

Koehler, J. J., Chia, A. & Lindsey, J. S. (1995). The random match probability (RMP) in DNA evidence: Irrelevant and prejudicial? Jurimetrics Journal, 35, 201-219.

**THE RANDOM MATCH PROBABILITY (RMP) IN DNA EVIDENCE:  
IRRELEVANT AND PREJUDICIAL?**

Jonathan J. Koehler, Audrey Chia, Samuel Lindsey<sup>1</sup>

January, 1995

---

<sup>1</sup> Jonathan J. Koehler, Behavioral Decision Making Faculty, Management Science and Information Systems Department, CBA 5.202, The University of Texas at Austin, Austin, TX 78712 (Phone: 512-471-7856; e-mail: Koehler@mail.utexas.edu). Audrey Chia, Department of Management, CBA 4.202, The University of Texas at Austin, Austin, TX 78712. Samuel Lindsey, Department of Psychology, University of Virginia, Charlottesville, VA 22903. This research was supported by NSF grant SES-8209544 to the first author. Parts of this paper were presented at the American Psychology-Law Conference in Santa Fe, NM in March 1994. We thank the Travis County jury office in Austin, Texas for its cooperation. We also thank Daniel Kahneman, David Kaye, Elizabeth Loftus, Barry Scheck, William Thompson, and Gary Wells for helpful comments on an earlier draft.

### **Abstract**

The significance of reported DNA matches between a suspect and genetic material recovered from a crime scene is usually represented at trial by the random match probability (RMP). The RMP identifies the frequency of the DNA profile in a reference population. This paper shows that RMPs contribute little to an assessment of the diagnostic significance of a reported DNA match beyond that given by the false positive laboratory error rate when RMPs are several orders of magnitude smaller than this error rate. Evidence that this principle is poorly understood by mock jurors is also presented in support of the argument that introduction of RMPs may be prejudicial in some cases. A new approach, based on Bayesian principles and proficiency test data, is offered for identifying the diagnostic significance of DNA and other forensic science evidence.

The influence of science on the law is nowhere more profound than in criminal proceedings that include DNA analyses of genetic traces (e.g., blood, semen, hair, saliva) recovered from violent crime scenes. Typically, legal factfinders are told that the DNA pattern from a recovered trace "matches" the DNA pattern of a suspected source (usually the defendant). The significance of the match is usually represented by a random match probability (RMP).<sup>2</sup> The RMP identifies the probability that the DNA profile of a randomly selected person from some reference population (e.g., a racial group) will match the profile of the trace evidence. For large populations,  $RMP = P(M|S)$  where M is true match<sup>3</sup> and S is the source. The RMPs reported at trial are often on the order of one in many millions or billions.<sup>4</sup>

---

<sup>2</sup> Some experts use the phrase "random match probability," others use "match probability," and yet others refer to the frequency of the profile. Regardless of the defining phrase, a statistic that corresponds to the RMP as defined above is usually introduced in cases involving DNA evidence.

<sup>3</sup> A "true match" describes a state of nature identified by an error-free test. In contrast, a "reported match" provides evidence of a true match to the extent that the test that produced the match report is error-free.

<sup>4</sup> *Knight v. State* 435 S.E.2d 276 (1993 Ga.App.) ("The expert further noted that the probability of selecting an unrelated individual of the population from the same race as Knight who had a genetic profile matching the semen taken from the victim was one in ten billion"). In many cases, jurors are presented with a variety of extremely small odds estimates: *Dubose v. State* CR-89-359 (1993 Ala.Crim. App.) ("there is a 1 in 500 million probability that the appellant's particular DNA pattern would appear in the North American black population and a 1 in 22 billion probability that the particular pattern would appear in the North American Caucasian population"); *Blige v. State* 440 S.E. 2d 521, 523 (1994 Ga.App.) (Dr. Baird ... stated that in the white population the frequency of occurrence, assuming Hardy-Weinberg equilibrium, was one in three trillion; with down-sizing, the probability was one in one hundred and thirty million"); *State v. Futch*, 860 P.2d 264 (1993 Ore.App.) ("one expert testified that the odds were one in 66 billion, with his most conservative estimate being one in 6.3 billion. ... Another testified that ... it was unlikely that the odds were better than one in 16 billion. Other experts testified that the probability of a random match were one in 127 million"); *People v. Soto* 30 Cal. App. 4th 340 (1994 Ct.App. Fourth Dist.) ("Under the fixed bin method ... the frequency with which one could expect to see the same four points of DNA would be once in 189 million Hispanics or once in 38 million Caucasians. However, if the floating bin method is used, the probability is one in 6.7 billion persons. And if the database is expanded to include the FBI's nationwide DNA database, the frequency probability escalates to one in 2.3 billion Caucasians, or one in 55 million Southwest Hispanics..."). In a 1994 Canadian case, *R. v. Love*, a figure of 1 in 230 billion was introduced, along with a confidence interval that ranged from 1 in hundreds of millions to 1 in several quadrillion.

Scientists disagree about the accuracy of various methods that have been used to compute RMPs.<sup>5</sup> A conservative computation recommended in a 1992 National Research Council (NRC) report has been criticized extensively.<sup>6</sup> As a result, the courts do not have the technical guidance they need to make sensible and consistent judgments concerning the admissibility of RMPs. A second expert panel has been convened to re-examine this issue.<sup>7</sup>

The new panel should adopt a broader frame of the statistical issue than that used by the first panel. In this paper, we show that RMPs (a) are often practically irrelevant to the diagnosticity of a reported DNA match, and (b) may be legally prejudicial. Regarding the first point, we note that a Bayesian likelihood ratio (LR) captures the evidentiary significance of DNA evidence. We show that this ratio is influenced far more by laboratory error<sup>8</sup> rates than by the RMP. Regarding the second point, we present evidence that jurors in DNA cases may be confused by, and overly impressed with, RMPs. We recommend that scientists focus their research efforts away from narrow RMP questions toward more important issues related to the probative value of forensic science reports in general, and DNA matches in particular.

#### A Bayesian Approach to Reported Matches

---

<sup>5</sup> N. J. Risch & B. Devlin, *On the probability of matching DNA fingerprints*, 255 *SCIENCE* 717 (1992); R. C. Lewontin & D. L. Hartl, *Population genetics in forensic DNA typing*, 254 *SCIENCE* 1745 (1991).

<sup>6</sup> Committee on DNA technology in forensic science, *DNA technology in forensic science* (1992) [NRC, 1992]. Second International Conference on Forensic Statistics. *JURIMETRICS* (1994). B. Devlin, N. Risch, & K. Roeder, *Statistical evaluation of DNA fingerprinting: A critique of the NRC's report*, 259 *SCIENCE* 748 (1993) [Devlin et al., 1993]; B. S. Weir, *Forensic population genetics and the National Research Council (NRC)*, 52 *AMERICAN JOURNAL OF HUMAN GENETICS* 437 (1993) [Weir, 1993].

<sup>7</sup> R. Sherman, *New scrutiny for DNA testing*, *THE NATIONAL LAW JOURNAL* 3, 52 (Oct. 18, 1993); R. Barbieri, *Jury still out on DNA evidence*, *THE RECORDER* (Nov. 29, 1993), p. 1. The new National Research Council Committee on DNA Forensic Science held a public hearing in Washington D. C. on November 18, 1994.

<sup>8</sup> As used here, "laboratory error" includes all relevant handling, human and technical errors. These include: mislabelings, misrecordings, misrepresentations, case mix-ups, contaminations, and various interpretive errors.

Forensic science analysis links a suspect to a crime as depicted in Figure 1.

-----  
 Insert Figure 1 About Here  
 -----

Potential for error exists at each step in the inferential chain. A reported match may not be a true match if a laboratory error has been made; a suspect who provides a true match may not be the source of the trace if the match is purely coincidental; the source of a trace may not have been at the crime scene if the real perpetrator deliberately left genetic material from the suspect at the scene; and, finally, the source of the trace may have been at the crime scene, but may have left the trace in a way that is consistent with his innocence. Notice that factors unrelated to the matching evidence--e.g., a suspect's opportunities, motives and alibis--are largely responsible for the strength of the inferences between stages 3 and 4 and between stages 4 and 5. We therefore focus our attention on estimating the odds that a suspect who reportedly matches a recovered trace ("M") is the source (S) of that trace,  $P(S | "M") / P(-S | "M")$ .

According to Bayes's theorem  $[P(S | "M") / P(-S | "M")]$  =  $[P(S) / P(-S)] \times [P("M" | S) / P("M" | -S)]$ . The prior odds ratio  $P(S) / P(-S)$  describes the relative odds that a suspect is and is not the source of the trace prior to a DNA match report. This ratio shows that  $P(S | "M") / P(-S | "M")$  cannot be estimated on the basis of genetic matching evidence alone; nongenetic evidence must also be considered.<sup>9</sup> The LR  $P("M" | S) / P("M" | -S)$  captures the diagnostic

---

<sup>9</sup> This important fact has been widely misunderstood. Many judges, attorneys and experts believe  $P(S | "M") = 1 - RMP$  and  $P(-S | "M") = RMP$ . See J. J. Koehler, *Error and exaggeration in the presentation of DNA evidence at trial*, 34 JURIMETRICS 21 (1993) (Koehler, 1993a); J. J. Koehler, *DNA matches and statistics: Important questions, surprising answers*, 76 JUDICATURE 222 (1993) (Koehler, 1993b). The FBI director in charge of DNA testing explicitly defended this error, J. W. Hicks, *The facts about DNA typing*, 77 JUDICATURE 5, 55 (1993) (Hicks, 1993). But some courts, including the Supreme courts of Connecticut, Minnesota, and New Jersey understand this principle and have warned about the dangers of committing this mistake. See *State v. Skipper* 637 A.2d 1101 (Conn. S. Ct. 1994), *State v. Bloom* 516 N.W. 2d 159 (MN S. Ct. 1994), and *State v. Spann* 617 A.2d 247 (NJ S. Ct. 1993).

value of a reported match. As this ratio increases, one should become increasingly convinced that the reportedly matching suspect is the source of the trace.<sup>10</sup>

The LR is rarely used to convey evidentiary significance in American courtrooms.<sup>11</sup> We urge a policy change where forensic science evidence is introduced. Others have recommended the use of LRs in DNA cases,<sup>12</sup> and some have called for further investigation of the LR calculations.<sup>13</sup> But few recognized the important distinction between a true match and a reported match.<sup>14</sup> This is no mere technicality. It goes straight to the heart of the evidentiary significance of DNA fingerprinting at trial. Below, we show that a properly construed LR--unlike the RMP--depends critically on this distinction. We also show that proficiency tests of practicing laboratories help identify the diagnostic value of reported matches.

#### Estimating the Likelihood Ratio

Recall that  $P("M" | S) / P("M" | -S)$  is the LR that captures the evidentiary significance

<sup>10</sup> In sequential data analyses, the Bayesian odds formulation may be transformed to achieve evidentiary additivity by computing the log to base 10 for each odds component. See L. D. Phillips, *BAYESIAN STATISTICS FOR SOCIAL SCIENTISTS*, 83 (1973).

<sup>11</sup> LRs are not used to describe the significance of reported DNA matches in criminal contexts involving trace evidence (Weir, 1993, p. 437). However they are often used to create paternity indexes in civil cases (D. H. Kaye, *The probability of an ultimate issue: The strange case of paternity testing*, 75 IOWA LAW REVIEW 75 (1989)) and in criminal cases where paternity is at issue (Davis & Davis v. State 476 N.E. 2d 127 (Ind.App.2 Dist. 1985); State v. Jackson 358 S.E.2d 679 (N.C. 1987); State v. Spann 563 A.2d 1145 (N.J.Super.A.D. 1989)).

<sup>12</sup> D. A. Berry, *Inferences using DNA profiling in forensic identification and paternity cases*, 6 STAT. SCI. 175 (1991); I. W. Evett, J. Scrange, & R. Pinchin, *An illustration of the advantages of efficient statistical methods for RFLP analysis in forensic science*, 52 AM. J. HUM. GENET 498 (1993); D. Jarjoura, J. Jamison, & S. Androulakakis, *Likelihood ratios for deoxyribonucleic acid (DNA) typing in criminal cases*, 39 JOURNAL OF FORENSIC SCIENCES 64 (1994); R. Lempert *DNA, science and the law: Two cheers for the ceiling principle*, 34 JURIMETRICS 41 (1993) [Lempert, 1993]; K. Roeder, *DNA fingerprinting: A review of the controversy* 9 STAT. SCI. 222 (1994) [Roeder, 1994]; Weir, 1993.

<sup>13</sup> Devlin et al., 1993, p. 748, fn 40.

<sup>14</sup> But see D. H. Kaye, 7 HARV. J. LAW & TECH 101, 156 (1993).

of a reported match. For DNA and many other highly accurate forensic science tests, the numerator of the LR will be high (e.g., .95, .99, .9999) and the denominator will be low (e.g., .05, .01, .0001). Notice that LR numerator variations in such tests do not substantially affect the magnitude of the LR.

The denominator  $P("M"|-S)$  may be treated as the sum of two "diagnosticity values,"  $P("M"&M|-S)$  and  $P("M"&-M|-S)$ . The first diagnosticity value,  $P("M"&M|-S)$ , may be regarded as one of two addends that comprise the RMP.<sup>15</sup> Thus the LR  $P("M"|S) / P("M"|-S)$  is, in principle, affected by the RMP. As the RMP becomes smaller,  $P("M"|-S)$  decreases, the overall LR increases, and we should be more convinced that a suspect who reportedly matches a DNA trace is the source of that trace.

The analysis below suggests that the magnitude of the second diagnosticity value,  $P("M"&-M|-S)$ , is of greater concern. Whereas the first diagnosticity value  $P("M"&M|-S)$  corresponds closely to the RMP, the second diagnosticity value  $P("M"&-M|-S)$  captures the impact of false positive laboratory error on the LR. The maximum value for the LR -- given when  $P("M"|S) = 1$  cannot be greater than the smaller of  $1 / P("M"&M|-S)$  or  $1 / P("M"&-M|-S)$ . That is, the maximum value for the LR is constrained by the inverse of either of the two diagnosticity values.

#### Estimates Based on Proficiency Test Data

A special case of  $P("M"&-M|-S)$  may be estimated from laboratory proficiency tests.<sup>16</sup>

---

<sup>15</sup>  $RMP = P(MI-S) = P("M"&MI-S) + P("-M"&MI-S)$ .

<sup>16</sup> **The idea that an industry-wide error rate, established through proficiency testing, can be of assistance in a particular case may be as counterintuitive to some legal minds as it is obvious to scientific minds. Some may reject industry-wide error rates on grounds that they are insufficiently relevant to the instant case. The scientist responds by noting that the industry-wide error rate provides a baseline from which an estimate of the probability of error in the instant case can be made. Case-specific circumstances may suggest that error is more or less likely to arise in the instant case than in industry-wide proficiency tests. To the extent these circumstances are diagnostic, the baseline error estimate should be adjusted accordingly. When no such**

These should be blind external tests that use realistic casework samples and require match/no match conclusions.<sup>17</sup> If test samples do not include any true matches that are not from identical sources (i.e.,  $P(M\&S) = 0$ , and  $P(-M\&S) = 0$ ), then  $P("M"\&-M|-S) = P("M"|-M)$ . The diagnosticity value  $P("M"\&-M|-S)$  may now be computed, provided that participating laboratories conduct  $\binom{N}{2}$  pairwise tests (where  $N$  = number of samples).<sup>18</sup>

To date, no DNA tests have satisfied all of these criteria.<sup>19</sup> However, proficiency tests

---

circumstances are presented, the baseline provides a best guess for the chance of error. That unique circumstances always exist does not itself provide reason to disregard industry-wide error rate estimates any more than the fact of human uniqueness provides physicians with reason to disregard medical studies about likely responses to various drugs and procedures. For a related argument favoring introduction of error rates for individual laboratories rather than individual technicians, see B. C. Scheck, *DNA and Daubert*, 15 CARDOZO LAW REVIEW 1959, 1984, fn. 93 (1994) [Scheck, 1994].

<sup>17</sup> Proficiency test results should be reported in dichotomous form so long as test results are reported as "matches" and "exclusions" in courtrooms. See B. Robertson & T. Vignaux UNDERSTANDING EXPERT EVIDENCE (in press) for an argument that match/exclusion terminology should be abandoned. In a talk given at the Second International Conference on Forensic Statistics, Tempe, AZ (March, 1993) ("*DNA Profiling: What's all the fuss about numbers?*"), I. W. Evett suggested that qualitative frequency terms (e.g., "rare") can be used to describe match quality.

<sup>18</sup> If laboratories do not conduct  $\binom{N}{2}$  pairwise tests, but instead type each of  $N$  samples individually and subsequently make  $\binom{N}{2}$  "recycled" pairwise comparisons,  $P("M"|-M)$  estimates may be inaccurate. For example, if the rate at which one sample contaminates the other on loading is 2%, far less than 2% of the pairwise comparisons on non-matching samples will be contaminated when a large proportion of test samples do not match (Personal communication, William C. Thompson, 3/2/94. See also W. C. Thompson, *Commentary on Kathryn Roeder, DNA fingerprinting: A review of the controversy*, 9 STAT. SCI. 263, 265 (1994)) [Thompson, 1994]. Likewise, when a large proportion of test samples match (e.g., eight of ten samples are from a common source), an error that occurs 2% of the time may lead to an excessively high proportion of reported matches on the relatively few non-matching pairs. These interpretive difficulties can be avoided by requiring pairwise tests rather than recycled comparisons.

<sup>19</sup> W. C. Thompson, *Evaluating the admissibility of new genetic identification tests : Lessons from the 'DNA War,'* 84 J. CRIM. LAW & CRIM. 22, 93. At present, most proficiency tests violate all of these criteria (i.e., they are nonblind, internal tests on large, pristine samples).

performed in 1987 and 1988 by the California Association of Crime Laboratory Directors (CACLD) satisfy more criteria than other tests, and therefore provide some basis for an initial estimate. We also provide estimates based on tests reported by Collaborative Testing Services (CTS) in 1992, 1993 and 1994.

CACLD Tests. In the CACLD tests, three commercial laboratories were provided with fairly realistic blood, semen, and hair samples and asked to determine which samples originated from common sources.<sup>20</sup> Results from the two laboratories that used RFLP techniques (as opposed to PCR) and provided match/no match conclusions are used here. Each laboratory was given 65 samples, yielding 2080 pairwise comparisons.<sup>21</sup> The two laboratories faced a combined total of 142 matching pairs and 4,018 non-matching pairs. Two (erroneous) match calls were made on non-matching pairs.<sup>22</sup> Therefore  $P("M"&-M|-S) = P("M"|-M) = 2/4,018$ . In a follow-up study, the laboratories faced a combined total of 2,450 pairwise comparisons, 2,370 of which were non-matching pairs. Three (erroneous) match calls were made.<sup>23</sup> Therefore  $P("M"&-M|-S) = 3/2,370$ . The precise diagnosticity value is less important than it's order of magnitude, which appears to be one in hundreds or several

---

<sup>20</sup> **California Association of Crime Laboratory Directors, DNA COMMITTEE REPORT #6.**, (October 1, 1988) [CACLD, 1988].

<sup>21</sup> **These pairwise comparisons are of the recycled variety cautioned against in fn 18. Because a large proportion of test samples were non-matching, certain errors may be underrepresented, and the resultant diagnosticity value may be too low.**

<sup>22</sup> **CACLD, 1988, p. 4-5; W. C. Thompson and E. S. Ford, *The meaning of a match: Sources of ambiguity in the interpretation of DNA prints*, in FORENSIC DNA TECHNOLOGY, M. A. Farley and J. J. Harrington, Eds. (Lewis, Chelsea, MI, 1991), chap. 7, pp. 115. Sample 59 was erroneously matched with samples 57 and 58 by one of the laboratories.**

<sup>23</sup> **California Association of Crime Laboratory Directors, DNA COMMITTEE--RESULTS OF BLIND TRIAL #2**, (March 29, 1990), pp. 6. Sample 142 was erroneously matched with samples 106, 134, and 140 by one of the laboratories.

thousand.

CTS Tests. Three sets of RFLP proficiency tests<sup>24</sup> conducted by CTS provide additional support for a diagnosticity value estimate of this magnitude.<sup>25</sup> We caution, however, that these tests, like most proficiency tests to date, were conducted under conditions that may not capture all of the difficulties associated with ordinary casework.<sup>26</sup>

We also caution that the use of one or two mixed samples in the CTS proficiency tests introduces some interpretive ambiguity into the computation of diagnosticity values.

Specifically, when sample A matched one of two fractions of a mixed sample B (e.g., A matches B<sub>1</sub>), it is not clear whether an erroneous match call between sample A and the second fraction of the mixed sample (B<sub>2</sub>) should be scored as an error. On the one hand, all match calls on non-matching samples and fractions of samples are errors. On the other hand, such "fraction errors" are unlikely to falsely incriminate an innocent suspect because the suspect matches the other fraction of the mixed sample. For this reason, diagnosticity values that include and exclude fraction comparisons and errors are computed for each of the three CTS reported below.

In the 1992 test, 34 laboratories were provided with 5 blood stains from different

---

<sup>24</sup> Although most participating laboratories in the CTS tests conducted RFLP analyses, some performed PCR analyses, and some performed both types of analyses. The data reported here pertain to RFLP analyses only.

<sup>25</sup> The three CTS tests are DNA PROFILING, REPORT No. 91-15 (1992), DNA PROFILING, REPORT No. 92-15 (1993), and DNA PROFILING, REPORT No. 93C (1994). The results discussed here are for laboratories that use RFLP techniques.

<sup>26</sup> Many of the stains were unusually large, supplied on new cotton cloth, and carefully preserved. Laboratories were told how many stains there would be, how and when the samples were prepared, and, in one case (Report No. 91-15), which stain contained a mixture of blood and semen. Finally, the source of some of the samples was identical to those that appeared in earlier CTS tests. One participating analyst in the 1992 test criticized the composition of test samples and wrote, "In the future, let's have samples that more realistically reflect casework materials" (p. 56). In response, CTS "acknowledge[s] that the mixed sample was not truly representative of case samples" (p. 1).

sources (A-E) along with a 6th stain (F) that was a mixture of blood from source A and semen from source D. When the 6th stain is separated into two fractions, there

were  $\binom{7}{2} = 21$  possible pairwise comparisons, two of which were matches, nineteen of which

were nonmatches. The analysts provided judgments about a combined total of 64 matching pairs and 617 non-matching pairs.<sup>27</sup> At least three (erroneous) match calls were made on non-matching pairs.<sup>28</sup> Therefore, based on this test,  $P("M" \& -M | -S) = P("M" | -M) \geq 33/617$ .

When fraction comparisons and errors are excluded from consideration, this test gives

$P("M" \& -M | -S) = P("M" | -M) \geq 40/537$ .<sup>29</sup>

In the 1993 test, 45 laboratories that reported RFLP results only were provided with 4 blood stains from different known sources (A-D) along with a 5th stain (E) that was a mixture

---

<sup>27</sup> Some analysts did not provide judgments about some of the pairwise comparisons. When the relevant data and conclusions were missing, some pairwise comparisons were excluded from consideration.

<sup>28</sup> Collaborative Testing Services, DNA PROFILING, REPORT 91-15 (1992). Laboratory 1504 erroneously concluded that the blood fraction of F matched sample D, and the sperm fraction of F matched sample A. Laboratory 1518 erroneously concluded that the blood fraction of F matched sample D (although their band sizes suggested otherwise).

In addition, we note three more ambiguous instances that were not counted as errors for our purposes. First, laboratory 1528 erroneously concluded that samples B and E had a common source based on a single probe. Because such an error is more relevant to  $P("M" \& MI - S)$  than to  $P("M" \& -MI - S)$ , it was not counted as an error here. Second, laboratory 1508 concluded "The female fraction of sample F continued profiles interpretable as a mixture of samples matching A and D." As this laboratory should have realized, the female fraction of sample F could not match D because the laboratory previously (and correctly) identified the D as the donor of the male fraction of sample F. Third, laboratory 1532 apparently reversed the band sizes for the blood fraction of sample F with the semen fraction of sample F in the RFLP analysis. This created two additional false positive errors (sperm fraction of F erroneously matched to A, blood fraction of F erroneously matched to D) across the two probes in which profiles were obtained for all samples. However, the conclusions reported by this laboratory apparently reflected the results of their PCR analysis rather than their RFLP analysis. The reversal did not occur on the PCR analyses. The laboratory explained the discrepancy by noting that the samples were not "handled according to the laboratory's regular quality control program; having a second person checking all transfers from one tube to another" (p. 56).

<sup>29</sup> Results exclude data from laboratory 1528, the laboratory that reported results from a single probe only. Inclusion of these results yields a diagnosticity value of 1/554 due to an incorrect match report for samples B and E.

of blood from source A ( $E_1$ ) and semen from source C ( $E_2$ ).<sup>30</sup> Because the laboratories were told that samples A, B, C, and D were from distinct sources, pairwise comparisons among them were not performed. Therefore eight pairwise comparisons were made, two involving matching pairs ( $AE_1$ ,  $CE_2$ ) and six involving non-matching pairs ( $AE_2$ ,  $BE_1$ ,  $BE_2$ ,  $CE_1$ ,  $DE_1$ ,  $DE_2$ ). The analysts provided judgments about a combined total of 85 matching pairs and 223 non-matching pairs.<sup>31</sup> Eighteen match calls were made on non-matching pairs.<sup>32</sup> Based on this test,  $P("M" \& -M | -S) = P("M" | -M) \geq 518/223$ . When fraction comparisons and errors are excluded, this test gives  $P("M" \& -M | -S) = 4/151$ .

In the 1994 test, 45 laboratories<sup>33</sup> were provided with 3 blood stains from different known sources (A-C). In addition, they received a mixture of blood from sources B and C (D),<sup>34</sup> and a mixture of blood from source A and semen from source B ( $E_1$  and  $E_2$ ). Nine pairwise comparisons were made, four involving matching pairs (BD, CD,  $AE_1$ ,  $BE_2$ ) and five involving non-matching pairs (AD,  $BE_1$ ,  $CE_1$ ,  $AE_2$ ,  $CE_2$ ). The analysts provided judgments

---

<sup>30</sup> Data from 12 laboratories that performed PCR analyses and 10 laboratories that performed both RFLP and PCR analyses have been excluded from consideration.

<sup>31</sup> Here too, missing values reduced the number of pairwise comparisons from what might be expected. There were five instances in which no conclusions were drawn about a matching pair (all five pertained to  $CE_2$ ), and 47 instances in which no conclusions were drawn about a non-matching pair (spread roughly equally among each of the 6 non-matching pairs).

<sup>32</sup> Collaborative Testing Services, DNA PROFILING, REPORT No. 92-15 (1993).

Fourteen of the eighteen false positive errors were fraction errors. Six laboratories erroneously concluded that the male fraction of E matched sample A, and eight laboratories erroneously concluded that the female fraction of E matched sample C. The non-fraction errors occurred when laboratories 1504 and 1537 erroneously concluded that sample D matched  $E_1$  (i.e., the female fraction of E) and that sample D matched  $E_2$  (i.e., the male fraction of E).

<sup>33</sup> Data from 14 laboratories that performed PCR analyses and 19 laboratories that performed both RFLP and PCR analyses have been excluded from consideration.

<sup>34</sup> Because laboratories were not told that D was a mixture, they were asked whether "any DNA from item D [as opposed to  $D_1$  and  $D_2$ ] can be matched to the DNA from items A-C."

about a combined total of 158 matching pairs and 177 non-matching pairs. Fourteen incorrect calls were made on non-matching pairs. This gives  $P("M"&-M|-S) = 14/177$ .<sup>35</sup> When fraction comparisons and errors are excluded, this test gives  $P("M"&-M|-S) = 0/115$ .

Conclusions and an Example. Despite some ambiguities in the interpretation of proficiency test data,<sup>36</sup> performance on the CACLD and CTS tests provides some basis for estimating a lower bound diagnosticity value. When the results from the 2 CACLD studies are combined,  $P("M"&-M|-S) = 5/6388$  (approximately 1/1300). When the results from the 3 CTS tests are combined (excluding fraction comparisons and errors),  $P("M"&-M|-S) = 4/803$  (approximately 1/200).<sup>37</sup> When the results from all of the CACLD and CTS studies are pooled,  $P("M"&-M|-S) = 9/7191$ , or approximately 1/800.

These proficiency test data suggest that the LR, which is less than or equal to  $1 / P("M"&-M|-S)$ , is not greater than several hundred or, perhaps, several thousand. Very small RMPs do not affect this ceiling. Indeed, when  $P("M"&M|-S)$  (the diagnosticity value that helps comprise the RMP) is several orders of magnitude smaller than  $P("M"&-M|-S)$  (the diagnosticity value that captures the impact of laboratory error), it has virtually no impact on the LR. The implication is that, in cases where the RMP is very small (e.g., 1 in many

---

<sup>35</sup> Collaborative Testing Services, DNA PROFILING, REPORT No. 93C (1994).

Fourteen of the eighteen false positive errors were fraction errors. Six laboratories erroneously concluded that the male fraction of E matched sample A, and eight laboratories erroneously concluded that the female fraction of E (E<sub>f</sub>) matched sample C.

<sup>36</sup> We acknowledge that there is room for disagreement about whether certain results should be scored as errors or not. For example, some may not wish to excuse the RFLP reversal errors made by laboratory 1532 on the 1992 CTS test. These disagreements do not substantially alter the diagnosticity value estimates provided in the text, nor do they undermine the central thesis of the paper.

<sup>37</sup> When fraction comparisons and errors are included, the results from the three CTS studies give  $P("M"&-M|-S) = 35/1017$  (approximately 1/29).

millions or billions), the diagnosticity of a reported match is reasonably conveyed as  $1 / P("M"&-M|-S)$ .<sup>38</sup> Thus, if  $RMP = .0000001$  and  $P("M"&-M|-S) = .0002$ , the maximum diagnosticity of a reported match ranges from 4997.5 to 5000.0.<sup>39</sup>

A nonforensic example illustrates the broader principle. Suppose a baseball infielder makes throwing errors fewer than one time in a million, but makes fielding errors 2% of the time. Assuming that errors are equally distributed across trials, the chance of an error on the infielder's next attempt (due to either throwing or fielding) is at least 2%. If an error occurs, it will almost surely be a fielding error. Further reductions in the infielder's throwing error rate to, say, one in a hundred million, will not be reflected in the overall error rate. Thus, a baseball talent scout should be no more impressed by improvements in the infielder's throwing ability than the legal factfinder should be upon hearing the vanishingly small RMPs that accompany reported DNA matches at trial. Just as the infielder's 2% fielding error rate sets a lower bound threshold for error estimates, the forensic scientist's laboratory error rate sets a

---

<sup>38</sup> Many others have reached a similar conclusion. See D. A. Berry, *Comment* [on DNA fingerprinting: A review of the controversy by K. Roeder], 9 STAT SCI. 252, 253 (1994); P. J. Hagerman, *DNA typing in the forensic arena*, 47 AM. J. HUM. GENET. 876 (1990); R. Lempert, *Some caveats concerning DNA as criminal identification evidence: With thanks to the Reverend Bayes*, 13 CARDOZO LAW REVIEW 303, 323-8 (1991); Lempert, 1993, p. 48; R. Lempert, *Comment: theory and practice in DNA fingerprinting* 9 STAT SCI. 255, 257 (1994) (laboratory error places the most serious limits on the evidentiary import of reported DNA matches. . . . The possibility of error must be honestly faced, and it must be incorporated into estimates of the incriminatory power of DNA matches"); L. D. Mueller, *The use of DNA typing in forensic science*, 3 ACCOUNTABILITY IN RESEARCH 1, 4 (1993); Thompson, 1994, at 266; B. S. Weir, *Population genetics in the forensic DNA debate*, 89 PROC. NATL. ACAD. SCI. 11654, 11658 (1992).

<sup>39</sup> If  $RMP = P(MI-S) = .0000001$ , then  $0 < P("M"&MI-S) < .0000001$ . So, if  $P("M"&MI-S) = .0002$ , and  $LR = P("M"IS) / P("M"|-S)$  is maximized when  $P("M"IS) = 1$ , then the maximum LR depends on  $P("M"&MI-S)$  as follows:  $1 / (.0002) < LR < 1 / (.0000001+.0002) = 5000 < LR < 4997.5$ . The LR will be smaller when  $P("M"IS) < 1$ .

lower bound for false positive match reports. But do jurors recognize this? The following two sections examine this empirical issue.

### University Study

In cases involving reported DNA matches, forensic scientists sometimes provide RMPs, along with misleading assurances that DNA typing is infallible.<sup>40</sup> The NRC cautioned against such hyperbole and emphasized the importance of providing jurors with accurate laboratory error rates in addition to RMPs.<sup>41</sup> Unfortunately, the NRC provided no guidance for combining RMPs and laboratory error rates to determine the diagnostic value of a reported match.

Left to their own devices, jurors may overweight extremely small RMPs for two reasons. First, vividness theory suggests that people give inferential weight to evidence in proportion to its vividness or memorability.<sup>42</sup> Very small statistics, such as 1 in millions or billions, may be quite vivid and memorable, and therefore exert a large impact on verdicts. Second, people often combine probabilistic items of evidence by averaging them.<sup>43</sup> When an

---

<sup>40</sup> *Kelly v. State*, 792 S.W. 2d 579 (Tex. App.—Fort Worth 1990), transcript, p. 919 (“There is no way to get a false positive with this technology”); *Cobey v. State*, 559 A. 2d 391 (Md. App. 1989) (An incorrect match is an “impossible” result); *People v. Fishback*, 829 P.2d 489 (Colo. App. 1991) (DNA analysis is “fallsafe”). For other examples, see Koehler, 1993a, p. 23 fn. 8; Koehler, 1993b, p. 228, fn. 16.

<sup>41</sup> “Interpretation of DNA typing results depend not only on population genetics, but also on laboratory error. . . . A laboratory’s overall rate of incorrect conclusions due to error should be reported with, but separately from, the probability of coincidental matches in the population. Both should be weighed in evaluating evidence” (NRC, 1992, p. 88, 94).

<sup>42</sup> R. E. Nisbett & L. Ross, *HUMAN INFERENCE: STRATEGIES AND SHORTCOMINGS OF SOCIAL JUDGMENT* (1980).

<sup>43</sup> N. H. Anderson, *Information integration theory: A brief survey*. In D. Krantz, R. C. Atkinson, R. D. Luce, P. Suppes (eds.), *CONTEMPORARY DEVELOPMENTS IN MATHEMATICAL PSYCHOLOGY*, VOL. 2. (1974); L. L. Lopes, *Procedural debiasing* 64 *ACTA PSYCHOLOGICA* 167 (1987); J. C. Shanteau, *Descriptive versus normative models of sequential inference judgment*, 93 *JOURNAL OF EXPERIMENTAL PSYCHOLOGY* 63 (1972).

averaging strategy is used to estimate the disjunctive probability of error from either of two sources, one of which is several orders of magnitude smaller than the other, it substantially overweights the contribution of the smaller error source. In this way, jurors who are provided with RMPs and laboratory error rates may attach great significance to very small--but minimally diagnostic--RMPs.<sup>44</sup>

In sum, vividness and averaging theories lead to a prediction that jurors who are provided with small RMPs may be substantially more willing to convict defendants than those who are not, regardless of whether or not they are provided with laboratory error rates. But when jurors are provided with an aggregated error rate that reflects both types of error--an error rate that is often fairly estimated by the laboratory error rate alone--they may be far less willing to convict.

To examine these hypotheses, from January, 1993 (??) to March, 1994 (??), we gave 259 jury-eligible subjects<sup>45</sup> at the University of Texas at Austin written summaries of a murder case. The evidence included a DNA match between the defendant and a blood trace recovered from the victim's fingernails. Aside from the DNA evidence, the case against the defendant was circumstantial and weak.<sup>46</sup> A typical case summary is presented in the Appendix. One hundred ninety-three subjects were randomly assigned both to one of three laboratory error rate groups (absent, .02, .001) and to one of two RMP groups (absent, .000000001). The laboratory error rates are representative of previous estimates,<sup>47</sup> and the RMP is representative

---

<sup>44</sup> The possibility that jurors will average RMPs and laboratory error rates was first raised by Lempert, 1993, p. 54.

<sup>45</sup> Subjects were U.S. citizens, 18 years of age or older, who had not been convicted of a felony.

<sup>46</sup> A strong case might have produced such a high proportion of guilty verdicts that the effect under investigation would be washed out.

<sup>47</sup> Hicks, 1993, p. 55 (.0008); Koehler, 1993b, p. 229 (.01-.04); Lempert, 1991, p. 325 (.01-.02); K. Roeder, 1994 (.0008); Thompson, 1994

of those heard at trial.<sup>48</sup> Those subjects who received two probability values received them in random order. Sixty-six subjects were presented with one of two aggregated error rates (.02, .001). They were told that this value reflected the combined possibility of error from coincidental matches and laboratory mistakes. Notice that these values roughly correspond to a normative aggregation of the RMP and laboratory error rate values provided to other subjects.

Subjects studied the evidence and provided verdicts (guilty, not guilty). By the likelihood ratio analysis above, the .000000001 RMP should have little effect on verdicts when paired with laboratory error rates (.02 and .001) that are many orders of magnitude larger. But by vividness and averaging, it was predicted that the RMP--when presented separately--would exert a substantial impact on jurors' verdicts.

Figure 2 shows that introduction of the RMP had a strong effect on mock jurors' verdicts both when laboratory error rates were absent and present.<sup>49</sup> Jurors in the RMP-absent groups (n=98) convicted only 12-18% of the time, whereas jurors in the RMP-present groups (n=95) convicted 44% of the time (loglinear partial chi-square(1, n=193) = 18.34,  $p < .0001$ ). Introduction of laboratory error rates (.1%, 2%) did not significantly affect conviction rates (loglinear partial chi-square(1, n=193) = 0.29, n.s.). Among the 66 jurors who received the aggregated error rate, the conviction rate was 21%. This conviction rate was not

---

[.00666].

<sup>48</sup> Although RMPs computed using the "ceiling principle" technique recommended by the 1992 NRC are usually larger than .000000001, smaller RMPs computed in other ways are commonly presented as well. ("the ceiling principle was not intended to be exclusive. Expert witnesses were still free to provide their statistical 'best estimate' of genotype frequencies ..." E. Lander & B. Budowle, *DNA fingerprinting dispute laid to rest*, NATURE, 736 (1994)).

<sup>49</sup> No statistically significant differences were detected between the 2% and .1% laboratory error rates (Main effect: loglinear partial chi-square(1, n=128) = 1.05, n.s.; two-way interaction: loglinear partial chi-square(1, n=128) = 0.09, n.s.). The .1% and 2% laboratory error rates were therefore collapsed into a single cell (laboratory error rate present). Likewise, the order in which the probabilities were presented produced no main effects or interactions and is therefore ignored.

significantly different from the laboratory error-present / RMP-absent groups.

-----  
 Insert Figure 2 About Here  
 -----

These data support predictions given by the vividness and averaging theories. Our mock jurors who received the  $RMP = .000000001$  statistic apparently were much more impressed with the DNA evidence than were those who did not receive an RMP statistic. The probability that a juror would convict in the hypothetical case doubled or tripled when the RMP was introduced. Separate introduction of the highly diagnostic laboratory error rate had little impact. But when the aggregated error rate was introduced, and jurors were not afforded separate access to a small RMP, convictions rates declined by nearly 50%. This suggests that the mock jurors who received separate RMP and laboratory error rate estimates did not combine these values according to the normative guidelines outlined here (which would have jurors weight the two estimates as they would a single aggregated value equal to the laboratory error rate).

#### Replication: Travis County Study

We substantially replicated this pattern of data on a sample of 269 Travis County jury-eligible subjects in Austin, Texas. The study was performed during three jury empaneling sessions in January, 1994.

The stimulus materials, design, and procedure were nearly identical to the University study. However, the aggregated error rate groups were not included in the Travis County replication. As in the University study, introduction of the RMP had a strong effect on mock jurors' willingness to convict at each of the three levels of laboratory error. Jurors in the RMP-absent groups ( $n = 144$ ) convicted 18-25% of the time, whereas jurors in the RMP-

present groups (n=125) convicted 44-54% of the time (loglinear partial chi-square(1, n=269) = 23.78, p < .0001).

-----  
 Insert Figure 3 About Here  
 -----

In sum, these data suggest that the current practice of providing jurors with RMPs is potentially prejudicial. Our mock jurors who were provided with a very small, but diagnostically irrelevant, RMP convicted twice as often as jurors who were provided with more diagnostic statistical information (laboratory error rate, aggregated error rate).<sup>50</sup> These results lend some credence to Professor Richard Lempert's suspicion that "jurors provided with a laboratory's false positive rate and with information about the likelihood, assuming no testing error of a match if the evidence DNA was not the defendant's, are likely to be hopelessly confused about the weight to accord the testimony."<sup>51</sup>

### Conclusion

This paper showed that the evidentiary significance of a reported DNA match is often

---

<sup>50</sup> Shortly before this paper went to press, we attempted to replicate this pattern of results in a "Post O.J. Simpson world." We wondered whether the relentless national publicity during the latter half of 1994 about the meaning and implications of DNA evidence in the O. J. Simpson murder case would alter the previously observed data pattern.

Although our sample size was relatively small (n=87), a substantially similar pattern emerged. Fifteen percent (n=40) of mock jurors in an aggregated error rate group voted to convict, whereas 38% (n=47) of jurors in the RMP-present group voted to convict. The aggregated error rate group heard that "the probability that a coincidental match or a human or laboratory error occurred which may have led to an incorrect match report was about 2%"; the RMP-present group heard that "(a) the probability that the DNA profile of a random man would match the DNA profile of the recovered hair as well as did Mr. Applegate's own hair was 1 in 1 billion, and (b) the probability that a laboratory or human error occurred which may have led to an incorrect match report was about 2%." We note that due to similarities between the hypothetical case used previously and certain well-publicized aspects the Simpson case, the hypothetical DNA case provided in this replication involved a convenience store robbery and was otherwise substantially different from the scenario presented in the Appendix.

<sup>51</sup> Lempert, 1991, at p. 325.

informed more by the possibility of laboratory error than by the possibility of a coincidental match. This is not because DNA typing procedures are sloppy or otherwise unreliable. Instead, it is a necessary consequence of a technology that has theoretically extreme discriminatory power. As the discriminatory power of a profiling technique increases, errors caused by coincidental circumstance become increasingly insignificant relative to errors that arise for other reasons.<sup>52</sup> At some point, increases in discriminatory power no longer increase the diagnostic value of a reported match. In the case of multiple probe DNA profiling, we have reached that point.<sup>53</sup>

This paper also presented evidence that jurors may not appreciate this principle. Mock jurors in a murder case that included DNA evidence attached great weight to a small RMP, even when it lacked diagnostic value relative to available laboratory error rates. We are concerned that this effect, in combination with misleading assurances from forensic science experts that laboratory errors are impossible or nearly impossible, could lead to convictions where acquittals might otherwise result. This concern is greatest in cases that rest largely or entirely on DNA evidence.<sup>54</sup>

An important issue for future research is whether the findings described here will persist even when corrective steps are taken. For example, expert testimony or judicial

---

<sup>52</sup> See also D. J. Balding and P. Donnelly 368 NATURE 285 (1994).

<sup>53</sup> These arguments apply with equal force to other forensic science techniques including fingerprinting. Fingerprinting is no less a probabilistic enterprise than DNA profiling. The fact that probabilities are not used to express the evidentiary significance of fingerprints is due more to historical circumstance than to sound scientific reasoning (M. J. Saks, SHEPARD'S EXPERT AND SCI. EVID. QUART., in press). We would argue, therefore, that fingerprint experts should not be permitted to testify that a particular person is certain to be the source of a particular print absent convincing evidence that errors do not occur.

<sup>54</sup> DNA data banks are now used in some states (e.g., Illinois, Minnesota, and Virginia) to identify potential suspects in cases where genetic material is recovered from the crime scene.

instructions explaining that laboratory error rates should be considered even in cases involving very small RMPs may be sufficient to sensitize jurors to the normative issue. We also caution that the studies presented here do not consider the effects of group deliberation.

From a forensic science standpoint, efforts should be made to measure the rates at which laboratory errors occur. We therefore add our voice to recent calls for the immediate implementation of a rigorous proficiency testing program for all DNA laboratories and applaud the progress that has been made thus far.<sup>55</sup> Error rate estimates obtained from such testing are critically important for identifying the chance of error in any given case.<sup>56</sup> Efforts should also be made to reduce or eliminate certain types of errors. Increases in the diagnosticity of reported matches are closely tied to error rate reductions.

From a legal policy standpoint, scientists should not be permitted to describe the significance of a reported DNA match via RMPs and vague comments about the improbability of laboratory error. Instead, they should carefully explain the difference between a reported

---

<sup>55</sup> NRC, 1992 ("Laboratory error rates should be measured with appropriate proficiency tests and should play a role in the interpretation of results of forensic DNA typing," p. 15 & p. 94; "Most important, there is no substitute for rigorous external proficiency testing via blind trials. Such proficiency testing constitutes scientific confirmation that a laboratory's implementation of a method is valid not only in theory but in practice. No laboratory should let its results with a new DNA typing method be used in court, unless it has undergone such proficiency testing via blind trials," p. 55); W. C. Thompson, *Evaluating the admissibility of new genetic identification tests: Lessons from the DNA war*, 84 JOURNAL OF CRIMINAL LAW AND CRIMINOLOGY 22 (1993); J. L. Peterson, *Recent findings from the forensic science foundations's proficiency testing program*, PROCEEDINGS OF THE SECOND INTERNATIONAL CONFERENCE ON FORENSIC STATISTICS, C3-C24 (1993); R. C. Lewontin, *Comment: The use of DNA profiles in forensic contexts*, 9 STAT SCI. 259, 261 (1994). The recently passed "DNA identification Act of 1994" provides guidelines for DNA quality assurance that include proficiency testing (see section 210303, Violent Crime Control and Law Enforcement Act of 1994, 103rd Congress, H.R. 3355, September 12, 1994).

<sup>56</sup> In cases where an individual technician or an individual laboratory has not performed enough proficiency tests to produce sufficiently reliable and informative error rate estimates (as given by confidence intervals), an industry-wide estimate should be used in its place.

match and a true match, and between a reported match and a true source. Unlike some,<sup>57</sup> we do not believe so-called "statistical probability" evidence should be inadmissible when it is highly probative. However, we do believe that a statistic whose evidentiary value is effectively trumped by another statistic should be excluded on grounds of insufficient relevance (i.e., Federal Rule of Evidence 401). Moreover, our data indicating that people may attach substantial weight to minimally diagnostic RMPs that are presented separate from error rate estimates, suggest that the prejudicial impact of RMPs may exceed their probative value. This raises the possibility of a Federal Rule of Evidence 403 challenge to the admissibility of RMPs as well.<sup>58</sup>

One alternative to present policy is to admit a LR that conveys the evidentiary significance of a reported match. Though less dramatic than some of the quantities currently presented in our courtrooms, properly understood LRs on the order of 100 or 1000 can and should be extremely persuasive.<sup>59</sup>

A LR approach has its own dangers. The appropriate interpretation of LRs is not intuitively obvious, and a body of literature indicates that they are easily confused with

---

<sup>57</sup> L. H. Tribe, *Trial by mathematics: precision and ritual in the legal process*, 84 HARVARD LAW REVIEW 1329 (1971).

<sup>58</sup> In a *Daubert* analysis of DNA evidence, Professor Barry Scheck goes a step further and concludes that because the reliability of DNA evidence is signaled by laboratory error rates, "without an estimate of error rate, a serious argument can be made that the DNA evidence should be excluded under Federal Rule of Evidence 403, if not Rule 401" Scheck, 1994, *supra* note 16, at 1988-9. When such error rates are available, Professor Scheck argues that *Daubert* mandates that "that error rate should be the only probability offered about the likelihood that the defendant was not the source of DNA trace evidence" at 1997.

<sup>59</sup> See M. Finkelstein, & W. Fairly, *A Bayesian approach to identification evidence*, 83 HARVARD LAW REVIEW 489 (1970).

posterior probabilities.<sup>60</sup> But there is also evidence that people can be trained to use quantitative information quickly and successfully.<sup>61</sup> Whether these findings can be extended to LR<sub>s</sub> in a legal context remains to be seen.

In sum, when verdicts hinge on the availability and interpretation of scientific evidence, scientists have a duty to present that evidence in an accurate and comprehensible form. The current practice of supplying jurors with extremely small RMPs is misleading when the risk of false positive laboratory errors is several orders of magnitude larger. In these cases, RMPs are diagnostically irrelevant and, perhaps, legally prejudicial.

---

<sup>60</sup> M. D. S. Braine, J. Connel, J. Freitag, & D. P. O'Brien, *Is the base-rate fallacy an instances of asserting the consequent?*, in K. L. Gilhooly, M. T. G. Leane, R. H. Logie, & G. Erdos (Eds.), *Lines of Thinking* (vol. 1) (1990); D. M. Eddy, *Probabilistic reasoning in clinical medicine: Problems and opportunities*, in D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases* (1982); R. L. Hamm, 72 *Psych. Rep.* 219 (1993); D. H. Kaye & J. J. Koehler, *Can jurors understand probabilistic evidence?*, 154 *Journal of the Royal Statistical Society Series A*, 75 (1991); W. C. Thompson, *Are juries competent to evaluate statistical evidence?*, 52 *Law and Contemporary Problems* 9 (1989); W. A. Wagenaar, *The proper seat: A Bayesian discussion of the position of expert witnesses*, 12 *Law and Human Behavior* 499 (1988).

<sup>61</sup> R. E. Nisbett, B. T. Fong, D. R. Lehman, & P. W. Cheng, *Teaching reasoning*, 238 *Science* 625 (1987); B. C. Smith, S. D. Penrod & A. L. Otto, *Bayesian presentations and juror use of probabilistic evidence*, Paper presented at American Psychology-Law Society Biennial Meeting San Diego, CA (March 1992).

**Appendix: Example of case used in Experiment\***

In the case of State v. Murphy, Steven Murphy was accused of strangling his wife, Sandra Murphy, to death on the night of January 8, 1987. The murder was first reported in a phone call to 911 by Mr. Murphy at 11:54 p.m. A tape recording of this call reveals that Mr. Murphy sobbed throughout the 4 minute conversation with the emergency operator.

The prosecution alleges that Mr. Murphy killed his wife in a fit of rage motivated by his displeasure over her decision to accept a part-time job at a local restaurant. Some evidence was introduced that Mr. Murphy did not want his wife to work and that this disagreement put a strain on their marriage. Evidence that Mr. Murphy threatened violence against his wife on a previous occasion was also introduced.

The defense contends that the Murphys were a loving family and that the police charged Steven Murphy with murdering his wife because they were too lazy to conduct a thorough investigation to find the real killers. The defense alleges that the killer or killers were most likely neighborhood drug addicts with a recent history of breaking into small family homes and robbing the owners.

Blood scrapings that were found beneath Ms. Murphy's fingernails were submitted to a DNA fingerprinting laboratory for analysis, along with a sample of Mr. Murphy's own blood. At trial, an expert testified that his tests could not rule out Mr. Murphy as a possible source of the DNA prints that were isolated from the blood scrapings beneath Ms. Murphy's fingernails.

Furthermore, the expert testified that (a) the probability that the DNA prints of a random man would match these scrapings as well as did Mr. Murphy's own DNA prints was 1 in 1,000,000,000, and (b) the probability that a human error occurred which may have led to an incorrect match report was about 1 in 1,000.

As a juror, you are instructed to consider all of the evidence in this case carefully. You are to find against Mr. Murphy only if the evidence convinces you "beyond a reasonable doubt" that Mr. Murphy is guilty of this crime.

Question

What verdict would you return: Not Guilty or Guilty? \_\_\_\_\_

\* RMP = 1 in 1,000,000,000, laboratory error rate = 1 in 1,000.