "Assessing Evidence: Proving Facts," in D. Carson & R. Bull, eds., *Handbook of Psychology in Legal Contexts*, pp. 329–45. Chichester, UK: John Wiley & Sons.

Schuller, R. A., A. Ryan, D. Krauss, & G. Jenkins (2013) "Mock Juror Sensitivity to Forensic Evidence in Drug Facilitated Sexual Assaults," 36 *International J. of Law & Psychiatry* 121.

Schweitzer, N. J., & M. J. Saks (2007) "The CSI Effect: Popular Fiction About Forensic Science Affects the Public's Expectations About Real Forensic Science," 47 *Jurimetrics* 357.
Shelton, D. E. (2009) "Twenty-First Century Forensic Science Challenges for Trial Judges in Criminal Cases: Where the 'Polybutadiene' Meets the 'Bitumen,'" 18 *Widener Law J.* 309.
Shelton, D. E., Y. S. Kim, & G. Barak (2006) "A Study of Juror Expectations and Demands Concerning Scientific Evidence: Does the 'CSI Effect' Exist?" 9 *Vanderbilt J. of Entertainment & Technology Law* 331.
Starrs, J. E. (2006) "The Absence of Evidence," 30 *Scientific Sleuthing Rev.* 14.
Tabachnick, B. G., & L. S. Fidell (2013) *Using Multivariate Statistics*, 6th ed. Upper Saddle River, NJ: Pearson.
Tribe, L. (1971) "Trial by Mathematics: Precision and Ritual in the Legal Process," 84 *Harvard Law Rev.* 1329.
Tversky, A., & D. Kahneman (1974) "Judgment Under Uncertainty: Heuristics and Biases," 185 *Science* 1124.
Tyler, T. (2006) "Viewing CSI and the Threshold of Guilt: Managing Truth and Justice in Reality and Fiction," 115 *Yale Law J.* 1050.
United States v. Hoffman (1992) 964 F.2d 21.
Ward, W. C., & H. M. Jenkins (1965) "The Display of Information and the Judgment of Contingency," 19 *Canadian J. of Psychology* 231.

Appendix A

Suppose that a forensic scientist conducted a study in which 10 participants fired a gun and were then tested for the presence of gunshot residue (GSR). Suppose that 50 percent of these participants tested positive for GSR. These figures appear in the center column of the table below:

	Did Fire Gun (G)	*Did Not Fire Gun (NG)*
Positive Test Result (+)	5	0
Negative Test Result (–)	5	10
	10	10

It is easy to see that if 10 people did fire a gun (G), the test for GSR would come back positive (+) for five of the 10 people. Formally, $p(+|G) = 5/10$ or 50 percent.

We make the simplifying assumption that the false positive rate is zero—in other words, not a single person who did not fire a gun will test positive for GSR. This assumption implies that if 10 people did not fire a gun (NG), the test would come back negative for all 10 people. In formal terms, $p(-|NG) = 10/10$ or 100 percent.

The likelihood ratio for a negative GSR test result is simply the ratio of these conditional probabilities: $(p(-|NG) = 100\%)/(p(+|G) = 50\%)$ or 2.

Since the numerator of the likelihood ratio is fixed (based on the assumption that the false positive rate is zero), it is clear that the likelihood ratio depends on the size of the denominator (i.e., $p(+|G)$): the ratio will decrease as the denominator becomes larger and increase as the denominator becomes smaller. For example, when $p(+|G) = 20$ percent, the likelihood ratio is 5, whereas when $p(+|G) = 80$ percent, the likelihood ratio is 1.25. Thus, the probative value of a negative test result directly depends on the rate at which GSR is detected among those who fired a gun.

Appendix B

The Crime

Mr. Lee Kim, a 58-year-old Korean-American, was shot and killed in front of his grocery store on the corner of Adams Street and 8th Avenue in the City of Astoria, New York. The crime occurred about 10:05 p.m. on July 3, 2004.

At 10:06 p.m., Marion Johnson, a 72-year-old retired school librarian, called a "911" operator and reported the shooting. Johnson lives in a second-floor apartment above Kim's store. She told the "911" operator that she had heard Mr. Kim arguing with someone. She heard Mr. Kim shout "Benjie, no" and then heard a gunshot. When she looked out the window she saw a tall blond man running away from the grocery store, heading west on 8th Avenue. He was wearing a gray shirt with the number "2" on the back. She then noticed Mr. Kim lying on the sidewalk in front of his store and called "911." The "911" operator alerted police to be on the lookout for a man matching the description provided by Mrs. Johnson.

At approximately 10:13 p.m., officer Brian Wilcox of the Astoria Police Department observed a tall blond man running along 9th Avenue, about a mile west of the scene of the crime. The man was wearing a New York Yankees baseball jersey and a pair of jeans. It was a "Derek Jeter" jersey with a large number 2 on the back. (Derek Jeter is a member of the New York Yankees baseball team). Officer Wilcox stopped the man for questioning.

The man identified himself as Benjamin Everett and produced a driver's license showing his address as a nearby apartment on 9th Avenue. He said was running for exercise and denied any involvement in the shooting. He was not carrying a gun. Because Everett matched the description of the shooter, Officer Wilcox took him into custody and escorted him, on foot, to a nearby police substation for questioning. At about 10:45 p.m., Herbert DiFonso, a technician from the Astoria Crime Laboratory, took samples from Everett's hands and clothing to be tested for the presence of gunshot residue (GSR).

Police and emergency personnel arrived at the scene of the shooting within five minutes of Mrs. Johnson's call. Mr. Kim had a fatal gunshot wound to his chest and was dead on arrival at a local hospital. A police officer found a handgun at the bottom of a trashcan just west of the crime scene.

The police asked the victim's wife whether her husband knew a blond man named Benjie. Mrs. Kim reported that a young man from the neighborhood named Benjie Everett had worked for her husband in the past, helping to unload deliveries and stock the store. She reported that her husband had fired Benjie Everett several weeks earlier due to suspicion that he was stealing money from the cash register. Mrs. Kim identified Benjamin Everett as the "Benjie" who had worked for her husband.

The next day the police had Benjamin Everett appear in a lineup with five other men of similar appearance. Marion Johnson observed the lineup through a one-way window. She was not told which person in the lineup was the suspect. However, she

immediately pointed to Benjamin Everett and said: "That's the man I saw. That's him." Based on this evidence, Everett was arrested and charged with murder.

The crime laboratory examined the handgun found near the crime scene. It was a 9 mm automatic. Because it had been partially immersed in oily water at the bottom of the trashcan, no fingerprints could be found on it. However, a firearms examiner was able to confirm that this gun had fired the fatal bullet that killed Mr. Kim. The results of the gunshot residue (GSR) test on Everett's hands and clothing were negative. No GSR was detected.

Key Testimony at Trial

Testimony of Mrs. Marion Johnson

Q: Mrs. Johnson, can you tell the jury what you were doing the evening of July 3rd, 2004?
Mrs. Johnson: I was sitting in my living room watching television.
Q: What did you hear while you were watching TV?
Mrs. Johnson: I remember hearing Mr. Kim locking up his store like he has for 10 years.
Q: What time did you hear Mr. Kim locking up?
Mrs. Johnson: He always closes the grocery story at 10 o'clock, so it was shortly after that. Also "Law & Order" had just come on, and I never miss "Law & Order". That's how I know it was just after 10.
Q: So just after 10 pm you heard Mr. Kim locking up his store for the evening. What did you hear next?
Mrs. Johnson: I heard Mr. Kim start talking to another man and then I heard some angry words and scuffling. And Mr. Kim said, "Benjie. No!" And then I heard that awful gun shot. It echoed right up to my apartment. It was just terrible.
Q: Did it sound like there was an altercation between Mr. Kim and the other man before the shot was fired.
Mrs. Johnson: That's right. Poor Mr. Kim.
Q: What did you do after you heard the gun shot?
Mrs. Johnson: I ran to my window to see what had happened. I saw Mr. Kim lying on the sidewalk and I saw that man [indicating defendant Benjamin Everett] running away.
Q: Are you certain that the defendant, Benjamin Everett was the man you saw?
Mrs. Johnson: Yes, I recognized him the next day in the police lineup. He is the same man. I'm sure of it.

Testimony of Herbert DiFonso, Criminalist from the Astoria Crime Laboratory

Q: I understand that you tested for the presence of gunshot residue or GSR on the defendant?
Mr. DiFonso: Yes. I collected samples from him shortly after the shooting. I tested those samples for GSR.
Q: Before we get into the details, can you tell the jury what GSR is?
Mr. DiFonso: When a firearm is discharged, hot gases created by the explosion of the primer tend to spray out, carrying with them inorganic residue from the primer. That residue can often be detected on the hands and clothing of someone who was near the gun when it was fired. We call that residue gunshot residue or GSR. It generally consists of round particles containing a combination of three elements: lead, barium and antimony.
Q: How do you test for GSR?
Mr. DiFonso: We collect samples using an adhesive tape device, sometimes called an adhesive stub or lift. We press the adhesive substance onto the skin and clothing and it lifts off GSR particles, if they are present. We then examine the material on the lifts microscopically to see whether any suspicious particles are present. If it appears that GSR is present, we confirm that it is GSR by testing for the presence of lead, barium and antimony. We use a procedure

known as SEM/EDS, which stands for scanning electron microscopy with energy dispersive x-ray spectrometry detection. It is the most accurate test available.
Q: Now in this case did you attempt to find GSR on Mr. Everett?
Mr. DiFonso: Yes, I took lifts from his hands and his clothing. But they did not contain any detectable amounts of GSR. So the findings were negative.
Q: What possible explanations might there be for a negative finding?
Mr. DiFonso: Well, there are a number of reasons you could get a negative finding. One possibility is that the individual was not near a firearm that was discharged. Another possibility is that the GSR particles were removed before the test through physical activity, hand washing, or weather conditions. Another possibility is that the firearm failed to produce a sufficient quantity of GSR to be detectable by the test. A final possibility could be a problem in the collection process, such as improper use of the collection kit.
Q: If you get GSR on you can you wipe it off?
A: You can remove GSR from hands with soap and water. To remove it from clothing you usually have to run it through the laundry.
Q: Did you do any tests of the murder weapon in this case to see how much GSR it produced?
Mr. DiFonso: Yes, we conducted a study in which we attempted to replicate the conditions of the shooting in this case to check our ability to detect GSR from that particular handgun.
Q: Can you describe the study?
Mr. DiFonso: We had ten different volunteers from the lab fire the murder weapon. Each was wearing a Yankees jersey of the same type the defendant was wearing. After each man fired the gun, we had him engage in vigorous exercise by running around a track for about eight minutes, which we estimate is the length of time between when Mr. Kim was shot and when the defendant was stopped by the officer. We then waited about another 20 minutes and attempted to collect GSR from the hands and clothing of each man.
Q: And what did you find?
[0% Condition:
Mr. DiFonso: We did not find detectable levels of GSR on any of the men. We think that the handgun in question has particularly tight construction and produces relatively little backwash of GSR, so relatively little was deposited in the first place. Then the motion of running combined with sweating probably removed what little was there, leaving an insufficient amount to detect.
Q: Out of ten men who fired the murder weapon, you didn't find GSR on any of them?
50%/60%/90% Conditions:
Mr. DiFonso: We found detectable levels of GSR on the hands and clothing of [five/six/nine] of the ten men. We think that the handgun in question has particularly tight construction and produces relatively little backwash of GSR, so relatively little was deposited in the first place. Then the motion of running combined with sweating probably knocked off some of what was there, so that not all of the men had enough left to detect
Q: So, out of ten men who fired the weapon, you found GSR on [five, six, nine]?]
[100% Condition:
Mr. DiFonso: We found detectable levels of GSR on the hands and clothing of all ten men.
Q: So, out of ten men who fired the weapon, you found GSR on all ten?]
Mr. DiFonso: That is correct.

Closing Arguments

Prosecutor

The evidence leaves no doubt about who committed this crime. Marion Johnson heard Mr. Kim arguing with a man he called "Benjie." She saw a tall blond man running away from the crime scene. She told police the killer was wearing a gray shirt with the number 2 on the back. And sure enough, a few minutes later Officer Wilcox saw the defendant running down the street just west of the crime scene wearing a gray Derek Jeter baseball jersey with the number

2 on the back. The next day Mrs. Johnson was able to pick the defendant out of a police line up. She recognized him immediately as the man she saw running from the crime scene.

The defense lawyer is going to get up here and talk about GSR because that's about all he has to talk about. Yes, it is true that the crime lab found no GSR on Benjamin Everett. But there could be lots of reasons for that. Maybe he wiped it off. Maybe there wasn't enough to detect. *Mr. DiFonso told you about the test they did in which ten men fired the murder weapon.*

[(0% Condition): When they tested those men they didn't find GSR on any of them. So there was no chance they would find GSR on the defendant after he fired the gun][(50% and 60% conditions): Mr. DiFonso told you about the test they did in which ten men fired the murder weapon. When they tested those ten men they found GSR on only five of them. So it was pretty much a toss up whether they would find GSR on the defendant after he fired the gun.] [(90% condition): Mr. DiFonso told you about the test they did in which ten men fired the murder weapon. When they tested those ten men they found GSR on only nine of them. [Note: the italicized section was omitted in the 100% condition].

So it was not certain that they would find GSR on the defendant after he fired the gun.

How many men do you think there are in Astoria who have blond hair, who were running in the street right after the murder wearing a Derek Jeter jersey, who knew Mr. Kim and probably had a grudge against him, who look so much like the killer that an eyewitness would pick them out of a line up, and who answer to the name Benjie? I'll tell you how many. There is only one and he is sitting right there in front of you.

Defense Lawyer

Mrs. Johnson is an elderly lady. She was on the second floor, looking down, during the night. How good a look could she have gotten at the man running from the crime scene? You can't trust an identification made under those circumstances. We know that eyewitnesses can be unreliable. It's not that they are lying. They are just honestly mistaken. It happens all the time.

Mrs. Johnson claims to have heard Mr. Kim say something that sounded like the name "Benjie". Well Benjie is a common name, if it was even a name Mr. Kim spoke. We'll never know for sure.

Fortunately, we have some solid scientific evidence in this case—the GSR test. If Benjamin Everett shot Mr. Kim there should have been gunshot residue on his hands and his clothing. But the crime lab found no GSR on his hands and no GSR on his clothing. That is because the police arrested the wrong man. There would have been no time for him to wash his hands or launder his clothes. Benjamin Everett did not shoot anyone. That's why there was no GSR.

[50% Condition/60% Condition/90% Condition/100% Condition: Mr. DiFonso told you about his experiment. He had ten men fire the murder weapon. When they tested those men [five/six/nine/all ten] of them had GSR. So if the defendant fired the gun there is a [50%/60%/90%100%] chance he would have GSR on him. But he didn't have any.][Note: this paragraph was omitted in the 0% condition]

Ben Everett just had the misfortune to be out jogging on the night of the murder and to be dressed in a manner similar to the killer. There are lots of young men in this

area who have blond hair and wear Derek Jeter shirts. The Yankees must have sold thousands of those shirts.

Judge's Instructions

If you are convinced beyond a reasonable doubt that the defendant committed this murder, then you must find him guilty. It is not required that the state prove guilt beyond all possible doubt. The test is one of reasonable doubt. A reasonable doubt is a doubt based upon reason and common sense—the kind of doubt that would make a reasonable person hesitate to act. Proof beyond a reasonable doubt, therefore, must be proof of such a convincing character that a reasonable person would not hesitate to rely and act upon it.