

## Cornell Law Library Scholarship@Cornell Law: A Digital Repository

---

Cornell Law Faculty Publications

Faculty Scholarship

---

12-1-2007

# Statistics in the Jury Box: How Jurors Respond to Mitochondrial DNA Match Probabilities

David H. Kaye

*Arizona State University Sandra Day O'Connor School of Law*, [kaye@alum.mit.edu](mailto:kaye@alum.mit.edu)

Valerie P. Hans

*Cornell Law School*, [valerie.hans@cornell.edu](mailto:valerie.hans@cornell.edu)

B. Michael Dann

*Maricopa County Superior Court, Arizona (retired)*, [m.dann@cox.net](mailto:m.dann@cox.net)


Erin J. Farley

*University of North Carolina, Wilmington*, [farley@uncw.edu](mailto:farley@uncw.edu)

Stephanie Albertson

*University of Delaware*, [estfania@udel.edu](mailto:estfania@udel.edu)

Follow this and additional works at: <http://scholarship.law.cornell.edu/facpub>

 Part of the [Applied Statistics Commons](#), [Criminal Law Commons](#), [Criminal Procedure Commons](#), [Evidence Commons](#), and the [Science and Technology Commons](#)

---

### Recommended Citation

Kaye, David H.; Hans, Valerie P.; Dann, B. Michael; Farley, Erin J.; and Albertson, Stephanie, "Statistics in the Jury Box: How Jurors Respond to Mitochondrial DNA Match Probabilities" (2007). *Cornell Law Faculty Publications*. Paper 363.  
<http://scholarship.law.cornell.edu/facpub/363>

This Article is brought to you for free and open access by the Faculty Scholarship at Scholarship@Cornell Law: A Digital Repository. It has been accepted for inclusion in Cornell Law Faculty Publications by an authorized administrator of Scholarship@Cornell Law: A Digital Repository. For more information, please contact [jmp8@cornell.edu](mailto:jmp8@cornell.edu).

# Statistics in the Jury Box: How Jurors Respond to Mitochondrial DNA Match Probabilities

*David H. Kaye, Valerie P. Hans, B. Michael Dann, Erin Farley, and Stephanie Albertson\**

This article describes parts of an unusually realistic experiment on the comprehension of expert testimony on mitochondrial DNA (mtDNA) sequencing in a criminal trial for robbery. Specifically, we examine how jurors who responded to summonses for jury duty evaluated portions of videotaped testimony involving probabilities and statistics. Although some jurors showed susceptibility to classic fallacies in interpreting conditional probabilities, the jurors as a whole were not overwhelmed by a 99.98 percent exclusion probability that the prosecution presented. Cognitive errors favoring the defense were more prevalent than ones favoring the prosecution. These findings lend scant support to the legal argument that mtDNA evidence (with modest exclusion probabilities) should be excluded because

---

\*Address correspondence to David H. Kaye, Center for the Study of Law, Science, and Technology, ASU College of Law, Tempe, AZ 85287-7906; email: k@asu.edu. Kaye is Regents' Professor, ASU Sandra Day O'Connor College of Law, Professor, School of Life Sciences, and Fellow, Center for the Study of Law, Science, and Technology, ASU College of Law; Hans is Professor of Law, Cornell Law School; Dann is Judge, Maricopa County Superior Court, Arizona (retired); Farley is Assistant Professor, Department of Sociology and Criminal Justice, University of North Carolina, Wilmington; Albertson is a Ph.D. candidate in the Department of Sociology and Criminal Justice, University of Delaware.

A version of this article was presented at the First Annual Conference on Empirical Legal Studies, Austin, Texas, 2006, at which Jay Koehler provided valuable commentary. The article also benefited from comments by Edward Cheng, Lynn Mather, Dale Nance, and an anonymous referee, as well as a review of previous mock-jury research on probabilistic evidence prepared by Bruce Wagar for a seminar at ASU. The authors also are grateful to the judges and staff of the Superior Court of Delaware in New Castle County, and the 480 jury-eligible citizens of the county who served on the mock trials and deliberations. The data described here come from research funded by Grant 2002-IJ-CX-0026, National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Points of view expressed in this article are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.

jurors are prone to overvalue such evidence. The article also introduces a new method for inferring the perceived probability of guilt that satisfies the burden of persuasion for most jurors.

## I. INTRODUCTION

This article describes parts of an experiment on the comprehension of expert testimony on mitochondrial DNA (mtDNA) sequencing in a criminal trial for robbery. The experiment was designed to ascertain whether members of a jury pool were able to understand this evidence, how they used it in deciding a mock jury trial, and whether their comprehension could be improved through various jury reform techniques.<sup>1</sup> Here, we examine how the jurors evaluated portions of the testimony involving probabilities and statistics. In short, we report that although some jurors showed susceptibility to classic fallacies in interpreting conditional probabilities, the jurors as a group were not overwhelmed by testimony from a prosecution expert that “more than 99.98% of all Caucasians would be excluded” by the DNA match. Most jurors accepted a defense criticism of this computation. Moreover, it appears that many jurors were inclined to agree with the defendant’s overstated argument that because dozens of men in the area might have mtDNA types consistent with those of the robber, the match with the defendant’s DNA was worthless. A smaller number seemed to commit what has been called the “prosecutor’s fallacy.” These jurors confused the proportion of the general population that would be excluded by the DNA test with the probability that the defendant was the source of the crime-scene DNA. On balance, these findings do not indicate that jurors generally were unduly impressed by the prosecution’s DNA evidence. Consequently, our results challenge the legal argument that mtDNA evidence should be excluded because jurors are prone to overvalue such evidence.

Section II of this article reviews the literature on jury comprehension of probability evidence. This research suggests that in some circumstances, jurors make systematic errors in estimating probabilities. Section III provides an overview of the methodology of our study. (Appendix A presents addi-

---

<sup>1</sup>See, e.g., B. Michael Dann, Valerie P. Hans & David H. Kaye, *Can Jury Trial Innovations Improve Juror Understanding of DNA Evidence?* 90 *Judicature* 152 (2007); B. Michael Dann, Valerie P. Hans & David H. Kaye, *Testing the Effects of Selected Jury Trial Innovations on Juror Comprehension of Contested mtDNA Evidence*, Final Technical Report (2004).

tional details.) Section IV examines juror responses to questions designed to probe their reasoning about the statistical aspects of the mtDNA match. Section V analyzes the jurors' stated probabilities of guilt. It describes a new method for quantifying the burden of persuasion actually applied by the jurors. Section VI summarizes the principal findings and their implications for trial practice and the law of evidence.

## II. PREVIOUS RESEARCH ON JURY COMPREHENSION OF DNA PROBABILITIES

When a DNA match is presented in the courtroom, it normally is accompanied by statistics intended to show how powerful the evidence is. The quantitative measures that typically are seen in court are the relative frequency of the matching DNA type and the probability that DNA from a randomly selected person from one or more major racial or ethnic groups would be of this type.<sup>2</sup> This latter statistic often is referred to as a random-match probability (RMP), and both the method for computing it and its likely impact on lay jurors have been the subject of academic debate and courtroom battles.<sup>3</sup> Indeed, even before DNA typing joined the forensic armamentarium, two schools of thought emerged regarding legal factfinders' abilities to evaluate such statistical information about trace evidence. Neither was particularly encouraging.<sup>4</sup> First, there was armchair speculation that jurors and judges are likely to be overwhelmed by the apparent precision of quantitative presentations and uncritically afford them too much weight.<sup>5</sup> An alternative view, discomfiting in another way, held that inasmuch as study after study had shown that human information processors are more at ease with

---

<sup>2</sup>See D.H. Kaye, *The Role of Race in DNA Evidence*, \_\_\_ Sw. L. Rev. \_\_\_ (forthcoming 2007); D.H. Kaye, *Logical Relevance: Problems with the Reference Population and DNA Mixtures in *People v. Pizarro**, 3 *Law, Probability & Risk* 211 (2004); David H. Kaye & George F. Sensabaugh, *DNA Typing*, in 4 *Modern Scientific Evidence: The Law and Science of Expert Testimony* § 32 (David L. Faigman et al. eds., 4th ed. 2006).

<sup>3</sup>See, e.g., Kaye & Sensabaugh, *supra* note 2.

<sup>4</sup>David L. Faigman & A.J. Baglioni, Jr., *Bayes' Theorem in the Trial Process: Instructing Jurors on the Value of Statistical Evidence*, 12 *Law & Hum. Behav.* 1, 2-3 (1988).

<sup>5</sup>The leading exponent of this view is Lawrence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 *Harv. L. Rev.* 1328 (1971).

qualitative information than with quantitative data, jurors are more likely to undervalue statistical evidence.<sup>6</sup>

A classic body of experimentation examines human inference,<sup>7</sup> and a subset of this work examines the processing of explicitly probabilistic or statistical evidence in the legal context.<sup>8</sup> Even before the advent of DNA evidence, a series of studies indicated that jurors may have trouble drawing appropriate inferences from moderately small random-match probabilities (on the order of 0.01 to 0.001 or so).<sup>9</sup> In the DNA era, several researchers have examined the response to much smaller RMPs for DNA evidence. Table 1 summarizes some features and findings of the principal studies examining how people treat DNA evidence.

In general, the subjects (mock jurors, university students, and jurists) tended to improperly evaluate the significance of DNA match statistics in comparison to the impact prescribed by probability theory.<sup>10</sup>

---

<sup>6</sup>Michael J. Saks & Robert F. Kidd, *Human Information Processing and Adjudication: Trial by Heuristics*, 15 *Law & Soc'y Rev.* 123, 149 (1981).

<sup>7</sup>See generally *Simple Heuristics that Make Us Smart* (Gerd Gigerenzer & P.M. Todd eds., 1999); *Behavioral Foundations of Economic Theory* (Robin M. Hogarth & Melvin W. Reder eds., 1987); *Judgment Under Uncertainty: Heuristics and Biases* (Daniel Kahneman, P. Slovic & A. Tversky eds., 1982); Richard E. Nisbett & Lee Ross, *Human Inference: Strategies and Shortcomings of Social Judgment* (1980); Scott Plous, *The Psychology of Judgment and Decision Making* (1993).

<sup>8</sup>See, e.g., David H. Kaye & Jonathan J. Koehler, *Can Jurors Understand Probabilistic Evidence?* 154 *J. Royal Stat. Soc'y Ann.* (1991); William C. Thompson, *Are Juries Competent to Evaluate Statistical Evidence?* 52 *Law & Contemp. Probs.* 9 (1989).

<sup>9</sup>See authorities cited, *supra* note 8.

<sup>10</sup>J.J. Koehler et al., *The Random Match Probability (RMP) in DNA Evidence: Irrelevant and Prejudicial?* 35 *Jurimetrics J.* 201 (1995); J.J. Koehler, *On Conveying the Probative Value of DNA Evidence: Frequencies, Likelihood Ratios and Error Rates*, 67 *U. Colo. L. Rev.* 859 (1996); Jason Schklar & Shari Seidman Diamond, *Juror Reactions to DNA Evidence: Errors and Expectancies*, 23 *Law & Hum. Behav.* 159 (1999); Jonathan J. Koehler, *When Are People Persuaded by DNA Match Statistics?* 25 *Law & Hum. Behav.* 493 (2001) [hereafter, Koehler I]; Jonathan J. Koehler, *The Psychology of Numbers in the Courtroom: How to Make DNA-Match Statistics Seem Impressive or Insufficient*, 74 *S. Cal. L. Rev.* 1275 (2001) [hereafter, Koehler II]; J.J. Koehler & L. Macchi, *Thinking about Low-Probability Events: An Exemplar Cuing Theory*, 15 *Psych. Sci.* 540 (2004); Samuel Lindsey et al., *Communicating Statistical DNA Evidence*, 43 *Jurimetrics J.* 147 (2003); Dale A. Nance & Scott B. Morris, *An Empirical Assessment of Presentation Formats for Trace Evidence with a Relatively Large and Quantifiable Random Match Probability*, 42 *Jurimetrics J.* 1 (2002) [hereafter, Nance & Morris I]; Dale A. Nance & Scott B. Morris, *Juror Understanding of DNA Evidence: An Empirical Assessment of Presentation Formats for Trace*

In this regard, many of the studies focused on Bayes's theorem, which describes how new information affects a previously established probability. Let  $D$  designate the hypothesis that the defendant is the source of a bloodstain found at the scene of a crime, and let  $Odds(D)$  be the (prior) odds in favor of  $D$ . These prior odds are subjectively determined on the basis of other evidence in the case. For example, based on a tentative eyewitness identification, a juror might assess the prior odds as two-to-one. Now suppose this juror learns that the crime-scene stain contains DNA that matches the defendant's DNA. This new information (which we can call  $M$ , for match) changes  $Odds(D)$  in a simple way:

$$Odds(D|M) = \frac{P(M|D)}{P(M|\bar{D})} Odds(D). \tag{1.1}$$

The prior odds are just multiplied by a "likelihood ratio" that is given by (1) the conditional probability of the data (the match  $M$ ) assuming that the defendant is the source ( $D$ ) to (2) the conditional probability assuming that he is not ( $\bar{D}$ ). If the DNA is certain to match when a suspect is the source,<sup>11</sup> then the numerator in the likelihood ratio is  $P(M|D) = 1$ . The denominator often is taken to be the probability of a random match—the chance that a single, randomly selected, unrelated individual would have the matching DNA type.<sup>12</sup> Hence,

---

Evidence with a Relatively Small Random Match Probability, 34 *J. Legal Stud.* 395 (2005) [hereafter, Nance & Morris II]; Franco Taroni & Colin G.G. Aitken, Probabilistic Reasoning in the Law, Part 1: Assessment of Probabilistic and Explanations Value of DNA Evidence, 38 *Sci. & Just.* 165 (1998).

Work with trace evidence other than DNA also suggests that large RMPs affect jurors less than Bayes's rule would suggest. See, e.g., Jane Goodman, Jurors' Comprehension and Assessment of Probabilistic Evidence, 16 *Am. J. Trial Advoc.* 361 (1992); William C. Thompson & Edward L. Schumann, Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy, 11 *Law & Hum. Behav.* 167, 180 (1987); Brian C. Smith et al., Jurors' Use of Probabilistic Evidence, 20 *Law & Hum. Behav.* 49 (1996) (using hypothetical blood types or enzymes with RMPs of 20 percent, 35 percent, and 80 percent, and finding (1) conservatism relative to Bayes's theorem, (2) no strong manifestations of transposition, and (3) that explaining Bayes's rule mitigated the irrelevance fallacy).

<sup>11</sup>This would be the case only if there is zero chance of a false negative (sensitivity = 1).

<sup>12</sup>For critical discussions of this interpretation, see David J. Balding, Weight-of-Evidence for Forensic DNA Profiles (2005); D.H. Kaye et al., The New Wigmore on Evidence: Expert Evidence § 12 (2004).

Table 1: Previous Studies of Jury Comprehension of DNA Evidence

<i>Study</i>	<i>RMP</i> <sup>a</sup>	<i>Decision Aids</i>	<i>Jury Valuation</i> <sup>b</sup>
Koehler et al. (1995)	1/(1 billion) <sup>c</sup>	None	Erroneous <sup>d</sup>
Koehler (1996)	1/100, 1/1000	None	Inconsistent <sup>e</sup>
Schklar & Diamond (1999)	2/100, <sup>f</sup> 1/(1 billion) <sup>g</sup>	Combination instructions	Undervalued
Koehler et al. (2001a, 2001b)	1/(1,000) <sup>h</sup> , 1/(1 million) <sup>i</sup> , 1/(1 billion) <sup>j</sup>	None (various methods of characterizing the statistic)	
Koehler et al. (2004)	1/1,000 <sup>k</sup> , 1/100,000 <sup>l</sup>	None; the statistic was presented as a relative frequency <sup>m</sup>	Undervalued
Nance & Morris (2002)	4%	(1) None vs. (2) Frequencies and expected numbers of true and false positives	More correct posterior probabilities in Condition (2)
Lindsey et al. (2003)	0.0001% <sup>n</sup>	None (various methods of characterizing the statistic)	Undervalued
Nance & Morris (2005)	1/40,000 <sup>o</sup>	Various methods of characterizing the statistics <sup>p</sup>	Undervalued
Taroni & Aitken (1998)	1/50, 1/10000		

<sup>a</sup>Random-match probability.

<sup>b</sup>Relative to a Bayesian norm.

<sup>c</sup>Along with a “laboratory error rate of” 0.02 or 0.001.

<sup>d</sup>In that “mock jurors who were provided with [only the] RMP convicted twice as often as jurors who were provided with more diagnostic statistical information (laboratory error rate, aggregated error rate).”

<sup>e</sup>In that mock jurors reacted differently to a frequency (“1 person out of every 100 (1000)”) as opposed to a likelihood ratio (“100 (1000) times more likely we would see this match if the defendant is the source of the semen than if the defendant is not”) or (unjustifiably) as posterior odds (“Given that we see this match, it is approximately 100 (1000) times more likely that the defendant is the source of the semen than that he is not the source”).

<sup>f</sup>Along with a true-positive probability (sensitivity) of 1 and a false-positive probability (1 – specificity) of 1/(1 billion).

<sup>g</sup>Along with a sensitivity of 1 and a false-positive probability of 2/100.

<sup>h</sup>Presented as “0.1%” and “one in 1,000.”

<sup>i</sup>Presented as a probability and a frequency.

<sup>j</sup>Presented as a probability and a frequency.

<sup>k</sup>Presented as “0.1 out of a hundred,” “1 out of 1,000,” and “2 out of 2,000.”

<sup>l</sup>Presented as “0.1 out of 10,000,” “1 out of 100,000,” “2 out of 200,000.”

<sup>m</sup>The percentage of people with matching types in the relevant population.

<sup>n</sup>Along with a sensitivity of 1 and a false-positive probability of 0.001 percent.

<sup>o</sup>Along with, in some groups, a false-positive probability of 1/1,000.

<sup>p</sup>Presented as the percentage excluded, a frequency, a likelihood ratio, and as a shift in prior odds (“Given prior odds of 3 : 2 of the suspect being the source of the DNA, the odds after considering the DNA match evidence are 75 : 1 that the suspect is the source”).

$$\text{Odds}(D|M) = \frac{\text{Odds}(D)}{\text{RMP}}. \quad (1.2)$$

The fairly consistent finding is that the jurors' estimates of  $\text{Odds}(D|M)$  are smaller than the number dictated by Bayes's rule.<sup>13</sup> In our example, if the prior odds are 2 (two to one), and the RMP is 2/100, then the posterior odds are 100 (100 to 1). Yet the research indicates that most people would report posterior odds of less than 100. Such conservatism is usually regarded as a manifestation of "misaggregation error" (improperly combining the new data with the prior information) or "misperception error" (improperly assessing the probative value of the new information).<sup>14</sup>

In addition to reporting the failure to update a prior probability to the extent dictated by probability theory, researchers have identified other common mistakes that both lay people and many professionals make in reasoning about the probabilities used to describe the significance of a DNA match. One is called the "fallacy of the transposed conditional" by statisticians and the "prosecutor's fallacy" in legal circles.<sup>15</sup> Consider a case in which the expert testifies that the RMP is 1 percent (meaning that there is only a 1 percent chance that the DNA from a randomly selected, unrelated person in the relevant general population would match the DNA profile in the crime-scene sample). The fallacy consists of concluding that because there is only a 1 percent chance that an innocent person would match, the chance that the defendant is innocent is 1 percent and, hence, that there is a 99 percent chance that the defendant is guilty. This transposes the conditional probability that a man would match given that he is innocent,  $P(M|I)$ , into the conditional probability that the defendant is innocent given that he matches,  $P(I|M)$ . This transposition is comparable to asserting that if the probability that a man can speak Spanish given that he is a citizen of Peru is 99 percent, then the probability that a man is Peruvian given that he speaks Spanish also is 99 percent. Used to prove a defendant's guilt, these transpositions exemplify the "prosecutor's fallacy."

---

<sup>13</sup>A more complicated expression for the posterior odds incorporates the sensitivity and specificity of the DNA test for identity.

<sup>14</sup>W. DuCharme, Response Bias Explanation of Conservative Human Inference, 85 *Experimental Psychol.* 66 (1970). Schklar & Diamond, *supra* note 10, consider other possible explanations for the apparent undervaluation.

<sup>15</sup>Kaye et al., *supra* note 12, at § 12.4.1(b), at 463–74 (2004); Kaye & Sensabaugh, *supra* note 2.



However, the “prosecutor’s fallacy” is not always common,<sup>16</sup> and its occurrence is relatively rare compared to that of the “defense attorney’s fallacy” (which we prefer to call the “relevance fallacy”).<sup>17</sup> The relevance fallacy occurs where, given the same RMP of 1 percent, the jury concludes that since only 1 percent of the relevant population could have contributed the sample and that number is, say, 100 people, then the odds that the defendant supplied the DNA sample are only 1 in 100 and, therefore, that the evidence has virtually no value in linking the defendant to the crime. Both fallacies misconstrue the probability statements commonly used in DNA evidence presentations at trials.<sup>18</sup>

The studies listed in Table 1 share some important limitations, many of which were originally acknowledged by the investigators. Several of the studies relied exclusively on college students as study participants; yet, education level is surely an important factor when it comes to comprehension and use of complex evidence. Furthermore, all the participants read the facts and applicable law of their cases from brief written summaries and the scientific presentations were quite limited. In actual trials, there can be extended presentation and cross-examination of scientific experts, and lawyers have additional opportunities in opening and closing statements to

---

<sup>16</sup>See *infra* Section IV.C.

<sup>17</sup>Committee on DNA Forensic Science: An Update, National Research Council, the Evaluation of Forensic DNA Evidence 198 (1996). Another study deserves mention in this regard, although it did not involve DNA evidence. Smith et al., *supra* note 10, presented mock jurors (mostly students from the University of Minnesota) with both nonstatistical and statistical evidence involving blood and semen. *Id.* at 54–55. The RMPs were large. Some jurors saw blood-enzyme evidence that matched 80 percent of the population, while others saw blood-enzyme evidence that matched 20 percent of the population, and all saw other blood-type evidence that matched 35 percent of the population. *Id.* at 57. Some groups were presented with the “Prosecutor’s Fallacy”—that is, expert testimony telling them to interpret RMP as probability of innocence—while others were exposed to the “Defense Attorney’s Fallacy.” *Id.* at 54. Finally, some groups were instructed in Bayes’s theorem and its application to the evidence. *Id.* Many of the subjects’ probabilistic assessments displayed slight undervaluation relative to the Bayesian norm. Presenting Bayes’s rule had no effect on average, *id.* at 70, but it did help overcome the defense attorney’s fallacy. *Id.* at 72. Exposure to the prosecutor’s fallacy did not appear to have had much effect.

<sup>18</sup>Kaye & Sensabaugh, *supra* note 2; Committee on DNA Forensic Science: An Update, *supra* note 17, at 198; Norman Fenton & Martin Neil, The “Jury Observation Fallacy” and the Use of Bayesian Networks to Present Probabilistic Legal Arguments, *Mathematics Today* 180, 181 (Dec. 2000); see also Thompson, *supra* note 8.

reinforce key portions of expert scientific evidence. Finally, unlike actual jury trials, most of these studies did not allow for group deliberation.

### III. THE MOCK TRIAL AND JURORS

The methodology of our study and the characteristics of the volunteers who participated in the experiment are described in detail elsewhere, but a brief summary will be helpful.<sup>19</sup> The subjects were 480 Wilmington, Delaware, residents who appeared at the New Castle County Courthouse in response to summonses for jury duty but were not needed to serve on a regular jury. Reflecting the county jury pool generally, about four-fifths were white, about half were female, and most had taken at least some mathematics and science courses in high school or college.<sup>20</sup>

These participants were taken to conference rooms in the courthouse where they watched a 70-minute videotaped trial. They then deliberated in juries of eight (there were 60 such mock juries) until they reached a unanimous verdict or declared themselves hung. Jurors completed three questionnaires: one at the beginning of the study, one after viewing the mock trial, and one after deliberation.<sup>21</sup>

The mock trial was based on *State v. Pappas*,<sup>22</sup> the first appeal in Connecticut from a prosecution relying on mtDNA evidence. Usually, forensic

---

<sup>19</sup>See Dann et al., *supra* note 1.

<sup>20</sup>Their views about science were quite similar to those reported in national surveys, with widespread positive views about the benefits of science along with a significant minority who expressed concerns about science. About a fifth of the sample had substantial mathematics or science experience on the job.

<sup>21</sup>Some of the juries experienced a very traditional trial format—no notetaking, no juror-generated questions of witnesses, and no written materials to guide the jurors. Others experienced different combinations of jury trial reform techniques (including notetaking, asking questions about the scientific evidence to experts (contacted by phone through the judge), using a checklist, and jury notebooks). The impact of the reform techniques is not a central concern of this article, so the specific way the reform techniques were implemented will not be described further. A detailed account may be found in the research project's technical report. See Dann et al., *supra* note 1.

<sup>22</sup>776 A.2d 1091 (Conn. 2001) (discussed in Marlan D. Walker, Note, Mitochondrial DNA Evidence in *State v. Pappas*, 43 *Jurimetrics J.* 427 (2003)).

testing uses the extremely long DNA strands coiled in cell nuclei.<sup>23</sup> When the quantity of nuclear DNA suitable for analysis is too small, however, mitochondrial DNA analysis might be used. MtDNA typing works with samples that are too small for nuclear-DNA testing because there are hundreds or thousands of mitochondria in a single cell, but only one nucleus. Hence, far fewer cells are needed for mtDNA testing. MtDNA sequencing can be conducted with DNA extracted from teeth, bones, and even a few strands of human hair. Because the mitochondria in all cells are copied from those of the zygote (the fertilized egg cell whose mitochondria come entirely from the mother), all individuals in the maternal line of descent have the same mtDNA sequence. Although mtDNA is therefore less individualizing than is nuclear DNA, it is a useful forensic identification tool that has been presented in many courts.<sup>24</sup>

We adapted material from the trial transcript and the reported decision of the Connecticut Supreme Court upholding both the admission of the mtDNA evidence and the defendant's conviction. The mock trial, filmed in a courtroom, included introductory instructions by a judge, opening statements by a prosecutor and a defense attorney, witness testimony, including experts for each side who discussed mtDNA evidence,<sup>25</sup> closing arguments, and legal instructions by the judge. This trial pertained to an armed robbery of a bank by a lone, masked gunman who fled from the bank, pursued by local police. The police found the robber's discarded blue sweatshirt and stolen currency. Laboratory examination of the sweatshirt revealed two human head hairs in the hood. An anonymous call directed the police to the defendant. The police collected a sample of his head hair. The two samples of hair were sent to the FBI crime laboratory for DNA analysis. The defendant was arrested and charged upon learning that his mtDNA matched that found in the sweatshirt hairs. Other evidence also incriminated the defendant, but it was relatively weak.

---

<sup>23</sup>See, e.g., Committee on DNA Technology in Forensic Science: An Update, *supra* note 17; Kaye & Sensabaugh, *supra* note 2.

<sup>24</sup>Cases involving mtDNA evidence, reflecting its growing use in criminal investigation and trial, are collected in Kaye & Sensabaugh, *supra* note 2.

<sup>25</sup>In the actual trial, the jury heard only from the prosecution's expert witness regarding the mtDNA evidence. The defendant did not present an expert, and the jury convicted him. In our film, the prosecution expert used PowerPoint slides adapted from ones used by the FBI; and we designed additional slides for the defense expert witness.

In our film, the FBI analyst testified on direct examination that only 1 in 5,072 Caucasian men have mtDNA types that match that of hairs from the sweatshirt, meaning that 99.98 percent of that population would be excluded as a possible source of the hairs. On cross-examination, he agreed that for the metropolitan area in which the robbery occurred, this would leave “six white males as the possible source of those hairs.” The defense expert, a genetics professor, questioned the FBI calculation on two prime grounds. First, the FBI’s database of 5,071 mtDNA types of Caucasian men was not a random sample. Second, the prosecution expert failed to account for possible heteroplasmy in its calculation of the likelihood of a match.<sup>26</sup> “Heteroplasmy” occurs when an individual has more than one type of mtDNA. A mutation, usually at one base pair, sometimes gives rise to two detectable types in the same individual. A small sample of cells might show one type but not the other. To avoid excluding a heteroplasmic individual (whose crime-scene sample might show one mtDNA type and whose reference sample might show the other type), the FBI does not count a difference at a single base pair as an “exclusion.”

Explaining this approach by the FBI, the defense expert observed that the random-match probability is affected by the number of near matches as well as by the number of exact matches. She found nine instances in which the crime-scene sample was one base pair off from a sample in the FBI database. The FBI would not have excluded the nine individuals who provided these samples in its database. This set of nonexclusions has to be taken into account in figuring the likelihood of a match, said the defense expert. She asserted that even granting the appropriateness of the FBI reference database, “some 10 out of every 5,072 white men in Middletown would have a DNA sequence that could not be excluded from that of the hairs in this case.” Applying a nonexclusion proportion of 10/5,072, or 0.2 percent, to the town’s white male population, she concluded that “about 57 Caucasian males in the Middletown metro area could not be excluded as the source of the hairs found on the sweatshirt.”

Unlike several previous studies, we had the prosecution expert and the prosecutor refrain from suggesting the “prosecutor’s fallacy” to the jury. This adheres to judicial admonitions.<sup>27</sup> We hoped to determine how often jurors

---

<sup>26</sup>T. Melton, *Mitochondrial DNA Heteroplasmy*, 16 *Forensic Sci. Rev.* 1 (2004). Consequently, the FBI does not consider a difference at a single base pair sufficient to exclude a suspect.

<sup>27</sup>See, e.g., *United States v. Morrow*, 374 F. Supp. 2d 51 (D.D.C. 2005) (“[T]he Government must be ‘careful to frame the DNA profiling statistics presented at trial as the probability of a

would, on their own, come to that mistaken conclusion. On the other hand, defense counsel explicitly argued the “defense attorney’s fallacy” in asking for an acquittal. The expert testimony and the arguments on these statistical issues are reproduced in Appendix A.

## IV. JURORS’ EVALUATIONS OF STATISTICAL ARGUMENTS

### A. *The Relevance of Heteroplasmy*

One of the more complicated issues presented by the mtDNA experts, and a major focus of the defense expert, was that the FBI’s estimate of the percentage of matching people ignores the fact that due to heteroplasmy, men who differ at a single base pair cannot be excluded as possible matches.<sup>28</sup> The FBI analyst ignored heteroplasmy when he testified that only 1 in 5,072 Caucasian men have mtDNA types that match that of hairs from the sweat-shirt. The prosecutor claimed, in his closing argument, that heteroplasmy is irrelevant because the defendant himself was not shown to be heteroplasmic. However, as the defense expert noted, and as we have described above, whether the defendant is heteroplasmic is beside the point, for one must still consider the possibility of heteroplasmy in calculating the probability of a match. Because the FBI does not count a difference at a single base pair as an exclusion, the random-match probability is affected by the number of near matches as well as the number of exact matches. By and large, the jurors seemed to appreciate this fact. About two-thirds of them, responding after deliberation, asserted it was necessary to consider heteroplasmy even though the defendant is not heteroplasmic.

### B. *Exclusion Probabilities*

The prosecution expert testified that the mtDNA evidence excluded 99.98 percent of the population as the source of the hairs. Because the defense

---

random match, not the probability of the defendant’s innocence that is the crux of the prosecutor’s fallacy.’ While this is a very real danger, the courts that have dealt with this potential problem have found that careful oversight by the district court and proper explanation can easily thwart this issue.” (internal citations omitted).

<sup>28</sup>See *supra* note 26.

expert included single base pair differences as matches, she spoke of an RMP of 10/5,072, which is 10 times larger than the prosecution's figure. She restated this as 0.2 percent, which equates to an exclusion proportion of 99.8 percent (although she did not make this last conversion). Thus, the expert testimony as a whole established that more than 99 percent of the population would have been excluded by mtDNA testing. Yet, only 69 percent of the mock jurors agreed with the proposition that "the mtDNA evidence in this case excludes at least 99% of the population as the source of the hairs."

Why did nearly a third of the jurors not accept the figure of "at least 99%" for the proportion excluded? Some of them may not have remembered or understood the figures given by the experts, but other possible explanations are consistent with greater juror comprehension. A juror could have understood that the mtDNA match, on its face, excluded "more than 99% of the population," but still believed that this is not what "the mtDNA evidence" as a whole "established." The defense expert questioned the value of the FBI database in estimating the frequency of Caucasian mtDNA types, and counsel's cross-examination of the FBI expert elicited the concession that the database was not a random sample and was not from the region in which the robbery took place. In deliberations, some jurors were skeptical of the extrapolation from the database to the town of Middletown.<sup>29</sup> For example, one juror pointed out that:

the FBI got 5,072 different individuals, got their DNA samples. We don't know if they were out in San Francisco and got 5,072 DNA samples from Latin Americans only and then decided to use that database to characterize the population in Middletown, or if it was a random sample across the country. We have no idea whether that database that they have is actually representative of the population in Middletown or not. For all we know there could be three major families in Middletown that have been there for hundreds of years that have lots of family members in the area and, consequently, you see a lot of the DNA pattern out of that particular city . . .<sup>30</sup>

In the same vein, jurors might also have worried that even if the national database were representative of Middletown, random sampling error could

---

<sup>29</sup>Cf. Ronald J. Allen & Michael S. Pardo, *The Problematic Value of Mathematical Models of Evidence*, 36 *J. Legal Stud.* 107 (2007) (suggesting that the use of a database to draw inferences in a specific case is always open to debate).

<sup>30</sup>Statements of jurors made in the course of deliberations are taken from Erin Jennifer Farley, *Deliberating Science: Juries, Scientific Evidence and Commonsense Justice* (2006) (Ph.D. dissertation, University of Delaware).

make the RMP in the Middletown population larger than that in the FBI's national sample. Although this type of statistical error was not mentioned explicitly in testimony or argument, during jury deliberations various jurors commented on the small size of the FBI sample. For instance, one juror exclaimed, "[a] database of 5,000 people—how many people are in the United States? 5,000 is nothing." Another invoked his professional knowledge to declare that "I'm a research scientist by trade, and when they started talking about mitochondrial DNA, I thought, 'Oh boy this is good stuff!' It's very timely, you know, but I can tell you that their percentages could change if they measured DNA samples from 10,000 people rather than 5,000 people."

Concern over laboratory contamination or other sources of laboratory error also could have led jurors to discount the 99+ percent figure. Some previous research indicates that undergraduate students believe that laboratories frequently err in declaring DNA matches.<sup>31</sup> Here, counsel questioned the FBI expert about possible contamination, but the defense made no overt argument that contamination occurred in this case. Other analyses reported elsewhere confirm that a significant minority of jurors were concerned about contamination of the mtDNA material in the present case.<sup>32</sup> Thus, some jurors probably rejected the uncontested 99 percent exclusion probability not because they failed to understand the testimony about the database and the exclusion rate, but because they had doubts about the reliability of the mitochondrial DNA evidence itself.

### C. Transposition

Even if the probability of exclusion were perceived as 99 percent or more, it would not necessarily follow that this same number is the probability that the hairs came from the defendant. As discussed in Section I of this article, using the same number for both probabilities confuses the conditional probability of a match in the mtDNA ( $M$ ) given an innocent suspect ( $I$ ) with the conditional probability of innocence ( $I$ ) given a match ( $M$ ). To restate the conclusion given earlier, if the suspect is innocent, then the probability that his DNA will match is  $P(M|I) = 1$  percent. The inverse probability,  $P(I|M)$ ,

---

<sup>31</sup>Schklar & Diamond, *supra* note 10, at 165 (reporting that students believed that nuclear-DNA matches result from laboratory error 10 percent of the time).

<sup>32</sup>Valerie P. Hans, David H. Kaye, B. Michael Dann, Erin J. Farley & Stephanie Albertson, *Science in the Jury Box: Juror Comprehension of the Biology of mtDNA Evidence*. Unpublished manuscript, Cornell Law School (May 8, 2007).

could be quite different. If the suspect matches, then all that can be said (from the fact of the match alone) is that he is included in the class of people who might have left DNA at the crime scene. This group might be large, or it might be small. If its size is denoted by  $N$ , and if every individual in the group is equally likely to have left the DNA, then  $P(I|M) = 1/N$ . For example, if five individuals have matching DNA and if there is no other evidence pointing to any one of them, then the chance that any named individual—including the suspect—is innocent is  $4/5 = 80$  percent. This is a far cry from the 1 percent figure that comes from transposing the conditional probability of a match given innocence.

In our study, nearly half (48 percent) the mock jurors agreed with the proposition that “the mtDNA evidence in this case shows there is about a 1% chance that someone else besides the defendant committed the crime.” This figure surely overstates the incidence of the transposition fallacy. We do not know how many of these jurors also believed that the probability of a random match was about 1 percent. We know that the bulk of the jurors (330 out of 480) agreed that the probability of exclusion, which is essentially  $P(\bar{M}|I)$ , was “at least 99%,” but we cannot say how many of them would translate “at least” into “about.” Moreover, jurors could have arrived at the figure of “about 1%” without transposing. For example, a juror who reacts as prescribed by Bayes’s rule could reason as follows.

Given the other evidence in the case, the prior odds of innocence are 1. There is essentially no chance that the laboratory mistyped the samples. Hence, the posterior odds of innocence are given by the likelihood ratio  $P(M|I)/1$ . Using the prosecutor’s figure of 1 percent for  $P(M|I)$ , the posterior odds are then  $1/100$ , corresponding to a posterior probability,  $P(I|M)$ , of  $1/101$ , which is “about 1%.”

Because we did not elicit subjective probabilities for other evidence from the jurors, we cannot directly estimate the incidence of the transposition fallacy from the answer to the question about the probability of innocence.

Nonetheless, some clue as to the incidence of the transposition fallacy might be teased out of the jurors’ answers. Of the 330 jurors who agreed that the exclusion probability was at least 0.99, 180 also stated that the mtDNA evidence showed about a 0.01 chance that the hairs were not the defendant’s. This suggests that, at the very most,  $180/330 = 55$  percent of the jurors were transposing. The actual rate is lower to the extent that some of the 55 percent arrived at the 0.01 figure by another route. A very rough indication of this “background rate” for 0.01 can be obtained by considering the fraction of the



jurors who did not accept the 0.99 figure and accepted the 0.01 figure. Among the 109 jurors not subscribing to 0.99 for the exclusion probability, 42 (39 percent) adopted the 0.01 figure. If this background rate can be applied to the group that did subscribe to 0.99 as well—admittedly a big “if”—then the rate for transposition is the total who could be transposing (55 percent) minus the background rate for arriving at 0.01 (39 percent), which is 16 percent.

This percentage is at the high end of the range from previous studies, but direct comparisons are clouded by the varying experimental conditions. Earlier studies involved different exclusion (or inclusion) probabilities and various ways of describing and presenting them to mock jurors or students. The first experiments involved forms of trace evidence other than DNA. Researchers reported that the overall incidence of apparent transposition was less than 2 percent (aggregated across frequencies of 1/10, 1/20, 1/100, and 1/1,000),<sup>33</sup> 5 percent (when the subjects were told both that “only two percent of people have hair that would be indistinguishable” and that “in a city of one million, there would be approximately 20,000 such individuals”),<sup>34</sup> 8 percent (reporting that “the scientific evidence established that there was a 99% likelihood that the fire was deliberate” when the random-match probability was framed as 1/1,000 and as one-tenth of 1 percent),<sup>35</sup> and 22 percent (when informed of “a two percent chance that the defendant’s hair would be indistinguishable . . . if he were innocent”).<sup>36</sup> For DNA evidence, the incidence has been estimated at a scant 0.3 percent (for a frequency of 0.0025 percent)<sup>37</sup> and 6 percent (for a frequency of 0.04, or one in 25).<sup>38</sup>

---

<sup>33</sup>Goodman, *supra* note 10, at 375.

<sup>34</sup>Thompson & Schumann, *supra* note 10, at 173–74.

<sup>35</sup>Goodman, *supra* note 10, at 384.

<sup>36</sup>Thompson & Schumann, *supra* note 10, at 173–74.

<sup>37</sup>Nance and Morris told their subjects that “the DNA profile in question can be found in only 25 ten-thousandths of one percent—.0025 percent—of the male population. In other words, on average we would expect to see the DNA profile . . . in one out of every 40,000 randomly selected men in the population.” Nance & Morris II, *supra* note 10, at 424 n.40. They observed that only 0.3 percent of their subjects apparently translated this into a probability of guilty of precisely 99.9975 percent.

<sup>38</sup>Further refinement of the measure indicated that the 6 percent figure likely overstated the incidence. Nance & Morris I, *supra* note 10, at 429, 431–34.

*D. Undervaluation and the Relevance Fallacy*

Without knowing the prior odds for each juror, we cannot evaluate the extent to which the jurors overrely or underrely on mtDNA match evidence.<sup>39</sup> Even so, there is reason to think that a considerable number of jurors succumbed to the relevance fallacy. The defense attorney here claimed that the mtDNA evidence is completely irrelevant because a substantial number of other people could also be the source of the hairs. In fact, the match is relevant, even though its probative value might be debatable or might depend on whether heteroplasmy is considered in assessing the likelihood of a match. Yet, when the jurors answered a question reiterating the defense attorney's argument that the mtDNA evidence is completely irrelevant, only half rejected the claim, and 40 percent agreed with the proposition that the mtDNA evidence in this case is entirely worthless because people other than the defendant could have contributed the hairs.<sup>40</sup>

This percentage is larger than previous reports of the incidence of the fallacy. In some of the studies mentioned in connection with the transposition fallacy, figures of 7 percent,<sup>41</sup> 8 percent,<sup>42</sup> and 17 percent<sup>43</sup> were given for the incidence of the relevance fallacy. Again, it is difficult to make direct comparisons in view of the differences in the design of the studies. Moreover, the fallacy rarely emerged unequivocally in deliberations. A few jurors advanced the argument, as when one explained: "Well, I'm not even relying on the DNA evidence because in my opinion its still circumstantial for the simple fact is you still have other people that could possibly match that. So I'm not even relying on the DNA at all." More commonly, however, jurors made the subtler point that the number of other men in the area with

---

<sup>39</sup>Other studies that have examined this question have asked about the participants' subjective probabilities of guilt based on different pieces of evidence prior to learning about DNA analysis results. Given our desire to measure the impact of trial innovations on juror assessments of the evidence, we did not use this artificial approach.

<sup>40</sup>The numbers did not change appreciably after deliberations. The question read as follows: "The mtDNA evidence in this case is completely irrelevant because a substantial number of other people could also be the source of the hairs."

<sup>41</sup>Nance & Morris I, *supra* note 10, at 433–34.

<sup>42</sup>Thompson & Schumann, *supra* note 10, at 174.

<sup>43</sup>*Id.*

matching mtDNA types established reasonable doubt. One juror explained his initial thinking as follows.

My opinion is based on the judge's instructions and that is the fact that the prosecution did not prove to me without a shadow of a doubt that he was guilty. Based on DNA evidence by both the . . . doctor and the FBI expert, there was a possibility, whether you look at it one in six or one in fifty-seven, that there could possibly be more than one person out there. That's my thought.

Although one might question this reasoning, it is not a claim that the match is logically irrelevant, and observations like these often generated discussion of how to combine the less-than-conclusive DNA match with the other evidence in the case. For example, one juror remarked:

I think the defense's expert concluded that there might be 57 other . . . males that would have that. But that is not 57 males of six foot tall, 170 pounds with a scar on their face. . . . [A]nd with that DNA match it really narrows it down. Now . . . if you two feel so strongly that there's not enough evidence, I think you ought to stick with your verdict. But I think the evidence is there myself . . .

Thus, the binary responses in the questionnaire could be giving an exaggerated impression of the extent to which the relevance fallacy affected the final verdicts.

#### *E. Relationship to Predeliberation Verdicts*

As one would expect, there was a connection between initial jurors' verdicts and their acceptance of various statements about the statistical import of the mtDNA testing.<sup>44</sup> Table 2 shows that more of the jurors initially inclined to vote guilty accepted the prosecution's position that at least 99 percent of the population was excluded, and fewer agreed with the defendant's position that the mtDNA evidence only established a 1 in 57 chance that he was the robber.<sup>45</sup>

---

<sup>44</sup>The average probability, the 1 percent chance, and the 1 in 57 items were all significantly related to verdict choice; the 99 percent exclusion-proportion question was related at the 0.055 level of statistical significance.

<sup>45</sup>At the same time, there was essentially no difference in the proportions who agree that the mtDNA evidence showed about a 1 percent chance that someone else was the robber—the statement that results from transposing the conditional probability related to the 99 percent figure in the first statement.

Table 2: Jurors’ Initial Verdict Choices and Their Views on mtDNA Statistics

<i>Percent Who Agree with Statement that</i>	<i>All</i>	<i>Guilty</i>	<i>Not Guilty</i>	<i>Unsure</i>
“The mtDNA evidence . . . excludes at least 99% of the population as the source of the hairs.”	69	74	62	72
“The mtDNA evidence . . . shows . . . only a 1 in 57 chance that the defendant committed the crime.”	47	38	57	54
“The mtDNA evidence . . . shows . . . about a 1% chance that someone else . . . committed the crime.”	49	46	47	60

## V. ESTIMATES OF THE PROBABILITY OF GUILT

### A. *The Effect of Deliberations*

A majority of jurors thought that there was a substantial probability that the defendant was the robber, but these estimates varied widely.<sup>46</sup> Figure 1A displays the frequency distribution of the jurors’ probability estimates elicited before the deliberations; Figure 1B displays the postdeliberation distribution.<sup>47</sup> Comparing the two figures, one can see the shift that occurs from the preliminary judgments to the postdeliberation judgments that the defendant is the robber. With deliberations, jurors shift toward greater certainty that the defendant is the robber. Before deliberations, there is a peak around 50 percent, which probably reflects “a default tendency to characterize a general state of uncertainty, when it is present, in terms like ‘50 : 50.’”<sup>48</sup> This peak is much smaller after deliberations. Since the distribution shifts to the right, this suggests that deliberations led the jurors who were uncertain to be more confident that the defendant is indeed guilty. The

<sup>46</sup>The precise question, posed both before and after deliberations, read as follows: “In your view, what is the likelihood that the defendant Kevin Jones is actually the person who committed the robbery? \_\_\_%.”

<sup>47</sup>The probability estimates and their percentiles are given in Appendix B.

<sup>48</sup>Nance & Morris I, *supra* note 10, at 417. The same “spike” is present in the juror assessments in Nance & Morris II, *supra* note 10, at 407. These studies did not include deliberations, which seem to be important in this regard.

Figure 1A: Distribution of jurors' probabilities of guilt elicited before deliberations.

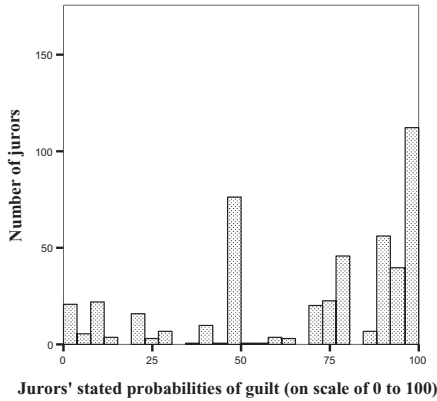
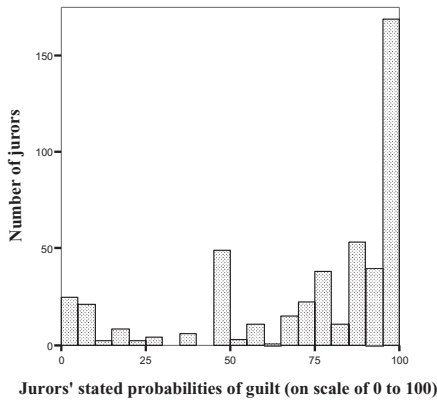


Figure 1B: Distribution of jurors' probabilities of guilt elicited after deliberations.



change from pre- to postdeliberation assessments could reflect the impact of juror discussions on how to combine the DNA evidence with the other information in the case.<sup>49</sup>

<sup>49</sup>See supra Section IV.D.

Table 3: Jurors’ Verdict Choices and Their Probability Estimates

	<i>Initial Verdict Probability:Final Verdict Probability</i>			
	<i>All</i>	<i>Guilty</i>	<i>Not Guilty</i>	<i>Unsure</i>
Mean probability that defendant is the robber	0.69 : 0.75	0.90 : 0.90	0.42 : 0.59.	0.64 : 0.84
Probability for the 25th percentile	0.50 : 0.55	0.85 : 0.90	0.19 : 0.40	0.50 : 0.68
Median probability	0.80 : 0.90	0.95 : 0.99	0.50 : 0.63	0.65 : 0.93
Probability for the 75th percentile	0.95 : 0.99	0.99 : 1.0	0.61 : 0.89	0.80 : 1.0

As Table 3 shows, the mean probability given before deliberation was 0.69, and the median was 0.80. After deliberation, the mean was 0.75, and the median was 0.90.<sup>50</sup>

*B. Stated Probabilities and Verdicts*

Not surprisingly, the stated probability of guilt was related to the initial verdict choice. Those who initially voted guilty on average judged the probability at 0.90. Those who initially voted not guilty rated the probability on average at 0.42; those who were unsure fell in between, at 0.64.<sup>51</sup> The pattern after deliberation was similar, with the mean probability shifted upward among the jurors who initially were inclined to convict, unsure, or inclined to acquit (0.90, 0.59, 0.84 for the three groups, respectively). To convey a more complete picture of the different distributions for the two final verdict choices, Figure 2 shows the distributions of source probabilities for those jurors voting to acquit and those voting to convict.<sup>52</sup>

<sup>50</sup>A similar mean shift following deliberations is reported in Francis C. Dane, In Search of Reasonable Doubt: A Systematic Examination of Selected Quantification Approaches, 9 Law & Hum. Behav. 141, 151 (1985) (reporting a simulation study with psychology students whose mean “probability of commission” as expressed on a “21-point percentage scale” rose from 0.66 to 0.70 following deliberations).

<sup>51</sup> $F(2, 478) = 211.49, p = 0.0001$ .

<sup>52</sup>Figure B1, in Appendix B, provides side-by-side histograms for the same data.



and 0.79 for mock jurors.<sup>54</sup> Kerr et al. found a mean threshold probability of 0.82 for subjects who received a “lax” definition of reasonable doubt (or none at all), and 0.87 for those who received a “stringent” instruction.<sup>55</sup> Nagel reported that students’ “desirable threshold probability came close[] to .80 to .90.”<sup>56</sup>

Of course, the stated values might not reflect what jurors do in practice. Accordingly, there have been efforts to infer the critical probability actually employed by examining the verdicts from mock jurors along with their stated probabilities of guilt. It is tempting to say that the threshold for “proof beyond a reasonable doubt” is simply the smallest stated probability for which a large fraction of jurors vote to convict. Reasoning that the percentage of jurors voting to convict should rise as their stated probability of guilt increases, Dane cumulated votes to convict as a function of the jurors’ subjective probabilities, and he adopted the perceived probability associated with a 75 percent cumulative level of guilty verdicts as the point that corresponds to proof beyond a reasonable doubt.<sup>57</sup> For his mock trial, this height of the “ogive function”<sup>58</sup> gave a critical probability of 0.88 both before and after deliberation.<sup>59</sup>

The probability inferred from the cumulative frequency of guilty votes, however, is a problematic measure of the jurors’ view of reasonable doubt. A numerical example shows why. Table 4 presents a hypothetical distribution in which equal numbers of mock jurors are situated at each subjective probability. The cumulative proportion of jurors voting for conviction reaches 75 percent when the subjective probability is between 0.75 and 0.80.

---

<sup>54</sup>Rita James Simon & Linda Mahan, *Quantifying Burdens of Proof: A View from the Bench, the Jury, and the Classroom*, 5 *Law & Soc’y Rev.* 319, 324 (1971). The undergraduate sociology students who deliberated as six-member juries for 30 minutes in a simulation of a homicide case had the same mean as individual judges (0.89). The median threshold probability was 0.88 for judges, 0.86 for jurors, and 0.91 for the students. *Id.*

<sup>55</sup>N.L. Kerr et al., *Guilty Beyond a Reasonable Doubt: Effects of Concept Definition and Assigned Decision Rule on the Judgments of Mock Jurors*, 34 *J. Personality & Social Psychol.* 282 (1976).

<sup>56</sup>Stuart Nagel, *Bringing the Values of Jurors in Line with the Law*, 63 *Judicature* 189 (1979).

<sup>57</sup>Dane, *supra* note 50.

<sup>58</sup>*Id.* at 145. The ogive is a cumulative frequency polygon.

<sup>59</sup>*Id.* at 152–53.



Table 4: Cumulative Proportion of Jurors Voting to Convict for Equal Numbers of Jurors at Each Estimated Probability of Guilt

<i>Juror's Probability of Guilt</i>	<i>Number of Jurors</i>	<i>Number Voting to Convict</i>	<i>Proportion Voting to Convict</i>	<i>Cumulative Proportion Voting to Convict</i>
0.99	100	98	0.98	0.980
0.95	100	75	0.75	0.865
0.90	100	60	0.60	0.776
0.85	100	40	0.40	0.683
0.80	100	20	0.20	0.586
0.75	100	5	0.05	0.496
0.70	100	1	0.01	0.427
0.65	100	0	0.00	0.374
0.60	100	0	0.00	0.332

NOTE: The resulting “critical probability” for a conviction rate of 75 percent is between 0.85 and 0.90.

Table 5: Cumulative Proportion of Jurors Voting to Convict When More Jurors Estimate a Probability of Guilt of 0.99

<i>Juror's Probability of Guilt</i>	<i>Number of Jurors</i>	<i>Number Voting to Convict</i>	<i>Proportion Voting to Convict</i>	<i>Cumulative Proportion Voting to Convict</i>
0.99	300	294	0.98	0.98
0.95	100	75	0.75	0.92
0.90	100	60	0.60	0.86
0.85	100	40	0.40	0.78
0.80	100	20	0.20	0.70
0.75	100	5	0.05	0.62
0.70	100	1	0.01	0.55
0.65	100	0	0.00	0.495
0.60	100	0	0.00	0.45

NOTE: The “critical probability” is lower although the proportion of jurors voting to convict is identical at every probability.

Table 5 presents a distribution that differs in only one respect—more jurors estimate the probability of guilt to be 0.99. This fact has no bearing on the jurors’ implicit burden of persuasion, since the rates of conviction at each probability are the same. Yet, the cumulative approach (for a 75 percent conviction rate) now yields a critical probability between 0.80 and 0.85.

A valid measure of the jurors’ implicit range for proof beyond a reasonable doubt must reveal the zone in which jurors confronted with cases that give rise to particular probabilities of guilt are sufficiently likely to vote to

convict. In other words, we must consider  $P(c|p)$ , the conditional probability of a vote to convict ( $c = 1$ ) given the subjective probability of guilt ( $p$ ). The idea here is that when the jurors' conviction probability  $P$  is large enough, most jurors, attempting to follow the judge's instructions, apparently feel that  $p$  is in the range that leaves no room for reasonable doubt. Operationalizing this notion to uncover the values of  $p$  that correspond to proof beyond a reasonable doubt therefore requires us to specify a value of  $P$  that is "large enough" and to ascertain the function  $P(c = 1|p)$  that indicates the juror's willingness to convict for a given subjective probability of guilt.

Nance and Morris used this concept to conclude that "the burden of persuasion implicitly employed by the jury as an institution, as distinct from the measure employed by individual jurors . . . lies somewhere between 75 and 85 percent . . . ." <sup>60</sup> They approximated  $P(c = 1|p)$  as the proportion of jurors voting to convict at various ranges for  $p$ —for  $p$  between 0.05 and 0.15, between 0.15 and 0.25, and so on. In the lower probability-of-guilt intervals, most jurors voted to acquit. When the self-reported probabilities of guilt fell between 0.75 and 0.85, however, a juror was more likely to convict than to acquit. Nance and Morris regarded this zone as "an interesting measure of the burden of proof."<sup>61</sup> For our predeliberation data, the majority of jurors in the intervals (centered at  $p = 0.10, 0.20, . . . , 0.90$ ) do not vote to convict until the subjective probability of guilt falls into the same range—between 0.75 to 0.85. After deliberation, the range for "proof beyond a reasonable doubt" rises to the next interval—0.85 to 0.95.

This approach is an improvement over Dane's ogive procedure because it is a way to ascertain the jurors' willingness to convict conditional on their subjective probability of guilt.<sup>62</sup> It is an instance of examining a dependent variable  $Y$  conditional on an independent variable  $X$  by grouping the  $X$  values into increasing intervals and finding the mean  $Y$  for each interval. This slice-and-average procedure can indicate how  $Y$  changes (on average) as a function of  $X$ . Its appeal is that it does not presuppose any particular functional form; however, this nonparametric procedure has a

---

<sup>60</sup>Nance & Morris II, *supra* note 10, at 407; see also Nance & Morris I, *supra* note 10, at 417 (reporting the same "crossover" point in this earlier study).

<sup>61</sup>Nance & Morris I, *supra* note 10, at 417.

<sup>62</sup>In technical terms, Nance and Morris used simple frequency polygons instead of the cumulative frequency polygon (the ogive) that Dane employed, and they used a cut-off point of 50 percent rather than 75 percent.

drawback. Only a small fraction of the data can be used to estimate the mean  $Y$  in each interval. This can make the estimated means for each  $X$  slice unreliable. If we have a good basis for believing that the functional relationship between  $X$  and  $Y$  is linear, then simple linear regression is better because only two parameters (the intercept and the slope of the true regression line) need to be estimated and this can be done in light of all the data on  $Y$  and  $X$ . Hence, if the linear relationship holds, the regression line provides more reliable estimates of  $Y$  conditioned on  $X$ .

The relationship between the two variables here— $c$ , a conviction, and  $p$ , the juror's perceived probability of guilt—is not linear, but the regression can be done with a functional form that accounts for the fact that the average of  $c$  (the proportion of convictions) tends to be near zero when  $p$  is small, to rise quickly in the midrange of  $p$ , and to level off again near one when  $p$  is large. One function that has this shape is the logistic curve:

$$c = 1/[1 + e^{-(\alpha + \beta p)}]. \quad (2.1)$$

The expression can be transformed to a linear form by putting  $c$  on a log-odds scale to obtain:

$$\ln(\text{odds}) = \alpha + \beta p, \quad (2.2)$$

where the odds are the proportion of convictions divided by its complement.

Equation (2.2) states that the natural logarithm of the odds of a conviction is directly proportional to  $p$ . When jurors report that the chance that the defendant is guilty is zero, the logarithm of the odds of a conviction is given by the parameter  $\alpha$  (which could be zero). When  $p$  is one (jurors report that they are certain of guilt), the log-odds of a vote to convict is  $\alpha + \beta$ . The parameter is the slope of log-odds line.

These two parameters can be estimated from the full range of data by regressing the log-odds form of  $c$  on  $p$ . As Table 6 indicates, the estimated logistic function for the postdeliberation data is such that:

$$\ln(\text{odds}) = -3.44 + 0.045p. \quad (3)$$

---

<sup>63</sup>About 92 percent and 81 percent of the pairs formed with the post- and pre-deliberation equations, respectively, are concordant.

Table 6: Implicit Burden of Persuasion as Ascertained by a Logistic Regression of Verdicts on Stated Probability of Guilt

	<i>Intercept</i>	<i>Slope</i>	<i>Burden of Persuasion (<math>p_{50}</math>)</i>
Predeliberation estimate s.e.	-5.31 (0.61)	0.078 (0.0077)	0.68
Postdeliberation estimate s.e.	-3.44 (0.44)	0.045 (0.0051)	0.76

Now let  $p_{50}$  denote the guilt probability  $p$  at which the odds of a vote to convict are one-to-one. (This is to say that the conditional probability  $P$  of convicting equals 0.50 when  $p = p_{50}$ .) A juror whose probability of voting to convict is 0.50 is right at the borderline between acquitting and convicting. On average, a higher subjective probability  $p$  makes a conviction more probable than not ( $P > 1/2$ ). Substituting odds of 1 into Equation (3) and solving for  $p$ , we conclude that when the stated probability is greater than 0.76, or 76 percent, the odds are that a juror will vote to convict after deliberating.<sup>64</sup>

Although we have followed Nance and Morris in using a conviction probability of  $P(c = 1|p) = 1/2$  to mark the point at which proof beyond a reasonable doubt exists in the minds of most jurors, one might be more demanding. For instance, if we say that it takes a value of  $P(c = 1|p) = 3/4$  to discern the  $p$  that is beyond reasonable doubt for enough jurors, then according to our postdeliberation regression equation, this subjective probability  $p_{75}$  is 0.997.

These estimates of the subjective probability at which a juror applying the beyond a reasonable doubt standard is sufficiently likely to convict for there to be proof beyond a reasonable doubt in that juror’s mind are hardly precise. The postdeliberation regression equation easily could depart from the estimate of  $\ln(\text{odds}) = -3.44 + 0.045p$ . The standard error (s.e.) of the intercept is 0.44, and the standard error of the slope is 0.0051. Suppose, then, that the true intercept were just one standard error below our point estimate ( $-3.44 - 0.44$ ), and the true slope were merely one standard error higher than our estimate ( $0.045 + 0.0051$ ). Setting  $\ln(\text{odds}) = 0$  and solving for  $p$  now yields a  $p_{50}$  of 0.59. Thus, a stated probability of 0.59, or 59 percent,

<sup>64</sup>When the probability of a vote to convict is one-half, the odds are 1, and the log-odds are 0. Solving  $0 = -3.44 + 0.045p$  gives  $p_{50} = 0.76$ . For predeliberation data,  $p_{50} = 0.68$ .

lies within a confidence interval for the probability that the jurors need to find guilt beyond a reasonable doubt. At the other end, when the estimate for the intercept is increased by just one standard error, and the estimate for the slope is reduced by one standard error, the threshold probability for a conviction rises to 96 percent. The true value for the subjective probability that is enough to satisfy more than half the jurors that there is proof beyond a reasonable doubt could be considerably larger or smaller than our estimate of 0.76. Hence, our study does not give a firm value for the probability that jurors implicitly treated as satisfying the prosecutor's burden of persuasion. Nevertheless, the logistic regression model can be a useful supplement to previous methods for quantifying "proof beyond a reasonable doubt."

#### *D. Implications for the Law of Evidence*

The fact that the jurors' estimates of the probabilities of guilt in Figures 1A and 1B are well distributed across the spectrum indicates that the mtDNA evidence did not overwhelm most jurors. A juror who perceived the DNA tests as essentially conclusive would be expected to provide a value close to 100 percent for the "probability that the defendant Kevin Jones is actually the person who committed the robbery." Yet, only about a third of the jurors placed this source probability in the range of 0.90 or higher, and only about one-quarter ventured above 95 percent.<sup>65</sup> These moderate percentages occurred notwithstanding the fact that nearly half the jurors also accepted as true the statement that the mtDNA evidence showed only about a 1 percent chance that someone else was the robber. These are the kinds of responses that would be expected of jurors who thought the DNA evidence was powerful standing alone but that the other evidence in the case raised substantial doubt. Jurors reasoning in this fashion cannot be accused of thinking that mtDNA evidence is so powerful that it disposes of the issue of identity.<sup>66</sup>

---

<sup>65</sup>See Appendix B.

<sup>66</sup>That only 48 percent of the jurors preferred verdicts of guilty before deliberating and only 50 percent did so following deliberations further attests to the limited impact of the mtDNA evidence in the mock juror case. Cf. Koehler et al., *supra* note 10 (reporting conviction rates of only 44 percent and 54 percent when jurors were given a random-match probability of one in a billion).

## VI. SUMMARY AND IMPLICATIONS

In assessing the statistical evidence relating to the match, most jurors did not make obvious mistakes. The jurors generally—and correctly—accepted the defense position that even though the defendant was not heteroplasmic, the exclusion probability needs to reflect the presence of heteroplasmy in the population. A more ambiguous finding is that even though the two experts and both counsel all agreed that the FBI's population database indicated that in excess of 99 percent of the male Caucasian population would be excluded, nearly a third of the jurors denied that “the mtDNA evidence in this case excludes at least 99% of the population as the source of the hairs.” Perhaps more jurors were not willing to subscribe to the 99+ percent figure because the defense challenged the validity of the database itself and many jurors harbored doubts about the “reliability” of mtDNA testing.

There are indications of both the transposition fallacy (which usually favors the prosecution) and the relevance fallacy (which usually favors the defense). Conceivably, more than half the jurors might have transposed the conditional probabilities, but the true fraction is more likely to have been closer to one-sixth or even less. As in previous studies, the relevance fallacy was more clearly evident, with 40 percent of the jurors rating a statement of it as true.

Whatever the incidence of the two commonly discussed fallacies, the jurors generally recognized the limitations of mtDNA evidence. The final probabilities of guilt supplied by the jurors suggest that they were far from overwhelmed by the mtDNA match and statistics, and during deliberations, discussion of the nature and extent of other evidence in the case was common.

These findings, together with previous research, offer some guidance in presenting probabilities in cases involving scientific evidence. As other studies have indicated, from the defense perspective, shifting the spotlight away from the defendant to some substantial number of other people who might share the incriminating genotypes can diminish the impact of the random-match probability.<sup>67</sup> This effect was apparent in some of the jury deliberations here. For example, one juror argued that:

---

<sup>67</sup>See Koehler & Macchi, *supra* note 10; Koehler II, *supra* note 10.

I think we all can see that, yeah this kid done it but you know, I really don't think there's rock solid [proof]. I certainly wouldn't want to be put away knowing that maybe 57 other people did what they're accusing me of doing, you know. I don't think anybody, you wouldn't want any of your friends you know. . . . I would want a little more iron clad than that.

Prosecutors, in turn, might consider strategies to inoculate against the more extreme “defense attorney’s fallacy” or to mitigate misaggregation or misperception error in several ways. They could argue that the evidence is very powerful because a match is so many times more probable when the source of the trace evidence is tested than when an innocent person is tested;<sup>68</sup> they could display a table that shows the impact of the match on a suitable range of prior probabilities;<sup>69</sup> or they could emphasize the probability of exclusion.<sup>70</sup>

Describing the probative value of evidence fairly and comprehensibly remains a challenge to the legal system. This challenge, however, is not so formidable as to warrant a global objection to “probability evidence.”<sup>71</sup> This study, at least, indicates that when the prosecution does not commit the transposition fallacy, when the defense urges that the match is irrelevant due to a substantial number of other people who also could match, when the scientific technique is less “reliable” than nuclear-DNA typing, and when there is a question as to the size and representativeness of the reference database, jurors do not treat an exclusion probability in excess of 99 percent as dispositive.

## APPENDIX A

### *I. Comments on Statistics in Prosecutor’s Opening Statement*

The other prosecution witness will be an expert on DNA analysis from the Federal Bureau of Investigation Crime Laboratory in Quantico, Virginia.

<sup>68</sup>See also Nance & Morris II, *supra* note 10, at 421; cf. J.J. Koehler, On Conveying the Probative Value of DNA Evidence: Frequencies, Likelihood Ratios and Error Rates, 67 U. Colo. L. Rev. 859, 880–81 (1996) (comparing the impact of presenting the RMP as a frequency and a likelihood ratio).

<sup>69</sup>See Nance & Morris II, *supra* note 10, at 421.

<sup>70</sup>See Taroni & Aitken, *supra* note 10, at 168.

<sup>71</sup>See Kaye et al., *supra* note 12, ch. 12; D.H. Kaye, Revisiting Dreyfus: A More Complete Account of a Trial by Mathematics, 90 Minn. L. Rev. 825 (2007).

Special Agent Jaye will tell you of a DNA analysis done on some human hair found on the discarded sweatshirt worn during the bank robbery and on samples of defendant's hair collected by the local police. The two DNA samples were found to match exactly. The FBI analyst will also tell you that the defendant's DNA profile is very rare, has never been observed in the FBI's database of such profiles and is different from the profiles of almost all other Caucasian people. He will explain that nearly 99.98 percent of the Caucasian population can be excluded as the possible contributors of the hairs found in the hood of the sweatshirt.

*II. Comments on Statistics in Defense Counsel's Opening Statement*

What the government has brought you is a relatively new form of DNA testing called mitochondrial DNA testing, or mtDNA for short. Both the FBI and our expert agree that mtDNA is not unique to each person and can only tell that the person that supplied that form of DNA is a member of a group, or class, of people who could have the same mtDNA. As a matter of fact, we will show that up to as many as 57 Caucasian men in the Middletown metro area could have been the source of the hair found on the prosecution's sweatshirt. A criminal conviction cannot rest on those kinds of odds and coincidences.

*III. Testimony of Agent David Jaye on Statistics*

A. Direct Examination

Q: Is there a way to tell what fraction of the population would be excluded?

A: Yes, the next slide shows how we determine how rare the sequence is. We assemble the sequences from a sample of the population into a database, and we compare the defendant's sequence with those in this database.

Q: What did you find when you made this comparison?

A: Our database consists of 5,071 Caucasian mtDNA samples collected over the past few years from hospitals, blood banks, paternity cases, and other sources. We could not find this particular sequence in our database. This was the first time we ever saw it.

Q: And what does this tell you about how rare this DNA type is?

A: Well, it was not present in the database. That means that 5,071 individuals did not have the type, and one (Mr. Jones) did. Therefore, our data indicate that this particular sequence arises in maybe 1 out of every 5,072 Caucasians.



Q: Expressed in percentage terms, Agent Jaye, what number of randomly selected Caucasians would you expect to have a different mtDNA profile than the defendant's?

A: That's the third line on the slide.  $1/5,072$  is less than two-hundredths of one percent. To put it another way, more than 99.98% of all Caucasians would be excluded. But, someone in the same maternal line as the defendant should have the same mtDNA profile and could not be excluded as the possible source.

#### B. Cross-Examination

Q: Going back to the fact that mtDNA is not as good an identifier as nuclear DNA, since mtDNA is not a unique identifier, the most you can say is that the defendant is merely one of a group, or class, of people who have the same mtDNA profile?

A: Yes, but a member of a relatively small group.

Q: Well, let's explore that a bit further. You used the database of five thousand or so individuals to arrive at this conclusion, didn't you?

A: That's right.

Q: Is the database a random sample?

A: It is a sample from different groups, like people who give blood at blood banks and people who are involved in paternity cases. So it is a convenience sample, but one that is representative of the population.

Q: Again, Agent Jaye, I would appreciate it if you would answer my questions yes or no, if you can. This is not a random sample, is it?

A: No, the FBI does not go door-to-door, picking people out at random to give DNA samples.

Q: It is not even a sample of people from Middletown, is it?

A: No, it is a national sample.

Q: Thank you. There will be testimony later that the Caucasian population of this immediate area is just under 58,000. Assume, for a moment, that about half of those are males. Isn't it true that, accepting your figures as correct, that .02% of the Caucasian population cannot be excluded as the source of the mtDNA found in the sweatshirt hairs, that leaves about 6 white males just in this metro area as the possible source of those hairs?

A: Let me see. Yes that seems right, but . . .

Q: Thank you, sir, you've answered my question. I also want to be clear about what you mean when you say that a person is not excluded. Does a defendant have to have every base pair in the same place as the questioned sample for you to declare a match?

A: Yes, they have to be identical, but from time to time, mutations occur. When this happens, an individual can have a mixture of mtDNA consisting of the original sequence and other, mutated ones that have been copied as the cell divides. If the mutated copies are numerous enough, we would find two different bases at a site. One base would be in the original sequence. The other would be in the mutated sequence. The presence of the two types in the same individual is known as “heteroplasmy.” Therefore, to be on the safe side, we would not report that a defendant is excluded if he differs from the questioned sample at only a single base pair. We regard that as inconclusive.

### C. Redirect Examination

Q: Just a few questions, your honor. Agent Jaye, have you heard anything in your cross-examination that would cause you to change the percentage of the white male population that would be excluded as the possible source of the mtDNA profile you found in the sweatshirt hairs and the hairs that were taken from the defendant?

A: No, the number is still over 99.98 percent.

Q: What about these mutations, this heteroplasmy, you were asked about?

A: That has nothing to do with this case. Both the questioned sample and the known sample had only one DNA sequence, and it was the same for each of them. I mean, there were no ambiguous base pairs at all.

## *IV. Testimony of Professor Elizabeth Allison on Statistics*

### A. Direct Examination

Q: Do you take issue with the FBI’s database of 5,071 mtDNA profiles that was used in this case?

A: I have some concerns about how representative it is, and I also disagree with some of the calculations.

Q: Could you explain about the calculations, please?

A: Of course. The most recent studies have shown that human hairs are especially prone to heteroplasmy, that is, where the mtDNA from a given person’s hair varies at least at one base pair. As many as 10 or 20 percent of white males could have heteroplasmy. Indeed, at some level, everyone probably is heteroplasmic. The FBI knows this, and that is why their protocol does not let them exclude an individual who differs at only one point in the

sequence. There is a small chance that a hair from these individuals would only show the sequence in the crime-scene sample. So rather than exclude these suspects, they report that the test is “inconclusive.” Because these individuals are not definitively excluded, they must be counted along with those who are clearly excluded. But the FBI did not do that in this case. It stopped when it found that the sequence here was not represented in its database. It should have counted all the sequences in the database that were only one base pair away from this sequence.

Q: Have you performed this computation?

A: Yes, I have an overhead that explains it. There were 9 sequences in the Caucasian database that were within one base pair of the sequence in the hairs from the sweatshirt. And there was the defendant’s sequence that also matched. Thus, the total number is 10.

Q: And what does this tell you about how many people in Middletown the DNA test would exclude?

A: Well, first, you have to make a big assumption—that the FBI’s database is representative of Caucasians in Middletown. Personally, I’d be hesitant to jump to that conclusion. But if we can make that assumption, then we could say that some 10 out of every 5,072 white men in Middletown would have a DNA sequence that could not be excluded from that of the hairs in this case. That’s almost 0.2% of the Caucasian population of Middletown.

Q: Dr. Allison, at my request, did you consult the most recent U.S. census for the Middletown area?

A: Yes, I did.

Q: And how many Caucasian males lived in the Middletown area, according to the most recent census?

A: Well, it says there was a total of about 58,000 Caucasians, just under half of whom, or about 29,000, were males.

Q: What do you conclude about the number of white men who could not be excluded by this new form of DNA testing?

A: The result is given on my second slide. Just under .2 percent of these 29,000—that’s about 57—Caucasian males in the Middletown metro area could not be excluded as the source of the hairs found on the sweatshirt.

Q: The FBI estimates that the number is only 6. In your opinion as a geneticist, is that a reasonable figure?

A: In my opinion, it is much too low. It ignores the high rate of heteroplasmy in human hairs.

B. Cross-Examination

Q: Do you agree that this defendant did not exhibit any heteroplasmy?

A: I agree that the FBI did not find that he was heteroplasmic. But the problem with the estimate of the number of people who would not be excluded has nothing to do with whether or not this particular defendant is heteroplasmic. It has to do with how the FBI counts people in its database—whether they should be counted as necessarily excluded as the possible contributor or whether they could have contributed the suspect sample. Because of the possible existence of heteroplasmy in a substantial percentage of white males, I am of the view that the FBI's figure for the percentage of the white male population that can be excluded as the source of the sweatshirt hairs is too high.

Q: You agree, do you not, that only two hundredths of a percent of the white male population has the same DNA sequence as the one that the hairs from the sweatshirt have?

A: Yes, to the extent that the FBI database is representative, that is true. But the point is that a lot more people than those can't be excluded. Their sequences are not identical, but they are so close that, when one considers the possibility of heteroplasmy, they cannot be excluded.

Q: But you do not deny that the database shows that only two hundredths of a percent have the sequence of the two hairs in this case, do you?

A: I accept the arithmetic.

Q: And you also agree that the precise DNA sequence found in both the defendant here and the sweatshirt hairs had never been recorded before in the FBI's database?

A: Yes, I agree with that.

*V. Comments on Statistics in Prosecutor's Closing Statement*

The DNA evidence also shows that the defendant's hair was found on the blue sweatshirt found along the railroad tracks just minutes after the robbery. The mtDNA found in the hair taken from him by the detective matched the DNA in the hairs found in the hood of the sweatshirt. The two samples were identical at all 610 points of comparison. His mtDNA profile is so rare that it has not been observed in the FBI's database of over 5,071 samples. Over 99.98 percent of white males can be excluded as the source of the hairs. Of course, there could be some other white males in this metro area with similar mtDNA. But the DNA evidence confirms what all the other evidence says—everything points to this defendant.

As far as the defendant's DNA witness is concerned, she simply wanted to count all DNA results that would be discarded as inconclusive as "nonexclusions." Since she agreed with the FBI analyst that the defendant is not heteroplasmic. Her conclusions and her calculations should be disregarded. They are simply not relevant to this case.

#### *VI. Comments on Statistics in Defense Counsel's Closing Statement*

First, let's take a look at the so-called "scientific" evidence in the case. The experts told us that mtDNA is not unique to an individual like nuclear DNA is—and nuclear DNA is the one we're used to hearing about. The prosecutor's DNA witness told us that he could not positively say that the defendant was the source of the DNA found in the hairs from the sweatshirt. Not only that. He also told us that anyone in the defendant's family would have the very same mtDNA.

Now, the DNA expert we called to testify for you had a very important point to make concerning the number of unrelated Caucasians one should expect to see with the same mtDNA profile as the two samples in the case. Her point is still valid even though she agrees that the defendant does not show signs of heteroplasmy. Doctor Allison told you that when you count samples in the FBI database that differ at only one base pair as "not excluded," then the percentage of Caucasians that could be excluded as the source of the mtDNA found in this case is lower than the number the state presented.

This means that about 57 other Caucasian males in the metro Middletown area—not to mention the much larger number of white men in the state or who might have been visiting in town at the time of the bank robbery—had the same mtDNA profile as the defendant and could have been the source of the hair on the sweatshirt. The defendant is just 1 of these 57 or so men who could have left the hairs on the sweatshirt. That doesn't add up to proof beyond a reasonable doubt. It is not even close.

## APPENDIX B: JURORS' PRE- AND POSTDELIBERATION PROBABILITY ESTIMATES

<i>Predeliberation</i>				<i>Postdeliberation</i>			
<i>Probability</i>	<i>Count</i>	<i>Percent</i>	<i>CumPct</i>	<i>Probability</i>	<i>Count</i>	<i>Percent</i>	<i>CumPct</i>
0.00%	6	1.25	1.25	0.00%	7	1.46	1.46
0.02	1	0.21	1.46	0.20	1	0.21	1.67
0.10	1	0.21	1.67	1.00	5	1.04	2.71
1.00	5	1.04	2.71	2.00	2	0.42	3.13
1.80	1	0.21	2.92	4.00	1	0.21	3.33
2.00	6	1.25	4.17	5.00	9	1.88	5.21
3.00	1	0.21	4.38	7.00	2	0.42	5.63
5.00	6	1.25	5.63	8.00	3	0.63	6.25
8.00	2	0.42	6.04	9.00	6	1.25	7.50
10.00	20	4.17	10.21	9.50	1	0.21	7.71
15.00	4	0.83	11.04	10.00	9	1.88	9.58
20.00	16	3.33	14.38	15.00	2	0.42	10.00
25.00	3	0.63	15.00	20.00	8	1.67	11.67
30.00	7	1.46	16.46	25.00	2	0.42	12.08
35.00	1	0.21	16.67	30.00	4	0.83	12.92
40.00	10	2.08	18.75	40.00	6	1.25	14.17
45.00	1	0.21	18.96	50.00	49	10.21	24.38
50.00	76	15.83	34.79	51.00	2	0.42	24.79
51.00	1	0.21	35.00	55.00	1	0.21	25.00
55.00	1	0.21	35.21	56.00	1	0.21	25.21
60.00	4	0.83	36.04	60.00	10	2.08	27.29
65.00	3	0.63	36.67	65.00	1	0.21	27.50
70.00	20	4.17	40.83	70.00	15	3.13	30.63
75.00	23	4.79	45.63	75.00	22	4.58	35.21
80.00	46	9.58	55.21	80.00	38	7.92	43.13
85.00	7	1.46	56.67	85.00	11	2.29	45.42
90.00	56	11.67	68.33	90.00	53	11.04	56.46
95.00	39	8.13	76.46	91.00	1	0.21	56.67
96.00	1	0.21	76.67	95.00	39	8.13	64.79
97.00	2	0.42	77.08	96.00	1	0.21	65.00
98.00	16	3.33	80.42	97.00	2	0.42	65.42
98.80	1	0.21	80.63	98.00	15	3.13	68.54
98.98	2	0.42	81.04	98.20	1	0.21	68.75
99.00	41	8.54	89.58	98.90	1	0.21	68.96
99.08	1	0.21	89.79	99.00	47	9.79	78.75
99.80	2	0.42	90.21	99.08	1	0.21	78.96
99.90	6	1.25	91.46	99.50	1	0.21	79.17
99.98	2	0.42	91.88	99.60	1	0.21	79.38
100.00	39	8.13	100.00	99.80	3	0.63	80.00
				99.90	10	2.08	82.08
				99.98	4	0.83	82.92
				99.99	2	0.42	83.33
				100.00	80	16.67	100.00

Figure B1 Distribution of guilty and not guilty jurors' probabilities of guilt.

